Digitisation as a Method of Preservation?*

Hartmut Weber
Marianne Dörr

Preface

Digitisation is no doubt the issue that more than anything else 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 media for preservation purposes – sometimes to the extent that funds are earmarked for digitisation that might previously have been allocated to microfilming or conservation.

Yet, for preservation managers digitisation is in a way a wolf in sheep's clothing. How to 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 are worrying about saving paper materials slowly degrading over 100 or 200 years? In the midst of all the excitement about the potential of the new media, it is not always easy to keep all the advantages and disadvantages firmly in mind.

The Deutsche Forchungsgemeinschaft (DFG, German Research Association) is actively involved in preservation of research materials. In allocating grant money to projects, they take the view that in preservation the enormous potential of digitisation for access should be combined with the stability of microfilm for long--term storage. The present study was commissioned to investigate

* *Digitisation as a Method of Preservation?*. Final report of a working group of the Deutsche Forschungsgemeinschaft (German Research Association), by Hartmut Weber, Landesarchivdirektion Baden-Württemberg, Stuttgart, and Marianne Dörr, Bayerische Staatsbibliotek, München. Published by the European Commis--sion on Preservation and Access (ECPA). Originally published in German in the *Zeitschrift für Biblioteksuesen und Bibliographie*, 44, Jg., Heft 1, 1997, S. 53-76.

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 digitisation, 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. The European Commission on Preservation and Access is happy to be able to publish an English translation of this study which will help to make the results widely available outside the German-speaking world. The Commission would like to thank the Deutsche Forschungsgemeinschaft for their 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 Commission on Preservation and Access hopes this publication will contribute to the development of balanced strategies for microfilming and digitisation. As always, comments are welcome and can be addressed to the Secretariat.

European Commission on Preservation and Access
Amsterdam, July 1997

Introduction

Newspapers, books, manuscripts and archives have for decades been filmed at public expense, in order to protect them from the endogenous deterioration of paper, or from other causes of damage which threaten books and archive material, and to ensure the permanence of the information they contain. Duplicate microfilms, rather than endangered originals, are produced to researchers. As deterioration progresses rapidly, a joint *Bund-Länder* (federal-state) working group in Germany has, in conjunction with a conference organised 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, opens 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 future be incorporated in the concept of any preservation programme. With this in mind, the sub-committee of the Deutsche Forschungsgemeinschaft (German Research Association) responsible for questions of preservation suggested in the spring of 1995 the setting up of a working group to discuss questions ot digitisation, in particular digitisation 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) in the 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); Dr Hartmut Storp (Dr Storp Consulting, Ahrensburg).

Investigation of the technical state of digitisation of microfilm and changing compatibilities of microforms and digital conversion forms were at the centre of the group's work. Filming and digitisation tests were carried out with standardised test materials and the results evaluated. The group prescribed minimum standards for the printout-quality of microforms (material, image quality and filming organisation) for problem-free digitisation. It also set requirements for high-quality digitisation, relying on the quality index for the reproduction quality of manuscripts, as this is used as a quality standard for microfilming. In addition to the widespread black and white film and bitonal digitisation, the possibilities of the digitisation of colour microfilm were also considered. In addition there were discussions on the processing of microfilm and on the hardware and software provision for quality control and use of data. The vital questions of data security and migration in digitisation 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 digitisation into preservation projects could be derived: microfilm has continuing priority as a recording and storage medium on grounds of quality and «future proofing». As a medium of production for required documents, the digital form, with its advantages of swift and remote access, in a quality depending on the intended use, should be used. Direct digitisation can achieve a result of higher quality in only a few cases. The following final report, under the title *Digitisation of endangered library and archival material* was compiled by the authors, with participation by Hartmut Haux and Martin Fock-Althaus, with the support of all members of the working group. It was finalised in the summer of 1996. It documents the state of development and offers recommendations to serve as technical and organisational 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 here considered cover the field of digitisation in all its complexity, and can thus serve where appropriate as a model cheeklist for the preparation of projects.

1 Should endangered books and archives be filmed or digitised?

Conversion of damaged or endangered books and archives is a effective and economic conservation measure. Moreover, in contrast to measures to preserve or restore originals, the transfer of information, the survival of which is in danger, to ageing-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 and/or for the permanent replacement of the original medium threatened by deterioration, requires systems which produce, over very long periods of time, the highest possible reproduction quality, availability and access, together with economy. Microfilm has the advantage, in comparison with other modern information media, that the material undergoes no fundamental technical transformation and is thus «future-proofed». The analogue-stored information is directly accessible, with relatively little effort, to the human eye. Increasing national and international compatibility of microfilming systems ensures acceptance across frontiers. Microform can be economically created, duplicated and distributed. Microfilm systems can be combined with EDP access systems. But microfilm call also be efficiently digitised with microfilm scanners. The capacity for economical digitisation of microfilm makes it a constantly upwards compatible medium with quality reserves, which will pay dividends in the future, given the expected improvements in reproduction quality and economic viability.

As an analogue and ageing-resisting storage medium, whose accessibility can be maintained with relatively small resources over long periods of time, and which remains available at all times for further processing in digital systems, microfilm has its place in the digital media world. Microfilm, as a high-quality intermediate storage medium, 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 prior to digitisation from the resulting microfilm. This remains true on financial grounds even when the only concern is digitisation of material for new levels of access and use. In such circumstanccs too, microfilm, as a compatible long-term storage medium, justifies over long periods of time the resources invested in its preparation and handling. Heavy expenditure for data migration and frequent technical and organisational measures to preserve readability in new system environments of material only available in digital form can thus be avoided.

When an original is to be digitised directly, it is important to bear in mind that the additional purpose of digital storage and processing must not be achieved at the cost of disadvantages in reproduction quality, loss of data after little use through perishability, or lack of compatibility or «future proofing» of the information medium or of the hardware. A programme specifying the technical and organisational steps involved in periodical migration, which can be constantly refined, should be part of the system design. Here too microfilm as a medium has a part to play. It is in principle possible to expose digital image data on microfilm. However, contrary to statements that now and then appear in the professional literature, printing out digitised data on microfilm, which can then be used as an analogue long-term storage medium, involves a notable reduction in quality. Microfilm which has been produced in this way cannot at present be used for digitisation with any guarantee of an acceptable result. Full compatibility of analogue and digital storage forms does therefore not yet exist.

2 Requirements for film quality and film organisation with reference to the option of film digitisation

2.1 *Choice and quality of film*

In both preparation and execution, planned filming projects should take into account the possibility of later digitisation. Whereas in the area of materials and technique there are only a few extra requirements that go somewhat further than the normal rules and requirements, there are additional points which should certainly be noted in the area of film organisation.

For the filming of strongly contrasted material, such as text, line drawings and engravings, the whole current range of pan-chromatic AHU-microfilm (Antihalation Undercoated) on polyester basis can be used. In recent years this has also been optimised in respect of digitisation. 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), such as books containing photographic reproductions, or coloured material which is to be filmed in black and white. This 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 digitisation, but filming nonetheless requires attention to the correct lighting and exposure, as well as to optimal readibility (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. Orientated 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 № 2 in line pairs per millimetre (lp/mm) and *h* the height of the small letter «e». Microfilm systems which give a value of 120 lp/mm and higher in the middle and at the edge of the image generally meet this standard. So far as the kind of microform is concerned, 35mm roll film for the master and as the starting point for digitisation is clearly to be preferred. Its image size guarantees sufficient quality reserves, even with problematic material, up to a size of 60 × 80 cm. As a rule, the normal commercial film lengths of 65m or 30.5m are to be recommended, the longer film being handled more easily because of the shorter preparation time. Far more successful results are obtained from digitisation of negative than positive film. A duplicate film of the lowest possible generation should be used for digitisation. As the preservation master exists for preservation purposes and cannot he 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 he digitised. In principle, however, it is also possible to digitise a diazo copy. Filming with the use of blips is in every case a requirement for an efficient working method with microfilm scanners.

Digitisation of microfiche is also possible. However, the smaller image field results in a lower reproduction quality in large-scale work.

Digitisation of microfiche also requires markedly greater staff resources. Microfiche digitisation also makes more demands on the software, which raises the time and thus the costs in comparison with work on roll film, which can be made largely automatic. Conversely, when selected extracts, as opposed to a single continuous run are required, the use of microfiche can be financially advisable.

A flaw-free film is a prerequisite for the best possible results in digitisation. 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 digitisation than 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 additional adjustment of the microfilm scanner raises resources and thus costs. The material to be filmed should, therefore, be presented so far as possible in a uniform way. These guidelines deal with:

a. *Reduction ratio* – One reduction ratio should ideally be selected for a complete filming project, but in any event for each single film. If need be, the material to be filmed should be arranged by size. In digitisation 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. Material should be placed in the middle of the front edge of the filming table. If this is not possible, it 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. Rotation by the digitising firm of otherwise aligned material is another cost factor. Books and documents should as a general rule 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 are to 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 digitisation leads to further costs, as this function is not normally included in current digitisation software, and therefore needs to be done later, as a separate 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 material from the whole digitised image. Elimination of the peripheral zones not only contributes to the optical image, but also to reduction of the storage quantity.

2.3 *Organisation and documentation of filming*

As usual with conservation 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 digitising firm the question of identifying the film in a way which is machine controllable and also allows for the possibility of data supply.

There are certain elements of the organisation of filming which with ordinary film projects are often – wrongly – ignored. They are more important in the case of film digitisation. They include take-counters, subdivision of films by indication sheets, placing 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 considerably

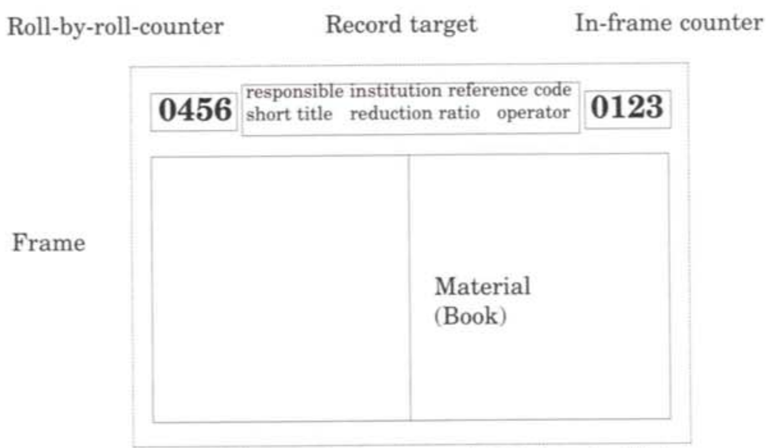

Figure 1. Placement of source materials for microfilming

easier and also leads to a reduction in costs. As well as 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 aspect is particularly important in relation to data organisation, *i.e.* accuracy of access and avoidance of superfluous page turning on the screen. Its high value justifies the small extra resources required at the preparation and filming stages.

The resources committed to structuring the information depend on the nature of the filmed material and the use to which it is to be put. It clearly makes no sense to put a 300-page book on the screen without any further subdivisions. At all events, material that is to be filmed requires *de facto* more extensive indexing which can also serve for consultation of the medium microfilm; the time and, therefore, personnel and economic implications of this must not be underestimated.

2.4 *Suggestions on choice of system*

The reproduction quality depends essentially on the installation of suitable filming equipment.The requirements which have heen described are as a rule met by modern planetary cameras, which guarantee resolution of at least 120 lp/mm over the entire screen. The equipment should also include the following: automatic focussing; lighting automatically adjustable, depending on the material being filmed; possibility of turning the camera head; adjustability of the lamps (*i.e.* for lighting book folds); image field projection; adjustable image masking; automatic lighting of blips; take counters. For filming books and archival material, the camera should produce optimal results with reduction ratios 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.

As second-generation films (duplicating masters) are normally used in digitisation, it should therefore be noted that the film should be silver-halogenide duplicate film of the same polarity (DDP film), and that the duplicating should be undertaken with high-quality duplicating equipment (working under vacuum on parallel-running films), in order to keep the loss of resolution by duplicating to a minimum.

It can in general be said that, for filming with a view to subsequent digitisation, the choice of system and the procedures in general will be dictated by the same criteria as apply in the case of good--quality microfilming. However, more attention needs to be paid to making the film form as unified as possible, to the organisation of the filming, the structure of the film, and to its documentation.

2.5 *Digitisation of existing films which, for one reason or another, do not meet these requirements*

It is possible to digitise existing films and film copies. In such cases it is of fundamental importance to work with films of the lowest possible generation. It is advisable in every case to undertake a thorough analysis of the films (material, state of preservation, reduction factor, reproduction quality, filming technique, nature of the material, organisation of the filming). This analysis is best undertaken in cooperation with an experienced service provider.

Before any award of a contract, digitisation 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 result of digitisation.

2.6 *Differing recommendations for colour microfilm*

The starting point for digitisation of film should be a high-resolution, permanent-colour bleach-fixing-process microfilm on a polyester film base. From this a reproduction quality, especially as regards resolution, which matches that of black and white microfilm can be expected.

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

The practice until now, almost without exception, has been to digitise colour film using a proprietary system developed for the amateur market. Cost has played a part here too. The cheapest version of this limits the area of the image which can be digitised to 24 × 36 mm. However, a «full-step» colour microfilm image occupies 32 × 45 mm. Filming using the maximum size of the «full-step», which is of advantage for reproduction quality and for further processing when dealing with larger or more difficult material, is not possible with 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. Processing 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 digitised 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 digitisation of colour film. For colour microfilm which meets the standards, however, it has only limited use, particularly with regard to format. However, failure to use the full-step format of unperforated 35mm microfilm would normally be associated with loss of quality, especially in the case of colour microfilm. But loss of quality should not be accepted. Film scanners which can digitise such films are available. Colour film scanners have been introduced in the reprographic field which can work with film up to a format of 6 × 9 cm and with filmstripes. It is possible with these to digitise full-step 35mm colour microfilm. These scanners have a resolution up to 2000 dots per inch (dpi). The colour digitisation is not limited to photographic CD, but may be produced in whichever is the preferred of several other possible formats. However, the current state of technology and the comparatively small demand for digitisation of colour microfilm make this a fairly expensive process. In view of developments we can expect in the future, however, 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.*

* In the meantime a digital camera has become available which is able to digitise also from colour microfilm full step with a resolution up to 3800 × 4600 pixels.

3 Recommendations for digitisation 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 the purpose to which it is to be applied. In other words, as a general rule, digitisation of microfilm should not aim at the best possible result, in the way that is mandatory for direct digitisation of endangered original material.

Bitonal digitisation is adequate for the reproduction of printed text, including line drawings, also for modern non-impact typescript (plastic carbon band, inkjet and laser printers) on pan-chromatic AHU microfilm. Gray scale must be used for digitisation of the following: manuscript; pencil and crayon drawings; typescript produced with a silk ribbon; coloured illustrations and drawings; other material with varying shades of gray; black and white and colour photographs. 16 gray scale (4 bit) is usually adequate for digitisation of contrast enhancing AHU film. For digitisation from half-tone film, 256 gray scale (8 bit) should be used. Digitisation with gray scale has serious implications for storage requirements, and thus on cost, at all stages of the process. It should thus be undertaken only where this reproduction quality is indispensable.

In digitisation from film, the necessary resolution is determined by the size of the smallest element which is to be clearly discernible. With printed texts this is the height of the small «e», with manuscripts the doubled letter width (as 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 digitisation 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 millimetres. For digitisation with gray scale the formula is: QI = (a × 0.039)/2.

With bitonal digitisation 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 quality reserves of the microfilm in the background, it will be sufficient for most purposes to aim for a digital secondary form 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 digitisation 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 digitisation, 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 digitised image data should be by digital audio tapes (DAT) or CD-R (Recordable). Readability independent of hardware is guaranteed for both media through standardisation (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, as the reliability of DAT rests essentially on the tension of the tape, which can undergo changes during transport. CD-R also offers the possibility, following supply of the digital conversion form – and as something of a quality control – of direct access with viewer software to the images, without the need of first storing them on the hard disk of the computer system.

It is important to reach a binding agreement with the company undertaking the digitisation that it will store the transferred material for at least as long as the results are to be controlled and safeguarded by the customer.

The digital conversion form is reliably secured when loss-free compressed or uncompressed image data have been secured on at least two data carriers, it has been checked that their contents are identical and that they are 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 readibility of the primary data carrier, working duplicates should be produced from it by means of multiple copies. A decompression test for every stored digital copy provides a further enhancement of 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.

Readability and further processing on differing equipment with differing systems and programmes is possible. It should however be noted that, despite thorough-going standardisation, the TIFF format allows variations, which may not be compatible with the installed software.

Here too careful discussion and possibly experimental runs with test data are to be 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 this should be used for data delivery in order to save storage space. However, as not all programmes 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 colour pictures, has variable compression ratios which are all lossy and thus not to be recommended.

As image data can be organised in different ways, it is advisable to agree with the service provider on the organisation 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 only possible with documents which consist of only a few pages.

For further work on the data on the Internet, it is advisable to convert data into platform-independent formats which allow inclusion of the wides 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 digitised images, various programmes for viewing and manipulation are available for PC and UNIX environments. These include «Viewer», obtainable as public domain software and also as financially attractive shareware programmes. It is recommended to install in each institution only one specific, standardised software, whose compatibility with the supply of digitised conversion formats can be rigorously tested in advance.

As a rule viewer-software should have the following features: page turning forwards and backwards; 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; turning the image; inverting the image; 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). In the PC area, we may mention as an example Imaging for Windows, an extra 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 control and display of digitised 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 which meets requirements for inspection and use of digitised images must be provided in each institution. The comparatively large quantities of data contained in digitised images as opposed to paper files lead to heavier demands on the data bus and the 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 colour 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 really 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 digitisation, the converted format must be preserved in the long term. If only on financial grounds, repeated digitisation is out of the question. Given the increasing importance of electronic information systems in research and teaching, the digitised images should be useable in the future for a multitude of 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 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 standardisation effort – that are 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 organisational 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 which 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 which produced it remains available.

3.7 *Financial viability*

Digitisation of microform should be a matter of service provision. The costs of digitising according to the foregoing recommendations a uniformly produced microfilm (35mm roll film) 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 readibility of the filmed material. As digitisation costs are also dependent on the market situation, it is not possible to give any general indication of prices which will have long-term validity.

The cost factors we have mentioned take account only of digitisation 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 filmscanner according to the customer's requirements must also be considered. Finally, there are the costs of downloading the data, operation of the CD-R, of the carrier medium and of packing and transport. In relevant cases, the relatively high costs of improvement in quality by individual work and image enhancement with special software play a part.

The choice between digitisation 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 can turn out more economic to digitise a second time from the microfilm rather than constantly migrate the data.

3.8 *Digitisation and optical character recognition*

Optical character recognition (OCR) is a machine process which turns visible alpha-numeric signs into coded data (codes corresponding to the alpha-numeric 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 which supports pattern recognition with dictionaries, linguistic methods and features of «artificial intelligence». The text recognition programmes increasingly integrate dictionaries and substitution lists which are adjustable according to the degrees of security. To prevent the substitution of inaccurate characters that were wrongly recognised 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 which are either not recognised, or not recognised with certainty, are retained as images and remain in that uncoded form, in position in the remaining – correctly recognised – text.

In addition to reliable text recognition, page segmenting is an essential performance feature of text recognition systems, *i.e.* interpretation of contextual information such as columns, blocks of text and graphics. Further features are deskew, segmenting of individual units, recognition of types of handwriting and signatures and 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 arises through 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 digitising can therefore improve the recognition rate, as with digitising 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. Digitising negative film avoids the disruption caused by dirt and scratches. In practice, there has not yet been sufficient experience of 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 in order to reach conclusions about practical compatibility between digital and analogue conversion forms. We used for this purpose a test surface in Format DIN A2 with standard test indications for resolution (reproduction sharpness), reproduction of gray scale and colour rcproduction. Further, samples of different printed text and handwriting, together with black and white and colour photographs, were mounted on the test surface. The surface was filmed in black and white and in colour on different microfilm material (35mm) and directly digitised, using different scanners, bitonally, in gray scale and in colour. Working paper copies were made of the film and of the digitised conversion formats. The results of the test surface were digitised with different film scanners. As a further step, transfer was made from film to microfiche. Finally, the digitised image data from the test surface was exposed by COM systems on microfilm. The results were then evaluated, the analogue test patterns microscopically, according to the appropriate standards, the digital formats with assistance of a high resolution colour 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 digitisation from original in gray scale. Transfers from microfilm, and in part also from original, onto microfiche and into digital conversion forms, retain readibility. Working copies on paper (printouts from reader-printers and laser printers) are at a similar level. A four-colour print from a digitised original is the only example to reach the test indication group 4.5. In this case digitisation of the test surface and its colour microfilm copies was carried out 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 surface 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.

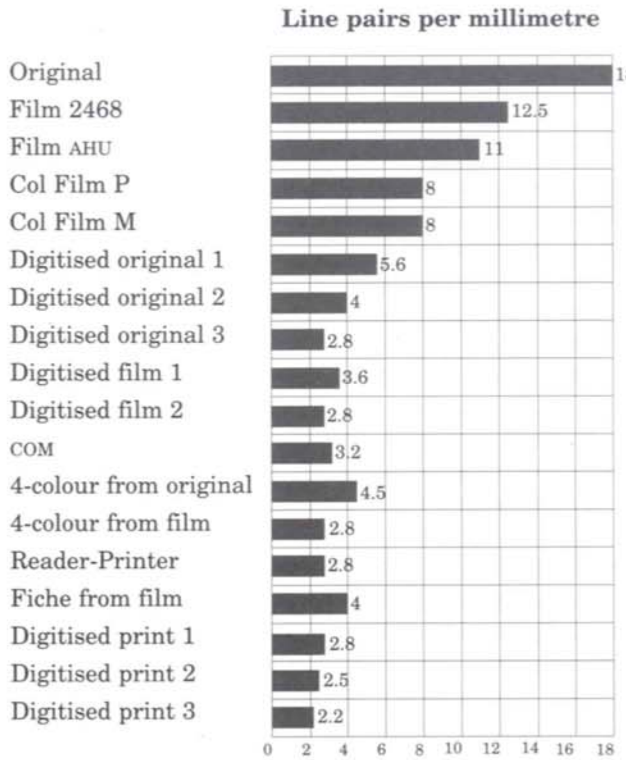

Figure 2. Sharpness of reproduction

Thus film digitising produced rather better results, as the contrast of the material in microfilming was already somewhat enhanced and generally evened out.

These results are of an experiment with a relatively large test surface, corresponding approximately to the size of a sheet of newspaper or an open large folio book. Better results can be expected from smaller material, in respect of which the relationships remain in principle the same.

An essential point to stress is that a lower quality must be expected from digitisation of film, as compared with the high value digitisation from original with gray scale. On the other hand, the quality reserves of the film used in this experiment proved themselves to be so superior, that they would be adequate if the digitising systems were to be further developed, and digitising with higher resolution were to be financially more viable. In the case of less resource-intensive bitonal digitisation, the results of digitising from film and from the original were similar.

4.2 *Printing out from digitised storage onto film (image)*

It proved difficult to find a company able to print out onto microfilm a TIFF data file of the digitised 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 the special equipment of one firm, which had been installed especially 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, 36-fold and 18-fold only in part. As the pixel number of the equipment was about 3200 x 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 carried out through a foreign firm, which managed to print out the whole of the test surface onto microfilm (35mm). However, the reproduction quality, with a resolution level of 1.8, was very low, undoubtedly due in part to the fact that a film with low resolution capacity had heen used instead of microfilm.

All that was available for this experiment was a digitised 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 digitising 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 analogue and digital media in the direction digital-to-analogue. 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 digitising the microfilm were identical with that used to produce the film via COM (see paragraph 3.1). Current experiences with optical systems throw doubt on whether such «film to film» compatibility can ever be achieved.**

** 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.

4.3 *Should filming or digitising come first?*

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

5 Digitising from the original

5.1 *Quality requirements*

In the current state of technology, digitising from the original gives a better reproduction quality for coloured material and material with weak contrasts than digitising from film. When endangered original material is digitised, 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 digitised secondary form exists only to improve access. A later, repeated digitisation of the endangered original, even if possible, is not consistent with the aim of preservation. This means that the first digitisation 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 digitisation 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 digitisation of printed text which includes line drawings. A resolution of 400 dpi is generally adequate for bitonal digitisation with texts which are clear, larger and in particular evenly spaced (10 point and above), which have been produced by modern, non-impact typewriters, such as plastic carbon band, or by inkjet or laser printer. 256 gray scale and a resolution of 400 dpi should be used for the following: manuscripts; drawings with pencil or crayon; typescript with silk ribbons; coloured illustrations and other drawings with varying gray shades; black and white and colour photographs. These recommendations also correspond to American quality requirements for digitising original material.

The suggestions on filming technique in paragraph 2.2 and on film organisation and documentation in paragraph 2.3 can contribute usefully to digitisation and to the further processing of the digitised conversion form.

5.2 *Criteria for the choice of system*

Scanners which work like a planetary camera, digitising the material from above, must always be used for sewn and bound volumes. Feeder scanners and flat bed scanners are not suitable for books and archives. It is especially important to follow the precautions for the protection of books and volumes described in paragraph 2.4. Equipment of this kind is indispensable for the digitisation of unique material which is in an impaired condition.

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 in case of deterioration on the part of the digital data carrier there is nomicroform to fall back on, additional quality tests are necessary for the storage of digitised image data on optical disk. The following procedure is suggested:

First the digitised 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 reproduceable 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 as the copy for preservation. If needs be, it serves for production of further duplicates. It is not absolutely necessary to subject the duplicates intended for daily use to the same quality test as the primary data carrier. If in the course of normal use it becomes apparent that individual copies are not correctly reproduceable, 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*

Organisational and technical measures are in any case advisable in the migration of digital conversion forms, to safeguard the process of transferral of information and for reasons of economic viability. However, they become indispensable where the digital form is the only form in addition to the original, or where it is expected that it will sooner or later replace the original. A repeated digitisation of the original is not only to be avoided on grounds of preservation, but would be prohibitively expensive.

The organisational 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 adequately to safeguard the data carrier which is created in each case.

5.8 *Financial viability*

Where books or archival documents are to be digitised as a whole, this should be done by commercial firms. Where only certain pages, or parts of a document, are to be digitised, this can be done by the institution itself. The cost of digitising 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 digitised with flat bed or feeder scanners than books or other bound volumes, for which special book scanners need to be installed.

When working out the cost of digitisation 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 digitise 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 digitise in parallel from the original at the same time. Paragraph 3.7 is relevant on other points.

5.9 *Differing recommendations on colour images*

It should be noted that in the current state of technology, digitisation of colour can only be done at relatively low resolution values, or for limited quantities of material, as very large quantities of data are involved. Test runs should always be carried out, to establish if the reproduction quality is acceptable.

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

6 Cooperation and exchange of information

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

7 Suggestions for further reading

GERTZ, Janet
Oversize Color Images Project, 1994-1995. Final Report on Phase 1. Washington, DC: Commission on Preservation and Access, August 1995.

GULBINS, Jürgen, SEYFRIED, Markus and STRACK-ZIMMERMANN, Hans
Elekronische Archivierungssysteme (Electronic Archive Systems). Berlin, Heidelberg and others: Springer 1993.

KENNEY, Anne R. and CHAPMAN, Stephen
Digital Imaging for Libraries and Archives. Ithaca, N.Y.: Dept. of Preservation and Conservation, Cornell University Library, 1996.
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 Association for Information and Image Management, 1993.

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

WATERS, Donald
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

HARTMUT WEBER
Landesarchivdirektion Baden-Württemberg, Stuttgart

MARIANNE DÖRR
Bayerische Staatsbibliothek, München