DOCUMENT RESUME

ED 417 749 IR 056 977

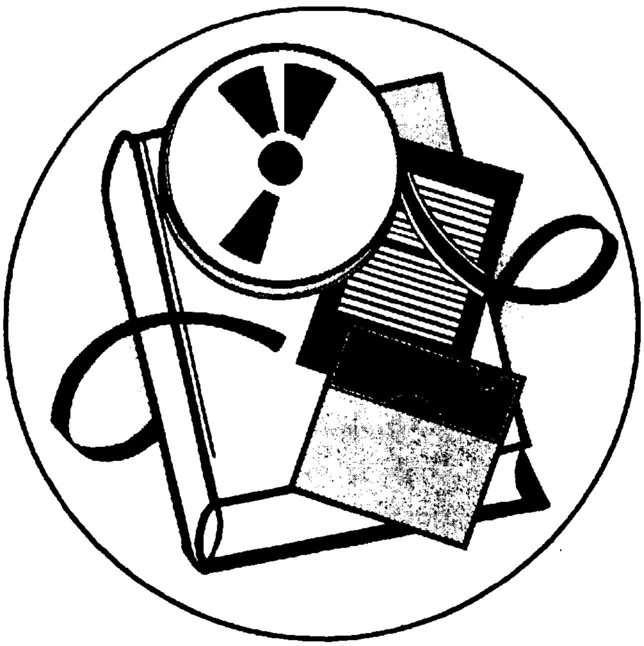
AUTHOR	Weber, Hartmut; Dorr, Marianne
TITLE	Digitization as a Method of Preservation? Final Report of a Working Group of the Deutsche Forschungsgemeinschaft (German Research Association).
INSTITUTION	Commission on Preservation and Access, Washington, DC.
ISBN	ISBN-1-887334-55-6
PUB DATE	1997-10-00
NOTE	33p.; Cover title: Digitization as a Means of Preservation? Translated from German to English by Andrew Medlicott.
AVAILABLE FROM	The Commission on Preservation and Access, A Program of the Council on Library and Information Resources, 1400 16th Street, NW, Suite 740, Washington, DC 30036-2217.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Access to Information; Archives; Foreign Countries; Information Management; *Information Storage; Information Technology; *Microfilm; *Microreproduction; Nonprint Media; *Preservation; *Reference Materials
IDENTIFIERS	*Digital Technology; *Germany

ABSTRACT

The German Research Association (DFG) is actively involved in preservation of research materials; it takes the view that in preservation, the enormous potential of digitization for access should be combined with the stability of microfilm for long-term storage. A working group was convened to investigate the technical state of digitization of microfilm and the changing compatibilities of microforms and digital conversion forms. This report documents the state of development and offers recommendations to serve as technical and organizational guidelines for filming and conversion projects (particularly those in the public domain). Chapters in this report are: (1) Should Endangered Books and Archives Be Filmed or Digitized?; (2) Requirements for Film Quality and Film Organization with Reference to the Option of Film Digitization; (3) Recommendations for the Digitization of Microfilm; (4) Microfilm and Digital Storage Formats as Compatible Media; (5) Digitizing from the Original; and (6) Cooperation and Exchange of Information; and (7) Suggestions for Further Reading. (AEF)

The Commission on Preservation & Access

Digitization as a Means of Preservation?



October 1997

A private, nonprofit organization acting on behalf of the nation's libraries, archives, and universities to develop and encourage collaborative strategies for preserving and providing access to the accumulated human record.

The Commission on Preservation and Access, a program of the Council on Library and Information Resources, supports the efforts of libraries and archives to save endangered portions of their paper-based collections and to meet the new preservation challenges of the digital environment. Working with institutions around the world, the Commission disseminates knowledge of best preservation practices and promotes a coordinated approach to preservation activity.

The European Commission on Preservation and Access (ECPA) was formally constituted as a non-profit foundation in March 1994 to foster, develop, and support collaboration among libraries, archives, and allied organizations in Europe, in order to ensure the preservation of the published and documentary record in all formats and to provide enhanced access to the cultural and intellectual heritage.

Translation of 'Digitalisierung als Mittel der Bestandserhaltung? Abschlußbericht einer Arbeitsgruppe der Deutschen Forschungsgemeinschaft.' *Zeitschrift für Bibliothekswesen und Bibliographie*, 44. Jg., Heft 1, 1997, S. 53-76.

Published July 1997 by the European Commission on Preservation and Access, Amsterdam (ISBN 90-6984-190-8). Republished September 1997 by the Commission on Preservation and Access (a program of the Council on Library and Information Resources), Washington, D.C. (ISBN 1-887334-55-6).

Distributed in Europe by the ECPA Secretariat, European Commission on Preservation and Access, Trippenhuis, Kloveniersburgwal 29, P.O. Box 19121, 1000 GC AMSTERDAM, The Netherlands. Distributed in all other countries by the Commission on Preservation and Access, 1755 Massachusetts Ave., NW, Suite 500, Washington, D.C. 20036.

♾ The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences-Permanence of Paper for Printed Library Materials ANSI Z39.48-1992.

Digitization as a Method of Preservation?

Final report of a working group of the Deutsche Forschungsgemeinschaft (German Research Association)

Hartmut Weber

Landesarchivdirektion Baden-Württemberg, Stuttgart

Marianne Dörr

Bayerische Staatsbibliothek, München

Translated by
Andrew Medlicott

Commission on Preservation and Access, Washington, D.C.

European Commission on Preservation and Access, Amsterdam

October 1997

The European Commission on Preservation and Access gratefully acknowledges the financial support of the following institutions and organizations

Bayerische Staatsbibliothek

Bibliothèque nationale de France

British Library

Bundesarchiv

Die Deutsche Bibliothek, the National Library of Germany

Koninklijke Bibliotheek, the National Library of the Netherlands

Marianne and Marcus Wallenberg Foundation

Ministry of Education, Culture and Science of the Netherlands

National Archives of Denmark

National Archives of Finland

National Archives of Ireland

National Archives of Norway

National Library of Finland

National Library of Ireland

National Library of Norway

National Library of Poland

National Library of Scotland

National Library of Wales

National and University Library of Iceland

National and University Library of the Republic of Slovenia

Royal Library of Denmark

Royal Netherlands Academy of Arts and Sciences

Swedish Research and Development Project on Paper Preservation

Swiss Federal Archives

Contents

Preface

Digitization is no doubt the issue that most fascinates and haunts preservation managers in archives and libraries at the moment. The possibilities seem limitless, the advantages are obvious, and from all sides there is pressure to exploit the new medium for preservation purposes—sometimes to the extent that funds are earmarked for digitization that might previously have been allocated to microfilming or conservation.

Yet, for preservation managers digitization is, in a way, a wolf in sheep's clothing. How does one deal, from a preservation point of view, with a medium that is notoriously unstable, for which 10 years is long term? What is the point of relying on such technology, when we worry about saving paper materials that are slowly degrading over 100 or 200 years? In the midst of all the excitement about the potential of the new medium, it is not always easy to keep all the advantages and disadvantages firmly in mind.

The Deutsche Forschungsgemeinschaft (DFG, German Research Association) is actively involved in preservation of research materials. In allocating grant money to projects, it takes the view that in preservation, the enormous potential of digitization for access should be combined with the stability of microfilm for long-term storage. The DFG thus commissioned a study to investigate the relationship between the two methods and to establish how the two could be profitably combined. The result was a detailed report on the technical requirements and advantages of using microfilm as the basis for digitization, which showed how one can have the best of both worlds and achieve both optimal access and maximum preservation.

The report was made available in German in the fall of 1996 on the Internet and was published in January 1997. In July 1997, the European Commission on Preservation and Access (CPA) published an English translation of the study to make the results widely available in the non-German-speaking world. The U.S. Commission on Preservation and Access is pleased to republish the English version for distribution outside of Europe. The CPA and European Commission would like to thank the Deutsche Forschungsgemeinschaft for its cooperation and the authors, Dr. Hartmut Weber and Dr. Marianne Dörr, as well as the translator, Andrew Medlicott, for their work on the English version.

The European and U.S. Commissions on Preservation and Access hope this publication will contribute to the development of balanced strategies for microfilming and digitization.

Commission on Preservation and Access
Washington, D.C., October 1997

Introduction

Newspapers, books, manuscripts, and archives have for decades been filmed at public expense to protect them from the endogenous deterioration of paper, or from other causes of damage that threaten books and archival material, and to ensure the permanence of the information they contain. Researchers use duplicate microfilms, rather than fragile originals, for their work.

Because printed materials continue to deteriorate rapidly, a joint *Bund-Länder* (federal-state) working group in Germany has, in conjunction with a conference organized by the Ministers of Culture of the German states, recommended a further extension of filming. The hectic developments in network and data technology, with their constantly improving capacity for the transmission of document images, open the way to new forms of use. The victory parade of the Internet and the vista of virtual digital libraries, offering ubiquitous and swift access of consistently high quality to documents, must in the future be incorporated into the concept of any preservation program. With this in mind, the subcommittee of the Deutsche Forschungsgemeinschaft (German Research Association) responsible for questions of preservation suggested in the spring of 1995 the establishment of a working group to discuss questions of digitization, in particular the digitization of microfilm. The group was to consist of librarians, archivists, and technical experts currently working in the field, and would explore the demands of quality assurance, and the possibilities and limits of the new techniques.

The working group was convened in November 1995, with Dr. Hartmut Weber (Landesarchivdirektion [State Archives Administration] Baden-Württemberg, Stuttgart) as chair. The other members were: Professor Dr. Hans Bohrmann (Institut für Zeitungsforschung [Institute for newspaper research], Dortmund; Werner Clausnitzer (MS-Mikrofilm Optical Disc GmbH, Wuppertal); Dr. Marianne Dörr (Bavarian State Library, Munich); Dipl. Kfm. Martin Fock-Althaus (SRZ Satz-Rechen-Zentrum, Berlin); Dipl. Ing. Hartmut Haux (Zeutschel GmbH, Tübingen); Leo Otte (Classen-Papertronics KG-Convertronics, Essen); and Dr. Hartmut Storp (Dr. Storp Consulting, Ahrensburg).

The group concentrated on investigating the technical state of digitization of microfilm and the changing compatibilities of microforms and digital conversion forms. Filming and digitization tests were carried out with standardized test materials and the results evaluated. The group prescribed minimum standards for the printout quality of microforms (material, image quality, and filming organization) for problem-free digitization. It also set requirements for high-quality digitization, relying on the quality index for the reproduction quality of manuscripts, as this is used as a quality standard for microfilming. In addition to black-and-white film and bitonal digitization, the possibilities for digitizing color microfilm were considered. There were also discussions on the processing of microfilm and on the hardware and software needed for quality control and use of data. The vital questions of data security and migration in digitization projects were a central theme. Aspects of financial viability were taken into account at all points. From the findings of the working group, a strategy for the introduction of digitization into preservation projects could be derived: microfilm has continuing priority as a recording and storage medium because of its quality and stability over time. As a medium for document delivery, the digital form, with its advantages of swift and remote access, in a quality depending on the intended use, should be employed. Direct digitization can achieve a result of higher quality in only a few cases.

The following final report, *Digitization as a Means of Preservation?*, was compiled by the authors, with participation by Hartmut Haux and Martin Fock-Althaus and the support of all members of the working group. It was finalized in the summer of 1996. It documents the state of development and offers recommendations to serve as technical and organizational guidelines for filming and conversion projects (particularly those in the public domain). The working group is aware that the speedy development of technology in this area means that conclusions will not have long-term validity. However, the problems considered here cover the field of digitization in all its complexity, and can thus serve, where appropriate, as a model checklist for the preparation of projects.

1 Should endangered books and archives be filmed or digitized?

The reformatting of damaged or endangered books and archives is an effective and economic conservation measure. Moreover, in contrast to measures to preserve or restore originals, the transfer of information to age-resistant media can also serve the objective of wider and better access.

Image conversion of endangered archive or library material to other media, for protection or for the permanent replacement of the original medium threatened by deterioration, requires systems that produce, over very long periods of time and economically, the highest possible reproduction quality, availability, and access. Compared with other modern information media, microfilm has the advantage that the material undergoes no fundamental technical transformation and is thus "future-proofed." The analog-stored information is directly accessible, with relatively little effort, to the human eye. Increasing national and international compatibility of microfilming systems ensures acceptance across national borders. Microforms can be economically created, duplicated, and distributed. Microfilm systems can be combined with electronic data processing (EDP) access systems. But microfilm can also be efficiently digitized with microfilm scanners. This will become more economical as the reproduction quality and financial viability of digital access systems improve.

Microfilm is an analog and age-resistant storage medium whose accessibility can be maintained with relatively few resources over long periods of time. Moreover, it remains available at all times for further processing in digital systems. Thus, it has a place in the digital media world. As a high-quality, intermediate storage medium, microfilm offers new and attractive methods and levels of access to books and archive material, with the help of digital access systems.

For the reasons given, it is advisable to film endangered material before digitizing, rather than microfilming from the digital medium. There is a financial rationale for this even when the only concern is digitization of material for new levels of access and use. Because microfilm is a long-term storage medium, it can minimize heavy expenditures for data migration and the frequent technical and organizational measures needed to preserve readability in new systems environments of material available only in digital form. Over the long-term, this justifies the resources invested in the preparation and handling of microfilm.

When an original is to be digitized directly, it is important to remember that the advantages of digital storage and processing must not be gained at the cost of reproduction quality, low durability, or lack of compatibility or "future proofing" of the information medium or of the hardware. A program specifying the technical and organizational steps involved in periodic migration, which can be constantly refined, should be part of the system design. Here, too, microfilm as a medium has a part to play. In principle, it is possible to transfer digital image data to microfilm.

However, contrary to statements that sometimes appear in the professional literature, converting digitized data to microfilm, which can then be used as an anolog long-term storage medium, involves a notable reduction in quality. Microfilm produced in this way cannot now be used for digitization with any guarantee of an acceptable result. Analog and digital storage forms are thus not yet fully compatible.

2 Requirements for film quality and film organization with reference to the option of film digitization

2.1 Choice and quality of film

In both their preparation and execution, filming projects should take into account the possibility of later digitization. In terms of materials and technique, there are only a few requirements that go beyond the normal rules and requirements. However, there are additional points that should certainly be noted in the area of film organization.

For the filming of high-contrast material, such as text, line drawings, and engravings, the current range of pan-chromatic antihalation undercoated (AHU) microfilm on polyester base can be used. In recent years, this has also been improved with a view to optimizing digitization. On the other hand, filming with a half-tone microfilm such as Kodak 2468 or 3468 will be best for material with a greater range of gray tones (continuous tones). This would include books containing photographic reproductions or color material that is to be filmed in black and white. Half-tone microfilm produces a film of positive polarity. A corresponding improvement in the continous-tone quality is achieved by putting AHU microfilm through a special developing process.

As a rule, reproduction quality and, especially, resolution capacity of microfilm systems far exceed those of image digitization, but filming nonetheless requires attention to the correct lighting and exposure, as well as to optimal readability (optical resolution). These depend on the optical characteristics of the camera and correct adjustment of the camera system. It is important to ensure the best possible quality of the master film, taking into account the deterioration between the master and the duplicate (second generation) film, which is produced for working purposes. The guideline here is the Quality Index (QI) 8 (higher quality) in Annex C of the international standard ISO 6199. Oriented to the height of the small "e" in printed material (corresponding in manuscripts to double the width of letters such as e, l, g, and f), we arrive at the formula QI = a × h, where *a* is the resolution number of the ISO resolution test pattern No 2 in line pairs per millimeter (lp/mm) and *h* the height of the small letter "e". Microfilm systems that give a value of 120 lp/mm and higher in the middle and at the edge of the image generally meet this standard.

Regarding the kind of microform, 35mm roll film for the master and as the starting point for digitization is best. Its image size guarantees sufficient quality, even with problematic material, up to a size of 60 × 80 cm. The normal commercial film lengths of 65m or 30.5m are recommended; the longer film is easier to handle because of the shorter preparation time. Far more successful results are obtained from digitization of negative than of positive film. A duplicate film of the lowest possible generation should be used for digitization. As the preservation master exists for preservation purposes and cannot be used directly, a silver halogen duplicate, produced from the preservation master with negative polarity with the help of a same-polarity duplicating film (DDP—Direct Duplicating Print Film), should be digitized. In principle, however, it is also possible to digitize a diazo copy. Filming with the use of blips is always necessary for an efficient working method with microfilm scanners.

It is also possible to digitize microfiche. However, the smaller image field results in a lower reproduction quality in large-scale work. Digitization of microfiche requires many more staff resources and makes more demands on the software, which increases the time and thus the costs compared with work on roll film, which can be made largely automatic. Conversely, when selected extracts, as opposed to a single continuous run, are required, it may be more economical to use microfiche.

A flaw-free film is a prerequisite for the best possible results in digitization. Density of the film, resolution, and background shadow should at least meet the ISO standard. Distortions should be avoided, as they can no more be corrected in digitization than can lack of clarity or other shortcomings in the master. Shadows in the book fold should also be avoided, as they can be corrected only to a limited extent, and with additional resources.

2.2 Filming technique

Every adjustment of the microfilm scanner requires additional resources and incurs greater costs. Therefore, the material to be filmed should be presented as uniformly as possible. These guidelines deal with the following:

a. *Reduction ratio*. Ideally, one reduction ratio should be selected for a complete filming project. If this is not possible, no more than one reduction ratio should be used for any single film. If need be, the material to be filmed should be arranged by size. In digitization, the image is scaled up to the size of the original. In most graphic formats, the image header can include details of the selected resolution and of the total number of pixels. If necessary, these can be used in the viewer software for reconstruction and indication of the original size.

b. *Positioning of the material*. The material should be placed on the filming table in a uniform way. This positioning must not be altered within a film. The optimal positioning is to place the material in the middle of the front edge of the filming table. If this is not possible, the material should be placed in the middle of the table, with pencil or other markings to indicate the correct position.

c. *Alignment of the material*. The alignment of the material should correspond to the desired appearance on the screen, and should thus be readable, i.e., horizontal. Otherwise, the digitizing service will have to rotate the material, which will add to the cost. Generally, books and documents should be filmed in half steps, in accordance with the image mode 2A of ISO 6199. Larger volumes and newspapers should be filmed in full steps, in accordance with the image mode 2B. Changes of the image position and of the film steps within one film should be avoided.

If only one page is to be shown on screen, this must be taken into account during filming. Later splicing of a filmed double page in digitization leads to further costs, as this function is not normally included in current digitization software, and needs to be done subsequently, as a manual task.

d. *Contrast between background and filmed material.* The contrast between the background and the material to be filmed should be increased by making the background uniform and dark.

Attention to the above points (b), (c), and (d) is a prerequisite for a largely automatic, and thereby economical, detachment of the background material from the whole digitized image. Elimination of the peripheral zones not only contributes to the optical image, but also reduces the amount of data to be stored.

2.3 Organization and documentation of filming

As with preservation filming, every film should start with an introductory sequence. This should clearly identify the film, including its unique number, relevant information about ownership, content, filming technique (reduction ratio and scale), and a test frame with information about readability and continuous tone reproduction according to DIN or ISO. It may also be appropriate to discuss with the digitizing firm the question of identifying the film in a way that is machine-controllable and facilitates the delivery of individual films or parts of films.

In ordinary film projects, certain elements of the organization of filming are often—wrongly—ignored. They are more important in the case of film digitization. They include take-counters, subdivision of films by indication sheets, placement of blips, and documentation on the filming procedure. Structuring a film by legible indications on the filmed material, running take numbers, blips, and appropriate indications on a take frame (see Figure 1), together with a consistent documentation of this structuring, makes indexing for retrieval and further processing much easier and reduces costs. Besides the single blip, which in conjunction with a take-counter usually suffices to identify individual frames, it is also possible to use group or sequence blips. This is particularly important in relation to data organization, i.e., accuracy of access and avoidance of superfluous page turning on the screen. Its high value justifies the extra effort required during the stages of preparation and filming.

The resources committed to structuring the information depend on the nature of the filmed material and how it will be used. It makes no sense to put a 300-page book on the screen with no markers to facilitate retrieval. In all cases, material to be filmed requires more extensive indexing that can also improve access to the original microfilm. The time required for this, and, therefore, the implications for personnel and costs, must not be underestimated.

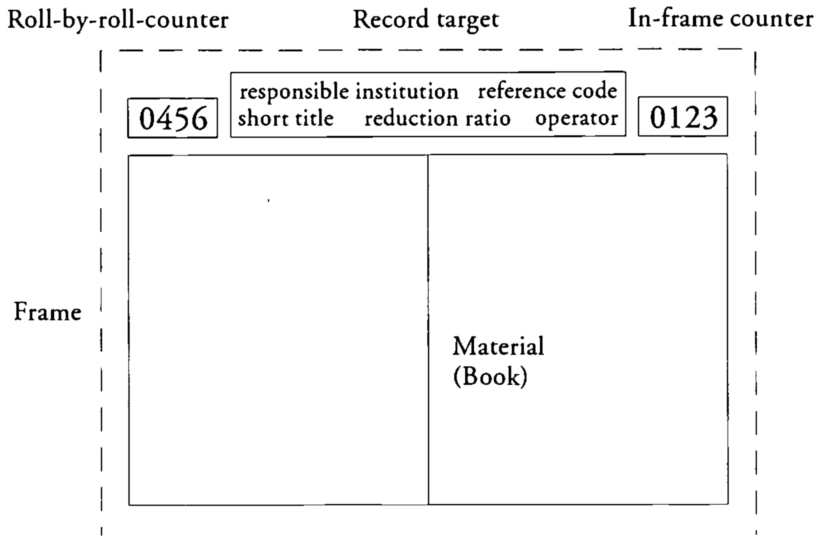


Figure 1. Placement of source materials for microfilming

2.4 Suggestions on choice of system

Reproduction quality depends essentially on the installation of suitable filming equipment. The requirements that have been described are met by most modern planetary cameras, which guarantee resolution of at least 120 lp/mm over the entire screen. The equipment should also include the following: automatic focusing, lighting that adjusts automatically to the material being filmed, a camera head that can turn, adjustable lamps (for lighting book folds), image field projection, adjustable image masking, automatic lighting of blips, and take counters. For filming books and archival material, the camera should produce optimal results with reduction ratios of between 8 and 24. For conservation reasons, the planetary camera should also have a device for protecting bindings and book covers, such as a two-part book-cradle with a sufficiently open glass plate with adjustable pressure. It should be possible to film heavier and oversize volumes without damaging them.

Since second-generation films (duplicating masters) are normally used in digitization, the film should be silver-halogenide duplicate film of the same polarity (DDP film). Duplicating should be undertaken with high-quality duplicating equipment (working under vacuum on parallel-running films) to minimize the loss of resolution.

In general, for filming with a view to subsequent digitization, the choice of system and the procedures usually will be dictated by the same criteria as apply in the case of good-quality microfilming. However, more attention must be paid to making the film form as unified as possible, and to the organization of the filming, the structure of the film, and its documentation.

2.5 Digitization of existing films that, for one reason or another, do not meet these requirements

It is possible to digitize existing films and film copies. In such cases, it is essential to work with films of the lowest possible generation. It is advisable in every case to thoroughly analyze the films in terms of material, state of preservation, reduction factor, reproduction quality, filming technique, nature of the material, and organization of the filming. This analysis is best undertaken in cooperation with an experienced service provider. Before a contract is awarded, digitization tests should be carried out with standard test material. It is only on such a basis that a firm can arrive at a realistic price, which will include the possibility of improvement through treatment of individual parts of the film and image enhancement. The intended use, in the context of cost, will determine agreement on the quality standard required. Any damage to the film, such as scratches, dirt, or fraying, will also influence the quality of digitization.

2.6 Differing recommendations for color microfilm

The starting point for digitization of color film should be a high-resolution, permanent-color bleach-fixing-process microfilm on a polyester film base. This should yield a high-resolution, reproduction quality microfilm that matches the quality of black-and-white microfilm.

In the past, duplicates of color microfilm have not proved entirely satisfactory. Exceptionally, therefore, and applying all the measures of film conservation, preservation masters are digitized. It is, for that reason, an advantage to be able, as is possible with some cameras, to produce two preservation masters in the same working run.

Because of its cost, the practice until now, almost without exception, has been to digitize color film using a proprietary system developed for the amateur market. The cheapest version limits the area of the image that can be digitized to 24 × 36 mm. However, a "full-step" color microfilm image occupies 32 × 45 mm. Film made using the maximum size of the full-step, which is best for reproduction quality and for further processing when dealing with larger or more difficult material, is not possible with a normal photographic CD. With half-step filming and smaller image areas, it is necessary to establish in advance whether the picture format can be carried by the system, as reels can only be wound in one direction. Transfer of uncut microfilm is certainly possible. However, as spool devices are not part of the film scanner, the film can be damaged. Under this system, the film material is digitized with differing resolutions and transferred in compressed form onto photographic CD. The lowest resolution of the five resolution steps is 128 lines × 192 pixels, the highest 2048 lines × 3072 pixels.

The photographic CD system was developed above all for the large amateur photographic market and is therefore widely and cheaply accessible for the digitization of color film. It has only limited use, however, in producing color microfilm, particularly with regard to format. Failure to use the full-step format of unperforated 35 mm microfilm normally leads to loss of quality, especially in the case of color microfilm. But loss of quality does not have to be accepted. Color film scanners have been introduced in the reprographic field that can work with film up to a format of 6 × 9 cm and with filmstrips. They are able to digitize full-step 35 mm color microfilm and have a resolution of up to 2000 dots per inch (dpi). The output format is not limited to a photographic CD; it may be produced in any of several other formats. However, the current state of technology and the comparatively low demand for digitization of color microfilm make this a fairly expensive process. Still, in view of developments we can expect in the future, it would be a mistake to sacrifice reproduction quality and standard compatibility or tested systems of working for the sake of a currently useable, producer-independent system, even if this were economically advantageous.[1]

[1] In the meantime, a digital camera has become available that is able to digitize also from color microfilm full step with a resolution of up to 3800 x 4600 pixels.

3 Recommendations for the digitization of microfilm

3.1 Picture quality

Where good-quality microfilm is available as a long-term storage medium, the reproduction quality of the digital conversion form will be determined by its intended purpose. In other words, as a general rule, digitization of microfilm should not aim at the best possible result in the way that is mandatory for direct digitization of endangered original material.

Bitonal digitization on pan-chromatic AHU microfilm is adequate for the reproduction of printed text, including line drawings, and for modern non-impact typescript (plastic carbon band, and inkjet and laser printers). Gray scale must be used to digitize manuscripts, pencil and crayon drawings, typescript produced with a silk ribbon, color illustrations and drawings, other material with varying shades of gray, and black-and-white and color photographs. Sixteen gray scale (4 bit) is usually adequate for digitizing contrast-enhancing AHU film. For digitization from halftone film, 256 gray scale (8 bit) should be used. Digitization with gray scale requires far more storage, which has serious implications for cost at all stages of the process. It should thus be undertaken only where such reproduction quality is indispensable.

In digitizing from film, the necessary resolution is determined by the size of the smallest element that is to be clearly discernible. With printed texts, this is the height of the small "e"; with manuscripts it is the doubled letter width described in paragraph 2.1. In applying the appropriate formulas of the quality index, resolution requirements are determined in relation to the size of these elements. For bitonal digitization, the quality index is calculated according to the following formula: QI = (0.039h)/3, where *a* is the resolution in dpi and *h* the height of the small "e" in millimeters. For digitization with gray scale, the formula is: QI = (a × 0.039)/2.

With bitonal digitization, a resolution of 615 dpi (for 256 gray scale 410 dpi) is necessary to reproduce the small "e" at a height of 1 mm at higher quality. Medium quality is achieved with 385 dpi (256 gray scale 256 dpi). Lower quality results from 277 dpi (256 gray scale 185 dpi).

Given the high quality of the microfilm, it will be sufficient for most purposes to aim for a digital copy of medium quality. The required resolution can then be calculated on the basis of the quality index QI = 5 for medium quality as follows: resolution in dpi a = 3 × 5/0.039h, where *h* is the height of the small "e". Where the height of the small "e" is 1 mm, this gives a value of 384. For digitization with gray scale, the formula is a = 2 × 5/0.039h, which, for an "e" of the same height, gives a value of 256. Letters of this size (about 7 pt) are often used in footnotes.

As an indication, the aim should be 350-400 dpi for bitonal digitization, 250-300 for gray scale. Test runs with typical films should be used to decide the quality required for each purpose.

3.2 Storage form

Transfer of the digitized image data should be by digital audio tapes (DAT) or CD-R (recordable). Readability independent of hardware is guaranteed for both media through standardization (DIN 66211 for DAT, ISO 9660 for CD-R). The current storage capacity of 650 Mb per CD-R and 2 Gb per DAT tape will increase in the near future.

In practice, CD-R offers advantages for data security, since the reliability of DAT depends on the tension of the tape, which can undergo changes during transport. CD-R also allows images to be viewed on a screen immediately, without first having to be stored on the hard disk. This can be helpful for ensuring quality control.

It is important to reach a binding agreement with the company undertaking the digitization that it will store the transferred material for at least as long as it takes for the customer to check and secure the data.

The digital conversion form is reliably secured when loss-free compressed or uncompressed image data have been secured on at least two data carriers, and it has been verified that their contents are identical and readable with no difficulty. In the simplest case, the two data carriers with the same content, the "primary data carrier" and the "working duplicate," will be created by repeated successive transfer of the image data.

To ensure readability of the primary data carrier, multiple working duplicates should be produced from it. Performing a decompression test for every stored digital copy further enhances data security (see paragraph 5.3).

3.3 Format, compression

The image data should be supplied the right way up (readable without being turned) in a continuous format, suitable for the largest possible number of applications. The Tagged Image File Format (TIFF) has established itself widely as a model format for image data. The advantage of this format—in contrast, for example, to Windows-Bitmap—derives from the fact that it is largely platform-independent. TIFF files can be read and further processed on differing equipment with differing systems and programs. It should, however, be noted that, despite thoroughgoing standardization, the TIFF format allows variations that may not be compatible with the installed software. Here, too, careful discussion and, possibly, experimental runs with test data are recommended. TIFF provides for uncompressed and compressed data supply. TIFF G 4 is available for compression without loss of black-and-white material. If loss-free compression is possible, it should be used for data delivery to save storage space. However, since not all programs can work with compressed TIFF data, the compatibility of the application must be established in advance. In any case of doubt, uncompressed supply is to be recommended. The Joint Photographic Experts Group (JPEG) format, which is frequently used for the transfer of half-tone and color pictures, has variable compression ratios that are all lossy and thus not to be recommended.

Because image data can be organized in different ways, it is advisable to agree with the service provider on the organization of the material appropriate to each application. As a rule, each picture will be stored in a separate file. Gathering related pictures in one file (multiple TIFF) is possible only with documents that consist of no more than a few pages.

For additional use of the data on the Internet, it is advisable to convert data into platform-independent formats that allow inclusion of the widest variety of documents. Such conversions are part of the service offered today by most of the specialist companies. Where appropriate, this format should be added to the contract.

3.4 Software requirements for image viewing

For access to digitized images, various programs for viewing and manipulation are available for PC and UNIX environments. These include "Viewer" software, obtainable as public-domain software or shareware programs. It is recommended to install at each institution only one specific, standardized software, whose compatibility with the supply of digitized conversion formats can be rigorously tested in advance.

As a rule, viewer software should have the following features: page-turning forward and backward; use of the whole screen for display; magnification of the whole image and of selected parts of the image; reduction of the whole image; option of return to the original image; image rotation; image inversion; and display of technical information from the headers, such as picture size, resolution, format, bit depth, and print. It is also very useful to have the option of image conversion into other formats and of image compression.

For instance, in the UNIX world xv is available as shareware. Depending on the installed hardware, appropriate viewers are contained in the supply range of the operating systems (e.g., HP-UX imageview). For PCs, Imaging for Windows is a feature available at no extra charge with Windows 95. Other examples of suitable software are PixView 2.1 from Pixel Translation, ScanMos UVP from MS Electronic Service, or, with limits, Hijaak Pro 2.0 from North American Software.

Software for the control and display of digitized images and for rapid access should be chosen with a view to its specific applications. The requirements we have outlined serve as performance criteria for the viewer components of this application software.

3.5 Hardware requirements for image viewing

Hardware installation that meets requirements for inspection and use of digitized images must be provided at each institution. The relatively large quantities of data contained in digitized images as compared to text files leads to heavier demands on the data bus and RAM, if the picture recovery time is to remain within acceptable limits. The minimum requirements are met by PC systems based on processors of type 468 with 66 MHz or Pentium, with Windows 3.11 or higher, 16 Mb RAM and a hard disk in the gigabyte range.

In the context of ergonomic design of the work station, particular importance attaches to size of screen (at least 17 inches diagonally), speed, the graphic card, and the appropriate drive. Normal PC screens with 14 inches are unsuitable for image representation, quite apart from the question of resolution. The resolution capacity of normal PC color screens is about 75 dpi, so the image resolution has to be reduced for producing it on the screen. Large screens manufactured specially for image work can reach higher resolutions, up to 120 dpi. In principle, the digital conversion form offers a higher resolution, but this becomes apparent only with magnification of selected parts of the screen (zooming).

3.6 Long-term preservation of the digital conversion form (migration)

Even where a high-quality microform is available alongside the digital conversion form, and thus allows, if necessary, for repeated digitization, the converted format must be preserved in the long term. If only on financial grounds, repeated digitization is out of the question. Given the increasing importance of electronic information systems in research and teaching, the digitized images should be useable in the future for many possible applications. The complete data should therefore be preserved for the long term retaining as much of the information as possible, i.e., with loss-free compression or uncompressed, in a format that allows every conceivable use. Storage of data that have been compressed and formatted only for one specific application is not sufficient.

The loss-free compressed or uncompressed image data must therefore be migrated to new systems in a TIFF format or in a platform-independent TIFF consequential format. This adaptation must follow a planned concept, in line with technical progress, and must not omit any development steps. The regular adaptation must take into account not only the expected durability of the storage medium, but also the currency of the format and the availability of the hardware and software needed for reading. The rapid succession of innovations in hardware and software, which seldom respect standardization efforts (scarce in this area anyway), can produce problems of compatibility. Migration must be carried out with extreme care. The results must be checked image by image, as the loss of one bit in a graphic file can result in serious loss of data, even up to a whole image. Responsible migration calls for organizational and technical measures to be undertaken before systems are replaced. The object of migration is to hold the data in at least two long-lived storage mediums, secure against interference, in a platform-independent format that is compatible with the EDP system being used. Thus, the complete contents of the transferred image data can be checked against the data source of the earlier generation, as long as the EDP system that produced it remains available.

3.7 Financial viability

Generally, the digitization of microforms should be done by a service bureau. The costs of digitizing a uniformly produced 35 mm microfilm according to the foregoing recommendations depend essentially on the size of the task, the mode (bitonal or gray scale), and the resolution, but also on the quality of the film and the type and readability of the filmed material. Since digitization costs are also dependent on the market situation, it is not possible to give any general indication of prices that will have long-term validity.

The cost factors we have mentioned take account only of digitization itself. Experience has shown that further costs are incurred by manual turning, splicing images out of the general frame, and marking. Programming costs and the initial cost of programming the film scanner according to the customer's requirements must also be considered. Finally, there are the costs of downloading the data, operating the CD-R, the carrier medium, and packing and transport. In cases where individual work and image enhancement with special software are necessary to improve quality, such costs must also be included.

The choice between digitization with a general raising of the resolution on the one hand and with gray scale on the other has an indirect bearing on the cost of the conversion. Higher densities of data mean higher costs in data supply, storage, and handling. The consequential costs of any planned migration must also be taken into account. In cases of need, it may be more economical to digitize a second time from the microfilm rather than to constantly migrate the data.

3.8 Digitization and optical character recognition

Optical character recognition (OCR) is a machine process that turns visible alpha-numeric signs into coded data (codes corresponding to the alphanumeric signs and their context), according to a more or less standard pattern of recognition. There is here a fundamental difference between fully automatic text recognition and trainable recognition that supports pattern recognition with dictionaries, linguistic methods, and features of "artificial intelligence." The text recognition programs increasingly integrate dictionaries and substitution lists that are adjustable according to the degrees of certainty. To prevent the substitution of inaccurate characters that were wrongly recognized as accurate, systems work with fuzzy logic and probabilities. Some systems include an interesting further feature known as "mixed mode." Signs or groups of signs that are either not recognized, or not recognized with certainty, are retained as images and remain in that uncoded form, in position in the remaining—correctly recognized—text.

In addition to reliable text recognition, page segmenting is an essential performance feature of text recognition systems—that is, the interpretation of contextual information such as columns, blocks of text, and graphics. Further features are deskew, segmenting of individual units, and recognition of types of handwriting and signatures or of more than one language in the same document.

The economical cut-off point for machine text recognition is at 99.95%. In other words, if there are more than 4 or 5 mistakes per 1,000 units, processing by hand is more economical.

Reliability of text recognition depends essentially on the background, the kind and size of the writing, and the contrast between text and background. Disruption of text recognition occurs when there is dirt on the material and omissions from the image information caused by incomplete or irregularly printed letters. Reliability also depends on the density of the image information. The greater the amount of image information being processed, the higher the recognition rate. Higher resolution in digitizing can therefore improve the recognition rate, as with digitizing in gray scale.

In principle, the quality criteria we have mentioned also apply to microfilm. The correct standard background density and minimal ground shade are important to achieve high resolution and adequate contrast. Digitizing negative film avoids the disruption caused by dirt and scratches. In practice, there has not yet been enough experience with machine text recognition in conjunction with microfilm to allow the formulation of reliable views.

4 Microfilm and digital storage formats as compatible media

4.1 Tests on compatibility and reproduction quality

The working group conducted an experiment to reach conclusions about the compatibility between digital and analog conversion forms. We used for this purpose a test chart in Format DIN A2 with standard test indications for resolution (reproduction sharpness), reproduction of gray scale, and color reproduction. Further, samples of different printed text and handwriting, together with black-and-white and color photographs, were mounted on the test chart. The chart was filmed in black and white and in color on different microfilm material (35mm) and directly digitized, using different scanners, bitonally, in gray scale, and in color. Working paper copies were made of the film and of the digitized conversion formats. The results of the test chart were digitized with different film scanners. As a further step, transfer was made from film to microfiche. Finally, the digitized image data from the test chart was exposed by COM systems on microfilm. The analog test patterns were then evaluated microscopically, according to the appropriate standards, and the digital formats were evaluated with assistance of a high resolution color monitor and the zoom function with image-reading software.

The experiment (Figure 2) revealed that digital systems do not yet achieve the high resolution of microfilm (Test group 12.5, 11, or 8). The higher score of 8 on the Quality Index is achieved only by microfilm. Medium quality was achieved by digitization from the original in gray scale. Transfers from microfilm, and in part also from the original, onto microfiche and into digital conversion forms retain readability. Working copies on paper (printouts from reader-printers and laser printers) are at a similar level. A four-color print from a digitized original is the only example to reach the test indication group 4.5. In this case, digitization of the test chart and its color microfilm copies was done with a drum scanner. The result showed what is technically feasible if cost is not a consideration.

The differences between the half-tone reproductions were so obvious that a subjective judgement was all that was needed. Because of its wide exposure scope, the microfilm was able to reproduce all elements present in the test chart to an appropriate level of quality. The digital conversion required more exposures with different parameters in order to present these elements correctly in different images. Thus, film digitizing produced better results, since the contrast of the material in microfilming was already somewhat enhanced and generally evened out.

These are the results of an experiment with a relatively large test chart, corresponding roughly to the size of a sheet of newspaper or an open large folio book. Better results can be expected from material of smaller size, assuming that the relationships remain the same.

It is essential to stress that a lower quality must be expected from digitization of film, as compared with the high-value digitization from the original with gray scale. On the other hand, the quality of the film

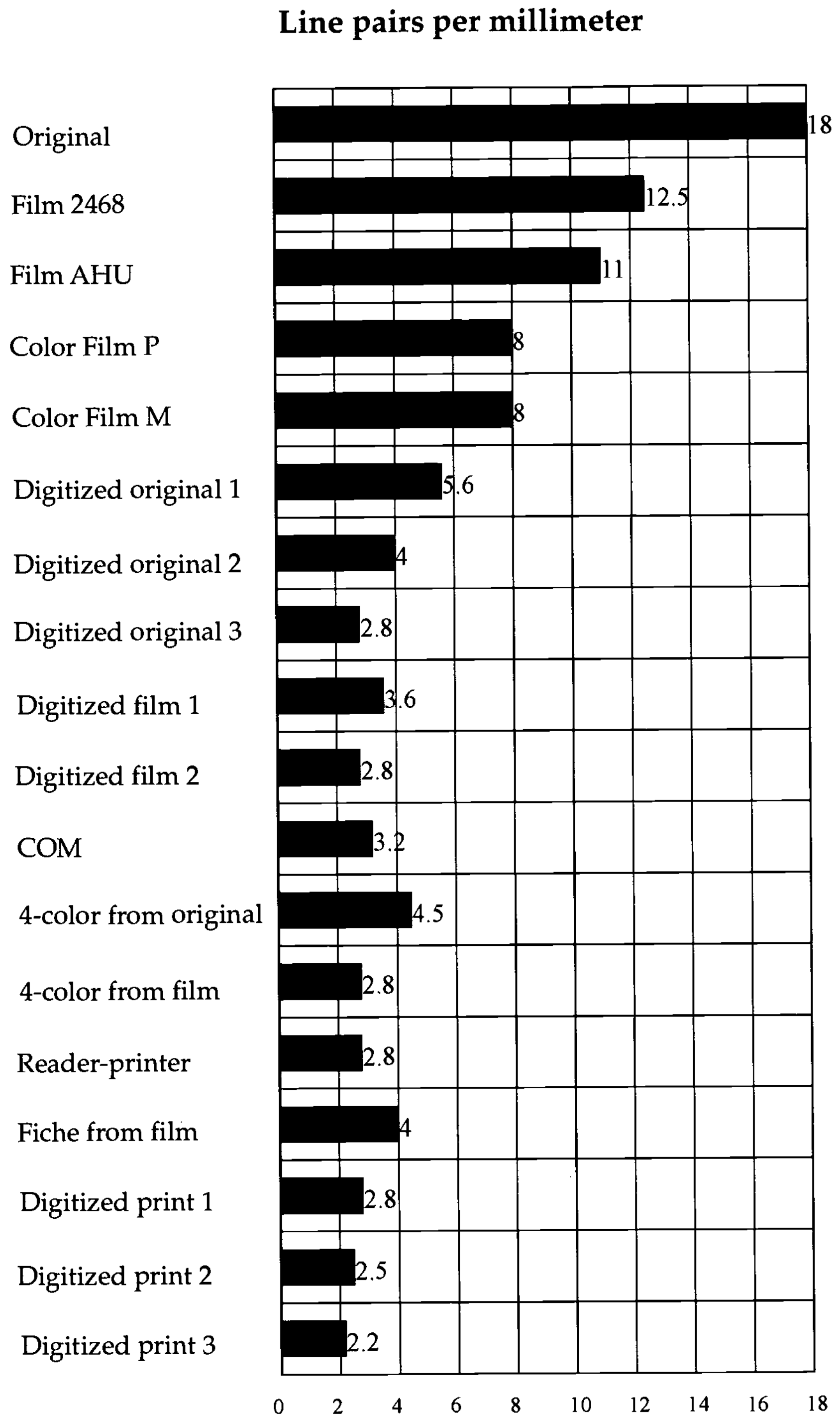


Figure 2. Sharpness of reproduction

used in this experiment was so superior that it would be adequate if the digitizing systems were to be further developed and digitizing with higher resolution were to be financially more viable. In the case of less resource-intensive bitonal digitization, the results of digitizing from film and from the original were similar.

4.2 Printing out from digitized storage onto film (image)

It proved difficult to find a company able to print out onto microfilm a TIFF data file of the digitized test surface. Printout of image data, in particular of material in the DIN A2 format, does not yet seem to be among the normal services of Computer Output Microfilm (COM) companies. We finally succeeded, thanks to one firm that had installed special equipment for work in gray scale and reproduction of detail for work on industrial technical drawings and black-and-white photographs.

The test material could be reduced 72-fold in full, and 36-fold and 18-fold only in part. Since the pixel number of the equipment was about 3200 × 2600, the smallest reduction (18-fold), which could only reproduce parts of the image, gave a resolution value of 3.2 on the ISO test scale, which more or less guaranteed readability.

A further exposure was done by a foreign firm, which managed to print out the whole of the test surface onto microfilm (35 mm). However, the reproduction quality, with a resolution level of 1.8, was very low, partly because a film with low-resolution capacity had been used instead of microfilm.

All that was available for this experiment was a digitized image data file in bitonal mode. Also, the detailed material proved too large in DIN A2 for current technical capacities. Both points mean that further experiments are necessary before any general verdict can be reached on the quality of the printout of digital image data onto microfilm, as well as on its availability and financial viability. The obvious prerequisites for digitizing for an optimal COM printout of image data must be included in these experiments. There is as yet no way we can speak of compatibility of analog and digital media in the direction digital-to-analog. Full compatibility would be achieved if the directly filmed microfilm and the film produced through the digital interim carrier were of comparable reproduction quality and if the image data produced by digitizing the microfilm were identical with that used to produce the film via COM (see paragraph 3.1). Current experiences with optical systems shed doubt on whether such 'film-to-film' compatibility can ever be achieved.[2]

4.3 Should filming or digitizing come first?

Printout onto microfilm of image data of archive and library material is obviously not yet generally available. The level of reproduction quality that can currently be achieved does not allow for further digitization through COM microfilm. For both these reasons, we cannot yet recommend digitizing from the original, followed by digitization of a film as an age-resistant storage medium. Again, we would advise producing a microfilm that meets the requirements we have described and then digitizing from that according to the intended use.

[2] Perhaps a new technique using electron beam exposure at resolutions up to 600 dpi (cf. URL: http://www.igraph.com/micro.htm) will produce better results in the future. The working group was not able to test this method in 1995/96.

5 Digitizing from the original

5.1 Quality requirements

In the current state of technology, digitizing from the original gives a better reproduction quality for color material and material with weak contrasts than digitizing from film. When endangered original material is digitized, the converted form acquires the status of a preservation master which, in an extreme case, will have to serve as a substitute for the lost original. In this case, of course, the reproduction quality must be higher than is necessary in cases where the digitized secondary form exists only to improve access. A later, repeated digitization of the endangered original, even if possible, is not consistent with the aim of preservation. This means that the first digitization must be of the highest possible standard.

It follows that, in applying the quality index (see paragraph 3.1), the highest quality (QI = 8) must be guaranteed. To reproduce the small "e" with a height of 1 mm at higher quality, bitonal digitization by that formula requires a resolution of 615 dpi (410 dpi for 256 gray scale).

A resolution of at least 600 dpi is recommended for bitonal digitization of printed text that includes line drawings. A resolution of 400 dpi is generally adequate for bitonal digitization with texts that are clear, larger, and, in particular, evenly spaced (10 point and above), and that have been produced by modern, non-impact typewriters, such as plastic carbon band, or by ink-jet or laser printer. Two-hundred-fifty-six gray scale and a resolution of 400 dpi should be used for the following: manuscripts, drawings with pencil or crayon, typescript with silk ribbons, color illustrations and other drawings with varying gray shades, and black-and-white and color photographs. These recommendations also correspond to American quality requirements for digitizing original material.

The suggestions on filming technique in paragraph 2.2 and on film organization and documentation in paragraph 2.3 can contribute usefully to digitization and to the further processing of the digitized conversion form.

5.2 Criteria for the choice of system

Scanners that work like a planetary camera, digitizing the material from above, must always be used for sewn and bound volumes. Feeder scanners and flatbed scanners are not suitable for books and archives. It is especially important to follow the precautions described in paragraph 2.4 for the protection of books and volumes. Equipment of this kind is indispensable for the digitization of unique material that is fragile.

5.3 Storage format

The comments in paragraph 3.2 are applicable here. If long-term storage of perhaps damaged original material is to be exclusively in digital form, and if, consequently, the digital data carrier deteriorates and there is no microform to fall back on, additional quality tests are necessary for the storage of digitized image data on optical disk. The following procedure is suggested:

First, digitized copies of the material are written to optical storage disks (the primary data holder). The data on the server's internal magnetic disk are not deleted but kept unaltered. After the image data have been stored as pages in TIFF data files in the primary data carrier, they are read back and a few of them are decompressed. The uncompressed or decompressed digital copy has a precisely defined number of image points, which can be calculated with reference to the format of the original material and the resolution chosen for the scanning. This size of the decompressed digital image (in Kb) is the product of the image-point number and the "bit-depth" with which each image-point is represented. A digital copy is thus correctly reproduced when its actual size equals the original value. This makes clear that the transferred copies have been securely stored in their correctly reproducible form. In the extremely rare cases where a digital copy cannot be perfectly reproduced in this test, the logical step is to erase it in the optical data carrier and immediately store it again.

The primary data carrier, created and quality-checked in this way, is the source of copies for data preservation. These working duplicates are for day-to-day use, while the primary data carrier remains the preservation master. If need be, it serves for production of further duplicates. It is not absolutely necessary to subject the working duplicates to the same quality test as the primary data carrier. If, in the course of normal use, it becomes apparent that individual copies can not be reproduced correctly, it is always possible to produce another duplicate, or go back to the primary data carrier for a further working duplicate.

5.4 Format and compression

As for paragraph 3.3

5.5 Requirements for image viewing software

As for paragraph 3.4

5.6 Requirements for image viewing hardware

As for paragraph 3.5

5.7 Migration

Organizational and technical measures are always advisable in the migration of digital conversion forms, to safeguard the transferral of information and for reasons of economy. However, they become indispensable when the digital form is the only form in addition to the original, or when it is expected that it will eventually replace the original. Repeated digitization of the original should be avoided on grounds of preservation and because it would be prohibitively expensive.

The organizational and technical measures for the safe migration of digital conversion forms must be included from the outset in planning, which must take account of the necessary resources. The recommendations in paragraph 3.6 apply to the planning and carrying out of migration, especially the requirement continually to adapt the lossless compressed or, as necessary, uncompressed data to new system environments, and to safeguard adequately the data carrier that is created in each case.

5.8 Financial viability

Where books or archival documents are to be digitized as a whole, this should be done by commercial firms. Where only certain pages, or parts of a document, are to be digitized, this can be done by the institution itself. The cost of digitizing books and documents (page size up to A4) depends on the amount of material, the mode (bitonal or gray scale), and the resolution, but also on the contrast values of the material, its type, and the way in which it is arranged. Simple, flat work, such as single sheets, can be more efficiently digitized with flatbed or feeder scanners than books or other bound volumes, for which special book scanners need to be installed.

When working out the cost of digitization from the original, it is essential to include in the calculation the further cost of migration. In particular, it will almost invariably prove financially more advantageous, when working with threatened originals, first to make a film and then to digitize from that, thus solving the problem of migration. In exceptional cases, with difficult material, it can be advisable, in the interest of reproduction quality, to film and digitize in parallel from the original at the same time. Paragraph 3.7 is relevant on other points.

5.9 Differing recommendations on color images

With current technology, digitization of color can be done only at relatively low resolution values, or for limited quantities of material, because very large quantities of data are involved. Test runs should always be carried out to establish whether the reproduction quality is acceptable.

In the interest of economical storage and processing of image data, compression processes play an even larger role in color digitization than in bitonal or gray scale digitization. At present, there is no compression process that does not involve a worsening of reproduction quality, in particular the distortion of color values.

6 Cooperation and exchange of information

Digitization projects pose new technical and organizational tasks for libraries and archives. Each institution must develop the expertise to plan and carry out digitization projects. Securing competent advice from qualified and experienced service providers is thus strongly recommended. At the same time, institutions involved in questions of digitization should exchange information. This will contribute to adequate market evaluation of the potential of the service providers and to a judgement about the financial vialibility of what they are offering. Moreover, in addition to the exchange of experiences, there should be early contact with other institutions that are planning or have carried out similar projects, to remedy one's own practical shortcomings. At the least, for the time being, institutions supporting digitization projects should insist on full reporting and ensure that the reports reach the professional public. Finally, a grounding in digitization should be a part of all library and archival training and development.

7 Suggestions for further reading

Janet Gertz: *Oversize Color Images Project, 1994-1995. Final Report on Phase I.* Washington, DC: Commission on Preservation and Access, August 1995.

Jürgen Gulbins, Markus Seyfried, and Hans Strack-Zimmermann: *Elektronische Archivierungssysteme* (Electronic Archive Systems). Berlin, Heidelberg and others: Springer 1993.

Anne R. Kenney and Stephen Chapman: *Digital Imaging for Libraries and Archives.* Ithaca, N.Y.: Dept. of Preservation and Conservation, Cornell University Library, 1996.

Anne R. Kenney and Stephen Chapman: *Digital Resolution Requirements for Replacing Text-Based Material: Methods for Benchmarking Image Quality.* Washington, DC: Commission on Preservation and Access, April 1995.

Resolution as it Relates to Photographic and Electronic Imaging. Technical Report. Silver Spring, MD: Association for Information and Image Management, 1993.

Donald Waters and John Garrett: *Preserving Digital Information. Report of the Task Force on Archiving of Digital Information.* Washington, DC: Commission on Preservation and Access, May 1996.

Donald J. Waters: *From Microfilm to Digital Imagery. On the feasibility of a project to study the means, costs and benefits of converting large quantities of preserved library materials from microfilm to digital images.* A Report to the Commission on Preservation and Access. Washington, DC: Commission on Preservation and Access, June 1991.

DIN 66211, DIN ISO 6199, ISO 9660

Publications of the European Commission on Preservation and Access

European Register of Microform Masters. Supporting International Cooperation
Werner Schwartz
1995, 10 pp.

Mass Deacidification. An Update of Possibilities and Limitations
Henk J. Porck
1996, 54 pp. ISBN 90-6984-162-2

Perservation Challenges in a Changing Political Climate. A Report from Russia
Galina Kislovskaya
1996, 20 pp. ISBN 90-6984-167-3

Choosing to Preserve. Towards a cooperative strategy for long-term access to the intellectual heritage
Papers of the international conference organized by the European Commission on Preservation and Access and Die Deutsche Bibliothek, Leipzig/Frankfurt am Main, March 29-30, 1996. 1997, vii + 165 pp.
ISBN 90-6984-166-5

The Commission on Preservation and Access
A program of the Council on Library and Information Resources
1755 Massachusetts Avenue, NW, Suite 500
Washington, DC 20036-2188

NOTICE

REPRODUCTION BASIS

- [x] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

- [ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket")..