

Universiteit Utrecht



# MEDOC

A Method for Digitising and Disclosing Special Collections  
Version 1.0

Frits Ankersmit and Hans Mulder



# Contents

|  |    |
|--|----|
| Preface .....  | 2  |
| 2. MEDOC VERSION 1.0 .....   | 4  |
| 2.1 Introduction .....   | 4  |
| 2.2 Outline.....   | 5  |
| 3. Composing the Project Group.....  | 6  |
| 4. What: Specifying the Collection .....   | 7  |
| 4.1 Introduction .....   | 7  |
| 4.2 Subject .....  | 7  |
| 4.3 Character and extent .....   | 7  |
| 4.4. Information about the collection .....  | 8  |
| 4.5 Results.....   | 8  |
| 5. Why: Reasons for Digitising and Disclosing the Collection .....                   | 9  |
| 5.1 Introduction .....   | 9  |
| 5.2 What is the social and cultural significance of the collection? .....            | 9  |
| 5.3 What is the present importance of the collection?.....                           | 10 |
| 5.4 What is the role of the collection in relation to research and education? .....  | 10 |
| 5.5 Is the collection accessible? .....  | 11 |
| 5.6 Remaining questions.....   | 11 |
| 5.7 Is the collection valuable in the sense of PR? .....                             | 11 |
| 5.8 Results and decisive discussions .....   | 12 |
| 6 How: Programme of requirements .....   | 13 |
| 6.1 Introduction .....   | 13 |
| 6.2 Determining present use and users, defining functionality.....                   | 15 |
| 6.3 Defining new audiences and new use, defining functionality.....                  | 18 |
| 6.4 Rendering functionality into quality requirements regarding digital sources..... | 18 |
| 6.4.1 What is quality? .....   | 18 |
| 6.5 Reiteration .....  | 20 |
| 6.6 Literature study.....  | 20 |
| 6.7 Representative selection from the collection .....                               | 21 |
| 6.8 Testing and benchmarking .....   | 21 |
| 7. How: Resources .....  | 23 |
| 7.1 Introduction .....   | 23 |
| 7.2 Determining the production elements.....   | 23 |
| 7.3 Analysis of technical preconditions.....   | 25 |
| 8. Evaluation .....  | 26 |
| 9 Acknowledgements .....   | 27 |
| 10 Appendices.....   | 29 |
| 10.1 Questionnaires .....  | 29 |
| 4.3 Character and extent of the collection .....                                     | 29 |
| 4.4. Information about the collection .....  | 29 |
| 5.1 Reasons for Digitising and Disclosing the Collection:Introduction.....           | 29 |
| 5.2 What is the social and cultural significance of the collection? .....            | 29 |
| 5.3 What is the present importance of the collection?.....                           | 30 |
| 5.4 What is the role of the collection in relation to research and education? .....  | 30 |
| 5.5 Is the collection accessible? .....  | 30 |
| 5.6 Remaining questions.....   | 30 |
| 5.7 Is the collection valuable in the sense of PR? .....                             | 31 |
| 6.2 Determining present use and users, defining functionality.....                   | 31 |

## Preface

Scientific libraries are the keepers of large collections comprising manuscripts as well as special and early printed books. Such libraries have an important role as guardians of our cultural heritage. Within this context, the collections usually have a significant function as material for education and research purposes.

The administration of these collections as valuable cultural objects usually collides with their practical use for education and research purposes. The public availability of a book does not quite enhance its life span. For this reason, there is a determined investigation of new ways for disclosing vulnerable and precious sources in order to make them accessible to the largest possible audience without endangering the original material itself.

Digital disclosure appears to be an adequate solution for this purpose. What is more, this solution also provides an extra value by enriching the original sources in various ways. This added value proves to be of great significance to researchers, tutors and students. For instance, the publication of text sources in a machine-readable version offers the possibility of browsing texts in a better, faster and different way.

Opposite to the large number of special collections in a library, usually stands a limited budget as well as variable interest from education and research departments. If it is considered to digitise a collection, it is necessary to come up with a sensible decision. MEDOC is specially designed to provide a basis for such a decision. It provides an instrument for obtaining this decision through a series of subsequent instructions.

MEDOC carries a version number. Constant application of the method in specifying requirements and (im)possibilities regarding envisaged projects should lead to revised versions containing modifications, improvements and additions. MEDOC has been primarily developed with regard to manuscripts and early printed books.

MEDOC distinguishes between digitising and enrichment. Digitisation involves a process during which a copy of the original sources is made. Enrichment deals with the organisation and addition of information. Both processes combined result in a digitally disclosed collection.

MEDOC focuses on the digitisation process. Naturally, there is an interactive relation between digitisation and enrichment. For this reason, MEDOC is also concerned with drawing up a programme of requirements regarding enrichment.

Realising a digital collection requires specific expertise. The added value of a digitised collection in the first place depends on input that is contributed by subject experts. Secondly, it also involves ICT skills. For a digital collection is in fact an information system that may have to be accessible via a network. Finally, understanding is required with regard to the interdependent factors like printed matter, colours, images and their representation on the computer screen. This knowledge is valuable in order to avoid distortion of the original collection.

MEDOC is in fact a manual for selecting and setting priorities with regard to the digitisation of entire collections as well as individual sources. In addition, MEDOC provides insight into the required resources. To this end, the method systematically specifies and arranges all knowledge, options and obstacles that relate to realising a digitally disclosed version of the collection. MEDOC gives a clear picture of the collection within its actual context. At the same time, it describes the nature and extent of the collection in detail. Furthermore, it delineates a method for disclosing and enriching material by means of a programme of requirements. Finally, it offers a realistic estimation of the resources that are required for this purpose. In short, MEDOC provides an answer to the questions *what*, *why* and *how*. The

application of the method to various collections implies that these collections can be compared with respect to aspects like significance, time, personnel and costs. Consequently, a well-considered decision can be made in selecting from various collections, even in case of a limited budget.

The essence of MEDOC comprises a series of questions that are put at the start of a digitisation project. For an outline of these questions, please refer to Appendix 1. In sections 4 to 8, the nature of the questions is explained in full.

MEDOC has been arranged for the circumstances at the Utrecht University Library. Nevertheless, the questions relating to Utrecht can easily be transferred to the context of other libraries and bodies that act as keepers of cultural objects.

## 2. MEDOC VERSION 1.0

### 2.1 Introduction

MEDOC provides an answer to the questions *what*, *why* and *how*. Each separate question consists of a number of related questions and examinations. MEDOC specifies these questions and reformulates them in the shape of various forms and steps. Determining the collection's characteristics and extent results in an answer to the *what* question. Within this context, special attention is given to the collection's physical aspects. Characteristics and condition of the collection determine various possibilities for developing digital copies.

The answer to the *why* question describes the cultural importance of the collection, its relation to other collections, its importance to research and education and, finally, to its disclosure.

The answer to the *how* question consists of a programme of requirements and of a listing of required resources.

The mode of operation is as follows. The three questions *what*, *why* and *how* are subdivided into relevant related questions and related examinations. These questions need to be answered successively and in a fixed sequence. Next, the answers and the examination results are listed in reports. These reports constitute the starting point for two decisive discussions on the question whether the examined and described collection should be actually disclosed.

The method involves a number of questionnaires and forms. These forms have been included into appendix 1. In a number of cases, a question may result in various answers. For instance, the quality of a digital copy is not determined by an absolute standard. This standard should meet the practical objectives that have been determined in advance. In case there are various answers, these can be indicated.

Applied to an extensive collection, the use of MEDOC requires a great deal of effort. Preferably, a single person does not conduct the activities. For this reason, the project usually starts with composing a group of experts and parties concerned. Together they are responsible for the specification stage, yet with respect to specific problems they each have their individual contributions. Organisation, objectives and activities of this project group are dealt with in section 3 *Composing the project group*.

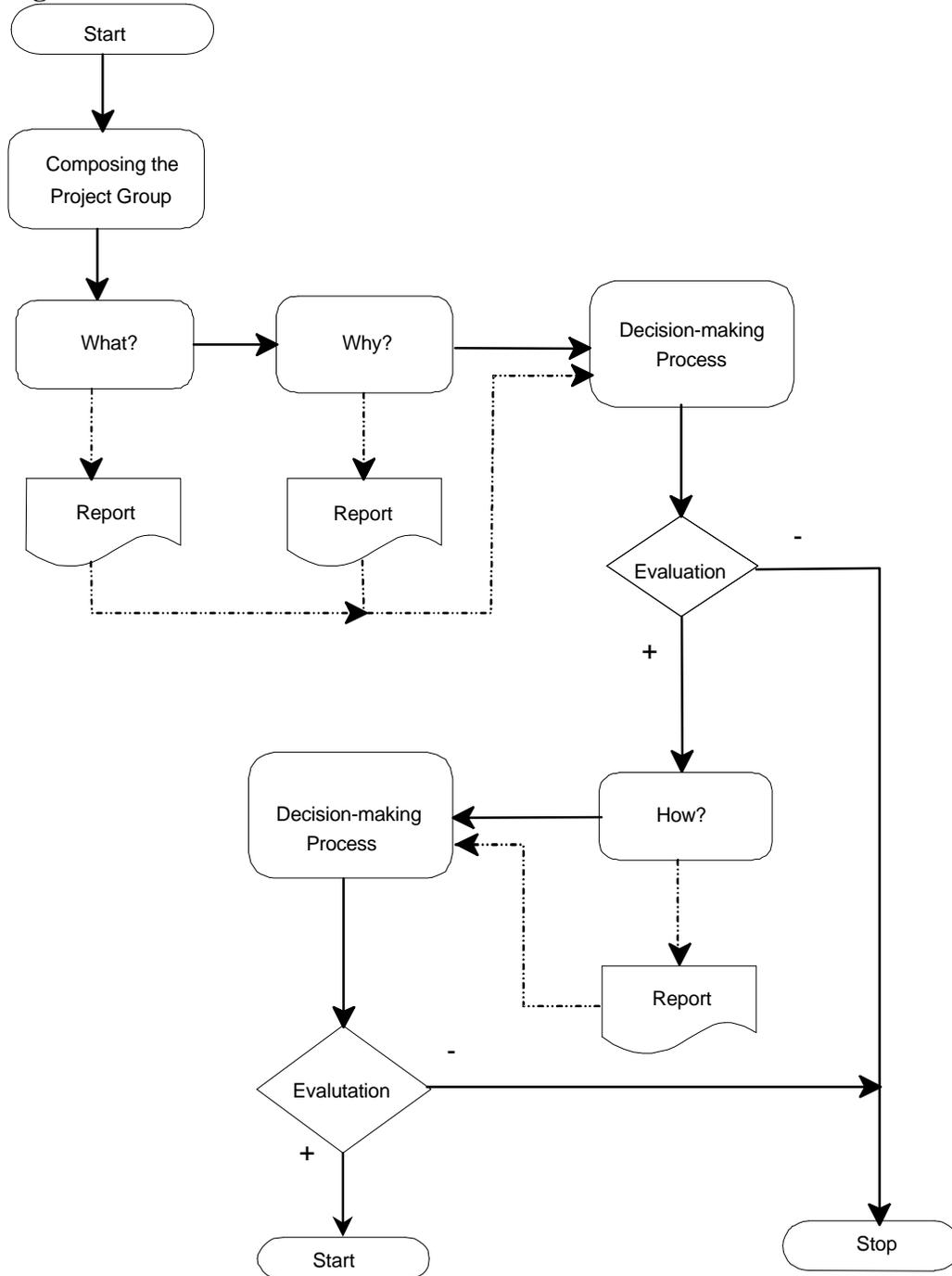
The answer to the *what* question is taken up in section 4 *What: Specifying the collection*. Section 5 *Why: Reasons for Digitisation and Disclosure* provides an answer to the *why* question.

The *how*-question has been split into various related questions and examinations. These are tackled in *Section 6 How: Programme of Requirements* and *Section 7 How: resources*. The last step of the method is concerned with evaluating the information that has been assembled by all the parties involved. This step is described in *Section 8 Evaluation*. The questions *what*, *why* and *how* are regarded successively and in this particular order. This order is determined by the fact that the answers to the former questions provide a starting point for answering the next series of questions.

## 2.2 Outline

Figure 1: Outline provides a graphical representation of the mode of operation within the MEDOC framework. The rectangles with rounded corners indicate the actions. The arrows designate the action sequence, the dotted lines the results. These results also serve as the starting points for various actions. The scheme also indicates which reports need to be made. Diamond shapes indicate decisive moments.

Figure 1 Outline



### 3. Composing the Project Group

The application of MEDOC is an extensive activity. An accurate and responsible start of digitisation projects requires a great deal of preparation. Preferably, a single person should not exclusively conduct the activities. Therefore, these projects usually start with organising a group of experts and parties concerned. Together they are responsible for specifying the collection. Yet, with respect to specific problems, they each have their individual contributions.

In the initial phase of the project, the following tasks need to be assigned. In case of a project on a smaller scale, a single person could carry out several tasks.

**Management:** organisation and communication must be clearly set up. Main and related items should be separated from each other. Control of the project budget must be in the hands of the management. The person in charge of this task does not necessarily have to be an expert in the field. Over-competence of the project leader in this respect may lead to possible over-focusing on content. This could affect negatively the controlling of the main issues and thus jeopardise the project's progress and objectives.

**Administration:** Keeping record of arrangements and describing the project's progress are imperative. On the one side, this is necessary for the sake of internal reporting, i.e. to the members of the project group. On the other side, this is required in view of external reporting that is to the authority who commissioned the project. In addition, these activities are indispensable for evaluating the project and for assessing whether and where correct or incorrect decisions have been made.

**Expert on the subject of the collection:** this person is to decide whether a collection is suitable for digitisation. In a library, this usually concerns a curator or reference librarian. Selection of the material involves physical criteria as well as criteria dealing with content. During the course of the project, the expert on the subject matter is in charge of the physical processing of the material. He needs to check regularly whether the digital version actually meets the predetermined requirements.

**Digitisation expert:** this person is to verify if a collection is suitable for digitisation. Amongst others, this person is able to determine whether Optical Character Recognition (OCR) can be applied. The digitisation expert is responsible for the digitisation process. He is also in charge with regularly checking the digitised material.

After assessing the *what* and *why* questions, the project group can be enlarged. If a tutorial component is added to the project, it is necessary to expand the project group by inviting an education expert.

Within a project group, the required expertise will never be fully available. It may therefore be necessary to consult external experts. These can be temporarily added to the team.

## 4. What: Specifying the Collection

### 4.1 Introduction

The questions of the collection's character and extent need to be answered accurately. Within MEDOC, the term 'collection' can be used in a broader sense. MEDOC can also be applied to a single object. The answer should then concentrate on:

- the subject
- the character and extent
- all information known about the collection
- all information known about the material with which the collection is to be disclosed.

From these answers a clear picture of the collection must arise. Meanwhile, information must also be gathered in order to answer the *how* question. The answers to the *what* questions provide the necessary input for the next steps in the MEDOC project.

### 4.2 Subject

First, the collection's subject must be determined. It is advisable that the description of the collection fits on a single A4. Nevertheless, such a concise description should evoke a clear picture of the collection for any outsider.

### 4.3 Character and extent

The second question that must be answered is as follows: 'What is the exact extent of the collection?' As to this aspect, being over-complete is virtually impossible. For each element of the collection, a number of questions need to be clarified. Determining the collection's actual extent is the first step towards accurately estimating the required amount of resources. By doing so, it becomes clear whether a collection has a heterogeneous nature. Such a heterogeneous collection contains both similar and non-similar elements. Heterogeneity can greatly affect the digitisation process. For instance, if a collection consists of a number of atlases that extremely vary in size, one constantly will have to enter new settings during the digitisation process. This will have an adverse effect on production time. Such unpleasant surprises can be avoided by gathering information on the collection's character and extent in the initial stage of the project. For each element of the collection, the following questions must be answered:

1. What is the overall description (amount of volumes, structure, table of contents, indices)?
2. What is the number of pages?
3. What is the page size (height and width in mm)?
4. What is the subject matter (characteristics and/or illustrations)?
5. Has any colour been used?
6. Which alphabet and/or scripture types are involved?
7. What type of material is involved (printed matter or manuscript)?
8. Are there any supplement pages of different size (number and position)?
9. Are any tables involved (number and position)?
10. Are any illustrations (number and position) involved?
11. Which language(s) are involved?

## **Table 1**

The answers to these questions should provide ample insight into the collection's character and extent. In this way, data are gathered that are necessary for drawing up both production plan and budget. Nevertheless, the required functionality of the digital collection must be clear first.

### ***4.4. Information about the collection***

The way in which a project is described is significant for preparing the programme of requirements as well as for planning the production. In many cases, catalogues, handbooks and publications can be used for enriching digital copies, thus adding extra value to them. Moreover, the disclosure mechanism can prove instrumental for organising digital copies. If a collection catalogue is available, it could serve as a (digital) entry to individual (digital) works. First, a listing must be drawn comprising the complete description of the collection. Subsequently, for all items in these listings, the questions in table 2 must be answered. Table 2 provides an overview of all questions that need to be answered.

- |  |
|--|
| <ol style="list-style-type: none"><li>1. Is the description complete and up to date?</li><li>2. What parts make up the title description?</li><li>3. Is a digital description available (on-line catalogue)?</li></ol> |
|--|

## **Table 2**

### ***4.5 Results***

The results of the specification are published in a report that accurately describes the collection's subject matter. Both character and extent are known and there is an overview indicating the method for disclosing the collection. The report serves as a reference for all parties concerned during every subsequent step. If in a later phase one has to determine how the digital collection is to be disclosed, this report provides all necessary information for making useful estimations about the cost and duration involved in its digital disclosure.

## 5. Why: Reasons for Digitising and Disclosing the Collection

### 5.1 Introduction

This stage in MEDOC results in the answer to the question *why* a particular collection should be digitally disclosed. Table 3 lists all the questions with respect to this aspect. The answers to these questions should be stated in a concise way in order to stress their argumentative nature. They are of notable interest if a selection is to be made from various digitisation projects.

The next sections contain a further explanation to these questions. They also include a number of additional questions providing an extension to the questions in table 3. The questions in table 3 have not been arranged to a particular order.

The significance of a collection is partly determined by its users. It is important that the users are identified in an early stage, in order to engage them into the process in some way, e.g. as a response group. Users generally have a different commitment to the material than the library. Therefore, they can probably add other positive arguments with respect to the collection's interest.

The first four questions in table 3 provide an opportunity for involving the users into the process. Questions 6 and 7 are specifically concerned with applying MEDOC to the Utrecht University Library.

- |  |
|--|
| <ol style="list-style-type: none"><li>1. What is the social and cultural significance of the collection?</li><li>2. What is the present importance of the collection?</li><li>3. What role has the collection in relation to research and education?</li><li>4. Is the collection accessible?</li><li>5. What is the physical condition of the collection?</li><li>6. Does the collection deal with Utrecht as a subject?</li><li>7. Does the collection's subject concern Utrecht University?</li><li>8. Is the collection valuable in the sense of PR aspects?</li></ol> |
|--|

**Table 3**

### 5.2 *What is the social and cultural significance of the collection?*

The cultural value of a collection is always relative. By means of specific questions, a collection can be clearly positioned as to other collections. In this respect, a special collection may distinguish itself from others or constitute a precious addition to similar collections.

The question concerning the cultural importance of a collection can therefore be differentiated into a number of more precise questions. These questions are listed in table 4.

The answers to these questions indicate the value of a particular collection in relation to other special collections in the library or to other or similar collections nation-wide or abroad. Simultaneously, as an additional advantage it should also become clear whether it is feasible to link the project to similar initiatives. This possibility would certainly enlarge the support for digitisation.

- |  |
|--|
| <ol style="list-style-type: none"><li>1. Can similar collections be found in the Netherlands?</li><li>2. Are there any similar collections abroad?</li></ol> |
|--|

- |  |
|--|
| <ol style="list-style-type: none"><li>3. In what way is this particular collection related to others managed by our library?</li><li>4. Have any initiatives been launched to digitally disclose these collections?</li><li>5. If so, is co-operation a feasible option?</li><li>6. Is the collection valuable as a work of reference?</li></ol> |
|--|

**Table 4**

### ***5.3 What is the present importance of the collection?***

Due to various reasons, library collections that have been stacked away for years in depositories can suddenly become the focal point of attention. For instance, the commemoration of a historic event could become a cause for placing a particular collection into the limelight. Linking the collection to topical aspects may facilitate fund-raising and increase its value in a PR sense. One of the topical aspects could be the fact that research and education are provided by the institute's proper facilities. In some cases, a department could request for a collection (or part of it) to be disclosed digitally. The questions that need to be answered in this case are:

- |   |
|---|
| <ol style="list-style-type: none"><li>1. Is the collection connected in a significant way to present education and research?</li><li>2. Is there any social interest that would support digitisation of a collection?</li></ol> |
|---|

**Table 5**

### ***5.4 What is the role of the collection in relation to research and education?***

The way the collection is put to use needs to be carefully examined. In this respect, the focus is on the collection's importance in relation to research and education and on its reference value. One of the major advantages of a digital document is that it offers simultaneous accessibility to multiple users at different places. Its extra value is also enhanced if various users from different disciplines use it.

However, the significance of a particular collection usually exceeds the limits of its actual audience. Although it is mainly intended for research and education at the university to which it belongs, an extraordinary collection usually also serves national and even international interest.

Vulnerable collections that are difficult to access can be very effectively used for education purposes if they are disclosed in digital form. Naturally, for the Utrecht University Library, research and education benefits for the university's proper staff and students are a decisive factor. With respect to this aspect, the following questions need to be addressed.

- |  |
|--|
| <ol style="list-style-type: none"><li>1. Is the collection of general importance as to education and research?</li><li>2. Does the collection relate to research and education at Utrecht University?</li><li>3. Does the collection acquire significant extra value by presenting it in digital form (accessibility, disclosure through information retrieval)?</li></ol> |
|--|

**Table 6**

### ***5.5 Is the collection accessible?***

The university library possesses rare collections, which in some cases contain unique material. These special collections can be studied under supervision. For valuable and vulnerable material, a policy of restricted disclosure must be observed. Digitising this material would certainly improve its availability. This is especially the case if the physical condition of the material does not allow for unlimited public examination. In this case, digital disclosure could be an alternative that would also serve conservation. Table 7 therefore lists the following questions:

- |  |
|--|
| <ol style="list-style-type: none"><li>1. What is the physical condition of the material?</li><li>2. Does the physical condition allow for public availability?</li><li>3. Does digital disclosure constitute an alternative for public availability of the original material?</li><li>4. Does the collection have a high priority in the conservation plans of the library?</li><li>5. Is digitising (part of) the material a useful alternative for other conservation methods?</li></ol> |
|--|

**Table 7**

### ***5.6 Remaining questions***

A number of *why* question specifically apply to a particular collection. For instance in the case of the Utrecht University Library, this especially goes for the collections that are concerned with the city and the province of Utrecht as a subject. A great deal of this material can only be found at the Utrecht University Library. These collections, which in many cases are unique, therefore make excellent candidates for digital disclosure. As to Utrecht, the next questions need to be answered:

- |  |
|--|
| <ol style="list-style-type: none"><li>1. Does the collection relate to the city and/or province of Utrecht?</li><li>2. Does the collection deal with the Utrecht University?</li><li>3. Did the collection belong to a (former) scholar of Utrecht University or does the material refer to this person?</li></ol> |
|--|

**Table 8**

### ***5.7 Is the collection valuable in the sense of PR?***

Section 5.3 mentioned the possibility of fund-raising in case the (digital) collection is valuable in a PR sense. Since digitisation projects are costly enterprises, external financing may prove necessary. If a digitisation projects fits into the PR policy of the university library and or university, fund-raising would probably become a less difficult matter. The question that needs to be addressed to here is:

- |  |
|--|
| <ol style="list-style-type: none"><li>1. Is digital disclosure in compliance with the PR policy of the Utrecht University Library and/or Utrecht University?</li></ol> |
|--|

**Table 9**

## ***5.8 Results and decisive discussions***

Answering all these questions results in a report that in the evaluation phase should serve as a basis for positioning the examined collection in relation to other collections that have also been nominated for digitisation and disclosure. Thus, it is possible to evaluate the digitisation and disclosure of this particular collection in relation to the general policy followed by the library. In addition, various arguments have been collected that specifically apply to the described collection. If the report needs to be useful in the decision-making process, it is important that the answers to the above-mentioned questions are concise. In this respect, a summary could also be a useful means for preparing the discussion.

On the basis of the available material, the assembled project group needs to determine if the project is to be continued or stopped. If there are various collections to be considered, it is possible to make an initial selection based on the same criteria and information. In this case, it could be decided to continue with a single collection, whilst excluding the other collections from the decision-making process.

In some cases, distinguishing between an internal and external context may enhance the discussion. Within a broader or external context, there could arise arguments that do not specifically apply to the institute in question. In this case, the collection is regarded within a wider framework. Then, the more general significance of digital disclosure is examined.

Within the external context, the collection is investigated for its special quality, taking into consideration its connection with other national and international collections, its topicality, its uniqueness, its use for research and education purposes and the possible necessity for conservation.

The internal, narrow context consists of arguments referring to collections that have a specific connection with the university proper. For instance, the Utrecht University Library has its roots in the region of Utrecht. Its special collections originate from ecclesiastical and convent libraries found in the city of Utrecht. For our library, the collections that are related to Utrecht hold an extra intrinsic value. This value could become a positive argument within the decision-making process.

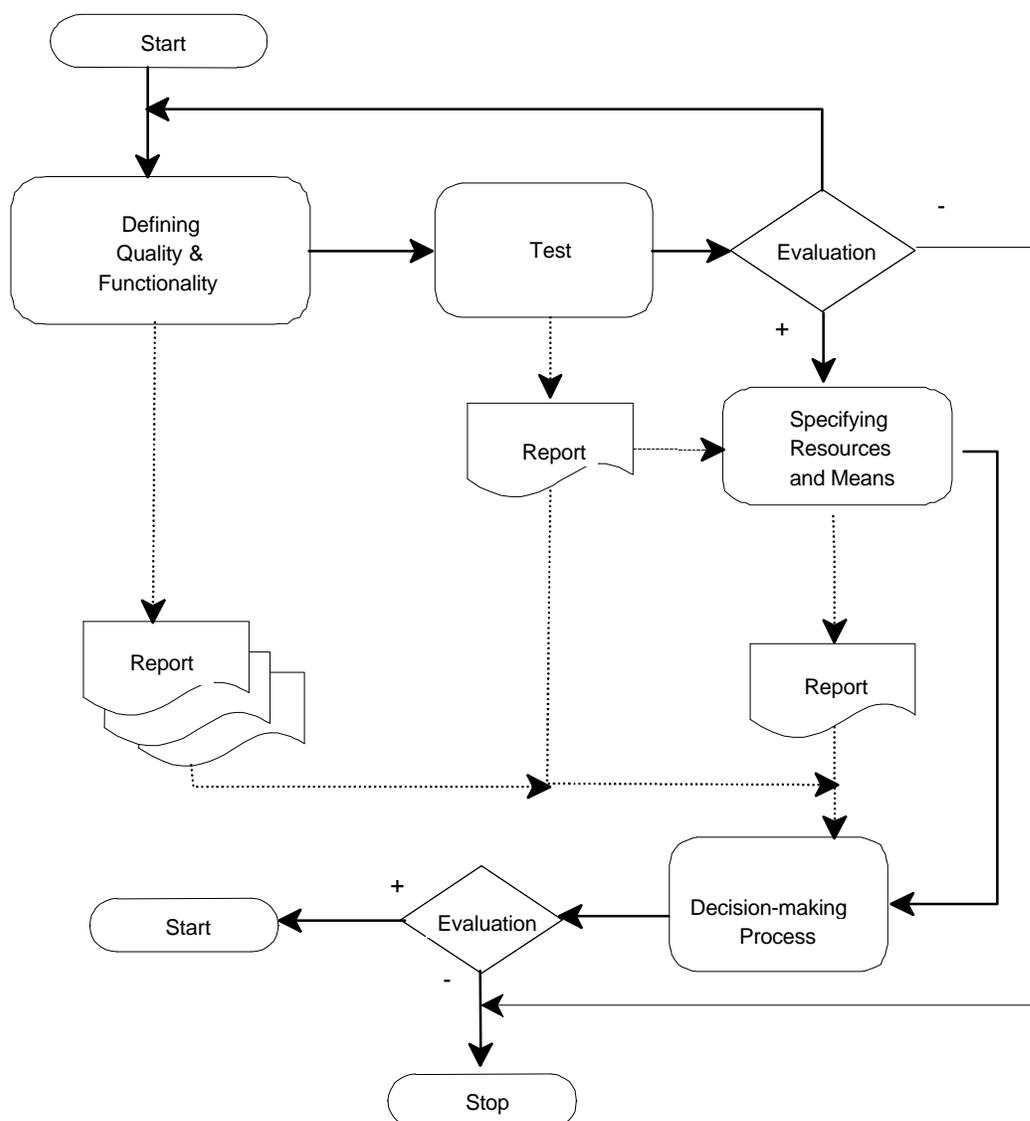
Nevertheless, during the discussion on priorities, it could prove difficult to compare the pro and counter arguments in relation to each other. In the end, the final decision could therefore prove arbitrary to a certain extent.

## 6 How: Programme of requirements

### 6.1 Introduction

The *how* question can be divided into two sub-questions. The first sub-question is restricted to the aspect of contents regarding the digital disclosure of the collection. The result of this question is a programme of requirements referring to quality and functionality. This programme of requirements serves as a foundation for the second sub-question. The second question is concerned with the required resources: which persons and which means are necessary for disclosing the material. The first sub-question is dealt with in the present section, the second one in section 7: *How: Resources*. The entire *how* question is summarised in figure 2 *How: Quality, Functionality, Resources*.

**Figure 2** *How: Quality, Functionality, Resources*



If a digitally disclosed source has to be an actual alternative for the original material, it must offer sufficient quality and functionality. In addition, both these aspects need to be adapted to different practical circumstances. The project group therefore has to examine the various ways in which the material is presently applied. Furthermore, future ways of its use and the presentation of the material on screen must be considered as well. This presentation has to provide the user with sufficient insight into the nature of the original.

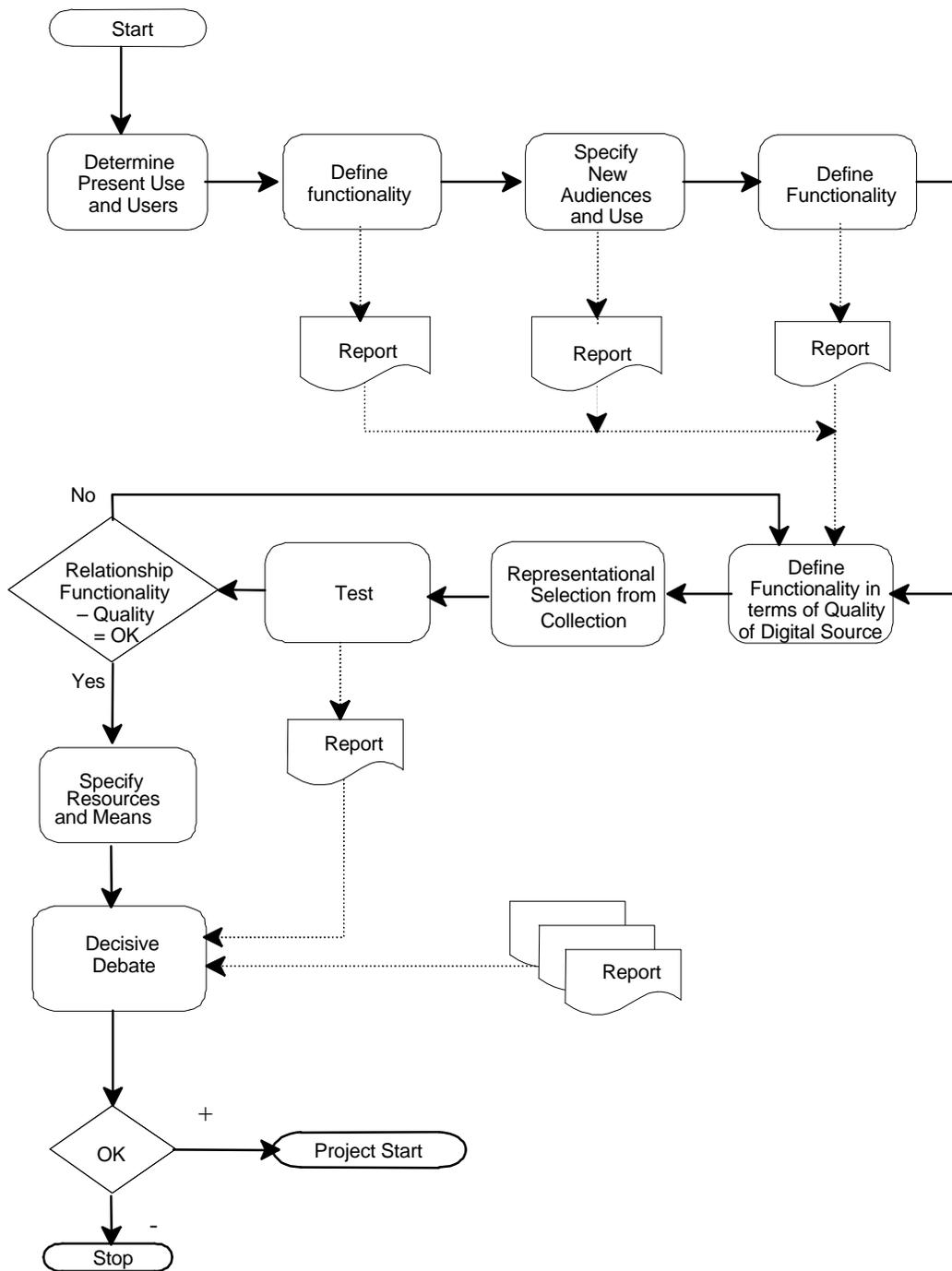
Quality and functionality are defined in interaction with each other. For instance, if the legibility of the disclosed source is the only functional criterion, black-and-white representation may be a reasonable option. If, however, colour and texture are crucial for studying purposes, an entirely different representation must be developed. In this case, natural colours and profound details are among the primary requirements. Additionally, other functionality must consequently be realised. The user will want to verify for instance the use of colours by means of calibrated colour samples. Within this context, the point of departure is the relationship between functionality and the user. If a collection, or a part of it, is to be digitally disclosed and subsequently presented, the involved requirements must be defined in connection with the public for whom the material is intended. Generally, this concerns an audience who will gain access to the information via a computer screen. In practice, they mostly are clients of the university library who need the material for education of research purposes. The model user serves as a reference for specifying the functionality and the quality of the sources in their digital form. Identifying this reference is the first step to be taken. Next, it must be examined how these requirements, specifications and functionality are to be translated into the quality standards for the digitisation of the collection. On the one side, these standards regard both colour and detail. On the other side, they concern structure and enrichment.

In this phase of MEDOC, one or more tests are performed in order to determine whether the required functionality can be linked to what is feasibly obtained with respect to the material. The importance of this test is stressed here in order to point out the unique character of the collection. However, a literature study could suffice if any test information on other, similar collections, is available. Nevertheless, one should be very accurate in this respect.

Figure 3: *How: Functionality and Quality* provides an overview of the various steps that are necessary for answering the *how* question. Figure 3 sums up the details that are contained in the *how* block in Figure 1: *Overview*. In the next sections, these steps are worked out in full.

**Figure 3: *How: Functionality and Quality***

+



## 6.2 Determining present use and users, defining functionality

Specifying present use and users can be disregarded, if we are dealing with a collection that no longer is looked into or is no longer public. If this is the case, section 6.2 can be skipped and we can proceed with section 6.3 *Defining New Audiences and New Use, Defining Functionality*.

There are various ways for specifying the present users of a collection. The most useful source of information usually consists of persons who are in direct touch with these users, i.e. reference librarians and collection curators. Nevertheless, it should also be possible to address

the users directly. This can be achieved by having them represented in the project or response group.

An indication for identifying the users could be provided by the answer to the question in section 5.4 *What role has the collection in relation to research and education?* This answer could actually name these users.

For an extensive insight into users and audience, the questions in table 10 must be answered first:

|  |
|--|
| <ol style="list-style-type: none"><li>1. Who are the users?</li><li>2. How large is the user group?</li><li>3. What are their objectives in using the material?</li><li>4. What sort of enrichment is required for optimally supporting the users in achieving their present goals?</li><li>5. Which problems do they experience in using the collection?</li><li>6. Have the users any special wishes with respect to the collection?</li><li>7. What kind of research and education involves the collection?</li></ol> |
|--|

**Table 10**

In this case, it is important to apply a generic mode of operation. The way in which a collection is used should never be determined by findings that merely relate to a single user. These findings should always relate to an entire user group.

The answers that are mentioned below are preferably worked out in an interview. In the interview, the questions and the subsequent remarks are used as indicators. Due to the unique character of every collection, an interview usually produces more details about practical use than a specified questionnaire. The specification as described in section 4.3 (*Character and Extent*) could serve as a basis for such an interview.

If possible, the extent of the potential user group should also be determined. This information is relevant if a commercial version of the digital collection is to be produced. It also matters for the decision-making process that focuses on the question whether or not to go ahead with the digitisation. It should be noted, though that special sources usually do not have a very broad audience.

Describing this part involves the specification of various users and their particular needs. This prevents an indication that is merely based on a single audience. In this phase of the project, the project group have an important task. They must take into account the fact that various user groups use the same material from individual perspectives. The description of the collection as produced in phase 1 of MEDOC may serve as an indication for drawing up a method of specifically putting questions. Information about the text, the illustrations, the tables and other material constitute a basis for retrieving helpful data about the way in which various audiences use the material. During the next step, the examination must involve the gathered information on the collection. Emphasis should be put on the use of descriptions, catalogues etc.

Section 5.4 *What is the role of the collection in relation to research and education?* describes the position of a collection within the context of research and education (see also Table 6.). First, the collection is examined on its importance to research and education. Subsequently, the collection is considered on its use in research and education. For instance, should the students conduct source inquiries? Do they use the collection as a reference? In this case, both education and research staff are excellent sources for providing the required information.

During the specification, it may become clear that not every part of the collection is used with the same frequency. In case the collection can or need not be entirely processed, this information may prove crucial for determining which parts need to be actually disclosed in a digital form.

What kind of enrichment is required for optimally supporting the users in achieving their present goals? This question is at the basis of defining various kinds of functionality and thus the realisation of extra value. Also important within this context is the specification of various problems and requirements concerning the collection(s). This specification constitutes the next basis for determining the specific extra value. The answers to the above-mentioned questions should result in a report describing the requirements regarding the digital collections with respect to its audience in a non-technical way. This report may include phrases like:

*'The collection contains volumes that mostly consist of text and tables with numerical data. Users are mainly interested in the contents of these tables. Therefore, in addition to a table format, these data should also be presented in spreadsheets.'*

### ***6.3 Defining new audiences and new use, defining functionality***

Developing an idea of the way in which the material may be used in the future is essential for achieving adequate disclosure. For the users, the digitally disclosed source must have added value in relation to the original material. This extra value must either support them in obtaining their goals or facilitate the definition of new targets.

Within this context, an important and indicative part is reserved for the specification of the collection during phase 1 of MEDOC as well as for the information collected in section 6.2 *Determining present use and users, defining functionality*. The answers to the *why* question are relevant for specifying new audiences. For instance, if a collection obtains renewed importance, it may also acquire a new public.

The report on defined functionality is set up in a similar way as described in section 6.2: *Determining present use and users, defining the functionality*. This report may contain the following sentences:

*'The users would like to compare the specific use of words within the context of a single sentence. Presently, this does not happen, since the collection is too large. This would be possible however by the introduction of searching by means of free text. To this end, the functional requirements are as follows: users can freely retrieve terms. These terms are presented to them as concordances. They themselves can indicate the context in terms of the number of words to the right and/or to the left.'*

Adequate quality and functionality of the digital representation do no amount to merely aiming at an exact and natural representation of the original source. Two examples may illustrate this. First, colour. If the material's legibility is the only requirement or the most important one, the representation does not necessarily have to be in the original colours. In processing a digital source, its legibility could be enhanced. Thus, a processed digital representation could attain improved clarity and legibility in comparison with the pallid original.

Facilities for browsing text are another example that illustrates the measurement of quality and functionality with respect to the user. Essentially, the digital representation of a collection amounts to a series of images. This is in spite of the fact that the contents of these images could consist of pure text. Text-browsing facilities become feasible, if the images are enriched. This is realised by presenting the text in machine-readable form. Thus, the researcher is provided with extra tools. In the original source, these were not available or, at best, just to a limited extent through indices. In this case, the text image on the screen may be very different from the paper original. Nevertheless, the presented quality is of a very high standard.

### ***6.4 Rendering functionality into quality requirements regarding digital sources***

#### **6.4.1 What is quality?**

After defining the requirements for a digital collection, it is now time to determine the way in which these (functional) requirements affect the quality of the collection. There are two alternatives for specifying this aspect.

First, a literature investigation into similar projects may indicate a number of possibilities and obstacles. Since each special collection has its own unique character, this could prove a limited option.

Secondly, necessary information can also be obtained by testing selected parts of the material. Such a test, for instance, could indicate whether the material is suitable for OCR. A test also tells if the quality and amount of detail required can be accomplished within the limits of present technology. In addition, such a test could establish conflicting requirements, should the users need increased details as well as site-independent disclosure (e.g. via WWW).

There is already a great deal of available knowledge about rendering disclosure functionality into quality requirements for a digital source. Most of this knowledge comes from measuring and defining the quality of printed matter. Within this context, the enrichment actions for producing the source are irrelevant. The factors that matter here are colour depth, resolution and the required export formats.

There is also a direct connection with the practical use of the collection. In case of doubt, the maxim is to select the highest possible quality. This step is necessary because the production of digital sources by means of scanning and OCR (which requires pre- and post processing) is a rather time-consuming and costly enterprise that, in addition, involves lots of expertise.

Quality, a clear idea of the production costs, accuracy in predicting storage and production as well as time involved, all these elements would certainly benefit from a precise specification with respect to the format and characteristics of the files that are at the basis of the digital source.

In the following section of this report, no distinction is made between on-screen representation and presentation in print. The computer screen offers only limited possibilities, whereas the printing of digital files requires special knowledge and skills. On screen, factors like size, font as well as colour models differ from their counterparts that are required for printed matter. If the digital sources also need to be reproduced in print, the standard that is required to this end is indicative for determining the quality of the digital files.

MEDOC is limited to the indispensable facets for producing a digital collection. Consequently, there is restricted focus on the aspect of collection enrichment. Within this context, therefore, no attention is paid to the functional requirements regarding those aspects of digital disclosure that are not directly connected with the quality of the digital sources, but rather with the level of enrichment. Enrichment in the shape of, for instance, references to secondary literature can be a crucial prerequisite. Nevertheless it not directly connected with the quality of the digital source.

Functionality must be assessed in association with the user. In order to perform a proper test, functionality must be translated into a quality standard for digital files. This quality is differentiated in terms of dynamics, colour depth, DPI (dots per inch). Colour depth is subdivided into bitonal colours (black/white), grey scales and colour. Grey scales are used for half-tone printing and black & white photographs.

An image is of adequate quality if all functional requirements are met in terms of dynamics, colour depth and DPI. For instance, a bitonal image may excellently serve its purposes if legibility is its only functional requirement.

On the other side, quality is also related to practical use. An accepted subdivision for this use is:

- Filing
- Publishing in print
- Publishing on screen

Within this context, an image for archival purposes meets the highest possible standard, whereas the quality for printing purposes is of a lower standard. Defining quality starts with relating this subdivision to DPI, colour depth and dynamics.

To a certain extent, there is a direct relationship between quality and DPI. This relationship is limited by the characteristics of the human eye. Determining the quality begins with SSD (i.e. ‘Smallest Significant Detail’), which is designated in millimetres. SSD concerns the smallest visible element in a document and it is determined by the functional user requirements. The smallest distinct lower case ‘e’ in a collection functions as SSD-standard for texts on screen.

In addition, there is a derived quality as well. If a document consists of text and this text has to be disclosed in a machine-readable form, it must be determined whether machine-readable text is feasible through OCR of scanned documents. In view of this option, the same document must be digitised several times in order to establish the required quality.

The question *What is quality?* eventually results in a specification with respect to each relevant item in the collection. Quality is specified in terms of DPI, resolution, colour depth and dynamics. Other factors can be considered as well, for instance the access time to the collection files. This aspect is relevant if the disclosure is offered by means of a network or the Internet. When defining quality, compression factors and export formats have to be regarded in this case as well.

### **6.5 Reiteration**

Quality and functionality are interdependent aspects. It may be necessary to revise various sorts of functionality as a result of literature studies, testing and benchmarking. Before continuing, re-testing is therefore essential. This reiteration is indicated by the diamond shape in figure 3 *How: Functionality and Quality*. The shape shows the decisive moment at which it may prove necessary to re-adapt and re-evaluate both functionality and quality.

### **6.6 Literature study**

A literature study could produce extra information on the feasibility of the programme of requirements, which is based on all the aforementioned questions. After confirming the fact that it concerns a similar collection, this study should focus on the following questions:

- |   |
|---|
| <ol style="list-style-type: none"><li>1. What are the requirements regarding quality and functionality?</li><li>2. How have these requirements been rendered into technical standards (DPI, colour depth, dynamics)?</li><li>3. Are these requirements similar or comparable to the ones relating to the present collection?</li><li>4. If not, can an extrapolation be made to the programme of requirements?</li><li>5. What are the results of the project with respect to the programme of requirements?</li><li>6. Have the set objectives been achieved?</li><li>7. If not, what recommendations have resulted?</li></ol> |
|---|

### **Table 11**

The literature study should result in a report. This report states whether the achieved outcome actually meets the programme requirements and, if so, to which extent. In many cases, a mere literature study does not provide sufficient ground for such a conclusion. Therefore, additional benchmarking tests are usually required.

### ***6.7 Representative selection from the collection***

The selection of representative material from the collection must take place in a close association with the users and other experts on the particular subject matter. This procedure will guarantee that the selected material actually represents the collection as a whole.

The extent of the selection depends on the collection's heterogeneity. Due to the combination of supplement folios and divergent formats, separate volumes in a collection may be multiform. The collection itself may be multiform because of the various characteristics that the volumes have in terms of format, age and composition. The selection must do justice to these multiform aspects.

The selection results in a number of digitally processed documents. During the selection, the following criteria in table 12 have to be observed.

- |  |
|--|
| <ol style="list-style-type: none"><li>1. The selection must represent all document types.</li><li>2. The material has to be characteristic for the average collection.</li><li>3. All selected documents must contain SSD.</li><li>4. If OCR is considered, the material has to be selected appropriately.</li></ol> |
|--|

**Table 12**

The selection must represent all the various document types that are contained in the collection, i.e. photographs, book pages, maps and charts etc. The selection must involve at least one item of each separate element from the collection. In addition, all these items must be selected proportionally to their frequency in the collection as a whole.

In this way, testing and benchmarking procedures should result in a fair idea of all the typical obstacles that could be encountered with various elements from the collections. For this reason, the material must be characteristic for the collection.

All the documents must contain SSD. For text, this commonly is the lower case 'e' in the running passages. For images, the experts on the subject matter must make the necessary recommendations.

### ***6.8 Testing and benchmarking***

The next step concerns the actual digitisation and storage of the selected material according to the criteria as described in the previous sections. This step may have variable forms. For instance, if the collection allows this option, the documents could be directly scanned.

It could prove necessary to remove the documents from their volume. It may also be inevitable to photograph all the documents and reproduce them on microfilm before the material can be digitally processed. If this is required, one should not forget that each intermediate re-rendering of the material brings about an interpretation causing the processed material to contain less information than the original. It is important therefore that the tests are carried out in various combinations of the described criteria, i.e. DPI, resolution, dynamics, compression and export formats. Only in this way, an optimum combination can be obtained.

The question of the quality of the files and whether the result lives up to the predetermined criteria needs to be verified by means of various tools as well as with the eye. The described functionality is simulated by means of image-processing programs. Such tools could be colour and/or grey scales. A control group that is not involved in the actual digitisation process but is still familiar with the quality standards should evaluate the images on screen and/or in other export media. Within this context, various applied export formats are

compared with respect to their image quality. The error percentage of machine-readable text is matched with the original.

This phase finds its completion in a set of files stored on disk as well as in a report stating on the results, the method and the problems. The report describes the way in which each file was processed, the settings involved, all physical characteristics as well as the final quality judgement. Additionally, it contains recommendations on how to resolve possible problems. Finally, the report advises on whether the required quality can be obtained for every item and if so, how this is to be accomplished.

## 7. How: Resources

### 7.1 Introduction

The penultimate phase in deciding on digital storage and disclosure is concerned with the determination of the required human resources and means.

In the first phase, the nature and extent of the collection are established. Next, it is decided which collections will be nominated. Subsequently, obstacles with respect to the collection's quality and availability are specified. In the fourth stage, this information is linked to the quality and functionality requirements, as determined by means of the questions in section 6. Finally, these findings must be related to the human resources, skills and means that are required for carrying out the actual digital disclosure of the collection.

This penultimate phase results in a report that specifies the required functionality and quality of the digital collection in terms of time, costs and necessary knowledge and skills. This specification is needed for determining whether the required resources are available within the organisation or whether they have to be acquired externally.

This phase is essential since the digitisation and disclosure of special collections is both a time- and labour-consuming enterprise. By means of the information gathered in the previous phase it is possible to accurately define the deployment of human resources and means necessary for each stage in the production process. These calculations are partly based on information that was previously obtained during various subsidiary projects within the framework of the *Electronic Library of Utrecht*, a major project of the Utrecht University library.

This specification involves the necessary resources, the knowledge as well as the worked-out requirements concerning functionality and quality. For instance, if collection texts must be presented in a machine-readable form, it is in this phase that we determine which resources are necessary for achieving this goal.

If OCR is applied, the scanned images must first be pre-processed. What is the amount of time and costs involved? Can we carry out those tasks ourselves or do we have to outsource them?

In the following phase, the text that has been processed in OCR must be checked and corrected. What kind of expertise is required from the persons who will be in charge of this task?

During the same phase, an estimate is made of the time and costs that are involved in all the subsequent steps leading to the digital disclosure of the collection. Simultaneously, storage and calculation capacity have to be thoroughly examined. The same goes for the technical preconditions.

### 7.2 Determining the production elements.

Consequently, it is possible to determine the various production elements. Nevertheless, a number of questions and uncertainties will remain unresolved. Additionally, it cannot be ruled out that various suggested options also require various production courses. All these aspects have to be worked out in the present phase. The digitisation of various elements must be differentiated. Generally, the production process involves the following phases:

- Preparation (pre-processing)
- Digitisation

- Verification and post-processing
- Enrichment
- Disclosure

During the preparation phase, the material is collected. In some cases, it must be prepared before it can be digitised. The material may have to be removed from the album. Occasionally, it first has to be conserved, photographed and put to microfilm. For each element of the collection, a working instruction is added, as well as a log. The latter is used for keeping record of the moment and nature of each action, as well as of the names of the involved persons.

During the processing of the collection it is essential that this log is accurately kept. In case of errors or difficulties, these can be more easily traced. The instructions consistently describe the way in which the elements should be processed.

The following questions regarding all elements in the production process need to be addressed. (Once more, it must be stressed that MEDOC is restricted to describing the phases that are concerned with the digitisation process):

- Preparation (pre-processing)
- Digitisation
- Verification and processing (post-processing)

In the evaluation process, the remaining steps must also be considered as well, for in the development process of smaller and medium large collections (enrichment and disclosure), the amount of hours and indeed the amount of costs tend to exceed the digitisation efforts. Table 13 contains a list of the corresponding questions.

|    |   |
|----|---|
| 1. | How many elements of the collection need to be pre-processed? (n) |
| 2. | What does the pre-processing phase comprise?                      |
| 3. | How many minutes are involved in pre-processing each element?     |
| 4. | What are the costs per hour? (C)                                  |
| 5. | What are the total costs? $([n \times B] / 60) \times C$          |
| 6. | How many working hours does a week comprise? (G)                  |
| 7. | What is the total time involved? $([n \times B] / 60 / G)$        |

**Table 13**

|   |
|---|
| <p><i>Explanation</i></p> <p>n = number of elements in a collection (e.g. pages)</p> <p>B = processing time per element in minutes</p> <p>C = personnel costs for processing per hour</p> <p>G = number of productive hours per week</p> <p>O = remaining costs</p> |
|---|

**Table 14**

A calculation can be made in terms of estimating the development time (indicated in weeks per element) and the costs that are related with this process. If the various elements have been designated correspondingly to Table 14, the total development time can be calculated with this formula:

Amount of development time in weeks =  $([n \times B] / 60) / G$

Costs =  $([n \times B] / 60 \times C + O$

Analogous formulas can be drawn for every element in the production process.

After pre-processing the collection is digitised according to the predetermined criteria. Following the completion of the digitisation procedure, the result is verified and perhaps worked out. In case of difficulties, it may be necessary to restart the process

In case of OCR, it could prove necessary to additionally pre-process the material. In order to calculate the costs of digitisation as well as to obtain an adequate idea of the development time, it is required to answer a number of questions. For this purpose, the data from the testing phase is used. Please refer to Table 15 for an overview:

|   |  |
|---|--|
| 1 | How many elements must be scanned? (n)   |
| 2 | What is the average amount of time needed for a scan? (B)                        |
| 3 | What are the costs per hour? (F)   |
| 4 | What is the average file size (in megabytes)? (D)                                |
| 5 | What are the costs involved for one megabyte? (E)                                |
| 6 | What are the total costs? $[n \times D] \times E + ([n \times B] / 60) \times F$ |
| 7 | How many working hours does a week comprise? (G)                                 |
| 8 | What is the total amount of time? $([n \times B] / 60) / G$                      |

**Table 15**

For post-processing, time and costs can be correspondingly calculated (see Table 13).

After digitising, the material must be enriched. The actual enrichment depends on the selected disclosure method. The digital sources must be contained into an information system that has to meet the programme of requirements. In terms of the method involved, the design and development of this information system do not essentially differ from the method that is used for realising other information systems. Existing methods and models for developing software can be applied here without objection. However, as was mentioned in the introduction, MEDOC is not concerned with the enrichment or with the implementation of the software involved.

### ***7.3 Analysis of technical preconditions***

An organisation that decides to digitally disclose a collection does not necessarily have the appropriate technical infrastructure for this purpose. Digital files and their enrichment could comprise huge amounts of megabytes. In addition, disclosure via the Internet also implies the availability of an appropriate technical infrastructure. An analysis of the technical preconditions results in the specification of hardware, software as well as network and management structures, which are required for successfully accomplishing the digital disclosure of the collection and its management.

## 8. Evaluation

Figure 1 *Overview of MEDOC* provides a scheme of the steps that are necessary for evaluating the questions on whether and how a specific collection can be disclosed in a digital fashion. The basis for this evaluation is constituted by reports that have been kept during the completion of the various stages in MEDOC. It also comprises a possible literature study including its findings as well as the outcome of the testing and benchmarking procedures.

All these reports serve as a foundation of the debate that ensues. Upon completion of MEDOC as described above, a clear image should arise of:

- |  |
|--|
| <ol style="list-style-type: none"><li>1. The nature and extent of the collection</li><li>2. The subject matter of the collection</li><li>3. The significance of the collection within various contexts</li><li>4. The requirements regarding quality and functionality</li><li>5. The necessary human resources and means</li><li>6. The project duration involved in digital disclosure</li><li>7. The various consequences for the organisation, management etc.</li></ol> |
|--|

### Table 16

All these aspects should be chronicled and recorded in various reports. The ensuing discussion should concentrate on comparing the requirements in order to achieve the predetermined objectives. This is the formal mode of operation. If no sufficient resources or means are available, subsequent steps are not likely to be taken.

Since accomplishing a digital collection of reasonable extent rather implies a medium long-term project, it must be appropriately supported by the organisation. Those who participate in the discussion should therefore be selected on the ground of their ability to establish these conditions.

The discussion should result in a recommendation. Theoretically, the project is subsequently carried out in accordance with the established specifications. More likely, however, is that a number of recommendations and conditions follow before the project actually starts. Yet, it also may be decided not to go ahead with the digitisation of the collection.

## 9 Acknowledgements

The preface to this paper states that MEDOC is based on the results of the project *Electronic Library Utrecht (ELU)*. Prime objective of the ELU project was to examine and specify the possibilities and consequences of applying ICT resources in a library organisation. With respect to the special collections, four subsidiary projects have been of immediate relevance.

1. The first subsidiary project concentrates on digitising and disclosing material from the Utrecht Organ Archive. In the mean time, the project has been extended to co-operation within a wider European context (EU *Raphael* Programme: European Organ Index).
2. The second project – under the name of *Thesaurus Musicarum Italicarum* – is concerned with partially digitising a collection of music-theoretical manuscripts (which comprises text, image as well as musical annotations) originating from 16<sup>th</sup> and 17<sup>th</sup> century Italy. The digital material will also be published in multimedia form. This project has been organised and conducted by the Utrecht Department of Computerisation and Literature in close co-operation with Utrecht University Library (<http://www.library.uu.nl/digiboeken/weyerman/weyerman.html>). *The Music Treatises* by Gioseffo Zarlino (1517-1590) have already been published on CD-ROM. Currently, a network version is being developed.
3. A third project comprises the *Van der Monde* and *Weyerman* collections. This project aimed at investigating the alternatives and limitations of various hardware, software, technologies and programs.
4. Meanwhile, the *Weyerman* project has been finished. The digital collection can now be accessed via the website of the Utrecht University Library. This ensuing project was conducted in close collaboration with the Centre of ICT and Media Applications (CIM) of Utrecht University.

Another project, which is no ELU project yet as an extensive enterprise has provided valuable expertise and results for MEDOC, is the digital version (CD-ROM) of the 9<sup>th</sup> century *Utrecht Psalter*.

All the accumulated knowledge and experience from these projects as well as the policy objectives behind the digitisation of special collections have been worked out in an instruction guide for selecting and conducting similar projects. This instruction book carries the name of MEDOC.

As you will notice we have not added a bibliography. However, there is one publication we need to mention : Anne R. Kenney and Stephen Chapman, *Digital Imaging for Libraries and Archives* (Ithaca: Cornell University Library, 1996). Much of our “digital knowledge” was based on the findings published in this book.

Many people have contributed to the realisation of MEDOC. In the first place, we would like to thank all members of the Advice Group: Margriet Blom, Natalia Grygierczyk and Pierre Pesch. In addition we want to name Marc van Gestel, who edited the present paper, as well as Antar El-Mecky, who was responsible for the English translation.

Finally, we would like to point out the interactive character of MEDOC. Version 1.0 is based on collective knowledge and experience. All knowledge involved in other projects not known

to us could constitute a valuable addition. Consequently, we wish to call upon all persons with new insights. We invite you to contact us so as to make possible the publishing of a second, revised version of MEDOC containing new information. Meanwhile, the present version can also be accessed via the Utrecht University Library homepage.

We sincerely hope that MEDOC will prove of practical use during the organisation and realisation of your digitisation projects.

Utrecht, October 1999

Utrecht University Library  
Department of Special Collections  
MEDOC  
PO Box 16007  
3500 DA Utrecht  
The Netherlands  
E-mail: [h.mulder@library.uu.nl](mailto:h.mulder@library.uu.nl)

## 10 Appendices

### 10.1 Questionnaires

#### 4.3 Character and extent of the collection

1. What is the overall description (amount of volumes, structure, table of contents, indices)?
2. What is the number of pages?
3. What is the page size (height and width in mm)?
4. What is the subject matter (characteristics and/or illustrations)?
5. Has any colour been used?
6. Which alphabet and/or scripture types are involved?
7. What type of material is involved (printed matter or manuscript)?
8. Are there any supplement pages of different size (number and position)?
9. Are any tables involved (number and position)?
10. Are any illustrations (number and position) involved?
11. Which language(s) are involved?

#### 4.4 Information about the collection

1. Is the description complete and up to date?
2. What parts make up the title description?
3. Is a digital description available (on-line catalogue)?

#### 5.1 Reasons for Digitising and Disclosing the Collection: Introduction

1. What is the social and cultural significance of the collection?
2. What is the present importance of the collection?
3. What role has the collection in relation to research and education?
4. Is the collection accessible?
5. What is the physical condition of the collection?
6. Does the collection deal with Utrecht as a subject?
7. Does the collection's subject concern Utrecht University?
8. Is the collection valuable in the sense of PR aspects?

#### 5.2 What is the social and cultural significance of the collection?

1. Can similar collections be found in the Netherlands?
2. Are there any similar collections abroad?
3. In what way is this particular collection related to others managed by our

library?

4. Have any initiatives been launched to digitally disclose these collections?
5. If so, is co-operation a feasible option?
6. Is the collection valuable as a work of reference?

### ***5.3 What is the present importance of the collection?***

1. Is the collection connected in a significant way to present education and research?
2. Is there any social interest that would support digitisation of a collection?

### ***5.4 What is the role of the collection in relation to research and education?***

1. Is the collection of general importance as to education and research?
2. Does the collection relate to research and education at Utrecht University?
3. Does the collection acquire significant extra value by presenting it in digital form (accessibility, disclosure through information retrieval)?

### ***5.5 Is the collection accessible?***

1. What is the physical condition of the material?
2. Does the physical condition allow for public availability?
3. Does digital disclosure constitute an alternative for public availability of the original material?
4. Does the collection have a high priority in the conservation plans of the library?
5. Is digitising (part of) the material a useful alternative for other conservation methods?

### ***5.6 Remaining questions***

1. Does the collection relate to the city and/or province of Utrecht?
2. Does the collection deal with the Utrecht University?

3. Did the collection belong to a (former) scholar of Utrecht University or does the material refer to this person?

### ***5.7 Is the collection valuable in the sense of PR?***

1. Is digital disclosure in compliance with the PR policy of the Utrecht University Library and/or Utrecht University?

### ***6.2 Determining present use and users, defining functionality***

1. Who are the users?
2. How large is the user group?
3. What are their objectives in using the material?
4. What sort of enrichment is required for optimally supporting the users in achieving their present goals?
5. Which problems do they experience in using the collection?
6. Have the users any special wishes with respect to the collection?
7. What kind of research and education involves the collection?