

B2
02/6522/4c

ISSN 0873-5670

PÁGINAS

ARQUIVOS & BIBLIOTECAS

a&b

MICROFILMAGEM
DIGITALIZAÇÃO

2

19
98



Edições Colibri

149347-20



UNIVERSIDADE DE AVEIRO
SERVIÇOS DE DOCUMENTAÇÃO

PÁGINAS

ARQUIVOS & BIBLIOTECAS

a & b

Microfilmagem



Digitalização

2

19
98

FICHA TÉCNICA

Director

Maria Luisa Cabral

Redacção

Maria Luisa Cabral
Fernanda Ribeiro
Armando Malheiro
Carlos Abreu

Revisão

Júlia Ordorica

Edição e Propriedade

Edições *Colibri*
Apartado 42001
1600 Lisboa

Periodicidade

Semestral

ISSN

0873-5670

Depósito Legal

118 811/97

Tiragem

1000 exemplares

Paginação

Albertino Calamote

Execução Gráfica

Colibri Artes Gráficas
Faculdade de Letras da Universidade de Lisboa
Alameda da Universidade
1699 Lisboa Codex
Tel/Fax 796 40 38

Internet: www.edi-colibri.pt

e-mail: colibri@edi-colibri.pt

*Os artigos assinados
são da exclusiva responsabilidade
do(s) seu(s) autor(es).*

Edição patrocinada por:
INSTITUTO PORTUGUÊS DO LIVRO E DAS BIBLIOTECAS
MINISTÉRIO DA CULTURA

Sumário

Editorial

5

HARTMUT WEBER, MARIANNE DÖRR
Digitisation as a method of preservation?

7

FRANCISCO BETHENCOURT
Bibliotecas digitais

35

MARIA LUÍSA CABRAL
Microfilmagem & digitalização

41

MANUEL LUÍS REAL
G.I.S.A. – Gestão Integrada de Sistemas de Arquivo

53

ALFREDO CALDEIRA
Projecto de digitalização do Arquivo Mário Soares

73

MARIA DA ASSUNÇÃO VASCONCELOS, CLARA MOREIRA
Projectos informáticos do Arquivo Distrital de Braga

93

Editorial

Aos responsáveis pelas *PÁGINAS a&b* pareceu oportuno organizar um número sobre a problemática em torno das bibliotecas digitais no seu sentido mais lato e flexível. Sem ter a pretensão de esgotar a temática, decidimos entrar frontalmente no debate. De certa forma até, provocá-lo. O que é isso de informação digital? Que experiências estão em curso entre nós? A digitalização constituirá uma alternativa válida para a preservação da herança cultural escrita? E se for assim, qual é o lugar da microfilmagem? Poderemos sem hesitação pôr esta de parte? Um conjunto de questões que vale a pena trazer à discussão. Não é admissível que se tente construir as nossas bibliotecas do futuro sem enfrentar os problemas actuais com um mínimo de discussão pública. Os artigos que se publicam revelam abordagens diversas mas também deixam muito claro que os profissionais estão perante grandes desafios a exigir muita formação. Bibliotecários e arquivistas têm de olhar pela sua formação com uma atenção constante e, se possível, ainda mais dirigida. As *PÁGINAS a&b* dão uma pequena contribuição na esperança de que os seus leitores exijam mais e mais de si próprios e às instituições onde trabalham.

Ainda não foi com este número que conseguimos levar à prática o modelo de revista defendido nas linhas programáticas que presidiram ao aparecimento das *PÁGINAS a&b*. Esta-

MARIA LUÍSA CABRAL

mos a trabalhar nesse sentido como também nos vamos coordenando com o objectivo de regularizar a nossa periodicidade. Ambas as metas são de difícilima concretização mas não fora o desafio e o nosso projecto acabaria em rotina, aquela mesma rotina que decididamente recusamos.

MARIA LUÍSA CABRAL

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 Forschungsgemeinschaft (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 Staatsbibliothek, München. Published by the European Commission on Preservation and Access (ECPA). Originally published in German in the *Zeitschrift für Bibliothekswesen 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 Medicott, 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 Forschungs-

gemeinschaft (German Research Association) responsible for questions of preservation suggested in the spring of 1995 the setting up of a working group to discuss questions of 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 checklist 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 can 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 circumstances 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 continuous-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 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. 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 $QI = a \times h$, where a is the resolution number of the ISO resolution test pattern N° 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 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 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

Roll-by-roll-counter Record target In-frame counter

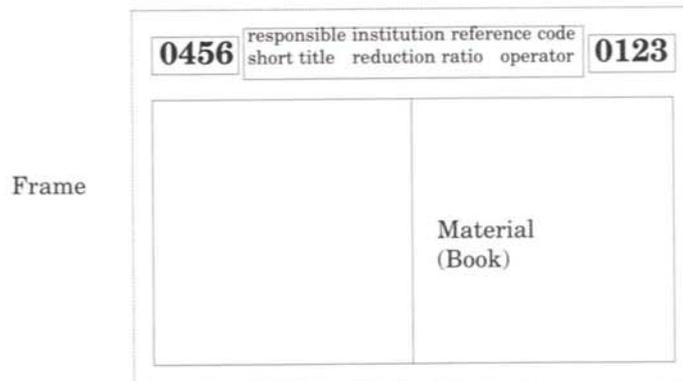


Figure 1. Placement of source materials for microfilming

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 been 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 over-size 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 \times 192 pixels, the highest 2048 lines \times 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 \times 9 cm and with filmstrips. 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 \times 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: $QI = (a \times 0.039)/2$.

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 $a = 3 \times 5/0.039h$, 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 $a = 2 \times 5/0.039h$, 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 readability 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 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 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 486 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 readability 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 film scanner 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 reproduction. 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 readability. 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.

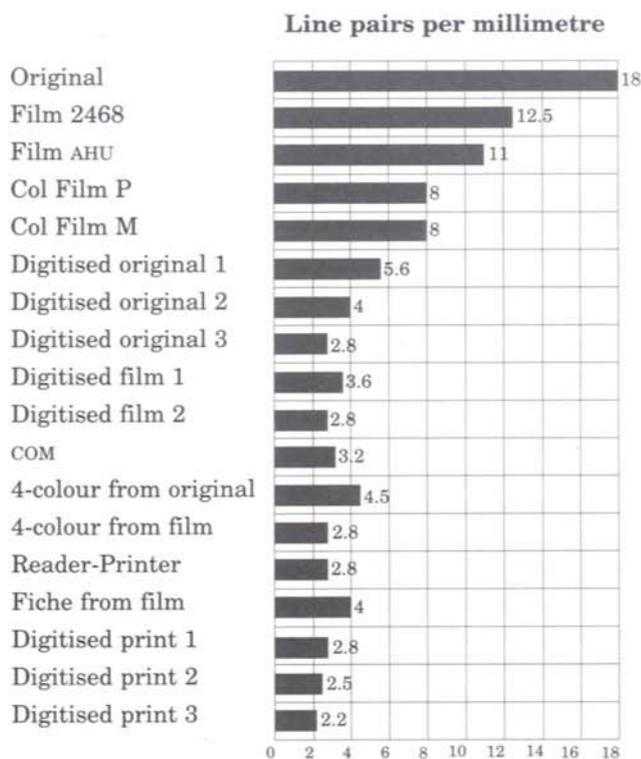


Figure 2. Sharpness of reproduction

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 been 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 no microform 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 digi-

tisation 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 I. Washington, DC: Commission on Preservation and Access, August 1995.

GULBINS, Jürgen, SEYFRIED, Markus and STRACK-ZIMMERMANN, Hans

Elektronische 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.

HARTMUT WEBER; MARIANNE DÖRR

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

Bibliotecas digitais

FRANCISCO BETHENCOURT

A digitalização de documentos avulsos e livros, se for feita de uma forma sistemática e coerente, por exemplo em torno de uma área temática, pode conduzir à constituição de uma biblioteca virtual. O suporte electrónico, dadas as suas características – capacidade de armazenamento, possibilidade de transferência, transmissão à distância, portabilidade e acesso –, pode disponibilizar centenas de milhares ou mesmo milhões de páginas de texto. As vantagens são evidentes:

- a) redução do espaço ocupado por essa informação, que equivaleria a quilómetros de estante em suporte de papel;
- b) democratização do acesso, dada a redução dos custos;
- c) ampliação do público-alvo, pois se essa biblioteca virtual for acessível directamente através da Internet pode ser consultada a partir de qualquer país no mundo.

Existem grandes projectos internacionais para a constituição deste tipo de bibliotecas, como a *Bibliotheca Universalis*, patrocinada pelos países do G7, na qual a nossa Biblioteca Nacional também participa. Esse projecto visa a constituição de uma espécie de colecção digital, em *full text*, das grandes obras produzidas pela humanidade ao longo dos séculos. Existem igualmente projectos nacionais, organizados em torno da literatura e da história, nomeadamente o *American Memory*, gerido pela Library of the Congress, que tem conhecido réplicas em diversos países, nomeadamente em França e em Espanha. Em Portugal lançámos em 1997 na Biblioteca Nacional, sob a coordenação do Prof. Ivo Castro, o projecto da *Biblioteca Virtual dos Autores Portugueses*, que veio a ser distribuída de forma massiva em Junho de 1998 aos leitores do *Diário de Notícias* em CD-ROM. Estão aí contidas centenas de milhares de

páginas de texto com mais de trinta obras básicas da literatura portuguesa, que custaram aos leitores apenas 1200\$00. Trata-se de um processo de democratização radical do acesso à cultura que não foi devidamente avaliado pelos conhecidos «analistas».

Estes grandes projectos têm vindo a ser desenvolvidos, em todos os países, através de uma cooperação estreita entre bibliotecas e universidades, embora com modalidades diversas. A Electronic Library, no Reino Unido, está a ser desenvolvida directamente pelas universidades, devendo ser disponibilizada através da rede de bibliotecas, enquanto a Digital Library, nos Estados Unidos, obedece a uma maior articulação entre os dois tipos de instituições. Neste último caso verifica-se uma maior intervenção das bibliotecas na organização de projectos, bem como na encomenda da produção intelectual e científica. Dir-se-ia que as bibliotecas assumiram plenamente o seu estatuto de intermediário privilegiado entre a criação de produtos electrónicos em diversas áreas do saber e o público consumidor. As bibliotecas constituem-se, neste caso, como agências que conhecem as tendências e os gostos do público leitor, orientando a própria pesquisa. A sua capacidade de disponibilização dos produtos electrónicos através das suas redes internas e externas, facilitada pela normalização biblioteconómica e pelo protocolo de comunicação entre bibliotecas a nível internacional, faz com que o agente distribuidor tenha um papel decisivo na própria definição dos conteúdos. Esta é uma nova tendência que seria impensável há trinta ou quarenta anos atrás, quando as bibliotecas se reduziam, na maior parte dos casos, ao estatuto de depósito e consulta de livros. Agora são estruturas que intervêm directamente na própria produção criativa.

As grandes bibliotecas têm agora condições para articular projectos ou conceber, de raiz, projectos conjuntos. O fenómeno da globalização tem o seu reflexo imediato no mundo das bibliotecas: não estamos a falar apenas da *Bibliotheca Universalis*, pois existem grandes projectos entre a Library of the Congress ou a New York Public Library e as grandes bibliotecas europeias, nomeadamente sobre a emigração para o Novo Mundo ou sobre o Plano Marshall. Para além disso, as bibliotecas com maiores recursos financeiros têm possibilidade de produzir projectos piloto em áreas inovadoras, como no caso da San Francisco Public Library, a qual desenvolveu bibliotecas virtuais em torno da literatura latino-americana, dos *Gay and Lesbian Studies* ou dos *Black American Studies*. Trata-se de uma resposta rápida a novas tendências de pesquisa e novas necessidades de informação do público leitor.

A transformação da natureza das bibliotecas, que passaram de depósitos de livros a entidades gestoras de informação, onde quer que ela exista no mundo, bem como a estruturas criativas de produção de conhecimentos, decorre do seu posicionamento na nova sociedade da informação. Quando a explosão do conhecimento ocorreu nas últimas décadas, as bibliotecas eram as únicas estruturas capazes de fazerem a catalogação, a classificação e a indexação dessa enorme massa de textos. Numa palavra, as bibliotecas puderam introduzir a ordem no caos informativo, com a vantagem de estarem integradas em associações com projectos comuns que tinham conduzido à normalização internacional dos procedimentos indicados. Foram estes organismos que permitiram tirar o máximo de rentabilidade do *boom* de conhecimento, tornando eficaz a multiplicação da informação em todo o mundo.

Esta é a vertente «cor-de-rosa» do papel das bibliotecas e das vantagens da sua intervenção na produção digital. Mas interessa abordarmos aqui os problemas, tanto técnicos como intelectuais, que essa intervenção levanta. Os problemas técnicos têm a ver com a possibilidade de proceder, ou não, a uma digitalização directa das colecções existentes nas bibliotecas, em lugar de se continuar com os projectos de microfilmagem. Como se sabe, as bibliotecas conhecem fortes problemas de preservação e conservação das espécies em suporte tradicional que possuem em depósito, cuja constante manipulação destrói as encadernações e o próprio papel. As colecções de periódicos do século XIX, por exemplo, constituem um quebra-cabeças, pois trata-se do papel mais frágil e quebradiço. Daí a necessidade de um forte investimento em microfilmagem, que deveria ser concentrado num projecto de âmbito nacional que envolvesse diversas bibliotecas. As vantagens são claras: reprodução dos microfímes e disponibilização de cópias em diferentes bibliotecas; conservação dos originais, cujo estado, em muitos casos, já está bastante deteriorado. A pergunta que se coloca é se a digitalização directa não seria mais económica. Infelizmente, ainda não é. A Library of the Congress tem uma estimativa dos custos médios dos três grandes processos de preservação e conservação disponíveis neste momento: desacidificação, microfilmagem e digitalização. Esses custos progridem na ordem de 1 para 6, sendo a microfilmagem ainda um processo bastante mais barato que a digitalização.

As razões dos elevados custos da digitalização são compreensíveis:

- a) o processo exige mais mão-de-obra e mais qualificada;
- b) ao contrário do que acontece com a microfilmagem, não se

trata apenas de «bater a chapa», a operação exige uma manipulação posterior, por *software*, das imagens, para que elas sejam reconhecidas e disponibilizadas;

- c) o espaço de armazenamento exigido aumenta exponencialmente quando estão em jogo imagens e não apenas texto.

Mas os problemas não são apenas de custos, são também técnicos:

- a) ainda não existe uma estabilização e normalização do *software* disponível;
- b) a capacidade de resolução das câmaras digitais, apesar dos desenvolvimentos recentes, ainda não atingiu a qualidade das câmaras convencionais;
- c) os suportes electrónicos existentes neste momento só garantem dez a quinze anos de vida, ao contrário dos microfilmes, que prevêm uma longevidade de centenas de anos, se forem respeitadas as condições requeridas de temperatura e humidade para a conservação dos originais em sais de prata.

Naturalmente que uma boa parte destes problemas técnicos poderá ser resolvida a curto prazo, mas é necessário ter em conta as experiências e as opções tomadas nos países mais desenvolvidos antes de se fazer um investimento massivo em programas de digitalização. A manutenção de vários sistemas para diferentes fins – desacidificação, encadernação, microfilmagem e digitalização – parece ser, a curto e médio prazos, a melhor aposta. A digitalização para projectos especiais, como aconteceu com a *Biblioteca Virtual dos Autores Portugueses*, constitui uma opção sólida, que permite criar *know how* e manter uma estrutura actualizada, capaz de dar o salto para um processo de massificação logo que existam as condições técnicas e materiais necessárias.

Os problemas intelectuais suscitados pelos projectos de digitalização têm a ver com o risco de fixação de conhecimentos e de cristalização de novos «cânones». Com efeito, a selecção de literatura latino-americana para a criação de uma biblioteca virtual nessa área pode concentrar as atenções num determinado conjunto de autores, deixando outros de lado que não são menos importantes. O poder de difusão deste «cânone» não é negligenciável, pois os estudantes e curiosos de determinadas áreas do saber dirigem-se de imediato aos *sites* disponíveis, procurando a informação mais acessível para trabalhar de imediato nos textos propostos. Poder-se-ia dizer que as antigas colectâneas de textos, nomeadamente aquelas que todos nós utilizámos no ensino secundário, tinham o

mesmo efeito. Mas a verdade é que o poder da informação na época digital é completamente diverso: nós desconfiávamos das ditas «selectas» e procurávamos as bibliotecas justamente para nos «perdermos» na busca de autores não canónicos. Era justamente esse o sentido de uma cultura humanista que existia subterrânea à escola, desconfiada da instituição. As bibliotecas constituíam então um espaço de liberdade, de conhecimento alternativo, embora o caso português não seja o mais significativo, dada a existência de censura durante dezenas de anos e dada a falta de recursos financeiros para apetrechar convenientemente essas mesmas estruturas. Hoje em dia temos bibliotecas que destroem uma parte dos seus fundos antigos para adquirir obras recentes em sectores «de ponta» ou seja, da moda, cuja perenidade é sempre impossível de prever. Isto aconteceu na San Francisco Public Library, fenómeno que levantou uma forte polémica na imprensa e conduziu à demissão do director. Esta situação constitui um elemento de encorajamento face ao risco de cristalização de novos conhecimentos: existe, como se verifica, uma opinião pública forte em certos países que evita ou castiga estes casos extremos. Para além disso, a velocidade de actualização das bibliotecas virtuais cresce de dia para dia, constituindo-se como espaços abertos a sugestões. Por último, as bibliotecas de livre acesso, que devem caracterizar a rede universitária de qualquer país, permitem a correcção de eventuais «cânones» que se queiram introduzir.

A última palavra é, por tudo isto, de optimismo face às virtualidades e à boa utilização das bibliotecas digitais. Os problemas técnicos terão tendência a ser resolvidos no futuro próximo, criando novas alternativas para as estratégias de preservação e conservação, bem como para a sistematização, disponibilização e produção do saber. Os problemas intelectuais serão igualmente ultrapassados, dado o crescimento exponencial das bibliotecas de livre acesso com fortes recursos financeiros, bem como o desenvolvimento de uma dinâmica de actualização permanente e ruptura dos conhecimentos constituídos. As bibliotecas verão o seu papel reforçado nesta sociedade da informação, embora no caso dos países que dedicam menos recursos a estas áreas, como o nosso, se torne indispensável um debate público sobre uma área essencial de desenvolvimento estratégico.

Microfilmagem & Digitalização A coexistência pacífica

MARIA LUÍSA CABRAL

O contexto geral

De há um tempo a esta parte, as bibliotecas e arquivos têm vindo a debater-se entre a necessidade e urgência de garantir a transferência de suporte do conteúdo intelectual dos documentos e o tipo de suporte para o qual se deverá fazer essa transferência. Este debate tem sido muito sério nos Estados Unidos, no Reino Unido, na Alemanha*, na Holanda ou na Austrália. Em Portugal nem tanto. É prova do que afirmo a vasta bibliografia original, toda resultante de investigação prolongada e séria, publicada, por exemplo, nos Estados Unidos pela Commission on Preservation and Access (Washington, DC). No Reino Unido ultimam-se agora relatórios sobre o problema cujo conteúdo só poderá vir, estou segura, a confirmar o que os americanos vêm dizendo. Diria que a situação em França é muito mais parecida com a nossa: talvez o debate tenha acontecido; talvez mesmo algumas instituições tenham pesado os prós e os contras de cada uma das alternativas existentes para transferir a mensagem impressa, mas as opções são tomadas numa perspectiva a curto prazo o que, inevitavelmente, ameaça a fiabilidade dessas mesmas opções. O projecto em curso na Bibliothèque nationale de France, em matéria de digitalização, é magnífico mas não deixa de ter o objectivo preciso e bem concreto de estar ao serviço do acesso. Mesmo com grande apoio financeiro (e político) o projecto pretende criar uma biblioteca de referência, enciclopédica para:

* O trabalho de Weber e Dörr que neste número se publica denota bem o investimento que se faz.

1. melhorar a comunicação da própria biblioteca;
2. permitir a consulta à distância;
3. melhorar a gestão das colecções evitando o manuseio das obras frágeis. Não se fala em substituir por outro suporte; refere-se, sim, em facilitar o acesso.

Entre os modelos saxónico e francófono, não hesito em afirmar que preferiria ver as instituições e os profissionais portugueses enveredarem pelo primeiro porque:

1. apesar do sucesso imediato da digitalização, sempre se acautela o longo prazo;
2. a abordagem saxónica revela uma atitude mais amadurecida, mais bem fundamentada, mais profissional. Não é a primeira vez que demonstro a minha inclinação saxónica e os anos apenas ajudam a reforçá-la.

No presente artigo, de bibliotecário para bibliotecário e arquivista, tentarei explicar do que trata a microfilmagem, e também a digitalização. Tentarei enumerar as características positivas e negativas de cada processo, como e quando se aplicam e, por último, tentarei provar como um e outro processo não se opõem, antes se complementam. Afinal, como coexistem.

O dilema a enfrentar

Antes de explicar os processos técnicos em causa, importa esclarecer as origens do nosso dilema. A razão tem sido muito divulgada: as bibliotecas e os arquivos do mundo debatem-se com o gravíssimo problema que constitui a deterioração do papel. Sob os nossos olhos o papel, sobretudo o papel do século XIX, amarelece, enfraquece e quebra. Dizem as fontes norte-americanas que 25% dos documentos à guarda das bibliotecas e arquivos de todo o mundo estão em vias de desaparecimento. Em Portugal não temos certezas tão assumidas, mas o conhecimento real das nossas colecções leva-nos a conclusões semelhantes. Transferir a informação para outro suporte surge como a resposta certa para atenuar o problema, uma solução temporária, nunca uma solução definitiva já que, bibliotecários e arquivistas, mais não podem ambicionar do que adiar aquela morte anunciada. Transferir o conteúdo intelectual não será a única alternativa mas, em todo o caso, a resposta que mais ao alcance das nossas mãos e capacidade está. Mesmo das mãos dos *déciseurs* e profissionais portugueses da informação. Se a

prática vem ao encontro das minhas palavras? Não vem. Desculpar-me-ão os colegas portugueses se os não mencionar mas, tanto quanto consegui recolher, apenas cinco instituições portuguesas, entre bibliotecas e arquivos, executam microfilme com regularidade; dessas cinco, diria que três começam a ter algumas preocupações com a qualidade do microfilme que executam procurando, de facto, produzir o chamado «microfilme de preservação».

Contudo, e apesar desta exiguidade, nem tudo está perdido se pensarmos que, mesmo nos Estados Unidos, é a publicação, em 1986, do relatório *Brittle Books* pelo Council on Library Resources (CLR) que desperta verdadeiramente as consciências para o problema geral de deterioração do papel. Para o CLR a prioridade são os livros, tendo ficado bem definido que o acesso àquilo que se preserva (os livros, os artefactos) é tão importante como a preservação da própria informação. O microfilme apresenta-se, então, como o processo técnico que permite estruturar e levar a cabo um vasto programa de âmbito federal (nacional) para salvaguarda duma herança que é colectiva¹.

É menos importante conhecer os motivos que levaram os norteamericanos a optar por lançar um programa para salvaguarda dos livros do que saber que:

1. existe um programa para transferência regular, sistemática e global de transferência de suporte;
2. o processo técnico seleccionado foi o microfilme;
3. ficou definido, como preocupação fundamental do programa, a garantia inalienável ao acesso, sendo as cópias dos microfilmes, quando encomendadas, feitas pelo simples custo de produção.

O relatório da Conferência de Bellagio², sete anos mais tarde, mais não fez do que sublinhar e reforçar a necessidade de preservar a memória colectiva, impressa ou manuscrita sobre papel, em perigo de desintegração. Também não interessa discutir se a opção deveria ter sido dirigida para os livros ou para os jornais. Os norteamericanos mantiveram o sentido pragmático, avançaram e hoje podem contar-nos da sua experiência, publicaram bibliografia q.b. sobre os procedimentos técnicos a ter em conta na produção do

¹ *Preservation microfilming: a guide for librarians and archivists*. Edited by Lisa Fox. 2nd ed. Chicago and London: American Library Association, 1996.

² *Preserving the intellectual heritage: a report of the Bellagio Conference, June 7-10, 1993 held at the Rockefeller Foundation Study and Conference Center in Bellagio, Italy*. – Washington, DC: CP&A, 1993.

microfilme de preservação, e, apoiados nesse trabalho contínuo e sistemático, procedem à indispensável investigação que lhes permitirá outras opções.

No virar do século, as bibliotecas e arquivos portugueses ainda não encararam com conhecimento de causa e seriedade (as colecções nunca foram diagnosticadas) a extensão do problema que têm entre mãos. Caminham mais depressa do que gostarão de admitir para a realidade virtual e, sem um desígnio colectivo ou nacional, acarinhos projectos de digitalização. Paradoxalmente há mais bibliotecas e arquivos envolvidos em projectos de digitalização do que em projectos de microfilmagem. Voltarei a esta situação contraditória tentando demonstrar como as bibliotecas e arquivos dispõem das ferramentas-base que permitiriam com solidez e sucesso a organização e execução dum Plano Nacional de Microfilmagem.

As conclusões da Conferência de Bellagio transmitem a experiência americana aos europeus em geral mas, no caso português, esse conhecimento prático ajusta-se como uma luva. E passo a citar em tradução livre:

[...] a cooperação entre bibliotecas autónomas e muito orgulhosas da sua história é particularmente difícil como é absolutamente imprescindível associar os objectivos e uma visão a longo prazo. A enorme força do hábito e da tradição fazem com que os investigadores e os bibliotecários resistam ao microfilme como uma forma de acesso e fá-los tentar a todo o custo manter os livros fragilizados nas estantes. Estes hábitos antigos persistem apesar de se ter tornado evidente que os investigadores exigem cada vez mais um maior acesso embora não possam viajar como o faziam de biblioteca para biblioteca e a falta de espaço nas bibliotecas seja cada vez mais agudo. [...] a estratégia assenta, portanto, em três princípios: 1º separar (os livros) em condições de tratar daqueles que já não o estão; 2º encorajar a cooperação e a partilha de responsabilidade e 3º assegurar um amplo acesso aos documentos preservados³.

Livros ou jornais, será que esta dicotomia interessa assim tanto? Talvez interesse, mas então precisamos de discutir e chegar a um consenso na base muito simples de que não poderemos fazer tudo ao mesmo tempo e que essa tentativa – estulta – corresponderia ao aborto deliberado de qualquer projecto.

Microfilmagem e digitalização como processos técnicos

Microfilmagem e digitalização são efectivamente as duas alternativas que se colocam habitualmente quando se discute a transferência do

³ Cf. *Bellagio Conference*.

conteúdo intelectual dos documentos. Regra geral, de forma autónoma. O seu cruzamento como solução alternativa, uma terceira via, é uma consideração recente e voltaremos a ela.

A microfilmagem é uma técnica de miniaturização sobre filme, de preferência de 35 mm, de qualquer documento existente em suporte de papel com o objectivo de preservar o seu conteúdo intelectual para o futuro. Trata-se duma técnica com largos anos de prática e investigação tendo em vista o seu próprio melhoramento, perfeitamente regulamentado, alicerçado num conjunto de normas internacionais que nos dão uma tremenda garantia de qualidade e fiabilidade. Sendo produzido e mantido em condições óptimas, o microfilme garante uma vida longa de vários séculos⁴.

A estes princípios gerais que definem o microfilme é absolutamente indispensável, para bibliotecas e arquivos, acrescentar o novo conceito de «microfilme de preservação». Um «microfilme de preservação» resulta do conjunto de actividades de selecção, preparação, produção, controlo de qualidade, controlo bibliográfico e armazenamento levadas a cabo para preservar o conteúdo intelectual de espécies deterioradas, instáveis ou estragadas, fornecendo uma cópia de segurança ou uma cópia para uso reduzindo o manuseamento das espécies originais. A produção do «microfilme de preservação» prevê a utilização de matérias-primas ou o recurso a métodos que têm uma longevidade máxima e prevê a produção duma matriz em filme negativo que será armazenada num depósito sob rigorosas condições de ambiente e apenas utilizado para produzir uma cópia de reprodução^{5, 6}.

As vantagens do microfilme facilmente identificáveis são os custos previsíveis com baixo custo de produção e de manutenção; o baixo custo do equipamento necessário à sua leitura; o facto de não oferecer riscos; o reconhecimento do seu valor legal e de arquivo; a enorme compactação que permite relativamente à documentação original (cerca de 98%); a segurança facilitada pelo pequeno volume; a normalização fixa. Como não há bela sem senão, eis uma lista de aspectos menos apelativos: a lentidão na pesquisa; a pesquisa sequencial; a utilização em monoposto; a rápida deterioração devido às sucessivas cópias; a consulta incómoda e fatigante.

⁴ Maria Luísa CABRAL, «Para uma estratégia em Preservação e Conservação na Biblioteca Nacional», *Leituras: Revista da Biblioteca Nacional*, Lisboa, (2) Primavera 1998, p. 199-208.

⁵ *Preservation microfilming: a guide for librarians and archivists*. Edited by Lisa Fox. 2nd ed. Chicago; London: American Library Association, 1996.

⁶ *RLG Preservation microfilming handbook*. Edited by Nancy E. Elkington. Mountain View, CA: RLG, 1992.

A microfilmagem, como ferramenta ao serviço da preservação, tem como objectivo reduzir o manuseamento do próprio documento e como substituto do documento tem papel relevante nas políticas de difusão e acesso à informação. Na execução do microfilme há que garantir a produção de três filmes com destinos diferentes: um filme original negativo (a matriz), em saís de prata, que ficará reservado para sempre, em casa forte climatizada em permanência e de preferência fora da instituição detentora das espécies bibliográficas microfilmadas; uma cópia (filme de 2ª geração) em prata, positivo, a considerar como *backup* a partir do qual se podem fazer reproduções a pedido e uma outra, em diazo positivo, que será o exemplar para consulta e leitura local. A cópia em diazo poderá estar guardada nas próprias salas de leitura e não requer os mesmos cuidados de climatização que a matriz em saís de prata.

Todos os documentos gráficos podem ser microfilmados. Para o serem, os de grande dimensão, como os jornais, têm de ser preparados: primeiro, «desmanchados» se estiverem encadernados, isto é, as pastas e a lombada têm de ser separadas do miolo; depois, os cadernos separados por folhas; e se o estado físico o exigir, será preciso uma intervenção laboratorial, a qual pode ser mais ou menos intensiva, antes da própria microfilmagem ou uma simples engomadela para obstar a eventuais vincos e dobras do papel. Assim, a microfilmagem não se apresenta como uma acção imediatista. A microfilmagem exige uma ponderação espécie a espécie e uma preparação individualizada. Também após a microfilmagem ainda se pode pôr o problema da reencadernação ou não dessas espécies. As mais das vezes, a opção é pelo armazenamento das espécies devidamente acondicionadas sem voltarem a ser reencadernadas (se era o caso) porque é economicamente mais rentável e, assim como assim, o documento de substituição já está garantido. Há bibliotecas que optam por eliminar os jornais uma vez microfilmados. Trata-se de solução extrema apenas a colocar quando o espaço é de todo inexistente ou quando a biblioteca não tem, de facto, obrigação patrimonial.

Uma vez que a microfilmagem das obras as danifica, e basta lembrar o esforço a que a encadernação dos livros é submetida mesmo se as máquinas estiverem providas de «berço», convém assegurar que o microfilme seja executado e guardado segundo todas as regras e cuidados de forma a que a espécie microfilmada possa ser arrumada com carácter definitivo já que um documento microfilmado, por princípio, nunca mais vai à leitura.

A outra alternativa técnica para a transferência de suporte do conteúdo intelectual dos documentos é a digitalização. Fugindo a

uma definição demasiado técnica, adiantaria que se trata de um processo que permite a «fotografia electrónica» capturada directamente do documento original. Uma imagem digital pode com exactidão proporcionar a informação, a disposição e a apresentação do documento original, incluindo os caracteres tipográficos, as anotações, os defeitos quaisquer que eles sejam, e as ilustrações. Uma imagem digital é um conjunto de pontos ou elementos picturais (*pixels*). Quanto maior for o número de pixéis, maior a resolução que se consegue. Cada pixel tem um valor tonal (preto, branco, tons de cinzento ou cor) e é representado digitalmente por um código binário (*bits*). Estes *bits* estão armazenados em computador numa ordem sequencial e a qualidade das imagens digitais depende da resolução na captura inicial. As imagens digitais armazenadas em ficheiro não podem ser pesquisadas só por si, exigindo a criação duma base de dados.

Bastará esta introdução para facilmente se perceber a sofisticação do processo e, sofisticado como é, faz-se acompanhar por um conjunto de vantagens, a saber:

- 1º fidelidade ao original;
- 2º manutenção da qualidade inicial independentemente do número de cópias;
- 3º distribuição em rede com uso múltiplo;
- 4º utilização da imagem «à medida».

Contrapondo a estas características positivas, a digitalização enfrenta:

- 1º a rápida evolução tecnológica;
- 2º as restrições legais incluindo as questões relacionadas com os direitos de autor;
- 3º a falta de normalização;
- 4º os requisitos para a migração;
- 5º a falta de apoios institucionais;
- 6º a dificuldade de armazenagem;
- 7º a falta de apoio e estabilidade por parte das firmas.

Lista pesada esta de desvantagens para instituições de carácter cultural! Mesmo que o único objectivo da digitalização seja o de facilitar e alargar o acesso à informação, a verdade é que muito trabalho de preservação fica feito pelo caminho. Enquanto passamos a suporte electrónico alguma informação com a preocupação de divulgar documentos desta biblioteca ou daquele arquivo, podemos ao mesmo tempo, executar o plano de microfilmagem cobrindo núcleos

igualmente importantes. A nossa atenção deverá, assim, ter em consideração esta simultaneidade temporal evitando sobreposições e mesmo duplicações no levar à prática uma estratégia de preservação. A consideração da digitalização apenas para preservar pode, no entanto, ser problemática. E cito:

[...] many university libraries have been creating or collecting digital information in a range of standard and proprietary formats. Each of these formats continues to evolve, becoming more complex as revised software versions add new features or functionality. It is not uncommon for software enhancements to «orphan», or leave unreadable, files generated by earlier versions. This threat to digital information of a certain age has surpassed the immediate danger of unstable media or obsolete hardware. The most pressing problem confronting managers of digital collections is now software and data-format obsolescence⁷.

Mas já não o é quando se pensa no acesso. Reside aí a enorme diferença entre a digitalização e o microfilme e a gigantesca vantagem daquela sobre este. Com os dados disponíveis em suporte electrónico a distribuição em rede é possível e, assim, recuar perante esta oferta pareceria uma aberração. É esta questão que deve ser particularmente tida em atenção e esta pode ser a questão decisiva.

Dois tempos, dois lugares

A enumeração das vantagens dum e doutro processo torna claro como eles se adaptam, igualmente, à transferência de suporte. Até aqui, os propósitos de cada solução são idênticos e não fossem outras as preocupações, o problema seria apenas de coexistência, possivelmente a desempatar por simples motivos financeiros.

A verdadeira questão que se equaciona, porém, prende-se com a durabilidade da informação nos novos suportes alternativos. Indissociável da durabilidade, põe-se o problema da utilização dessa informação, assim sob a nova roupagem. Trata-se de preocupações mais ou menos disfarçadas pelo enunciado das características respectivas. Assim, o reconhecimento do valor legal e de arquivo do microfilme indica implicitamente um longo tempo de vida; considerar o microfilme como um processo técnico normalizado é atribuir-lhe um nível de amadurecimento técnico logo, uma grande

⁷ James M. MORRIS, «Cornell Project will assess risks of migration strategy», *CLIR ISSUES*, (4), Jul-Aug. 1998, p. 1, 4.

estabilidade. Por seu lado, a digitalização garante uma fidelidade invejável ao original – na definição, na cor, nas manchas, na visualização das características do papel. Pelo contrário, a sua componente electrónica, amarra a digitalização à rápida evolução tecnológica, e a turbulência inevitável instala-se. De repente, a cisão. Em todos os aspectos relacionados com a armazenagem, o microfilme adianta-se, senhor da situação; por oposição, em todos os aspectos relacionados com a difusão, a digitalização recolhe maior votação. Qual ciclo de vida, a digitalização denota uma dinâmica e energia impensáveis na microfilmagem em plena idade madura, tranquila para quem a mudança não é muito querida.

A conclusão para a qual vamos caminhando coloca, dum lado, a microfilmagem como solução eleita para a preservação e, do outro lado, a digitalização como forma privilegiada para a comunicação e o acesso. Esta dicotomia, pesando custos e equipamento, durabilidade e transportabilidade da informação, flexibilidade de manuseamento e qualidade de imagem, vai-se tornando muito nítida. Sem mágoa nem ressentimento. Como, aliás, foi dito e repetido recentemente nas intervenções proferidas por Peter Fox (Cambridge University Library) e Daniel Renoult (Bibliothèque nationale de France) na Conferência da LIBER⁸.

Uma solução mista

As bibliotecas e arquivos estão aí para guardar (preservar) e para difundir (facultar o acesso). Embora se reconheça um certo conflito de objectivos, está-se longe de uma real incompatibilidade. Curioso constatar que idêntica complementaridade se aplica à microfilmagem e digitalização. Contudo, mais do que a simples coexistência, aquilo que a microfilmagem começa pode, com benefício, ser terminado pela digitalização.

Os microfilmes podem servir de base para a digitalização. Ou seja, podem executar-se microfilmes para preservar e, a partir deles, executar a digitalização para o acesso em rede. Esta parece ser uma solução consensual pelo menos enquanto não se tem uma certeza absoluta sobre a durabilidade do suporte electrónico.

Designando-a como «solução híbrida», Willis analisa pormenorizadamente as soluções em jogo, confronta vantagens e desvantagens, para concluir de forma inapelável que:

⁸ «Short-term, long-term strategies: conflict or harmony?», *LIBER Annual Conference*, Paris, 30 June - 4 July 1998.

*[...] however filming first, and creating digital images by selectively scanning the film seems to be the least risky current preservation option provided that appropriate attention is paid to indexing the filmed collection*⁹.

E mais adiante, no mesmo relatório:

*[...] today, according to a traditional definition film is the only truly archival medium. It will not become obsolete in the foreseeable future. Optical disk will be viewed as the permanent, low-cost, removable, random access storage media. Magnetic products (tape and disk) will continue to increase in storage capacity and reliability while decreasing in cost. Magnetic disk will provide temporary working storage for all work-in-process on all future image systems. Optical tape, too, bears storage for all work-in-process on all future image systems. Optical tape, too, bears watching. In configuring the ideal image storage system, the knowledgeable designer will construct a hierarchy of storage that takes advantage of the strength, access characteristics, longevity, and cost of each storage product to produce the greatest benefit at the least cost*¹⁰.

O desafio está, pois, lançado. Dum lado, o microfilme com o seu registo exacto e sequencial; do outro, o suporte electrónico a permitir uma informação mais à medida do utilizador, liberto da própria ordem documental. Enquanto aguardamos pela indispensável normalização tecnológica, podemos criar os suportes electrónicos que entendermos de acordo com as audiências a que se destinam. Estes suportes electrónicos assim concebidos têm um tempo de vida limitado, mas importa isso perante a mais valia que uma informação rica e variada representa junto do público? Sem esquecer evidentemente que durante esse tempo, se pôs, de facto, em marcha um verdadeiro programa de preservação preventiva.

*While most librarians will agree that digital surrogates play a role in preventive preservation by allowing patrons to use fragile or poorly preserved items to which they might not normally have unrestricted access in a reading room, there is no consensus about the role of digital files in preserving the information contained in the original. And that is likely to remain the case as long as standards have not been agreed upon and the technology continues to evolve*¹¹.

Porque, afinal:

The challenge of the future tense for preservation professionals is to continue to look beyond the object to the medium, and beyond the medium to the creator and the user, and embrace responsibility for longterm custody of all forms of recorded information to ensure continued access to them. The challenge for library and

⁹ Don WILLIS, *A hybrid systems approach to preservation of printed materials*. Washington, DC: The Commission on Preservation and Access, 1992, p. 16.

¹⁰ Don WILLIS, *A hybrid systems approach to preservation of printed materials*. Washington, DC: The Commission on Preservation and Access, 1992, p. 29.

¹¹ Abby SMITH, «Digitization prompts new preservation-and-access strategies», *CLIR ISSUES*, (1), Jan-Feb. 1998, p. 1, 7.

*archival managers is to understand the changing nature of their institutions' assets and provide the appropriate training and support for the staff who are charged with the survival of those assets*¹².

Um Plano Nacional de Microfilmagem: problema colectivo

No âmbito da transferência de suporte, para as bibliotecas e arquivos portugueses falta fazer muita coisa. A primeira que aqui se pode enunciar prende-se com o conhecimento exacto da condição física das colecções de documentos gráficos: livros, jornais, desenhos, cartografia ou iconografia. Não basta ter uma noção de que a situação é grave e perigosa. Há que saber com clareza afirmar qual a percentagem de documentos a precisar de intervenção e saber distinguir esta parte daquela que já não está em condições de sofrer uma intervenção. Saber quantos documentos para propor intervenções quantificadas do ponto de vista financeiro. Saberiam os cidadãos, saberiam as entidades públicas. Na etapa seguinte, a responsabilidade passa a ser muito mais de quem decide e distribui os orçamentos do que daqueles que diariamente têm de gerir as colecções ou o que resta delas. Para além dos números, urge também saber destrinçar os documentos: os que valendo como artefactos exigem ser preservados, os que mesmo indo ao laboratório de restauro pelo seu significado exigem que o conteúdo intelectual seja transferido, e os outros que deverão, ou poderão, sofrer um tratamento de desacidificação em massa.

Cumprir um programa assim tão global parece impensável pelos recursos humanos, técnicos e financeiros subjacentes. Então, há que escolher e essa terrível opção cabe aos bibliotecários e arquivistas.

Entre as muitas etapas a percorrer aquela que parece mais urgente, logo a seguir à caracterização das nossas colecções, tem a ver, na minha opinião, com a cooperação a estabelecer entre as várias bibliotecas portuguesas (omito aqui deliberadamente os arquivos porque sendo a natureza dos fundos diferente a cooperação deverá colocar-se noutros termos). A que colecções é que vamos garantir a transferência do conteúdo intelectual? Quem tem o quê? Quem faz o quê? Estas são questões de política de gestão de

¹² Abby SMITH, «Preservation in the future tense», *CLIR ISSUES*, (3), May-Jun. 1998, p. 1, 6

colecções e, possivelmente, serão necessárias muitas reuniões e encontros para se conseguir um entendimento terminológico, um consenso quanto aos fins, uma partilha de responsabilidades, uma divisão de funções.

Sentimos todos que é urgente avançar para uma reunião e discutir o que queremos fazer, se vamos recorrer à digitalização ou à microfilmagem. Como é que vamos fazer, onde é que arranjam os financiamento? Estas são questões difíceis mas não nos devem desmoralizar porque também temos factos positivos neste cenário. Quais são eles? A existência duma base de dados com implantação nacional, reflectindo já um bom quinhão do património bibliográfico que por aí existe; um inventário do património bibliográfico em marcha e que pode muito bem acrescentar-se às existências da PORBASE; uma legislação em fase de ultimate e que, a médio prazo, poderá vir a contemplar estas preocupações sobre a transferência do conteúdo intelectual dos documentos. A seguir, e não menos importante, as questões técnicas. Como se faz? Quem faz? Quem sabe fazer? Claro que são todas questões ponderosas que vão envolver técnicos e políticos por largo tempo mas esta não pode ser a desculpa para ignorarmos o problema. Se o primeiro passo não partir dos profissionais directamente relacionados com as colecções, quem é que vai sugerir uma acção imediata, concertada e sistemática que contorne o problema da deterioração do papel? A mim parece-me que podíamos, nós profissionais, fazer daquilo que nos enriquece, ainda que fragmentado, um bom motivo para nos fortalecermos. Já vai sendo tempo.

G.I.S.A.
Gestão Integrada de Sistemas de Arquivo
Um programa de gestão
da cadeia de procedimentos arquivísticos

MANUEL LUÍS REAL

O Arquivo como (sub)-Sistema de Informação

Qualquer estrutura organizacional gere informação, seja no desenvolvimento da sua actividade, seja na própria administração dos recursos informacionais que lhe servem de suporte. O circuito administrativo dos documentos, o arquivo, a biblioteca ou centro de documentação, as bases de dados de gestão e produção, a rede de comunicações, os postos de atendimento, etc. poderão ser partes integrantes de todo um (*macro*)-*Sistema de Informação* ao serviço de determinado organismo, independentemente da natureza e dos fins a que este se destina. Qualquer instrumento daquele tipo constituiu-se, por conseguinte, como um (*sub*)-*Sistema* da própria organização e é dotado de finalidade específica.

A gestão da actividade de uma empresa ou serviço recorre a documentos – tanto de origem interna, como externa – os quais circulam e se estruturam de acordo com as necessidades de testemunho, de prova ou de mera difusão informativa. A informação neles contida é assim sistematizada para seu próprio uso (ou para benefício de terceiros a quem parte da sua actividade se destina), dando formalmente origem ao respectivo *Arquivo*. Um arquivo desenvolve-se em simultâneo com a actividade do organismo produtor e, por conseguinte, vai-se estruturando naturalmente, tendo em conta as características funcionais e a complexidade progressiva do criador. Ele é apenas uma das unidades que compõem a rede de serviços do produtor, mas, ao contrário da generalidade dos seus outros sub-sistemas (informacionais ou não) o arquivo não pode deixar de reflectir o próprio universo sistémico da entidade que o criou. De facto, incumbe-lhe conservar e controlar toda a documentação

reunida durante o ciclo vital da entidade produtora e que foi sendo estruturada de acordo com a orgânica dos seus serviços e as funções por estes desempenhadas.

Numa perspectiva integrada dos arquivos há que distinguir, por sua vez, duas componentes sistémicas, através das quais podemos encará-los e que não devem ser confundidas:

- 1 – a estrutura da documentação e o seu sistema informacional;
- 2 – a estrutura do(s) serviço(s) de arquivo e o seu sistema funcional.

Uma das propriedades do arquivo, associada à *função* do organismo que o produziu, faz dele o espelho do próprio sistema de funcionamento do criador, que aí concentrou os dados que no decurso do tempo fundamentara a sua acção. A outra característica do arquivo, ligada especificamente ao *serviço* de organização dos documentos, tem antes a ver com o próprio controlo do sistema e o acesso à informação. Contudo, «função» e «serviço» não podem entrar em contradição com os princípios gerais do sistema informacional do criador, pelo que terão de se submeter a regras de gestão integradas.

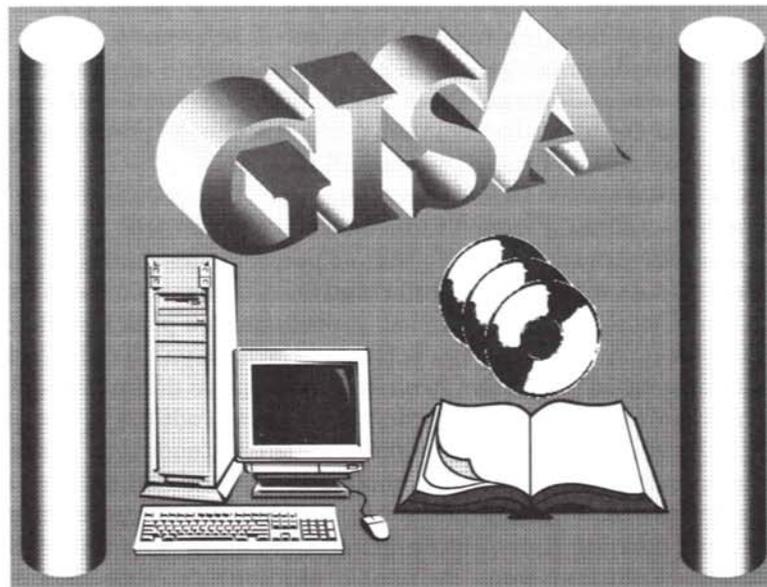


Fig. 1 – Logotipo do programa de Gestão Integrada de Sistemas de Arquivo (GISA)

A concepção de um modelo de Gestão Integrada de Sistemas de Arquivo

O desenvolvimento do programa de Gestão Integrada de Sistemas de Arquivo (GISA) só se tornou verdadeiramente possível com os recentes progressos experimentados no âmbito da teoria arquivística, a nível internacional.

A consistência de uma aplicação destinada a gerir um qualquer sistema organizacional, mais do que pelas ferramentas que utiliza, avalia-se sobretudo pela qualidade da relação que estabelece com a natureza do objecto. Ora, só mediante uma sólida base teórica é possível conhecer a estrutura íntima dos sistemas e, daí, partir para um modelo capaz de gerir a sua própria dinâmica. Nas duas últimas décadas, a Arquivística conheceu grandes progressos, que a levaram a uma nova consciência sobre as suas raízes e sobre a necessidade de corrigir os erros passados, que conduziram à fragmentação dos sistemas de arquivo. Na sequência deste movimento multiplicaram-se os estudos teóricos e foi já possível conceber uma metodologia, aceite internacionalmente, para analisar e representar qualquer unidade arquivística, caracterizando-a nos seus diversos aspectos e referindo-a ao contexto de origem.

A *normalização* dos procedimentos encontra-se, de facto, na primeira linha das preocupações dos arquivistas. Trata-se de um processo ainda em fase de amadurecimento, mas é desde já possível experimentar aplicações coerentes destinadas a descrever os arquivos e a recuperar a informação. A criação de uma norma geral foi um passo decisivo. Hoje trabalha-se no seu aperfeiçoamento e na preparação de directivas para que ela possa ser universalmente aplicada, mesmo tendo em conta as especificidades de alguns tipos de suportes e de documentos.

Mas estas questões são apenas parte do problema. Uma faceta essencial do trabalho arquivístico será aquela que diz respeito à interpretação dos próprios acervos, quando deslocados do seu *habitat* original, assim como a correcta articulação entre as categorias orgânico-funcionais identificadas e os níveis de descrição previstos pela norma.

Outro aspecto, não menos importante, é o que concerne à própria gestão do sistema informacional representado pelo fluxo de documentos que é necessário controlar. O problema da *gestão* de um Arquivo não se limita ao da mera pesquisa da representação descritiva e das formas de acesso do respectivo acervo. De facto, ele tem uma historicidade própria e, tanto pelo contexto da produção, como

pelos mecanismos de controlo funcional, um arquivo está longe de ser uma realidade estática. Hoje em dia, numa organização com certo grau de desenvolvimento, a sua gestão passa pela responsabilidade de vários serviços. Entendido como sistema de informação e um recurso do organismo produtor, o Arquivo é uma *entidade unívoca e integrada*, independentemente da maior ou menor complexidade dos serviços que o suportam. A organização, exploração e manutenção dos sistemas de arquivo depende frequentemente de uma rede diversificada de serviços e de uma cadeia de operações com responsabilidades repartidas, se bem que todas elas orientadas pelo objectivo comum de controlar o próprio sistema de informação. A experiência mostra-nos que há muitas vezes deficiente comunicação entre serviços e que as suas práticas não se encontram articuladas, dando origem a instrumentos de controlo nem sempre compatíveis entre si. Isto tem originado graves consequências para a quebra de unidade dos sistemas, com prejuízo das próprias organizações em fase activa e do entendimento que delas se fará, *a posteriori*, ao consultar a respectiva documentação.

O aprofundamento do conceito de Arquivo torna cada vez mais evidente a necessidade de *controlo integrado do sistema de informação* que lhe está subjacente. Mesmo que um arquivo se torne em subsistema de um novo organismo, em consequência da sua incorporação num centro de recolha, após o término do respectivo ciclo vital, é indispensável respeitar a sua estrutura operativa interna ou funcionalidade original, facto que só acontecerá se houver uma transferência simultânea de dados que garantam a reconstituição do respectivo fluxo informacional. Todos os procedimentos de controlo e de transferência de dados num arquivo – seja enquanto sistema de informação ao serviço do produtor, seja como sistema desactivado – podem ser convertidos num modelo adaptável ao que chamaremos a *cadeia arquivística* de gestão.

Na nossa perspectiva – e porque temos responsabilidades na gestão de um arquivo municipal com certa dimensão, onde estão representadas as diversas fases do ciclo vital dos documentos – acaba por ser insuficiente a escolha de um modelo destinado apenas para descrever massas documentais acumuladas ou, meramente, para gerir sistemas desactivados. Para os serviços em actividade plena é indispensável criar uma metodologia de gestão que seja capaz de controlar todas as fases de funcionamento da cadeia arquivística.

A actividade global de um serviço de arquivos pode, de alguma forma, ser entendida à semelhança da cadeia de produção de uma



Fig. 2 – Cadeia Arquivística. Operações Principais

fábrica, onde existem diversas operações com especificidade própria e que se completam entre si (fig. 2). Tal como numa empresa robotizada será possível encontrar um modelo em que a gestão do trabalho produtivo possa efectuar-se através da transferência automática de dados, de uma fase para outra. Estes serão manipulados em ordem ao cumprimento das funções específicas de determinada fase da cadeia arquivística, cabendo a cada uma seleccionar informação ou acrescentar-lhe mais-valia informativa sobre o sistema, de acordo com a operação desencadeada.

No estado actual de desenvolvimento normativo e, sobretudo, tendo em conta a situação concreta dos acervos e da própria administração portuguesa, achamos que só seria possível conceber um modelo consistente e eficaz, se aplicado aos serviços centrais de arquivo. São os serviços vulgarmente designados por «arquivo central administrativo» ou «arquivo geral», «arquivo intermédio» e «arquivo histórico», embora para o presente caso a designação seja pouco relevante, dado que o que está em causa são as operações arquivísticas em si e não o grau de usabilidade ou o local de instalação dos documentos.

Numa situação teoricamente ideal poder-se-ia pensar num modelo informático que controlasse também a documentação ao nível dos arquivos correntes e cuja aplicação já pudesse estar integrada na estrutura global do sistema. Mas não deixa de ser uma posição

geradora de alguma controvérsia. Apesar do avanço da perspectiva sistémica dos arquivos continua a ser praticamente impossível preparar uma aplicação «universal» ao nível da gestão administrativa corrente – isto é, aplicável a todos os serviços ou empresas – posto que aqui há grande especificidade de objectivos, de circuitos e de instrumentos de controlo. Não se pode também esquecer que de administração para administração – inclusive até entre organizações do mesmo ramo – existem grandes diferenças quanto ao nível de desenvolvimento e quanto aos meios disponíveis. Assim, é difícil encontrar um modelo capaz de ser adaptado a qualquer tipo de organismo em fase activa. Quando a documentação produzida se estabiliza e entra num circuito centralizado, cujo paradigma é extensível a qualquer sistema de informação, já é possível introduzir uma filosofia de gestão integradora e universalizante.

Atendendo ainda a que, em muitos casos, a produção e circulação de documentos administrativos se encontra deficientemente planificada, achamos preferível deixar maior flexibilidade à fase administrativa e recomendar, sempre que possível, aplicações compatíveis. Aos arquivos centrais cabe criar condições para poderem absorver com facilidade e eficiência a documentação corrente, e não geri-la, enquanto tal. O modelo aqui concebido não deixará de fazer a devida articulação com os arquivos correntes, através das operações de recenseamento e de incorporação. Por isso mesmo, o GISA tem em conta a gestão dos dados a partir do momento em que se inicia o processo de recenseamento dos documentos. Nesta ocasião estar-se-á em condições de produzir listas de incorporação e, logo a seguir, completar o seu conhecimento com a avaliação e controlo das espécies incorporadas. A execução da triagem e dos autos de eliminação assenta necessariamente naquelas primeiras operações, sendo possível a partir de então controlar o destino dos documentos. Seguir-se-á a fase de aperfeiçoamento dos instrumentos de acesso, através da inventariação e catalogação das unidades arquivísticas, numa escala de prioridades devidamente definida. A recuperação e difusão da informação rematam o ciclo arquivístico, embora o sistema necessite ainda do apoio de operações de controlo da qualidade, no domínio do tratamento dos dados e dos documentos.

Havendo-se chegado internacionalmente a um consenso quanto às regras básicas de descrição dos documentos é possível agora chegar a um modelo de gestão da cadeia arquivística, assente no controlo transversal dos dados que a norma aponta como estruturantes para a identificação das unidades arquivísticas.

O nascimento da norma ISAD(G) e as consequências para o GISA

Os arquivistas, durante muito tempo, tiveram dificuldade em sistematizar categorias de representação e de lhes dar uma forma capaz de descrever homoganeamente qualquer arquivo e seus componentes. Lidando com conjuntos de documentos dotados de individualidade própria, os profissionais do sector balanceavam entre descrições casuísticas e listagens analíticas, sem conseguir verdadeiramente afinar critérios para a caracterização das unidades arquivísticas. A relação com a estrutura da entidade produtora nem sempre era estabelecida e, de um modo geral, não existiam meios de representar consensualmente o universo sistémico de que tais documentos faziam parte.

Será Michael Cook, em 1985, quem pela primeira vez defende a ideia da introdução de um modelo único de descrição hierárquica, capaz de identificar qualquer conjunto documental, articulando-o com os demais níveis do sistema a que ele mesmo se reporta¹. Partindo desta base e, também, de experiências realizadas em diversos países no domínio do tratamento da informação arquivística², foi criada no âmbito do Conselho Internacional de Arquivos, em 1990, uma Comissão *Ad Hoc* incumbida de preparar uma norma de descrição multinível.

¹ A divulgação do seu modelo em Portugal foi feita pelo próprio, no 1º Congresso da BAD, logo em 1985: COOK, Michael – «The Manual of Archival Description: recent research in archival description in the United Kingdom», CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 1º, Porto, 1985 – *A Informação em tempo de mudança: actas*. Porto, BAD, 1985. Vol. 1, p. 409-415. O *Manual of Archival Description* surgirá em 1986. A edição revista do M.A.D. data de 1989 e conta já com a colaboração de Margaret Procter.

Embora as propostas de Michael Cook constituam o ponto de partida para a actual metodologia de descrição arquivística, pode-se considerar que teve um precursor no americano Oliver W. Holmes, o qual vinte anos antes publicava já no *American Archivist* (vol. 27, Jan. 1964) um artigo intitulado «Archival arrangement: five different operations at five different levels».

² Referimo-nos nomeadamente aos contributos do americano Steven Hensen e do Grupo de Trabalho Canadano sobre as Normas de Descrição: HENSEN, Steven – *Archives, personal papers and manuscripts: a cataloguing manual for archival repositories, historical societies and manuscripts libraries*. 2ª ed. Chicago: Society of American Archivists, 1989; BUREAU CANADIEN DES ARCHIVISTES. Groupe de Travail sur les Normes de Description en Archivistique – *Les normes de description en Archivistique: une nécessité*. Ottawa: B.C.A., 1986.

Para a vizinha Espanha, salientaremos também o ensaio de Antonia Heredia para sistematizar alguns princípios básicos de descrição arquivística: HEREDIA HERRERA, Antonia – *Manual de instrumentos de descripción documental*. Sevilha: Diputación Provincial, 1982.

Em 1994 acabaria por ser finalmente aprovada a *Norma ISAD(G)*, a qual significou um salto qualitativo de repercussão incalculável³. Pela primeira vez, os organismos produtores de documentos – assim como os serviços receptores de arquivos – passaram a dispor de um instrumento normativo capaz de representar qualquer unidade arquivística, seja qual for a sua posição dentro do sistema de que faz parte.

No Congresso de Pequim, o grupo que havia preparado a norma transformou-se em Comité especializado do C.I.A., ficando encarregado de concluir um outro documento, sobre o controlo de autoridade para nomes de entidades colectivas, pessoas e famílias: *Norma ISAAR (CPF)*⁴.

Estas duas normas encontram-se hoje em revisão e com tendência para se fundir num único texto regulador. Contrariamente à intenção inicial, há também a ideia de não desenvolver normas específicas para documentos em suporte especial. Atendendo a que um mesmo arquivo pode congrega unidades arquivísticas com documentos de tipo diverso, achou-se mais correcto aperfeiçoar uma norma universal, dotada de instruções sobre o preenchimento dos campos descritivos, de modo que estes possam cobrir as especificidades de cada tipo de suporte ou documento.

A tendência será, cada vez mais, para alargar o âmbito de aplicação da norma. Na sessão do Sub-Comité do Conselho Internacional de Arquivos para a elaboração de Instrumentos de Pesquisa, que se realizou em Lisboa a 15 de Maio de 1998, um dos seus membros, o sueco Jan Dahlin, sublinhou a relação entre «gestão de documentos» e «arquivos definitivos», adiantando que a norma visou sobretudo a descrição destes últimos⁵. Acrescentou ainda que a tradição do seu país vai no sentido de haver uma única profissão, sendo necessário que a norma se possa aplicar às duas vertentes.

Ora, é esta precisamente a filosofia do projecto GISA. Ela foi apresentada pela primeira vez em 1995, em concurso para financiamento através de fundos comunitários, no quadro do *Programa PRAXIS XXI*. Esta candidatura terá sido a única que se inseria no

³ INTERNATIONAL COUNCIL ON ARCHIVES. Ad Hoc Commission on Descriptive Standards – ISAD(G): General International Standard Archival Description. Ottawa: I.C.A., 1994.

⁴ INTERNATIONAL COUNCIL ON ARCHIVES. Ad Hoc Commission on Descriptive Standards – ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families. Draft. Ottawa: I.C.A., 1994.

⁵ O autor querer-se-ia referir à dicotomia existente, em certos países, entre os «archives» e os chamados «records». Nesta perspectiva, o «records management» incluiria também a gestão dos próprios arquivos centrais administrativos.

plano da investigação pura, no âmbito da disciplina da Arquivística. Porém, não veio a ser contemplada com qualquer financiamento⁶. O júri terá sido mais sensível a temas ligados com a pesquisa histórica e a projectos dirigidos preferencialmente à recolha de fontes, ao *scanner* de documentos, etc.

O GISA não se assumia como um projecto meramente auxiliar da investigação histórica e, por isso, foi na altura penalizado. Como projecto arquivístico autónomo, não podia deixar de se inserir também na área da administração e, conseqüentemente, os benefícios para o desenvolvimento da historiografia são apenas aqui indirectos e a médio prazo.

Esta aparente fragilidade da candidatura não deixou de se revelar, contudo, como um dos aspectos mais sólidos do projecto e a razão de ser da sua sobrevivência para além do concurso. Partindo do interior de entidades ligadas à prática da Arquivística – mas também com actividade de pesquisa no âmbito da disciplina – o GISA vinha ao encontro das preocupações teóricas do momento e correspondia às necessidades elementares dos serviços, em vista à sua própria modernização. Quando se tomou conhecimento do resultado do concurso já o projecto prosseguia em velocidade de cruzeiro. A experiência acumulada no seio das instituições promotoras permitiu-lhes também alguma auto-suficiência, não apenas na fase decisiva de arranque e de concepção global do sistema, mas também no desenvolvimento de uma aplicação informática. A ausência de financiamento externo obrigou a que o processo tivesse de ser um pouco mais lento. Porém, ele encontra-se actualmente numa fase bastante amadurecida, permitindo em breve dar início a uma nova etapa, destinada à sua difusão e, desta vez, com o auxílio de parceiros exteriores na área comercial e no sector da pesquisa em sistemas de comunicações multimédia.

Antecedentes do projecto GISA e seu desenvolvimento

A informatização do Arquivo da Câmara Municipal do Porto⁷ está a ser desenvolvida desde 1988, após um período preparatório

⁶ Por sermos parte interessada, declinamos o convite que nos foi dirigido para integrar as comissões de avaliação do PRAXIS XXI, afastando-nos deliberadamente de qualquer iniciativa que pudesse conduzir a situações menos claras.

⁷ Actualmente, os seus serviços centrais estão integrados num único Departamento de Arquivos, do qual fazem parte a Divisão de Arquivo Histórico e a Divisão de Arquivo Geral.

que conduziu à assinatura de um protocolo de cooperação entre a CMP e o INESC (Instituto de Engenharia de Sistemas de Computadores).

Os objectivos deste primeiro protocolo foram cumpridos, tendo dado origem ao projecto CALÍOPE, cujas características divulgámos em devido tempo⁸. Tratou-se de uma aplicação que visava articular bases de dados de estrutura mais simples com um banco de imagens em disco óptico. Devido ao peso específico que têm para a administração municipal os «Processos de Licenças para Obras Particulares» foi dada prioridade ao tratamento desta série. Foram descritos mais de 20 000 processos e fizeram-se ensaios de digitalização de imagens que se revelaram de grande qualidade. Avançou-se também no domínio da fotografia, ficando descritas e digitalizadas mais de 3000 imagens. Seguindo o mesmo princípio foram ainda concebidas bases de dados para outras séries de interesse prioritário, se bem que, nestes casos, não se tenha passado da fase de teste.

Enquanto se trabalhava no desenvolvimento destas aplicações, o Arquivo Histórico viria a adoptar o programa ARQBASE para prosseguir o *Inventário* geral do seu acervo, já que o CALÍOPE tinha outro objectivo. Aquela nova ferramenta fora concebida pelo então Instituto Português de Arquivos, segundo um avançado modelo de descrição multinível, inspirado no M.A.D. de Michael Cook e Margaret Procter⁹.

Apesar de concretizados os objectivos da 1ª fase do CALÍOPE e de ter ficado concluído o inventário de toda a documentação municipal já incorporada no Arquivo Histórico, o processo de informatização do Departamento de Arquivos da CMP conheceu algumas dificuldades, que vieram a determinar uma mudança de rumo e, inclusive, a definição de objectivos ainda mais ambiciosos do ponto de vista arquivístico.

A digitalização da série de Licenças de Obras Particulares começou a levantar problemas devido ao facto de os processos estarem encadernados e, sobretudo, em consequência de existirem

⁸ Manuel Luís REAL – «O Projecto CALÍOPE (Catálogo e Imagens Ópticas de Pesquisa Expedita)», *Cadernos BAD*. Lisboa, 2 (1992) p. 151-170.

⁹ Ana FRANQUEIRA; Madalena GARCIA – *ARQBASE: metodologia de descrição arquivística para tratamento automatizado de documentação histórica*. [Lisboa]: Instituto Português de Arquivos, 1991.

Para a biblioteca de referência do Arquivo Histórico Municipal do Porto adoptou-se ainda o programa PORBASE, desenvolvido pela Biblioteca Nacional. À semelhança do ARQBASE, esta aplicação corre em ambiente Mini Micro CDS/ISIS.

imagens de grande formato, regra geral em folhas dobradas. O *hardware* disponível então no mercado não respondia directamente a esta questão e a única saída possível seria a digitalização a partir de microfilme. Tal método é eficaz, mas exigiria a encomenda de serviços ao exterior e tornar-se-ia muito dispendioso, dado que a série em causa tem uma dimensão superior a 700 metros lineares de prateleiras.

Problemas de outra ordem começaram a surgir também com a aplicação do ARQBASE. A extinção do IPA fez cair o programa num impasse, com o levantamento de dúvidas a respeito dos respectivos direitos de autor e a completa paralização da pesquisa, na altura ainda em curso, para a melhoria do *software*. No mercado internacional começavam a vulgarizar-se ambientes mais amigáveis e eficazes que o Mini Micro CDS/ISIS, enquanto que, no plano arquivístico, surgia finalmente a norma de descrição ISAD(G). Tudo isto levou à rápida desactualização do ARQBASE, não obstante os serviços que ainda continua a prestar em alguns Arquivos.

Do ponto de vista interno, a situação do Departamento de Arquivos evoluiu também, desta feita para melhor. A Câmara Municipal do Porto havia avançado com o seu Plano Director da Informática (PDI), com o qual a actividade do Departamento passou a articular-se. Por outro lado, a Divisão do Arquivo Geral começara a desenvolver um trabalho sistemático de avaliação de documentos, cuja metodologia havia sido concebida em estreita colaboração com o Arquivo da Câmara Municipal de Vila Nova de Gaia¹⁰. Os dois municípios têm assinado, desde 1991, um protocolo de cooperação arquivística, o qual lhes está a proporcionar condições muito favoráveis para o desenvolvimento de iniciativas de interesse comum e, sobretudo, para troca de informação e de experiências. Os conhecimentos acumulados nesta área estratégica da avaliação de documentos e o trabalho desenvolvido ao nível da fase semi-activa dos arquivos de ambas as instituições foram decisivos para o passo dado em direcção ao GISA.

Perante as dificuldades enunciadas e dispondo agora das orientações de uma norma internacional, os Municípios do Porto e de Vila Nova de Gaia decidiram avançar em Março de 1995 para a já

¹⁰ Maria João Lopes Calheiros de CARVALHO; Alda Maria Pereira Padrão TEMUDO – *Metodologia de Avaliação Documental para as Câmaras Municipais*. Comunicação apresentada numa reunião de trabalho de Arquivistas, promovida pela direcção da BAD no Arquivo Histórico Municipal do Porto em Março de 1995.
– «A Avaliação Documental nos Arquivos Municipais». Comunicação apresentada no IV Encontro Nacional de Arquivos Municipais, Loulé, Fevereiro de 1997.

referida candidatura ao programa PRAXIS XXI, destinado ao financiamento de projectos de investigação científica e tecnológica de pequena e média dimensão.

A candidatura visava criar um produto que permitisse a gestão integrada de um Arquivo, coordenando os métodos de trabalho e a transferência de informação entre todas as fases da cadeia operativa. Assim, procurar-se-ia elaborar um programa que compatibilizasse as diversas operações arquivísticas, desde o recenseamento dos documentos em arquivo corrente até ao controlo do acesso às unidades de instalação pelos usuários do sistema, passando pelas transferências, pela avaliação e triagem dos documentos, pela inventariação do acervo, etc. O projecto viria a receber o nome de GISA, por se referir à *Gestão Integrada de Sistemas de Arquivo*.

Para efeitos do PRAXIS XXI, a candidatura contou ainda com duas outras parcerias: a do INESC (Instituto de Engenharia de Sistemas e Computadores) e da SCANIMAGEM – Tecnologias Digitais, Lda. Esta última firma, que está associada à Microfil, é uma das poucas empresas nacionais especializadas em digitalização a partir de microfilme e a sua participação seria sobretudo ligada ao fornecimento de um gestor de imagens e à execução de serviços. Quanto ao INESC estava-lhe reservado um papel inovador na área do acesso remoto a informação multimédia.

Como a candidatura não veio a ser contemplada dentro do Programa PRAXIS XXI, tais parcerias desfizeram-se. Os Arquivos Municipais do Porto e de Vila Nova de Gaia, que já estavam a desenvolver o produto no âmbito do mencionado protocolo de cooperação, decidiram então continuar o projecto pelos seus próprios meios, dada a importância de que ele se revestia para o desenvolvimento dos respectivos serviços. A sua relativa autonomia, inclusivamente no campo informático, representava uma grande diminuição dos custos de investimento. O projecto tornava-se assim viável, mesmo que por razões operacionais viesse a demorar um pouco mais de tempo.

De qualquer forma, houve necessidade de alterar a estratégia para o desenvolvimento e finalização do GISA. A concepção geral do programa passou a ser feita por uma equipa mais restrita. Quanto ao acordo para o desenvolvimento do *software* de comunicações e para a selecção de um parceiro comercial ficariam diferidos para o momento em que a concepção do produto estivesse amadurecida, no plano arquivístico e na sua aplicação informática. É nessa fase que o projecto se encontra, estando a ser preparado um concurso para a

escolha de uma empresa que finalize e garanta a distribuição do produto¹¹.

Para o desenvolvimento do produto-base, a CMP e a CMVNG foram ainda encontrar novos parceiros com conhecimentos em áreas específicas, que figurarão aí como co-autores. A Universidade do Porto, através do seu Curso de Especialização em Ciências Documentais, assinou um protocolo de cooperação com os promotores do projecto, tendo em vista, nomeadamente, o apoio no âmbito da Indexação e Controlo de Autoridade. Além disso, depois de criado o Centro Português de Fotografia este mostrou-se desde logo interessado em cooperar na pesquisa de um modelo descritivo para os documentos fotográficos. O facto de se tratar do Instituto Nacional que tutela o sector da Fotografia vem atribuir ao produto um valor acrescentado, o que nos leva a salientar, com simpatia, a disponibilidade desde logo assumida pelo CPF¹².

Embora com uma ligeira derrapagem no tempo, devido à impossibilidade de uma dedicação exclusiva ao projecto, apresentamos o Plano de Execução a que nos propusemos para 1998 (fig. 3). Pensamos poder concluir ainda este ano a versão 1.0 e esperamos que o produto possa estar disponível no mercado em 1999. Nesta ocasião estarão já operacionais os módulos básicos do sistema e algumas aplicações para catálogos específicos, que serão opcionais.

A estrutura do GISA

O projecto GISA obedece a um modelo teórico que configura a organização dos serviços de arquivo como um *sistema* coerente de funções e de relações. A cadeia de operações em arquivo, como já se disse acima, pode ser concebida à semelhança da linha de produção de uma empresa e nos seus instrumentos de controlo deve, por sua vez, respeitar o carácter relacional dos documentos, seja qual for a sua idade. No caso de coexistirem diversos arquivos reunidos num

¹¹ O concurso começou a ser estudado a partir de um esboço enviado pelo INESC, já em 1997, e poderá vir a ser conduzido pela SPI – Sociedade Portuguesa de Inovação.

¹² Para além das duas entidades promotoras e dos co-autores de módulos específicos, existem ainda parceiros privilegiados entre arquivos municipais da Área Metropolitana do Porto. À cabeça encontra-se o de Espinho, com o qual já existe um protocolo de cooperação técnica, de efeitos já muito sensíveis ao nível da avaliação de documentos e da formação de quadros administrativos. Este arquivo e muito provavelmente o de Vila do Conde, que também já dispõe de arquivistas profissionais, irão servir de campo experimental para testar o GISA, em paralelo com os autores do projecto.

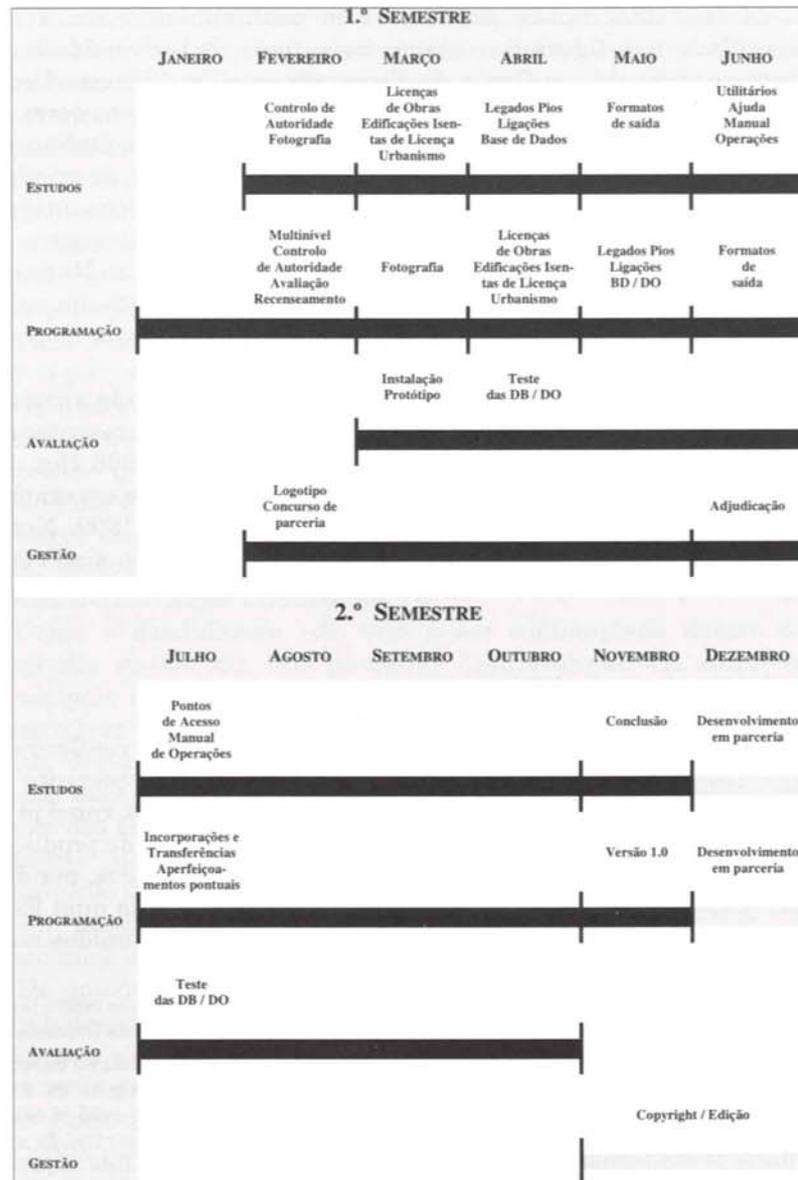


Fig. 3 – Plano de Execução do GISA, para 1998

único serviço, embora sendo de proveniência distinta, o sistema deve estar também preparado para delimitar o respectivo âmbito e para os descrever autonomamente.

O GISA encontra-se adaptado para a *descrição de sistemas desactivados* (arquivos «encerrados»), mas tem a particularidade acrescida de se destinar à *gestão de sistemas activos* (arquivos de administrações em exercício). Para o efeito, foi caracterizada a cadeia de operações arquivísticas ao nível dos serviços centrais, tanto na fase do chamado «arquivo» intermédio ou geral, como no estágio vulgarmente conhecido por «arquivo» histórico. Desta análise surgiu um *modelo relacional* para o conjunto das operações e instrumentos de trabalho em serviços de arquivo (fig. 4).

A ideia-base do sistema de gestão assenta na adaptabilidade da norma ISAD às operações nucleares de análise e de representação das unidades arquivísticas: a *Avaliação* e o *Inventário*. Em princípio, estas poderão corresponder também a dois grandes momentos do ciclo vital da documentação, significando isto que as BD (Bases de Dados) relativas à avaliação de documentos e ao inventário do arquivo irão constituir os principais instrumentos de trabalho,

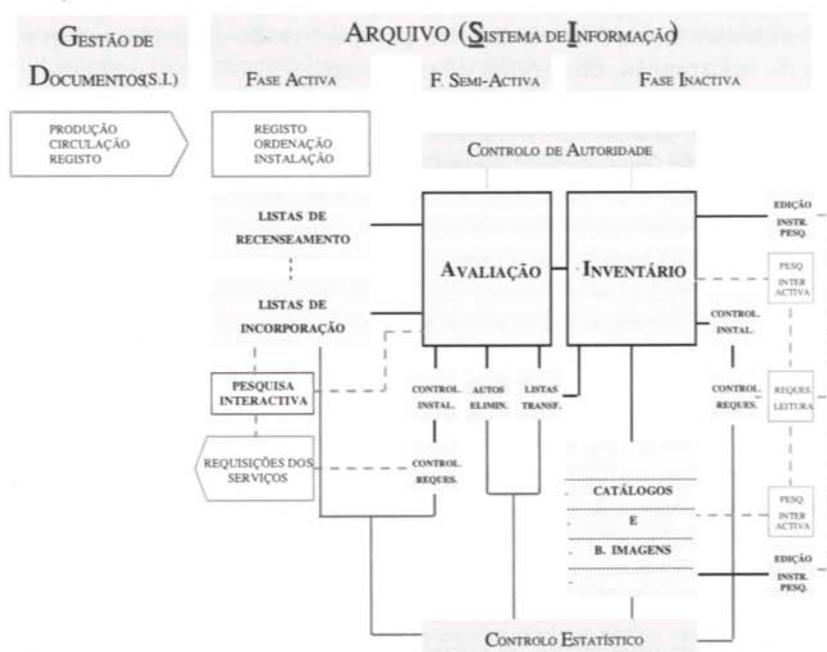


Fig. 4 – Estrutura Relacional das Operações Arquivísticas

respectivamente, do «arquivo» intermédio e do «arquivo» histórico¹³. Em ambos os casos se prevê a introdução da metodologia de descrição multinível. As operações arquivísticas situadas, seja a montante, seja a jusante destes dois módulos – e ainda outras em fase intermédia – irão ser igualmente apoiadas por instrumentos de controlo com informação seleccionada de acordo com a sua própria especificidade, mas, na medida do possível, sempre obedecendo à ordem sugerida pelas grandes zonas e campos descritivos da norma ISAD. Esta norma encontra-se configurada em seis grandes zonas (fig. 5) e diversos campos dentro de cada uma, num total de vinte e seis¹⁴. Nalgumas operações, a distância face ao esquema geral da norma é flagrante – eventualmente com a eliminação de uns quantos elementos e o acréscimo de novos dados – mas o respeito pelos critérios de ordenação sugeridos pela norma irá permitir o estabelecimento de uma estrutura estável, que facilitará as relações entre os diversos momentos da cadeia. A migração de dados será pois muito facilitada, com grandes vantagens em economia de tempo e em rigor. E não está apenas em causa evitar a duplicação de esforços ou diminuir os riscos de erro. O que mais importa salientar neste processo é que, só assim, os arquivos passarão a ser geridos coerentemente, na sua verdadeira perspectiva de sistemas integrados de informação, ao serviço do produtor.

O programa informático¹⁵, prevê duas grandes áreas de abordagem: a da Gestão do Sistema e a da Acessibilidade ao Sistema. No ficheiro de *Gestão do Sistema* enquadram-se as diversas operações de análise e de tratamento arquivístico, como: o recenseamento dos arquivos correntes e respectiva incorporação; a avaliação, triagem e eliminação dos documentos; a classificação, descrição, e indexação de dados sobre as unidades arquivísticas; o acesso à informação, a requisição das espécies documentais, a edição de instrumentos de pesquisa, etc. Diversos utilitários ou extensões do sistema poderão ainda permitir o controlo de autoridade na escolha dos termos de acesso, o controlo estatístico das diversas operações e,

¹³ As listas de incorporação não poderão também deixar de ser consideradas importantes instrumentos de acesso no âmbito do «arquivo» intermédio; o mesmo se passa com os catálogos, ao nível do chamado «arquivo» histórico.

¹⁴ Na base de dados de descrição multinível, através da qual se poderá elaborar o Inventário e outros instrumentos de acesso, foi prevista a sétima zona, destinada à Indexação.

¹⁵ O trabalho de programação foi iniciado por informáticos do Departamento de Arquivos da CMP que, neste momento, estão a ser também coadjuvados por técnicos da Câmara Municipal de Vila Nova de Gaia. A coordenação sectorial do projecto está a cargo de Vitor Mesquita.

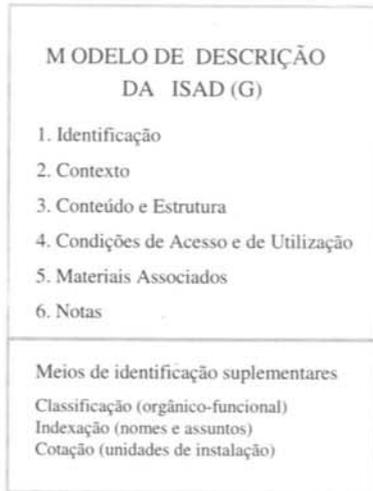


Fig. 5 – Identificação dos campos informativos num Instrumento de Acesso

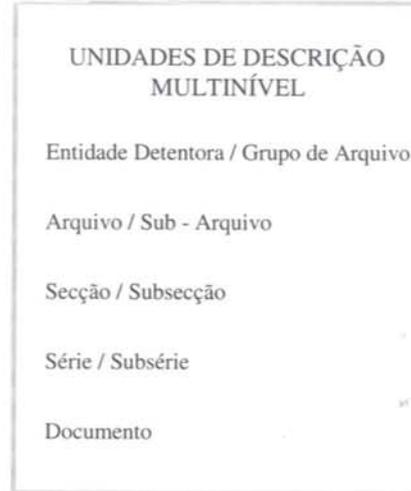


Fig. 6 – Níveis de Descrição das Unidades Arquivísticas

mesmo, o controlo das intervenções de restauro (por unidade de instalação), dos empréstimos para exposições, etc.

Os utilizadores, por sua vez, só terão direito de entrar na área da *Acessibilidade ao Sistema*, sendo logo orientados para um menu com a lista dos arquivos e colecções disponíveis. A descrição das unidades documentais aparece estruturada por *níveis orgânicos*: grupo de arquivos, arquivo, sub-arquivo, secção, subsecção, série, subsérie e documento (simples ou composto). Está também prevista a possível necessidade de descrever colecções, entendidas estas como grupos de documentos reunidos com base na afinidade tipológica dos suportes ou, meramente, por acumulação sem critério arquivístico. No caso da formação de colecções tipológicas, por conveniência de instalação especializada de certos tipos de suportes documentais, o GISA garante a possibilidade de endereçamentos para as respectivas unidades orgânicas. A pesquisa poderá sempre fazer-se pelos respectivos níveis orgânicos (fig. 6). Mas o acesso às unidades arquivísticas é igualmente possível através de outras *opções de pesquisa*, seja pela entidade produtora, seja por nomes de pessoas (singulares ou colectivas), nomes geográficos, topónimos citadinos, assuntos, datas, tipos documentais e cotas de instalação (fig. 7). Recomenda-se o uso de linguagem controlada dentro dos principais pontos de acesso, podendo a pesquisa ser ainda enrique-

UNIDADES ARQUIVÍSTICAS
TIPOS DOCUMENTAIS
ENTIDADES PRODUTORAS
NOMES DE PESSOAS
NOMES GEOGRÁFICOS
TOPÓNIMOS CIDADINOS
ASSUNTOS
DATAS
COTA
TERMOS SIMPLES
TERMOS RELACIONADOS
TEXTO LIVRE

Fig. 7 – Alternativas de Pesquisa

cida pelo recurso a termos relacionados. Todavia, está também contemplada a busca em texto livre, usando campos pré-seleccionados.

O uso do *controlo de autoridade* seguirá o esquema proposto pela norma ISAAR(CPF). O seu âmbito, contudo, limita-se à referência normalizada e caracterização das entidades produtoras. O problema do controlo de autoridade na área da informação documental é, porém, mais complexo. Ele deverá incidir também, tanto quanto possível, na definição de conceitos e na representação formal dos nomes utilizados nos principais pontos de acesso do

sistema. Assim, o GISA prevê um desenvolvimento especial nesta matéria, articulando os princípios da ISAAR com a modalidade usual – igualmente rigorosa, mas menos pesada – de controlo de autoridade para termos de indexação, nominais ou de assunto.

Definido o modelo teórico do GISA foram, em seguida, planificadas as *bases de dados* (BD) e respectivas interconexões. Caso a caso, tem-se procedido à respectiva parametrização (menus, zonas e campos descritivos, pontos de acesso, alternativas de pesquisa) e desenvolvimento. Estão já criadas as correspondentes *folhas de recolha de dados* (FRD), mas encontram-se ainda em estudo os *formatos de impressão* (FI).

Como se disse mais acima, além das BD indispensáveis para controlar o fluxo de operações da cadeia arquivística, foram definidos alguns *catálogos especiais*, alternativos à descrição multinível do tipo geral, quando se trate de peças ou documentos. A sua inserção no sistema respeita os princípios da ISAD(G) – quanto à relação vertical das unidades arquivísticas e ao formato genérico da representação – mas os elementos descritivos são adaptados à tipologia dos documentos a que tais catálogos se destinam.

O desenvolvimento do *software* tem sido efectuado por segmentos operativos do sistema, os quais, dentro em breve, irão ser ligados entre si¹⁶.

¹⁶ A programação foi iniciada em *Access*. Actualmente, o trabalho está a ser desenvolvido em *Visual Basic*.

O GISA inclui ainda a possibilidade, ao nível de catálogo, de coordenar a descrição dos documentos – simples ou compostos – com a visualização das respectivas imagens, gravadas em *disco óptico*¹⁷. Tratar-se-á, de alguma forma, o prolongamento da metodologia que presidiu à criação do projecto CALÍOPE, mas agora enriquecida com novos contributos para a contextualização e para a dinâmica do acesso.

Perspectivas para o futuro

Como objectivo inicial de experimentação do GISA foi escolhido o *Arquivo de João Martins Ferreira*, um autarca e importante homem de negócios da cidade do Porto, nos séculos XV e XVI. A Câmara Municipal do Porto havia recentemente adquirido 96 pergaminhos respeitantes a este cidadão e seus herdeiros imediatos. O número relativamente modesto de peças e a profundidade da análise efectuada do seu conteúdo tornaram bastante prático este conjunto para os primeiros ensaios do programa. Os resultados preliminares foram apresentados numa mostra sobre «Novas Tecnologias da Informação», organizada na Faculdade de Letras da Universidade de Coimbra, em 1996, durante as comemorações do 60º Aniversário do Curso de Especialização em Ciências Documentais. A mesma aplicação esteve também patente ao público na exposição dos referidos pergaminhos organizada pelo Arquivo Histórico Municipal do Porto, em Dezembro de 1996¹⁸.

Entretanto, o programa foi substancialmente melhorado, estando em vias de se iniciar a *fase de testes* nos módulos da primeira versão. A análise das aplicações, para eventuais melhorias, vai realizar-se nas entidades que elaboraram o produto e contam, também, com a colaboração de outros parceiros com os quais existem firmados protocolos de cooperação. As aplicações que vão ser primeiramente testadas dizem respeito ao recenseamento de arquivos correntes, à avaliação de documentos, à descrição multinível e ao controlo de autoridade.

¹⁷ A primeira experiência foi feita através do programa *AROP*, por amabilidade da Scanimagem. Entretanto, o sistema evoluiu para uma aplicação globalmente em *Visual Basic*, estando-se a utilizar um *Activ X* da Wang. Trata-se do *W.I.E.C.* (Wang Image Edit Control), que já permite controlar com relativa facilidade a digitalização e a edição de imagens.

¹⁸ ARQUIVO HISTÓRICO MUNICIPAL DO PORTO – *Um mercador e autarca dos séculos XV e XVI: o arquivo de João Martins Ferreira*. Porto: Câmara Municipal do Porto, 1996.

Estão ainda a ser desenvolvidas mais algumas bases de dados que constituirão o cerne da *versão 1.0*. A sua difusão comercial vai decorrer do resultado do concurso preparado para a escolha de uma empresa com as características para garantir a divulgação e comercialização do produto, assim como a necessária assistência, de usuário.

Espera-se que o programa possa também evoluir positivamente na área dos *sistemas de comunicação*. Com o apoio de especialistas neste sector pensamos que será possível promover a gestão dos sistemas de arquivo em modo remoto e, ainda, desenvolver um módulo para consulta diferida de informação, através da utilização de correio electrónico e acesso ao serviço público de comunicação de dados. A difusão pública de uma parte dos conteúdos poderá no futuro ser efectuada por intermédio da rede Internet, mediante um módulo adicional que feche a gestão das páginas da World Wide Web. Através de uma interface com o utilizador, poderão vir a ser inclusive visualizadas imagens dos próprios documentos e ser facultada a respectiva impressão.

Para o efeito, será importante que ao especificar-se as diferentes *formas de acesso* (local ou remoto), a informação possa ser escalonada consoante a sua importância e o seu nível de comunicabilidade. Este mecanismo facilitaria a coordenação do acesso de diferentes categorias de utilizadores do sistema, inclusivamente para determinar eventuais taxas de utilização, em função da natureza do serviço.

No âmbito da *Rede de Arquivos*, cremos que há todas as condições de desenvolver futuramente um projecto integrado ao nível da Área Metropolitana do Porto. Contudo, o modelo é suficientemente aberto, para poder ser utilizado por qualquer arquivo, em parceria ou isoladamente. Revelou-se também determinante a decisão anunciada pelo Instituto dos Arquivos Nacionais / Torre do Tombo, de vir a criar uma *Rede Nacional de Informação Arquivística*. É preocupação do IAN/TT instituir uma rede onde possam correr aplicações diversificadas, que obedeçam a naturais critérios de normalização. Tal medida deixa entrever a possibilidade de o GISA se transformar num instrumento complementar, dirigido integradamente à gestão e à difusão dos sistemas de arquivo.

Projecto de Digitalização do Arquivo Mário Soares

ALFREDO CALDEIRA

Constituição do Arquivo Mário Soares

O Arquivo Mário Soares (A.M.S.) foi formalmente constituído em 26 de Fevereiro de 1996, através do depósito que o Dr. Mário Soares fez do seu arquivo pessoal na Fundação Mário Soares.

O Protocolo que define as condições desse depósito estabelece:

- que a documentação permanece propriedade do Dr. Mário Soares;
- que compete à Fundação assegurar a sua conservação, organização, preservação e divulgação;
- que a Fundação assume a condução do Projecto de Digitalização do Arquivo Mário Soares¹, tendo em vista transferir toda a documentação para suporte óptico, acompanhando-a dos instrumentos adequados de descrição e recuperação da informação;
- que o Arquivo será progressivamente aberto à consulta pública através de meios informáticos, quer locais, quer remotos;
- que o Dr. Mário Soares depositará oportunamente na Fundação a documentação superveniente do seu arquivo;
- que será constituída uma Comissão de Acompanhamento, a quem competirá, em especial, elaborar e propor o Regulamento interno do Arquivo e as respectivas Normas de Acesso.

Em Março de 1996, em execução daquele Protocolo de Depósito e do Projecto de Digitalização, procedeu-se à transferência da documentação para as instalações da Sede da Fundação Mário Soares

¹ O Projecto de Digitalização do Arquivo Mário Soares, da autoria do Prof. Doutor Fernando Rosas e do signatário, foi elaborado e entregue em Setembro de 1995.

em Lisboa, que para o efeito haviam sido objecto das obras requeridas, aí incluída a respectiva climatização e a colocação da estanteira necessária.

Seguidamente, iniciaram-se as tarefas de arrumação e inventariação da documentação, ao mesmo tempo que se adquiriu o equipamento informático considerado necessário ao arranque das operações de digitalização do arquivo.

Foram, entretanto, desenvolvidos os procedimentos de descrição sumária da documentação do arquivo, tendo em vista a recuperação informática da informação nela contida, ensaiando-se, também, o estabelecimento das bases de dados mais adequadas a esse fim.

Em 19 de Junho de 1996, começou a digitalização intensiva do arquivo, enquanto ainda decorriam testes do equipamento informático.

Organização do Arquivo Mário Soares

O Arquivo Mário Soares estava organizado em pastas, tendo sido respeitada na sua transferência e arrumação a respectiva *ordem original*.

Registe-se, no entanto, que o arquivo pessoal do Dr. Mário Soares conheceu já diversas intervenções, quer ao nível da conservação, quer ao nível da sua descrição hierárquica, que provocaram sucessivos afastamentos da realidade da sua constituição e organização, suscitando problemas que vieram a marcar profundamente algumas soluções encontradas no decurso do próprio Projecto.

Em termos cronológicos, o A.M.S. estava dividido em duas grandes séries:

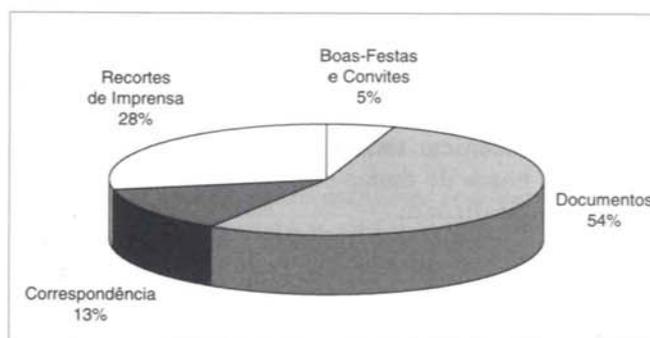
- documentação anterior a 25 de Abril de 1974;
- documentação posterior a 25 de Abril de 1974.

Sendo também possível distinguir os seguintes tipos principais de documentação:

1821	pastas de Documentos *
477	pastas de Correspondência
1000	pastas de Recortes de Imprensa
184	pastas de Boas-Festas e Convites

* Trata-se de uma designação genérica adoptada essencialmente para efeitos de organização das operações de digitalização. Esta designação envolveu, por outro

Considerada esta classificação, verificava-se a seguinte organização geral do Arquivo Mário Soares, em Março de 1996:



Os quadros que antecedem referem-se a documentação produzida até 9 de Março de 1996, data em que o Dr. Mário Soares terminou o segundo mandato como Presidente da República.

Objectivos gerais do Projecto de Digitalização do Arquivo Mário Soares

O Arquivo Mário Soares é, reconhecidamente, um acervo documental da maior importância para o conhecimento e os estudos da História portuguesa do século XX.

Reúne documentação de extrema variedade, que vai desde as actividades de oposição ao Estado Novo por parte das diferentes correntes anti-salazaristas a toda a vida política do titular desenvolvida desde 25 de Abril de 1974 – cobrindo, designadamente, a actividade do Partido Socialista, a participação nos Governos Provisórios e nos Governos Constitucionais, a Presidência da República e as relações externas com governantes e personalidades estrangeiras e, bem assim, com organizações internacionais.

O projecto de tratamento desse vasto espólio visou três objectivos fundamentais:

- transferir para as instalações da Fundação Mário Soares toda a documentação, aí criando as adequadas condições de conservação;

lado, conjuntos de documentação avulsa que não oferecia relevância para ser incluída no Projecto de Digitalização e cujas pastas, por conseguinte, vieram a ser anuladas.

- transferir para suporte óptico todo o Arquivo Mário Soares;
- abrir progressivamente à consulta pública o arquivo digital assim criado.

Para o efeito, definiram-se as seguintes metas de trabalho:

- digitalizar toda a documentação do A.M.S.;
- transferir para suporte óptico² toda a documentação do A.M.S.;
- descrever e classificar toda a documentação do A.M.S.;
- indexar nas bases de dados criadas para o efeito toda a documentação digitalizada;
- abrir progressivamente à consulta o A.M.S. digitalizado;
- colocar o A.M.S. em rede, designadamente através da Internet;
- editar em CD-ROM documentos seleccionados do A.M.S.;
- proceder a experiências de tratamento digital de fotografias e vídeos;
- procurar a incorporação de espólios similares ou complementares no A.M.S.

Tendo em vista o cumprimento desses objectivos, foi definido o seguinte calendário geral, cobrindo um período de 30 meses (Março de 1996 a Agosto de 1998):

- Março 1996 Transferência da documentação para as instalações da Fundação;
- Junho 1996 Início da digitalização e da descrição documental;
- Abril 1997 Abertura gradual à consulta e colocação na Internet;
- Agosto 1998 Conclusão das operações de digitalização do A.M.S.

Equipamento de digitalização do Arquivo Mário Soares

De acordo com o Projecto de Digitalização do Arquivo Mário Soares, e uma vez reconhecidas com maior detalhe as características da documentação que o integra, foi adquirido o seguinte equipamento informático, tendo em vista a constituição de um sistema coerente de digitalização e arquivo em disco óptico:

- 1 posto de Administração;
- 4 postos de Digitalização;
- 3 postos de Classificação;

² No Projecto estão a ser utilizados discos ópticos regraváveis de 1.3 Gb, prevendo-se que a informação neles armazenada seja posteriormente transferida para discos ópticos WORM.

- 3 postos de Consulta;
- 1 posto de Apoio à Consulta;
- todos ligados em rede;
- e funcionando sobre um servidor;
- a que está igualmente ligada uma *juke-box*, com capacidade para 101 Gb;
- 1 servidor afecto exclusivamente ao funcionamento da página na Internet.

Os postos de Digitalização utilizam *scanners*³ de diferentes características, tendo em vista a sua máxima rentabilização: dois para formatos A4 e dois para formatos A3. No início do Projecto foi utilizado um *scanner* rotativo de alta velocidade (50 imagens/minuto), para formatos A4 e A3 – uma vez concluído o tratamento da documentação normalizada que o justificava, foi trocado por um *scanner* A3.

O posto de Digitalização inicialmente equipado com o *scanner* rotativo e, bem assim, dois postos de Classificação foram dotados de *drives* de disco óptico de 1.3 Gb, de modo a tornar mais rápidas as respectivas operações, garantindo a armazenagem directa das imagens digitalizadas em discos ópticos.

Todo o sistema funciona sobre *Windows'95*, recorrendo a um *software* específico de digitalização, tratamento de imagem, compressão e indexação designado *File Magic Plus*.

O sistema dispõe ainda de duas capacidades específicas de tratamento das imagens digitalizadas: por um lado, *OCR* (Reconhecimento Óptico de Caracteres) com dicionário de português e inglês, que permite a conversão de imagens de textos dactilografados para texto electrónico (processador de texto); por outro lado, *Image Enhancement* (Melhoria de Imagem), que permite intervir sobre a própria qualidade das imagens digitalizadas, designadamente limpando ou corrigindo manchas, defeitos, etc.

A gestão da *juke-box* é feita através de um *software* apropriado, que permite o acesso rápido e coerente à informação aí armazenada e denominado *Cheyenne HSM 2522B Net*⁴.

³ Vistas as características gerais da documentação do A.M.S., optou-se por adquirir digitalizadores Fujitsu a preto e branco, sem prejuízo de se poderem considerar algumas, excepcionais, operações de digitalização com cor. Registe-se que o *software* utilizado gere indiferentemente imagens digitalizadas a preto e branco ou a cores e, bem assim, diferentes formatos, incluindo de vídeo.

⁴ O *software* inicialmente utilizado na gestão da *juke-box* não se revelou inteiramente frutuoso (designadamente em matéria de acesso em rede, local e externa, aos documentos arquivados), tendo sido substituído pelo programa referido, depois de um período de experimentação considerado concludente.

Utilizam-se, como já se disse, discos ópticos regraváveis de dupla face com a capacidade global de 1.3 Gb, sendo de assinalar que servem para guardar directamente a informação digitalizada e, bem assim, para obter cópias de segurança.

O sistema foi posteriormente dotado do equipamento susceptível de permitir o seu funcionamento em rede, designadamente via Internet⁵ e, bem assim, equipamento de gravação de CD-ROM – utilizado, designadamente, para produção de cópias de segurança⁶ e edições restritas de séries documentais.

Organização da digitalização do Arquivo Mário Soares

Uma vez adquirido, instalado e testado o equipamento informático, foi possível organizar as diferentes operações de digitalização e classificação do A.M.S. em termos que possibilitassem o cumprimento do calendário estabelecido no Projecto.

Para o efeito, definiram-se as seguintes prioridades:

- iniciar a digitalização dos documentos anteriores a 25 de Abril de 1974⁷, utilizando nessas operações três *scanners*;
- iniciar também a digitalização dos Recortes de Imprensa anteriores a 25 de Abril de 1974 que, pelas suas características (na sua maioria, fotocópias em formato A4)⁸, podiam ser digitalizados com maior rapidez num *scanner* rotativo de alta velocidade;
- simultaneamente, iniciar a descrição de todos os documentos⁹.

⁵ Nos termos de um Protocolo celebrado com a Fundação para a Computação Científica Nacional (FCCN), o Arquivo disponibiliza via Internet uma página da Fundação Mário Soares.

⁶ As cópias de segurança em CD-ROM pressupõem, em geral, uma fase já estabilizada da constituição do arquivo em suporte digital, evitando a produção de cópias incompletas e/ou não descritas.

⁷ Esta opção, embora tecnicamente mais difícil, quer pelas características de grande parte dos documentos, quer pelo seu próprio estado de conservação, encontrou justificação no especial interesse dessa documentação e na sua maior comunicabilidade.

⁸ Esses Recortes de Imprensa em fotocópia ocupavam cerca de 42% do total dos recortes. Importa ainda referir que existem no A.M.S. algumas centenas de páginas (geralmente de jornais) com dimensões superiores a A3, que não são susceptíveis de digitalização com o equipamento disponível, motivo pelo qual se recorreu à sua redução por fotocópia à dimensão A3, sem perda de legibilidade, possibilitando assim a sua posterior digitalização.

⁹ Importa referir que, vista a muito diversificada organização dos documentos nas respectivas pastas, não foi possível elaborar instrumentos de descrição que se

A definição dessas prioridades permitiu que as operações de digitalização intensiva do A.M.S. fossem iniciadas em 19 de Junho de 1996, arrancando simultaneamente a classificação da respectiva documentação, de acordo com a seguinte organização funcional:

<i>Função</i>	<i>Conteúdo funcional</i>
<i>Administração do Arquivo</i>	Dirigir o Projecto de Digitalização do A.M.S.
<i>Gestão do Arquivo</i>	Gerir os recursos humanos e materiais disponíveis, de acordo com as prioridades estabelecidas.
<i>Gestão das bases de dados</i>	Gerir os sistemas de classificação e indexação do A.M.S, de acordo com as normas definidas.
<i>Classificação</i>	Executar a descrição e classificação dos documentos do A.M.S.
<i>Digitalização</i>	Proceder à digitalização dos documentos, de acordo com as prioridades e metodologias estabelecidas.

NOTA:

O pessoal chamado a colaborar nas tarefas de organização e digitalização do A.M.S. tem sido recrutado a dois níveis principais, de acordo com as qualificações requeridas:

- por um lado, recorreu-se, para as tarefas de digitalização, a pessoal com o 12º ano de escolaridade ou em vias de ingresso no ensino superior;
- por outro, e no que respeita às tarefas de classificação documental, privilegiou-se pessoal com formação universitária, quase sempre em fase de frequência de Mestrado.

A esta estrutura acresceu a Comissão de Acompanhamento¹⁰, com a competência de supervisão da execução do projecto e, em especial, elaborar e propor o Regulamento interno do Arquivo e as respectivas Normas de Acesso.

O Regulamento e as Normas de Acesso foram ainda sujeitos a parecer do Ministério da Ciência e da Tecnologia, nos termos do

referissem apenas à unidade de instalação-pasta, antes se tornando necessário proceder a essa descrição ao nível de cada documento, o que, obviamente, tornou mais difícil e morosa toda a operação de descrição.

¹⁰ A Comissão de Acompanhamento foi constituída pelo Administrador e pela Gestora do Arquivo, por um Consultor Histórico (supervisão, de acordo com critérios histórico-científicos, do processo de digitalização do Arquivo) e um Consultor Arquivístico (classificação documental do Arquivo e estabelecimento dos critérios de indexação dos documentos).

Convénio de cooperação entretanto assinado (ver adiante) e tendo em especial atenção o carácter exemplar e de demonstração reconhecido ao Projecto.

Algumas questões metodológicas

Conforme já se referiu, considerou-se prioritária a digitalização da documentação anterior a 25 de Abril de 1974 e, bem assim, pela digitalização dos Recortes de Imprensa constituídos por fotocópias.

Estas opções, claramente justificadas pela natureza da documentação e também pela maior facilidade que se depararia na sua abertura à consulta, criaram algumas dificuldades técnicas iniciais, hoje ultrapassadas.

Por outro lado, e pela própria diversidade da documentação tratada, essas opções permitiram testar não só os equipamentos, como a metodologia de digitalização e, em momento posterior, permitiram apurar os modos de descrição dos documentos.

Esta última questão, amplamente debatida no âmbito do Projecto de Digitalização do A.M.S., apresentava especial relevância:

- por o A.M.S. ser um arquivo pessoal, ainda para mais objecto de diversas e nem sempre coincidentes classificações ao longo da sua produção e arrumação;
- e por se pretender disponibilizá-lo à consulta através de meios informáticos e em prazo breve.

Ora, a descrição dos documentos do A.M.S., devendo, por um lado, obedecer a princípios arquivísticos correctos, devia também, por outro, oferecer instrumentos de recuperação da informação suficientemente flexíveis para deles se poder tirar o máximo proveito no âmbito de um sistema de arquivo electrónico como o que se pretende implantar.

Verificou-se, a curto prazo, que não seria suficiente criar bases de dados histórica e arquivisticamente adequadas.

De facto, e uma vez que a descrição do Arquivo era referida a cada documento, estudou-se a possibilidade de criar instrumentos de recuperação da informação mais completos e que pudessem ultrapassar a necessária rigidez de uma base de dados (ou, pelo contrário, o seu crescimento exponencial).

Neste sentido, e explorando as próprias capacidades do sistema informático instalado, concluiu-se ser possível adoptar a seguinte metodologia:

- cada Pasta consta de uma base de dados informática, com uma designação genérica;
- uma vez digitalizada essa Pasta, registam-se na mesma base de dados todos os elementos descritivos dessa operação (incluindo os discos ópticos para que são transferidas as respectivas imagens digitalizadas);
- procede-se seguidamente à conferência das imagens arquivadas com os documentos da Pasta (controlo de qualidade)¹¹;
- todos os documentos (digitalizados) são objecto de uma descrição na base de dados criada para o efeito;
- essas descrições, uma vez corrigidas, são posteriormente transferidas, de modo automatizado, para um processador de texto, sendo arquivadas sob a designação de Folha de Rosto dessa Pasta;
- cada Folha de Rosto inclui, além da descrição dos documentos, a indicação da imagem referente a cada um deles, constituindo assim um verdadeiro índice dessa Pasta;
- e, por outro lado, considera-se em cada Folha de Rosto as possíveis articulações dos documentos nela descritos com as classificações constantes das bases de dados¹²;
- prevê-se assim, também, a divisão electrónica de cada Pasta em diversas sub-pastas que melhor possam adequar-se a cada um dos temas da referida base de dados (naturalmente que, no caso de divisão electrónica de uma pasta, também a respectiva folha de rosto será dividida em consequência);
- finalmente, a imagem de cada Folha de Rosto é acrescentada à respectiva Pasta (ou sub-pasta), ficando a fazer parte integrante dessa colecção de documentos electrónicos (como 1º documento).

Assim, a indexação das pastas (ou sub-pastas) na Base de Dados criada para o efeito no *File Magic* refere-se não apenas aos documentos, mas também à respectiva Folha de Rosto.

Uma vez indexadas as pastas (e as respectivas folhas de rosto), o sistema permite pesquisar em texto livre (*full text*) qualquer

¹¹ É também nesta fase que se repete a digitalização de documentos que possam ter ficado deficientemente reproduzidos e, bem assim, se produzem cópias de segurança em disco óptico.

¹² Importa aqui referir que se consideraram à partida bases de dados diferenciadas para os documentos anteriores e posteriores a 25 de Abril de 1974, vistas as características de cada uma dessas colecções. Gradualmente, está a ser preparada a uniformização da descrição de todos os documentos do Arquivo, de acordo com os respectivos níveis.

palavra incluída em todas as folhas de rosto. O sistema disponibiliza, por outro lado, um conjunto de instrumentos (*Data Base Queries*) susceptíveis de organizar essas pesquisas, de modo a devolver ao leitor, com capacidades relativamente sofisticadas, a indicação das pastas em que a expressão escolhida apareça – o que é realizado sem prejuízo da possibilidade de pesquisa directa nas bases de dados – que, obviamente, fornecerá uma resposta muito menos precisa do que aquela que pode ser obtida através da pesquisa em texto livre, mais direccionada.

Esta metodologia apresenta, no entanto, limitações decorrentes da própria *ordem original* dos documentos. De facto, e vista a arrumação existente nas pastas (nem sempre coerente ou sequer sequencial), torna-se necessário cruzar a respectiva informação, na medida em que uma mesma pasta contém quase sempre documentos referentes a mais de uma das classificações previstas na base de dados.

Para o efeito, previu-se a criação, no próprio documento (electrónico), de sistemas de encaminhamento directo (também electrónicos) para o(s) documento(s) e/ou temas com que se relacione. Esta solução permite aos leitores explorar em cada documento as suas diferentes possíveis relações com outros temas.

Há que assinalar, no entanto, que a digitalização do A.M.S. no prazo definido e a sua rápida abertura à consulta não permitiram o apuramento imediato de todas essas ligações, que serão progressivamente construídas.

Por isso, tem sido feito um esforço permanente de actualização das normas aplicadas, adequando-as às características específicas da documentação e às possibilidades de tratamento informático disponibilizadas pela programação instalada.

Esta realidade tem provocado, necessariamente, atrasos na execução das tarefas de classificação, mas julga-se preferível insistir no mesmo esquema geral de descrição detalhada dos documentos – solução que se considera mais segura e transparente, oferecendo também aos leitores a máxima informação sobre o arquivo.

Considera-se igualmente que o sistema que tem vindo a ser adoptado (descrição dos documentos em *Folhas de Rosto* pesquisáveis em texto livre) se insere na ajustada linha de tratamento informático de um arquivo digital – não tendo sido detectado qualquer problema de «lixo» ou «ruído» informático que fosse provocado por excesso de informação disponibilizada, argumento que aparece com frequência utilizado no âmbito da discussão de projectos similares.

Aliás, o próprio programa utilizado disponibiliza um recurso (*full text noise word list*) que permite definir as palavras ou frases que o sistema de pesquisa em texto livre deve ignorar e que pode ser definido pela administração do sistema – podendo cada uma dessas «listas de palavras a ignorar» ser configurada para os diferentes níveis de pesquisa (uma lista por cada *grupo* ou *super grupo*)¹³.

É importante assinalar, por outro lado, que a classificação de grande parte dos documentos posteriores a 25 de Abril de 1974 está já a ser feita com recurso a meios informáticos mais desenvolvidos (designadamente, com preenchimento de novos campos na base de dados), o que permite enriquecer os elementos de recuperação da informação, em articulação com o sistema adoptado de *Folhas de Rosto*.

Em diversos conjuntos desta documentação, aliás, em que é maior o número de originais dactilografados ou impressos, é possível utilizar as ferramentas de reconhecimento óptico de caracteres (OCR) com inscrição automática nos campos pré-definidos na base de dados – o que torna o trabalho de descrição mais rápido, criando imediatamente palavras-chave na base de dados. Este sistema viabiliza, por outro lado, a simultaneidade das operações de classificação e de *indexação*.

Convénio entre a Fundação Mário Soares e o Ministério da Ciência e da Tecnologia

Em 26 de Setembro de 1996, foi celebrado um Convénio entre o Ministério da Ciência e da Tecnologia (M.C.T.) e a Fundação Mário Soares, tendo como objectivo central «a reprodução digital do Arquivo Mário Soares e de outra documentação de interesse histórico e cultural doada ou depositada, ainda que transitoriamente, na Fundação Mário Soares».

Esse Convénio substituiu o que havia sido celebrado em 18 de Outubro de 1995 com a Junta Nacional de Investigação Científica e Tecnológica e o Instituto de Investigação Científica e Tropical, tendo em atenção a evolução verificada na sua execução e havendo, entretanto, a Fundação Mário Soares desenvolvido as acções neces-

¹³ O programa de gestão documental utilizado prevê a existência de super grupos e de grupos, correspondendo *grosso modo*, respectivamente, aos níveis da gaveta e do maço, cabendo neste as pastas ou sub-pastas com ele relacionadas ou nele contidas. Vista a capacidade de fácil criação de campos em cada grupo, é possível obter por essa via a definição de outros diferentes níveis de descrição documental.

sárias ao cumprimento do Projecto de Digitalização do A.M.S., a que agora foi reconhecido «carácter exemplar e de demonstração», «tendo em vista, designadamente, a modernização dos arquivos portugueses, instrumentos indispensáveis para a investigação científica».

Pelo Convénio assinado, a Fundação assumiu a responsabilidade pela execução do referido Projecto no prazo de 30 meses, contados a partir de Março de 1996, comprometendo-se a concretizar as seguintes acções principais:

- criação da página do Arquivo Mário Soares na Internet;
- abertura gradual do A.M.S. à consulta gratuita a partir do primeiro semestre de 1997;
- criação de condições para a realização de trabalhos de investigação que tenham por fonte principal a documentação que integra o A.M.S.;
- concretização, no decurso de 1997, da possibilidade de acessos em rede às bases de dados já constituídas do A.M.S.;
- edição em CD-ROM, ou outro suporte electrónico, de documentos seleccionados do Arquivo.

O Convénio estabeleceu, por outro lado, que o M.C.T. e a Fundação organizariam sessões de demonstração do Projecto, tendo em vista, especialmente, a sua divulgação junto de entidades interessadas em projectos similares – o que tem sido feito com enorme frequência, quer para interessados nacionais, quer estrangeiros.

Mais estabeleceu o referido Convénio que o M.C.T. apoiaria financeiramente o Projecto, sem prejuízo de outros apoios de entidades terceiras.

Outros protocolos de cooperação e de incorporação de acervos documentais

A Fundação Mário Soares entende que a gestão do A.M.S. deve incluir modos de relacionamento preferencial com instituições congéneres ou que prossigam fins que possam ser articulados com os seus.

Assim, e designadamente, foi prevista a realização de acordos de cooperação com o Instituto de História Contemporânea da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa e com o Centro de Documentação 25 de Abril da Universidade de Coimbra.

No mesmo sentido, foi estabelecido um acordo com o Instituto dos Arquivos Nacionais/Torre do Tombo no sentido de ser incorpo-

rada no A.M.S., sob forma digital, toda a documentação referente ao Dr. Mário Soares existente nos arquivos da extinta polícia política¹⁴. Este importante espólio, totalizando cerca de 70 000 imagens, foi digitalizado nas instalações dos AN/TT e prevê-se a sua progressiva abertura à consulta, sob forma electrónica, na Fundação – a abertura à consulta no âmbito do A.M.S. dessa documentação obrigará, evidentemente, a especiais cuidados na sua comunicação, designadamente nos termos do artigo 17º do Decreto-Lei nº 16/93, de 23 de Janeiro.

Entendeu-se, por outro lado, desejável que a Fundação estabelecesse protocolos com entidades congéneres estrangeiras, não apenas tendo em vista a troca de experiências em matéria das tecnologias utilizadas, como também para eventual intercâmbio de informação (disponibilizando mutuamente documentação arquivada em cada uma das instituições) – tendo vindo a ser desenvolvidos esforços de colaboração com diversas entidades possuidoras de fundos especialmente relevantes para o estudo da História Contemporânea portuguesa.

Idêntica orientação está a ser seguida em matéria de tratamento digital de documentação registada em suportes outros que não o papel, como é o caso do microfilme, da fotografia, do som e do vídeo – em relação a estas matérias, e considerada a especificidade técnica de alguns dos procedimentos e, bem assim, o custo dos equipamentos necessários, os estudos até agora realizados apontam para a sua progressiva experimentação, designadamente no âmbito do novo edifício que a Fundação projecta construir e que será dotado de instalações adequadas, preferentemente em colaboração com instituições vocacionadas para a utilização dessas tecnologias.

Refira-se, finalmente, que a Fundação tem vindo a celebrar diversos protocolos com detentores de acervos privados, tendo por objectivo a sua integração no A.M.S., seja sob forma de doação, seja em regime de depósito. Destaquem-se, entre outras entregas de documentos avulsos, os casos da correspondência de Francisco Ramos da Costa e da documentação da Comissão Promotora das Comemorações do 50º aniversário da criação do MUD Juvenil, de Alfredo Ribeiro dos Santos, de Joaquim Catanho de Menezes, de Manuel Maria Sarmiento Rodrigues, de Mário Pinto de Andrade e de Bento de Jesus Caraça.

¹⁴ Na solicitação endereçada aos AN/TT, foi igualmente incluída a documentação que fosse referenciada nos arquivos da extinta Legião Portuguesa. No entanto, e por razões de incapacidade actual de resposta dos Arquivos Nacionais, não se prevê para já a possibilidade de incorporação desta documentação.

A incorporação dessas colecções permite abrir esses acervos à investigação, salvaguardando em muitos casos a sua própria sobrevivência e, por outro lado, garantindo o progressivo alargamento dos horizontes históricos do fundo inicial.

Pretende-se assim, também, que o Arquivo Mário Soares se constitua progressivamente em local de consulta e investigação de documentos fundamentais para a compreensão do século XX português e, simultaneamente, vista a tecnologia utilizada no seu tratamento, se assuma também como centro de acesso em rede à informação disponibilizada. De facto, o recurso às novas tecnologias de informação é, sem dúvida, um suporte especialmente apto para a concretização de tais operações, permitindo em especial vir a ultrapassar de modo simplificado algumas das mais difíceis questões sobre a *propriedade* e a *localização* de tais acervos – uma vez digitalizados, a sua circulação em suporte electrónico poderá ser um importante apoio ao desenvolvimento da investigação.

Importará, aliás, a propósito, sublinhar a concepção expressa no *Livro Verde para a Sociedade de Informação em Portugal*, aprovado no Conselho de Ministros em 17 de Abril de 1997 – data em que o próprio Arquivo Mário Soares foi aberto ao público – sobre a utilização das novas tecnologias no âmbito dos arquivos, administrativos ou históricos:

Os arquivos administrativos são apenas uma das facetas de um processo mais amplo de digitalização de informação que o desenvolvimento da tecnologia dos suportes electrónicos, designadamente dos discos ópticos, veio proporcionar. A possibilidade de registar e de aceder instantaneamente a centenas ou mesmo milhares de *gigabytes* de informação, que se mantém inalterável durante elevados períodos de tempo e que pode ser arquivada em condições de grande segurança, abriu campo para a informatização e digitalização de arquivos históricos, culturais e de património artístico e arquitectónico. O desenvolvimento de bibliotecas digitais deve igualmente ser encarado numa óptica de acesso generalizado à informação sob formato digital.

Entretanto, e já no decurso do corrente ano, foi também celebrado um protocolo com a *MSTF-Software* para Microcomputadores, subsidiária em Portugal da *Microsoft Corporation*, tendo como objectivo principal o estabelecimento das bases de cooperação cultural, científica e educativa no âmbito do Arquivo Mário Soares e abrangendo condições de licenciamento do *software* necessário à manutenção do sistema informático da Fundação e consultadoria em matérias de interesse para o desenvolvimento do sistema instalado, com vista a reforçar as suas potencialidades actuais, especialmente no que respeita ao funcionamento em rede do Arquivo Mário

Soares e à sua gradual disponibilização à consulta através da Internet, formação de pessoal e estudo da organização informática do novo edifício da Fundação, tendo por objectivo garantir a coerência e a capacidade do sistema projectado.

Novo edifício

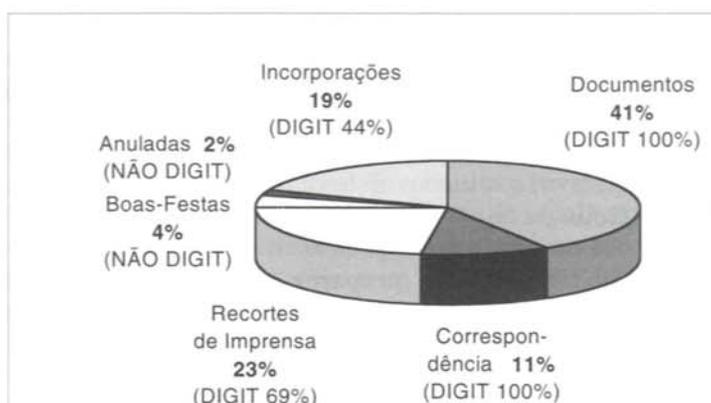
A Fundação tem vindo a preparar o alargamento das suas actuais instalações, tendo em vista, a criação de uma nova infra-estrutura que melhor permita realizar as suas actividades de âmbito cultural, educativo e científico, aí instalando os equipamentos necessários ao seu cabal desenvolvimento e criando áreas destinadas a serviços de interesse público.

O modelo elaborado configura a criação de um verdadeiro pólo cultural da Fundação Mário Soares, especialmente vocacionado para a investigação da história contemporânea portuguesa e da ciência política, viabilizando ainda uma estreita articulação com acções de formação a desenvolver no âmbito ou com a intervenção da Fundação, prevendo-se a criação de áreas de acesso público para consulta dos documentos que integram o Arquivo Mário Soares e dos livros que virão a constituir uma Biblioteca, com a implantação de depósitos destinados à conservação de documentos, fotografias, vídeos e livros e das necessárias áreas técnicas de apoio, bem como a criação de uma sala multiusos.

No âmbito deste projecto prevê-se o tratamento integrado do arquivo e da biblioteca, permitindo aos leitores o acesso indiferenciado e simultâneo a ambos, através de meios informáticos que rentabilizem a sua exploração. O projecto contempla também a conservação e progressiva disponibilização *on line* do espólio de registos sonoros, fotográficos e videográficos.

Situação geral da digitalização do Arquivo Mário Soares

Em 30 de Junho de 1998, a situação geral do Arquivo Mário Soares é a que se enuncia no quadro seguinte:



Este quadro merece algumas notas: antes do mais, sublinhe-se que cerca de 19% do Arquivo é já hoje constituído por documentação incorporada; e que, por outro lado, foi decidido não digitalizar cerca de 6% do fundo original (Boas-Festas/Convites e pastas consideradas anuladas).

No que respeita aos documentos entretanto incorporados no A.M.S., e que somam 835 pastas, já se encontram digitalizadas cerca de 44% – a sua digitalização prioritária relativamente a pastas de recortes constantes do fundo inicial foi uma opção óbvia face ao interesse documental que representavam. Registe-se, ainda assim, que apenas faltam digitalizar 314 pastas de recortes, mas já foram digitalizadas 365 pastas de documentação incorporada.

Por outro lado, em relação ao fundo original, foi decidido não digitalizar 184 pastas de Boas-Festas e Convites – tendo, no entanto, sido elaborada uma base de dados de descrição de toda essa documentação – e, bem assim, foram anuladas 77 pastas que, pelo seu conteúdo, não justificavam a digitalização.

Pode pois concluir-se que, em termos quantitativos, os objectivos do Projecto de Digitalização do Arquivo Mário Soares já foram ultrapassados, estando digitalizadas, nesta data, 3 358 pastas, de um total inicial de 3 221 pastas (descontadas as de Boas-Festas e Convites e anuladas).

Abertura do Arquivo Mário Soares à consulta

Desde a sua inauguração, em 17 de Abril de 1997, o perfil médio dos utilizadores do Arquivo é, conforme se esperava, inequívoca-

mente associado à investigação – cerca de 2/3 dos leitores entraram no arquivo para pesquisas no âmbito de investigações específicas e, sensivelmente na mesma percentagem, são estudantes ou docentes do ensino superior, com idades médias compreendidas entre os 25 e os 30 anos.

A frequência do arquivo digital tem correspondido, por outro lado, ao que se pretendia, evitando uma utilização excessiva dos meios técnicos disponibilizados, que ainda estão a ser aproveitados, parcialmente, pelo próprio pessoal do arquivo – não tendo sido também, por isso mesmo, desencadeadas acções específicas de promoção.

Convirá sublinhar, também, que muitos dos utilizadores que chegam ao Arquivo Mário Soares apresentam uma característica singular: já consultaram através da Internet as descrições dos documentos disponibilizados e, portanto, sabem melhor o que podem encontrar, rentabilizando a consulta da documentação digital. No mesmo sentido, tem sido patente a facilidade de quase todos os leitores na exploração do sistema de arquivo instalado.

Página na Internet

Sempre esteve prevista no Projecto de Digitalização do Arquivo Mário Soares a colocação em rede de uma página na Internet.

Nesse sentido, assinou-se um Protocolo com a Fundação para a Computação Científica Nacional (F.C.C.N.), tendo em vista criar as condições técnicas para a disponibilização dessa página com a seguinte localização: www.fmsoares.pt.

A colocação da Fundação Mário Soares na Internet, concretizada em 17 de Abril de 1997, foi outra expressão do conceito subjacente a este projecto de digitalização e, sobretudo, da noção de que, através dela, se pretende disponibilizar o máximo de informação à comunidade científica e ao público em geral, e não apenas dar conta da existência da instituição.

A página da Fundação Mário Soares na Internet coloca em realce, para além de informações gerais sobre a instituição e as suas realizações, o acesso gratuito a bases de dados do A.M.S. e à Casa-Museu Centro Cultural João Soares¹⁵:

¹⁵ A Casa-Museu Centro Cultural João Soares foi inaugurada em 8 de Dezembro de 1996.

- no que respeita ao Arquivo Mário Soares, pretendeu-se disponibilizar as bases de dados com a informação aberta à consulta, de tal modo que os leitores aí a possam colher previamente e sem necessidade de se deslocarem ao arquivo; a título exemplificativo, disponibilizam-se igualmente imagens digitalizadas de documentos seleccionados; oportunamente, prevê-se também a criação de um serviço de consulta ao arquivo via Internet, podendo o leitor requisitar, mediante condições a estabelecer, o acesso à documentação pretendida;
- no que respeita à Casa-Museu Centro Cultural João Soares, situada em Cortes, Leiria, a página disponibiliza o acesso às imagens das inúmeras ofertas que Mário Soares recebeu durante a sua vida pública e que se encontram fotografadas e digitalizadas, de modo a poderem ser visitadas através de meios informáticos, com textos de enquadramento que melhor permitam compreender as circunstâncias, lugares e épocas em que foram recebidas.

Permitiu, portanto, este projecto colocar em rede (através da rede científica nacional) uma enorme quantidade e variedade de informação, sucessivamente acrescentada e actualizada, que preenche uma notória ausência, no nosso país, de instrumentos de pesquisa em matéria de história contemporânea.

Creemos ser do maior interesse reforçar esta componente do projecto, designadamente através da sua articulação com os principais arquivos e bibliotecas nacionais e os diferentes centros de investigação, sem prejuízo da sua necessária internacionalização – em especial com instituições congéneres ou de cuja actividade possam decorrer resultados complementares.

De facto, não pretendeu a Fundação Mário Soares com este projecto criar uma estrutura isolada ou exclusivamente centrada no seu espólio documental.

Este projecto visa, fundamentalmente, criar um ponto de partida, tecnologicamente avançado, que possa servir, em conjugação com outras instituições, públicas e privadas, para desenvolver as potencialidades das novas tecnologias no sentido da disponibilização alargada da informação, reforçando os meios de conhecimento e de investigação.

A informação assim veiculada através da rede científica nacional permite, por outro lado, disponibilizar instrumentos *democratizados* de interactividade com a Fundação e, em especial com o Arquivo Mário Soares, alargando a capacidade de informação dos

cidadãos e, em especial, dos estudiosos da história contemporânea portuguesa.

Algumas notas finais

A ditadura que perdurou em Portugal de 1926 a 1974 afastou, obviamente, dos arquivos públicos o essencial da memória de quantos se lhe opuseram, ocultando a simples existência dos acontecimentos e movimentações que não eram constitutivos da imagem oficial. Mas não é menos verdade que esses mesmos arquivos públicos – e, sobretudo, as entidades que é previsto superintenderem a política arquivística nacional – não definiram, entretanto, qualquer programa coerente de incorporação sistemática dos acervos que, de modo mais ou menos disperso, são a expressão dessa outra memória nacional. Antes pelo contrário, as suas poucas recentes incursões na História Contemporânea quase se referem exclusivamente a arquivos daquele regime, como é, em especial, o caso dos de António de Oliveira Salazar e Marcelo Caetano ou da PIDE/DGS e LP – (aliás, fundamentais para a compreensão dos modos de agir do Estado Novo) – e, ainda assim, através de decisões legislativas casuísticas.

Mas o certo é que, por essa via, e em termos objectivos, se tem vindo a perpetuar a visão «oficial» do que se passou entre nós nessa metade do século XX.

A sociedade, no entanto, move-se... e mais rápida que a memória oficial. E move-se no sentido inequívoco da diversidade e pluralidade de memórias, que não se esgotam numa «memória nacional» unipolar e de pensamento único. Por isso mesmo, se tem assistido nos últimos anos em Portugal – como em todo o mundo – ao aparecimento de centros de documentação e arquivos que giram, precisamente, em volta de arquivos privados e que se constituem, simultaneamente, em centros de investigação e de inovação.

Esta realidade – em si mesma, positiva – suscita múltiplas questões, designadamente no que respeita aos (desejáveis) modos de articulação com os arquivos públicos e as universidades. Não será esta, no entanto, a ocasião de analisar esses problemas.

A constituição do Arquivo Mário Soares insere-se, precisamente, nessa vertente nova da nossa realidade arquivística, a que acresce o carácter inovador das metodologias nele utilizadas. Do mesmo modo que a abertura gradual à consulta de um arquivo pessoal em vida do seu titular representa, no nosso país, um passo de inegável importância. E, bem assim, a capacidade já concretizada de reunir

ALFREDO CALDEIRA

em sua volta outros acervos documentais que, se tal não acontecesse, quase seguramente desapareceriam ou, pelo menos, não seriam abertos à investigação.

Não quer isto dizer, evidentemente, que o Projecto de Digitalização do Arquivo Mário Soares se apresente como verdade definitiva. Antes pelo contrário, tem pretendido suscitar permanentemente uma discussão séria em seu redor, que possibilite não só a melhoria do seu funcionamento, como, sobretudo, o aproveitamento por todos os interessados dos ensinamentos que o seu próprio carácter inovador pode viabilizar.

ALFREDO CALDEIRA
Fundação Mário Soares

Projectos Informáticos do Arquivo Distrital de Braga

MARIA DA ASSUNÇÃO JÁCOME DE VASCONCELOS
CLARA SOFIA DA ROCHA PINTO MOREIRA

Introdução

Com o presente artigo pretendemos descrever os objectivos e motivação do HITEX, um sistema informático desenvolvido no Arquivo Distrital de Braga (A.D.B./U.M.), a partir de 1988, em colaboração com o Departamento de Informática da Universidade do Minho e com o apoio da companhia IBM-Portuguesa e da Fundação Calouste Gulbenkian, para registar, organizar e gerir informação de natureza histórica.

Pretende-se, igualmente, dar a conhecer a colaboração do Arquivo Distrital de Braga em outros projectos informáticos com destaque para o HINET e, actualmente, no Geira.

HITEX: um sistema em desenvolvimento

O *software* a que se designou de HITEX teve como objectivo a melhoria do acesso dos utentes do Arquivo Distrital de Braga aos fundos documentais nele existentes, bem como a necessidade do Arquivo em melhorar a sua tecnologia de publicação de fontes.

Assim, a filosofia de base do HITEX foi criar um sistema com «uma arquitectura aberta capaz de trocar informação com ferramentas em uso» no Arquivo (processadores de texto e bases de dados já existentes) e em outros ambientes futuros, bem como a criação de um módulo para compilação automática de índices (onomástico, toponímico, ideográfico, sistemático, etc.), por se tratar de uma tarefa trabalhosa na edição crítica de fontes históricas e arquivísticas.

A primeira fase do projecto concentrou-se na utilização de uma linguagem fácil para transcrição documental cuja descrição poderá ser encontrada em «HITEX. Um sistema para transcrição documental em larga escala» [PEREIRA, 1991]. De salientar que, na preparação do texto optou-se como ferramenta de base do sistema HITEX, o sistema LATEX, pelas seguintes razões:

- o LATEX baseia-se numa linguagem gráfica estruturada, de grande capacidade expressiva;
- o LATEX está disponível em praticamente todos os recursos computacionais, desde IBM-PC compatíveis aos ambientes Macintosh e Unix;
- o LATEX produz resultados gráficos (em TEX) de muito boa qualidade.

Escolhido o texto base passou-se à segunda fase do projecto, ou seja, à tarefa da *representação do conhecimento*, que neste sistema resultará nos índices, sendo estes uma representação (textual) de conhecimento histórico e/ou arquivístico. Os índices são, pois, produtos de um sistema de representação do conhecimento em que a sua geração automática será feita a partir de relatórios textuais de uma base de conhecimento previamente construída. Quer isto dizer, que a arquitectura da base de conhecimento do HITEX assemelha-se a uma rede semântica, com classes e subclasses, ou seja, a uma taxonomia de classes. Na sua feitura, a classe é caracterizada por atributos (propriedades), que serão dados a arquivar sobre os componentes dessa classe, também designados por instâncias. Nesta estrutura interligada, ou em rede, que é a base do HITEX, reside uma grande economia e flexibilidade do módulo, em virtude de todo o tipo de informação ficar automaticamente disponível, podendo, se for caso disso, ser acrescentado pelo seu autor/produtor.

Do ponto de vista do utilizador, a interface do protótipo permite basicamente realizar as seguintes tarefas:

- transcrição de fontes documentais;
- acesso ao *facsimile* do documento;
- marcação do texto com a identificação de entradas no índice e descrição das ditas marcas.

As marcas a que nos referimos correspondem aos elementos (*instâncias*) da base de conhecimento e a sua descrição corresponde ao preenchimento das suas características (*atributos*), pelo que

todas estas tarefas visam a construção e enriquecimento da base de conhecimento.

Posteriormente, o utilizador poderá gerir esta base de conhecimento através de actividades como:

- navegação sobre o diagrama da taxonomia do sistema;
- procura de componentes;
- criação de um novo componente da classe correntemente seleccionada na taxonomia;
- preenchimento de novos dados referentes ao componente correntemente seleccionado;
- especialização de um dado componente;
- adição de uma nova classe à taxonomia;
- adição de mais atributos a uma dada classe;
- interpelação assistida da base de conhecimentos; etc.

As duas figuras seguintes mostram ecrãs do protótipo disponível em MS-DOS. A janela superior esquerda mostra a taxonomia do sistema. Nas janelas laterais faz-se o acesso e a gestão de atributos e instâncias. Toda a conversação com o utilizador é feita através de menus. A janela inferior tem a ver com transcrição de fontes. Esta janela não é mais do que um simples editor de texto, capaz de comunicar com a base de conhecimentos sempre que se quiser. É de notar que uma fonte pode ser transcrita usando outros editores de texto, sendo posteriormente importada para esta janela com facilidade desde que não contenha caracteres de controlo obscuros. Este editor tem a vantagem de permitir referências cruzadas em modo semi-automático, ou seja, deixando o cursor na posição pretendida o utilizador pode pesquisar a base procurando os componentes que lhe interessarem. Um simples toque no rato será suficiente para que os identificadores dos componentes seleccionados sejam inseridos automaticamente na janela de texto, imediatamente após a posição do cursor. *Relembra-se que tais identificadores têm por objectivo marcar a ocorrência de itens que se destinam a índices.*



Figura 1 – Aspecto do ecrã no tratamento de uma fonte documental

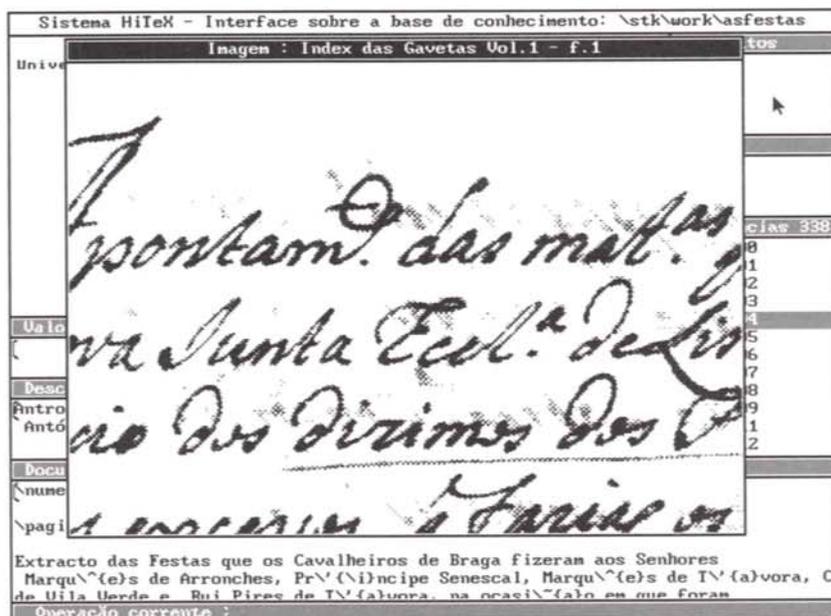
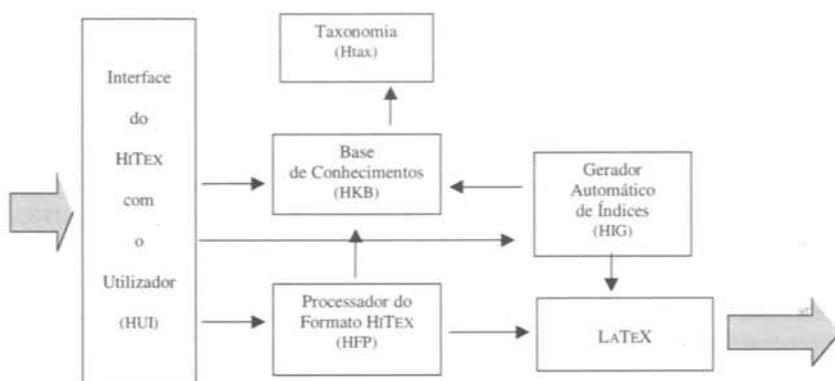


Figura 2 – Ecrã após ter recebido a imagem do índice das Gavetas do Cabido

Após a explanação da interface com o utilizador poder-se-á fazer a seguinte análise de *como* é o motor do sistema HITEX.



A figura representa o bloco estruturado do actual sistema HITEX. O bloco de Interface do HITEX com o Utilizador (HUI) é responsável pela gestão da interacção com o utilizador. Grande parte dessa interacção tem a ver com a criação e manutenção da Base de Conhecimentos (HKB) e correspondente taxonomia (HTAX) onde, como se viu, toda a informação é registada.

Os outros blocos estruturais do HITEX têm a ver com o tratamento de fontes (transcrição documental). HFP é o Processador do Formato HITEX atrás referido, produzindo código. Finalmente, o bloco Gerador Automático de Índices (HIG) pode ser activado para se gerarem automaticamente os índices toponímico, cronológico, etc. Mais uma vez, o texto dos índices é produzido em LATEX, e «colado» – se pretendido – ao texto já processado, para tudo ser impresso.

Em 1994 o projecto HITEX chegou ao fim da sua segunda fase de desenvolvimento após a construção de um protótipo escrito em SMALLTALK e C. Embora a interface com o utilizador esteja bastante mais aperfeiçoada do que a de um protótipo anterior em XMETOO, actualmente, com as novas tecnologias informáticas com grande incidência da interface gráfica, torna-se indispensável a readaptação do projecto, tanto no que diz respeito ao tratamento das imagens digitalizadas como no tamanho da base de conhecimentos.

Uma terceira fase do projecto está actualmente em desenvolvimento após o teste extensivo e avaliação pela equipa de historiadores do projecto.

Prevê-se a necessidade de aumentar a capacidade de dedução automática do sistema, assim como melhoramentos na produção de

índices textuais (a geração automática de linguagem natural requerá investimento adicional em linguística computacional).

Uma novidade desta integração consistirá na aplicação sistemática de um cálculo de implementação, desenvolvido independentemente pelo autor, com o objectivo de colmatar o desafio implementacional do HITEX.

Em 1995 desenvolveu-se o projecto HINET de divulgação do Arquivo Distrital de Braga na Internet, através de protocolo de colaboração com o Centro de Comunicações da Universidade do Minho e a PORTUGAL TELECOM, que visava a interligação do HITEX com a Internet.

Este projecto impulsionou o Arquivo Distrital de Braga a nível nacional e internacional, levando a um novo envolvimento no recente projecto GeIRA.

GeIRA: um sistema para divulgação cultural e científica

Duas grandes linhas orientadoras motivaram o nascimento do projecto GeIRA: a sociedade de informação, e a realidade social, económica e cultural da Euro-região Norte de Portugal – Galiza. Nesta perspectiva, o GeIRA é a ciência e tecnologia ao serviço da sociedade, afirmando-se como um sistema de difusão de informação apoiado na Internet – advento da sociedade de informação – e em outros meios de difusão de informação multimédia.

O projecto GeIRA arrancou em Janeiro de 1997 com financiamento atribuído pelo programa INTERREG II, o qual permitiu viabilizar, na zona de intervenção Portuguesa (Norte de Portugal), uma parte significativa da instalação de uma infra-estrutura distribuída de serviços, apoiada em equipas pluridisciplinares de investigadores de duas Universidades do Norte de Portugal: a Universidade do Minho (UM) e a Universidade de Trás-os-Montes e Alto Douro (UTAD).

O GeIRA é um projecto de desenvolvimento de serviços de informação multimédia e interactivos relativos ao património Português, sendo a área geográfica seleccionada para o seu arranque o Norte de Portugal. As principais vertentes de actuação do GeIRA são as seguintes:

- divulgar o potencial científico e tecnológico da região, em termos de pessoas, projectos de ensino e de I&D, e Instituições, e em especial ao nível das Instituições de Ensino Superior;

- utilizar as ciências – da informação e da comunicação – e as tecnologias – da multimédia e da interactividade – na valorização do património cultural e no estímulo à protecção e conservação do ambiente.

Os formatos de divulgação multimédia previstos no projecto baseiam-se essencialmente na utilização de servidores Web na Internet e na produção de CD-ROM's. Os servidores de informação na Internet são baseados em bases de dados integradas ou em cooperação, que suportarão formas de pesquisa múltipla por toda a informação existente no GeIRA, incluindo a pesquisa por texto livre, por intervalos de tempo, ou ainda por coordenadas geográficas (informação geo-referenciada).

No sentido de garantir adequados níveis de qualidade na execução do projecto, este foi estruturado em várias áreas de intervenção temática – de acordo com os objectivos e com as necessidades de funcionamento – nomeadamente Ciência e Tecnologia, Arqueologia, Arquivos, Bibliotecas, Museus e Natura.

GeIRA-Arquivos, HITEX e o futuro...

No âmbito do projecto GeIRA, o Arquivo Distrital de Braga assumiu a organização da equipa de trabalho da Área de Intervenção GeIRA-Arquivos na Universidade do Minho, através de protocolo com a Comissão Executiva do GeIRA.

Partindo das infra-estruturas, experiência e programas de pesquisa iniciados com o projecto HITEX e já utilizados no projecto HINET, o Arquivo Distrital de Braga propôs-se rentabilizar ainda mais o esforço dispendido, sobretudo, na caracterização e estruturação das bases de dados de informação arquivística, visando a concepção de normas nacionais de estruturação da informação via Internet e CD-ROM.

Um dos objectivos principais deste projecto que visa rentabilizar o estudo efectuado pela investigação anterior relacionada com o projecto HITEX é a concepção do sistema «HITEX++», uma nova versão para funcionamento em suporte www, como projecto de estágio de licenciatura.

Para além de:

- conceber um sistema de reconhecimento e cruzamento de informação retida nas diversas bases de dados já construídas

- no ADB e/ou Arquivos Municipais, com os quais o ADB tem protocolo;
- conceber uma forma de enriquecer toda a informação disponível através da associação de fac-símiles da documentação microfilmada;
 - elaborar um sistema de *compatibilização* da informação estruturada já existente com o acesso via Internet e/ou CD-ROM.

Após um primeiro ano de trabalho poder-se-á fazer um balanço positivo, tendo sido feita a reestruturação da informação arquivística existente sobre o Arquivo Distrital de Braga, tendo-se criado uma identidade própria www.adb.pt e novos serviços na Internet, com especial relevo na edição gráfica de livros digitais (o livro *Memórias Particulares* de José Inácio Peixoto editado em 1992 pelo ADB) e no tratamento de imagens em microfimes.

No que diz respeito a este último ponto – o tratamento de imagens em microfimes – a Área de Intervenção GeIRA-Arquivos adquiriu uma máquina digitalizadora de microfimes – ABR 3000D da BELL+HOWELL – com a qual prevê a digitalização integral dos microfimes do Registo Paroquial e de outros.

Estas imagens serão futuramente integradas no sistema «HITEX++» para pesquisa via Internet e CD-ROM.

BIBLIOGRAFIA

- AMADO, J. P.; CARDOSO, J. C.; NEVES, A. M.
Heródoto 1.0 Estação de Trabalho Informática em História e Arquivística. Lisboa: Fac. de Ciências Sociais e Humanas da U. N., 1990.
- ANDRADE, Pedro Luís
Implementação da Base de Conhecimento do HITEX em SMALLTALK. Braga: U. Minho, 1994. Relatório de Estágio, LESI.
- BARROCA, L.
Object-oriented Database Design in Archaeology. [Braga]: Dept. of Computer Science, University of Minho, 1991.
- GULAMHUSSEN, Asha
Desenvolvimento de uma base de dados relacional para o projecto HITEX. Braga: U. Minho, 1993. Relatório de Estágio, LMCC.
- MOREIRA, Clara Sofia
HITEX: A Integração e Teste do Sistema. Braga: U. Minho, 1994. Relatório de Estágio, LMCC.
- OLIVEIRA, J. N.; ARAÚJO, A. S.; SILVA, A. M.
«Historical Records Processing in the HITEX System». To appear in *Proc. of AHC'91 6th International Conference of the Association for History & Computing*. Odense, Denmark, Aug. 28-30, 1991.

Projectos Informáticos do Arquivo Distrital de Braga

OLIVEIRA, José Nuno

«HITEX: um sistema em desenvolvimento para historiadores e arquivistas», *Forum* (15/16), Braga, 1994, pp. 123-137.

Projecto HITEX. 4.º Relatório de progresso. Arquivo Distrital de Braga/Universidade do Minho, Agosto 1993.

«A Reification Calculus for Model-Oriented Software Specification», *Formal Aspects of Computing*, V. 2, (1-23) Springer-Verlag, 1990.

PEREIRA, António Manuel

HITEX: Um Sistema para Transcrição Documental em Larga Escala. Braga: U. Minho, 1991. Relatório de Estágio, LESI.

THALLER, M.

The Need for Standards: Data Modelling and Exchange in Modelling Historical Data: Towards a Standard for Encoding and Exchanging Machine Readable Texts, 1-18, Greenstein D.I. (Ed.), Max-Planck-Institut für Geschichte, St. Katharinen, 1991.

MARIA DA ASSUNÇÃO JÁCOME DE VASCONCELOS
CLARA SOFIA DA ROCHA PINTO MOREIRA
Arquivo Distrital de Braga

SANTOS JÚNIOR & FILHOS, LDA.

Ao serviço de bibliotecas, arquivos e museus



A primeira firma do país fornecedora de material
e equipamento para conservação, preservação e
restauro de obras de arte.

* * *

Rua da República da Bolívia, 22 – 5.º F
Telfs. (01) 778 89 51 / 715 51 57 – Fax (01) 715 51 57
1500 Lisboa – Portugal

Normas para publicação de artigos

PÁGINAS a&b aceitam artigos para publicação escritos em Português, Inglês, Francês ou Espanhol, de preferência originais, e desde que os mesmos não tenham sido submetidos simultaneamente para apreciação junto de outro Editor.

Qualquer que seja a decisão do Conselho Redactorial de *PÁGINAS a&b*, sobre o artigo submetido para apreciação, o(s) autor(es) será(ão) devidamente informado(s).

Os direitos de reprodução dos artigos publicados passarão a ser propriedade de *PÁGINAS a&b*.

Os textos submetidos a apreciação deverão ser acompanhados por disquete e, se o texto contiver gráficos, tabelas, quadros, etc., estes deverão ser fornecidos à parte.

As referências bibliográficas deverão seguir a NP 405-1 e, no caso de não estarem conforme esta, *PÁGINAS a&b* reservam-se o direito de introduzir as alterações necessárias.

A revisão de provas é da responsabilidade de *PÁGINAS a&b* e só em caso de solicitação específica aquelas serão entregues ao(s) autor(es).

A colaboração com *PÁGINAS a&b* é sempre a título gratuito mas o(s) autor(es) receberá(ão) 15 separatas do(s) artigo(s) publicado(s).

UA/SD	
N.º	6523/20
Data	02.05.9
Cota	BT

de

Publicação patrocinada pelo Instituto Português do Livro e das Bibliotecas



INSTITUTO PORTUGUÊS DO
LIVRO E DAS BIBLIOTECAS



MINISTÉRIO DA CULTURA