

# RLG WORKSHEET FOR ESTIMATING DIGITAL REFORMATTING COSTS

This worksheet is a guide to the preparation of a budget for activities involving digitization. It can be used for in-house scanning projects or for those utilizing an outside vendor. The activities are organized in eleven steps:

- A. Select materials
- B. Determine the size of the collection
- C. Prepare documents
- D. Determine imaging requirements (benchmarking)
- E. Determine requirements for and create metadata
- F. Determine imaging costs
- G. Determine text conversion costs
- H. Determine SGML encoding costs
- I. Determine Finding Aid Conversion Costs
- J. Post-process digital files
- K. Estimate additional local costs

The accompanying appendix provides a methodology for determining scanning requirements based on an assessment of document attributes and current scanning technologies. Endnotes follow Summary of Projected Expenses.

## A. SELECT MATERIALS

This step of the process includes:

- \* Identifying materials
- \* Determining legal restrictions (copyright, privacy rights, donor rights, etc.)
- \* Investigating the availability of digital and other versions, such as microforms (online search)
- \* Eliminating items due to condition or completeness
- \* Determining appropriate conversion process (e.g., film then scan, disbind originals, etc).

Estimate staff time required for selection, then the cost based on the level of staff performing each process.

*Name/Title:*  
 # of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <A<sub>1</sub>>

*Name/Title:*  
 # of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <A<sub>2</sub>>

*Name/Title:*  
 # of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <A<sub>3</sub>>

*Name/Title:*  
 # of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <A<sub>4</sub>>  
 AND/OR

*Name/Title:*  
 % FTE \_\_\_\_\_ at \_\_\_\_\_ salary/annual + \_\_\_\_\_ benefits/annual = \_\_\_\_\_ <A<sub>5</sub>>

Note: If various levels of staff are used in the process, calculate a blended monetary rate or repeat the formula for each staff member.

**TOTAL SELECTION COST = <A> = \_\_\_\_\_ <1>**

## B. DETERMINE THE SIZE OF THE COLLECTION

This information will be crucial in determining staff time for pre- and post-processing functions as well as costs associated with digital conversion.

### BOUND DOCUMENTS:

# of titles (if appropriate) _____		<B <sub>1</sub> >
# of volumes _____		<B <sub>2</sub> >
total # of pages to be imaged _____		<B <sub>3</sub> >
total # of pages slated for text conversion _____		<B <sub>4</sub> >
average # of characters per page <sup>2</sup> _____		<B <sub>5</sub> >

### SINGLE-LEAF OR UNBOUND DOCUMENTS:

# of titles (if appropriate) _____		<B <sub>6</sub> >
# of base units <sup>3</sup> _____		<B <sub>7</sub> >
total # of pages to be imaged _____		<B <sub>8</sub> >
total # of pages slated for text conversion _____		<B <sub>9</sub> >
average # of characters per page _____		<B <sub>10</sub> >

### MICROIMAGES:

# of titles (if appropriate) _____		<B <sub>11</sub> >
# of volumes (if appropriate) _____		<B <sub>12</sub> >
# of fiche or reels <sup>4</sup> _____		<B <sub>13</sub> >
# of frames <sup>5</sup> _____		<B <sub>14</sub> >
total # of pages to be imaged _____		<B <sub>15</sub> >
total # of pages slated for text conversion _____		<B <sub>16</sub> >
average # of characters per page _____		<B <sub>17</sub> >

### FINDING AIDS

# of finding aids _____		<B <sub>18</sub> >
# of pages _____		<B <sub>19</sub> >
avg # of characters per page _____		<B <sub>20</sub> >

**TOTAL # OF TITLES**

**TOTAL # OF VOLUMES/BASE UNITS**

**TOTAL # OF PAGES OR IMAGES**

**TOTAL # OF CHARACTERS**

**TOTAL # OF FINDING AIDS**

**TOTAL # OF PAGES OF FINDING AIDS**

## C. PREPARE DOCUMENTS

This step of the process involves the physical and intellectual preparation of documents. It includes:

- \* Retrieving documents from storage
- \* Checking out the documents from circulation
- \* Recording the physical condition of the documents
- \* Collating, identifying missing pages and damage
- \* Repairing pages, replacing missing or illegible pages
- \* Preparing intermediates (e.g., microforms, transparencies, photocopies, etc.)
- \* Disbinding originals (when appropriate)
- \* Creating documentation/instructions for bibliographic control, indexing, tagging, and encoding information (where appropriate)

### PERSONNEL COSTS

Estimate staff time required for preparing the collection defined in Section B based on the level of staff performing each process.

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <C<sub>1a</sub>>

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <C<sub>1b</sub>>

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <C<sub>1c</sub>>

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <C<sub>1d</sub>>

AND/OR

*Name/Title:*  
% FTE \_\_\_\_\_ at \_\_\_\_\_ salary/annual + \_\_\_\_\_ benefits/annual = \_\_\_\_\_ <C<sub>1e</sub>>

Note: If various levels of staff are used in the process, calculate a blended monetary rate or repeat the formula for each staff member.

### NON-PERSONNEL COSTS (conservation supplies, vendor charges for intermediates, etc.)

Description = \_\_\_\_\_ Cost = \_\_\_\_\_ <C<sub>2</sub>>  
 Description = \_\_\_\_\_ Cost = \_\_\_\_\_ <C<sub>3</sub>>  
 Description = \_\_\_\_\_ Cost = \_\_\_\_\_ <C<sub>4</sub>>  
 Description = \_\_\_\_\_ Cost = \_\_\_\_\_ <C<sub>5</sub>>

Total personnel cost (<C<sub>1</sub>>) \_\_\_\_\_ <C<sub>6</sub>>

Total non-personnel cost (<C<sub>2</sub>> + <C<sub>3</sub>> + <C<sub>4</sub>> + <C<sub>5</sub>>) \_\_\_\_\_ <C<sub>7</sub>>

**TOTAL PREPARATION COST = <C<sub>6</sub>> + <C<sub>7</sub>> = \_\_\_\_\_ <2>**

## D. DETERMINE IMAGING REQUIREMENTS (BENCHMARKING)

This step involves:

- \* Assessing essential document attributes to determine scanning requirements (resolution, bit depth, enhancements, file format, compression)
- \* Confirming results by scanning a subsample of documents
- \* Performing inspection of subsample on screen and in print

Note that the decision to perform text conversion and encoding, including links to page images, may affect scanning requirements. If text conversion/encoding are to be conducted, the sub-sample test should include an assessment of the accuracy of text conversion from digital images. Refer to Appendix A for information on how to benchmark conversion requirements for different document types. This step may or may not be a project activity.

*Name/Title:*

# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <D<sub>1</sub>>

*Name/Title:*

# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <D<sub>2</sub>>

*Name/Title:*

# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <D<sub>3</sub>>

*Name/Title:*

# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <D<sub>4</sub>>

AND/OR

*Name/Title:*

% FTE \_\_\_\_\_ at \_\_\_\_\_ salary/annual + \_\_\_\_\_ benefits/annual = \_\_\_\_\_ <D<sub>5</sub>>

Note: If various levels of staff are used in the process, calculate a blended monetary rate or repeat the formula for each staff member.

**TOTAL BENCHMARKING COST = <D> = \_\_\_\_\_ <3>**

## E. DETERMINE REQUIREMENTS FOR AND CREATE METADATA

Estimate costs associated with creating the requisite metadata for searching and retrieving digital image files. Refer to emerging metadata standards<sup>6</sup> for cataloging and file naming. This step includes:

- \* Creating catalog entries for digital resources
- \* Determining file naming and structuring strategies (for individual images and groups of images)
- \* Creating additional indexes (e.g., indexing at the article level for journal literature ) or revising/enhancing existing finding aids.

Estimate staff time required for this process as a whole, then the cost based on level of staff performing each process.

*Name/Title:*  
 # of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <E<sub>1</sub>>

*Name/Title:*  
 # of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <E<sub>2</sub>>

*Name/Title:*  
 # of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <E<sub>3</sub>>

*Name/Title:*  
 # of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <E<sub>4</sub>>

AND/OR

*Name/Title:*  
 % FTE \_\_\_\_\_ at \_\_\_\_\_ salary/annual + \_\_\_\_\_ benefits/annual = \_\_\_\_\_ <E<sub>5</sub>>

Note: If various levels of staff are used in the process, calculate a blended monetary rate or repeat the formula for each staff member.

**TOTAL METADATA COST = <E> = \_\_\_\_\_ <4>**

## F. DETERMINE IMAGING COSTS

Determine conversion costs based on the size and nature of the collection (Section B) and the imaging requirements (Section D). Secure firm bids from two service providers (internal or external) for the per image cost, including:

- \* Set-up and supplies
- \* Scanning
- \* Special handling
- \* Inspection
- \* Rescans
- \* Basic structuring and labeling
- \* Additional manipulation (e.g., cropping, despeckling, splitting frame images into page images, etc.)
- \* Storage media
- \* Shipping and insurance

Additionally, conversion bureaus can provide other services. Note that costs for additional services will depend on your resolution and bit-depth requirements and any special requirements. These costs must be determined in advance. Secure firm bids on a per image basis.

- \* Printing (for quality control and text conversion)
- \* Creating derivatives
- \* Creating backups

Obtain samples of work based on documents to be scanned in the project. Choose the most appropriate bid based on quality, cost, and other considerations. Fill in only the appropriate categories and repeat formulas as necessary (e.g., if different scanning resolutions are used for various documents in the same category).

### IMAGING COSTS

#### Regular size and single-leaf and unbound (<=8.5" x 14"):<sup>7</sup>

Check the scanning category (black and white, grayscale, color) and indicate any special treatment<sup>8</sup> that applies to your documents. If more than one applies, repeat the formula for each scanning category:

TYPED/PRINTED TEXT/LINE ART		<input type="checkbox"/> black and white	<input type="checkbox"/> grayscale	<input type="checkbox"/> color			
# of pages _____	X	_____ cost per page	=	TOTAL COST	=		<F <sub>1</sub> >
MANUSCRIPT		<input type="checkbox"/> black and white	<input type="checkbox"/> grayscale	<input type="checkbox"/> color			
# of pages _____	X	_____ cost per page	=	TOTAL COST	=		<F <sub>2</sub> >
CONTINUOUS TONE		<input type="checkbox"/> black and white	<input type="checkbox"/> grayscale	<input type="checkbox"/> color			
# of pages _____	X	_____ cost per page	=	TOTAL COST	=		<F <sub>3</sub> >
HALFTONE		<input type="checkbox"/> black and white	<input type="checkbox"/> grayscale	<input type="checkbox"/> color			
# of pages _____	X	_____ cost per page	=	TOTAL COST	=		<F <sub>4</sub> >
MIXED		<input type="checkbox"/> black and white	<input type="checkbox"/> grayscale	<input type="checkbox"/> color			
# of pages _____	X	_____ cost per page	=	TOTAL COST	=		<F <sub>5</sub> >

**Regular size and bound (<=8.5" x 14"): <sup>9</sup>**

Check the scanning category (black and white, grayscale, color, special treatment) that applies to your documents.

If more than one applies, repeat the formula for each scanning category:

TYPED/PRINTED TEXT/LINE ART	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>6</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
MANUSCRIPT	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>7</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
CONTINUOUS TONE	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>8</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
HALFTONE	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>9</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
MIXED	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>10</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							

**Oversize and bound (>8.5" x 14"):**

Check the scanning category (black and white, grayscale, color, special treatment) that applies to your documents.

If more than one applies, repeat the formula for each scanning category:

TYPED/PRINTED TEXT/LINE ART	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>11</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
MANUSCRIPT	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>12</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
CONTINUOUS TONE	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>13</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
HALFTONE	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>14</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
MIXED	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>15</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							

**Oversize and single-leaf or unbound (>8.5" x 14"):**

Check the scanning category (black and white, grayscale, color, special treatment) that applies to your documents.

If more than one applies, repeat the formula for each scanning category:

TYPED/PRINTED TEXT/LINE ART	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>16</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
MANUSCRIPT	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>17</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
CONTINUOUS TONE	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>18</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
HALFTONE	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>19</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							
MIXED	<input type="checkbox"/>	black and white	<input type="checkbox"/>	grayscale	<input type="checkbox"/>	color	=		=		<F <sub>20</sub> >
# of pages _____ X _____		cost per page		TOTAL COST							

**Microimages:**

Check the scanning category (black and white, grayscale, color) that applies to your documents. If more than one applies, repeat the formula for each scanning category. For microfilm and microfiche, be sure to indicate to the vendor whether the whole film/fiche or only single/selected frames need to be scanned, as the price will vary accordingly.

16 mm MICROFILM  black and white  grayscale  color  
 # of frames \_\_\_\_\_ X \_\_\_\_\_ cost per page = TOTAL COST = \_\_\_\_\_ <F<sub>21</sub>>

35 mm MICROFILM  black and white  grayscale  color  
 # of frames \_\_\_\_\_ X \_\_\_\_\_ cost per page = TOTAL COST = \_\_\_\_\_ <F<sub>22</sub>>

REGULAR MICROFICHE  black and white  grayscale  color  
 # of frames \_\_\_\_\_ X \_\_\_\_\_ cost per page = TOTAL COST = \_\_\_\_\_ <F<sub>23</sub>>

FULL-FRAME MICROFICHE  black and white  grayscale  color  
 # of frames \_\_\_\_\_ X \_\_\_\_\_ cost per page = TOTAL COST = \_\_\_\_\_ <F<sub>24</sub>>

JACKETED MICROFICHE  black and white  grayscale  color  
 # of frames \_\_\_\_\_ X \_\_\_\_\_ cost per page = TOTAL COST = \_\_\_\_\_ <F<sub>25</sub>>

APERTURE CARDS  black and white  grayscale  color  
 # of frames \_\_\_\_\_ X \_\_\_\_\_ cost per page = TOTAL COST = \_\_\_\_\_ <F<sub>26</sub>>

35 mm TRANSPARENCIES & NEGATIVES<sup>10</sup>  black and white  grayscale  color  
 # of frames \_\_\_\_\_ X \_\_\_\_\_ cost per page = TOTAL COST = \_\_\_\_\_ <F<sub>27</sub>>

3 1/4 x 4 1/4 TRANSPARENCIES & NEGATIVES  black and white  grayscale  color  
 # of frames \_\_\_\_\_ X \_\_\_\_\_ cost per page = TOTAL COST = \_\_\_\_\_ <F<sub>28</sub>>

4 x 5 TRANSPARENCIES & NEGATIVES  black and white  grayscale  color  
 # of frames \_\_\_\_\_ X \_\_\_\_\_ cost per page = TOTAL COST = \_\_\_\_\_ <F<sub>29</sub>>

**ADDITIONAL CONVERSION-RELATED COSTS**

Repeat the following formulas as necessary ( e.g., creating backup on different media).

Printing:

# of pages \_\_\_\_\_ X \_\_\_\_\_ cost per page = TOTAL COST = \_\_\_\_\_ <F<sub>30</sub>>

Derivatives:

# of images \_\_\_\_\_ X \_\_\_\_\_ cost per image = TOTAL COST = \_\_\_\_\_ <F<sub>31</sub>>

Backups:

average image file size (in megabytes) \_\_\_\_\_ X total number of images  
 X \_\_\_\_\_ /megabyte for backup \_\_\_\_\_ <F<sub>32</sub>>

Other :

\_\_\_\_\_ <F<sub>33</sub>>

**Name/Title:**

# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <F<sub>34</sub>>

Total imaging costs (<F<sub>1</sub>> + <F<sub>2</sub>> + ... + <F<sub>29</sub>>) = \_\_\_\_\_ <F<sub>35</sub>>

Total additional costs (<F<sub>30</sub>> + <F<sub>31</sub>> + ... + <F<sub>34</sub>>) = \_\_\_\_\_ <F<sub>36</sub>>

**TOTAL IMAGING COST = <F<sub>35</sub>> + <F<sub>36</sub>>** = \_\_\_\_\_ <5>

## G. DETERMINE TEXT CONVERSION COSTS (if appropriate)

Define nature and extent of text conversion (e.g., full-text of all documents, full-text of specific documents, table of contents only) and accuracy requirements. Staff time creating conversion instructions should be included under Document Preparation (Section C). Conversion costs for Finding Aids should be included under Finding Aid Conversion Costs (Section I).

Conversion may occur by processing images through Optical Character Recognition (OCR) programs, or by re-keying from the image or from a printed version (directly from original documents or printouts from digital images). The level of accuracy will be based on the purpose of text conversion. Your requirements for accuracy will vary depending if you are converting for encoding or if you are creating text files for indexing or searching purposes only. If you are creating text files for indexing or searching purposes only, a lower level of accuracy may suffice. Secure firm bids from two service providers (whether internal or external) for the per 1,000 characters conversion cost. These bids should be based on a sample test of materials so that vendors can determine whether to re-key or OCR. Note that estimates may differ if text conversion is considered as part of a larger bid for imaging and/or encoding. Vendor estimates should include fixed price per 1,000 characters for the following:

- \* Re-keying or OCR processing
- \* Inspection
- \* Corrections
- \* Supplies
- \* Special handling
- \* Media
- \* Shipping and insurance

Obtain sample of work based on documents to be converted in the project. Choose the most appropriate bid based on quality, accuracy, cost, and other considerations.

### PRINTED/PUBLISHED/TYPED DOCUMENTS

$$\begin{array}{l} \# \text{ of pages } \underline{\hspace{2cm}} \times \underline{\hspace{2cm}} \text{ average characters/page} = \# \text{ of total characters} \\ \underline{\hspace{2cm}} \text{ cost/1,000 characters} \quad \text{Total} = ((\text{cost per 1,000} \times \text{total characters})/1000) = \underline{\hspace{2cm}} \end{array} \begin{array}{l} \langle G_1 \rangle \\ \langle G_2 \rangle \end{array}$$

### UNPUBLISHED DOCUMENTS AND MANUSCRIPTS

$$\begin{array}{l} \# \text{ of pages } \underline{\hspace{2cm}} \times \underline{\hspace{2cm}} \text{ average characters/page} = \# \text{ of total characters} \\ \underline{\hspace{2cm}} \text{ cost/1,000 characters} \quad \text{Total} = ((\text{cost per 1,000} \times \text{total characters})/1000) = \underline{\hspace{2cm}} \end{array} \begin{array}{l} \langle G_3 \rangle \\ \langle G_4 \rangle \end{array}$$

$$\text{TOTAL TEXT CONVERSION COST} = \langle G_2 \rangle + \langle G_4 \rangle = \underline{\hspace{2cm}} \langle G \rangle$$

## H. DETERMINE SGML ENCODING COSTS (if appropriate)

Define nature and extent of coding and accuracy requirements. Staff time creating encoding instructions (including how to handle links to images) should be included under Document Preparation (Section C). It is assumed that SGML tagging will occur from ASCII (costs for creating ASCII files should be included in Section G). Secure firm bids from two service providers (whether internal or external) for the per 1,000 characters encoding/tagging cost.<sup>11</sup> These bids should be based on a sample test set of ASCII files. Note that estimates may be different if encoding is performed as a part of a larger text conversion effort (e.g., rekeying and encoding simultaneously). Vendor estimate should include fixed price per 1,000 characters for the following:

- \* Encoding/tagging
- \* Inspection
- \* Corrections
- \* Validation
- \* Supplies
- \* Special handling
- \* Media
- \* Shipping and insurance

Choose most appropriate bid based on quality, accuracy, cost, and other considerations.

To determine total SGML Encoding costs, first determine approximate number of characters, post-processing. Then divide that number by 1,000 and multiply by the quoted cost per 1,000 characters.

$$\text{Then, } \frac{((\langle G1 \rangle + \langle G3 \rangle) \times 1.5)}{(\underline{\hspace{2cm}} + \underline{\hspace{2cm}})} \times 1.5 = \underline{\hspace{2cm}} \langle H_1 \rangle$$

$$\frac{(\text{Total Characters} \div 1,000)}{(\underline{\hspace{2cm}} \div 1000)} \times \underline{\hspace{2cm}} \text{ cost/1,000 characters} = \underline{\hspace{2cm}} \langle H_2 \rangle$$

$$\text{TOTAL TEXT ENCODING COST} = \langle H_2 \rangle = \underline{\hspace{2cm}} \langle 7 \rangle$$

## I. DETERMINE FINDING AID CONVERSION COSTS (if appropriate)

Define nature and extent of finding aid conversion and encoding (e.g., Encoded Archival Description, HTML, ASCII). Staff time associated with creating conversion instructions or revising/enhancing Finding Aids should be included under Document Preparation (Section C).

Conversion may occur by processing images through Optical Character Recognition (OCR) programs, or by re-keying from the image or from a printed version (directly from original documents or printouts from digital images). Encoding costs will be based on the type of encoding selected. Use of SGML, and specifically the EAD (Encoded Archival Description) Document Type Definition, is recommended though some institutions may choose to encode finding aids in HTML (Hypertext Markup Language). Secure firm bids from at least two service providers - whether internal or external - for the conversion *and* encoding cost. In general, these bids will be made on a per 1,000 character (post-conversion) basis or per page basis and must include a guaranteed accuracy rate.<sup>12</sup> These bids should be based on a sample test of materials so that vendors can determine whether to re-key, OCR or convert existing electronic files.

Vendor estimates should include a fixed price per 1,000 characters (post-conversion) or per page for the following:

- \* Re-keying, OCR processing, or conversion of existing electronic files
- \* Encoding/tagging
- \* Inspection
- \* Corrections
- \* Supplies
- \* Special handling
- \* Media
- \* Shipping and insurance

**Estimating on a per 1,000 character basis:**

$$\begin{array}{l} \# \text{ of pages } \underline{\hspace{2cm}} \times \underline{\hspace{2cm}} \text{ average characters/page } \times 2.6 = \# \text{ of total characters } \underline{\hspace{2cm}} \langle I_1 \rangle \\ \langle I1 \rangle \times \underline{\hspace{2cm}} \text{ cost/1,000 characters } = \text{ Total conversion cost } = \underline{\hspace{2cm}} \langle I_2 \rangle \end{array}$$

OR

**Estimating on a per page basis**

$$\# \text{ of pages } \underline{\hspace{2cm}} \times \underline{\hspace{2cm}} \text{ cost/page } = \text{ Total Conversion cost } = \underline{\hspace{2cm}} \langle I_3 \rangle$$

$$\text{TOTAL FINDING AID CONVERSION COST} = \langle I_2 \rangle \text{ OR } \langle I_3 \rangle = \underline{\hspace{2cm}} \langle 8 \rangle$$

## J. POST-PROCESS DIGITAL FILES

This process involves manual and automated methods for quality review of digital images and accompanying structuring and labeling, as well as of text conversion and encoding. Note that this process may occur in two steps. First, the quality of digital image files should be reviewed prior to OCR and encoding. Second, it will be necessary to inspect text conversion and encoding for accuracy. An institution must establish and document inspection procedures. For example, a threshold for accuracy should be set and made clear to the vendor up front, and means for implementing any re-scanning or reprocessing devised. The level and extent of quality review will depend on the nature of the project, quality requirements (e.g., 100% accuracy), the disposition of the source documents, and the level of consistency achieved by the vendor. Inspection can include review of on-screen and printed versions of the digital files. Staff may be needed to:

- \* Load the digital files
- \* Print paper copies (if not performed as batch-process by the vendor)
- \* Conduct data integrity checks
- \* Perform on-screen and paper inspection
- \* Ascertain the accuracy and consistency of file naming, structuring, text conversion, and encoding
- \* Integrate corrections into the digital file sequence (e.g., re-scans, corrected OCR)
- \* Create derivatives for network access

Remember to factor in a percentage for re-inspection of corrected work.

### PERSONNEL COSTS

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <J<sub>1a</sub>>

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <J<sub>1b</sub>>

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <J<sub>1c</sub>>

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <J<sub>1d</sub>>

AND/OR

*Name/Title:*  
% FTE \_\_\_\_\_ at \_\_\_\_\_ salary/annual + \_\_\_\_\_ benefits/annual = \_\_\_\_\_ <J<sub>1e</sub>>

Note: If various levels of staff are used in the process, calculate a blended monetary rate or repeat the formula for each staff member.

### NON-PERSONNEL COSTS

Description = \_\_\_\_\_ Cost = \_\_\_\_\_ <J<sub>2</sub>>

Description = \_\_\_\_\_ Cost = \_\_\_\_\_ <J<sub>3</sub>>

Description = \_\_\_\_\_ Cost = \_\_\_\_\_ <J<sub>4</sub>>

Description = \_\_\_\_\_ Cost = \_\_\_\_\_ <J<sub>5</sub>>

Note: Post processing digital files may require purchase of appropriate hardware, and imaging and SGML software.

Total personnel cost (<J<sub>1</sub>>) = \_\_\_\_\_ <J<sub>6</sub>>

Total non-personnel cost (<J<sub>2</sub>> + <J<sub>3</sub>> + <J<sub>4</sub>> + <J<sub>5</sub>>) = \_\_\_\_\_ <J<sub>7</sub>>

**TOTAL POST PROCESSING COST = <J<sub>6</sub>> + <J<sub>7</sub>>** = \_\_\_\_\_ <9>

## K. ESTIMATE ADDITIONAL LOCAL COSTS

Additional local costs include:

- \* Project management and tracking
- \* Programming/systems support
- \* Shipping and insurance
- \* Purchasing storage devices, media and software
- \* Other

Travel to project meetings will be handled in overall budget. Note that future costs incurred in providing access to images are not covered.

### PERSONNEL COSTS

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <K<sub>1a</sub>>

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <K<sub>1b</sub>>

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <K<sub>1c</sub>>

*Name/Title:*  
# of hours \_\_\_\_\_ at \_\_\_\_\_ /hour + \_\_\_\_\_ benefits/hour = \_\_\_\_\_ <K<sub>1d</sub>>

AND/OR

*Name/Title:*  
% FTE \_\_\_\_\_ at \_\_\_\_\_ salary/annual + \_\_\_\_\_ benefits/annual = \_\_\_\_\_ <K<sub>1e</sub>>

Note: If various levels of staff are used in the process, calculate a blended monetary rate or repeat the formula for each staff member.

### NON-PERSONNEL COSTS

Description = _____	Cost	=	_____ <K <sub>2</sub> >
Description = _____	Cost	=	_____ <K <sub>3</sub> >
Description = _____	Cost	=	_____ <K <sub>4</sub> >
Description = _____	Cost	=	_____ <K <sub>5</sub> >

Total personnel cost (<K<sub>1</sub>>) = \_\_\_\_\_ <K<sub>6</sub>>

Total non-personnel cost (<K<sub>2</sub>> + <K<sub>3</sub>> + <K<sub>4</sub>> + <K<sub>5</sub>>) = \_\_\_\_\_ <K<sub>7</sub>>

**TOTAL ADDITIONAL COST = <K<sub>6</sub>> + <K<sub>7</sub>> = \_\_\_\_\_ <10>**

## SUMMARY OF PROJECTED EXPENSES

Direct Cost = <1> + <2> + <3> + <4> + <5> + <6> + <7> + <8> + <9> + <10> = \_\_\_\_\_ <11>

Indirect Cost<sup>10</sup> = <11> X % (indirect cost rate)  
 = \_\_\_\_\_ X \_\_\_\_\_ indirect cost rate = \_\_\_\_\_ <12>

Overall Project Cost = <11> + <12> = \_\_\_\_\_ <13>

Project Reimbursement Requested = <13> x % (reimbursement rate)  
 \_\_\_\_\_ X \_\_\_\_\_ (reimbursement rate) = \_\_\_\_\_ <14>

Cost Sharing = <13> - <14> = \_\_\_\_\_ <15>

Cost Per Image = <13> / Total number of images/pages created <from p. 2> = \_\_\_\_\_

## ENDNOTES

<sup>1</sup> Note that, in this worksheet, pound sign (#) indicates "number."

<sup>2</sup> For materials slated for text conversion.

<sup>3</sup> A unit refers to the physical or bibliographic organization of individual items, such as folders, volumes, bundles, etc.

<sup>4</sup> Indicate whether full reel/fiche or just selected frames will be scanned.

<sup>5</sup> In case of multiple pages per frame, indicate whether the full frame will be scanned as one image or subdivided into individual images.

<sup>6</sup> The following web sites provide information on emerging cataloging, naming, and structural metadata schemes:  
 General Information - <http://www.nlc-bnc.ca/ifla/II/metadata.htm>  
 Dublin Core - [http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin\\_core\\_report.html](http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin_core_report.html)  
 Warwick - <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>  
 Encoded Archival Description (EAD) - <http://lcweb.loc.gov/ead/>  
 Handle System - <http://www.handle.net/>

<sup>7</sup> Although the majority of flatbed scanners can handle documents no greater than 8.5" x 14", some high-end flatbed scanners can accommodate documents up to 11" x 17".

<sup>8</sup> Special treatment can include segmenting a page containing both illustrations and text, and treating each segment optimally. This type of treatment is usually more expensive than either bitonal or grayscale scanning.

<sup>9</sup> Regular size bound volumes can be scanned using right angle, prism or overhead flatbed scanners whose scanning platen may accommodate sizes up to 8.5" x 14". Some scanners can accommodate volumes up to 11" x 17" (e.g., Minolta PS3000) or greater (e.g., Bookeye Planetary scanner, which accommodates up to 16.5" x 24.4").

<sup>10</sup> This category includes mounted slides, film strips, and roll film. Be sure to indicate to vendors the particular formats of transparencies and negatives.

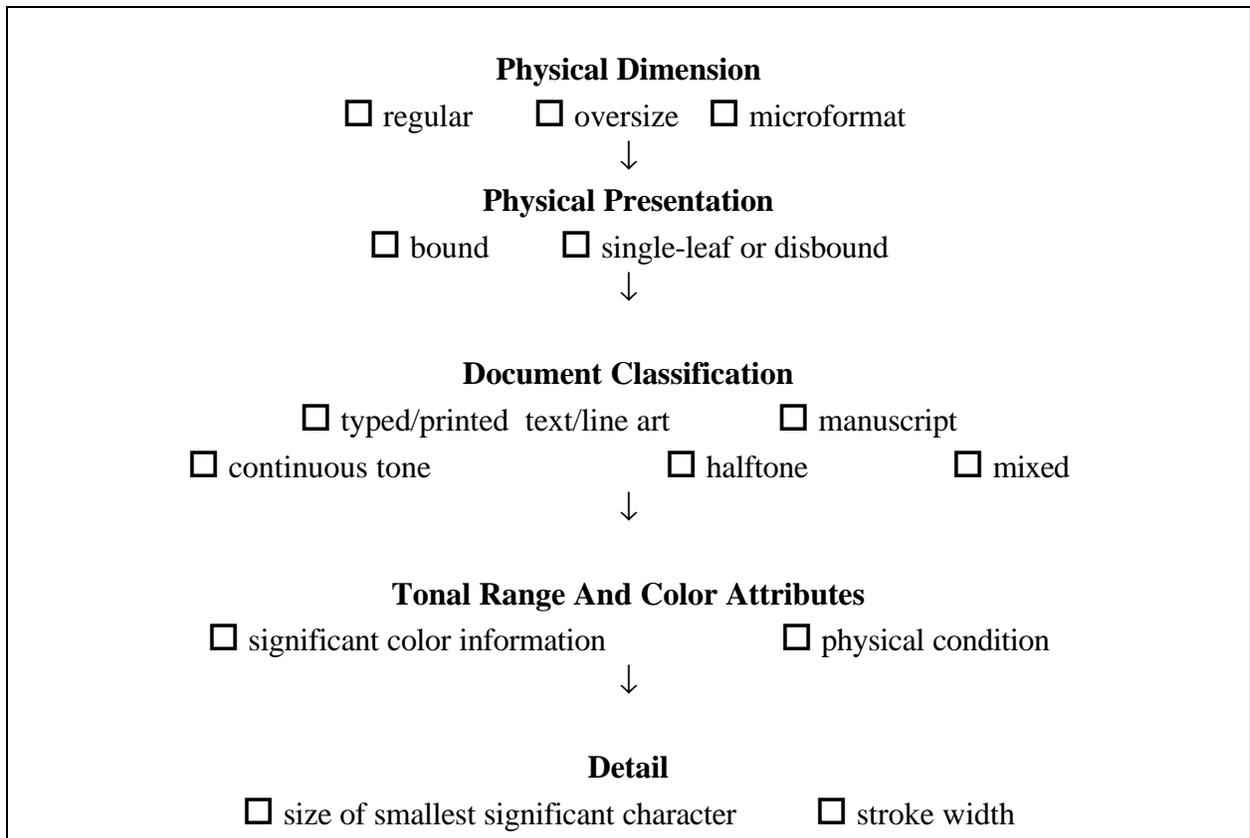
<sup>11</sup> A survey of businesses offering text encoding services found that costs (and bids) are based on a per 1,000 character, *post-processing* cost (e.g., a count of the total number of characters, including characters which comprise the tagging). To account for the tagging "overhead," it is recommended to multiply the total number of characters in your document by 1.5 in order to more accurately estimate the cost of encoding. \*\*This calculation is for general text encoding only. It does not apply to Finding Aids.

<sup>12</sup> The same survey of businesses offering encoding services for finding aids revealed that findings aids demand a much higher ratio of tags than encoding a "regular" text document does, especially when using the SGML EAD DTD. To account for the greater "tag overhead" of finding aids, it is recommended to multiply the total number of characters in the finding aid by 2.6 in order to more accurately estimate the cost of encoding.

**APPENDIX A: ASSESSING DOCUMENT ATTRIBUTES AND SCANNING REQUIREMENTS**

Digital image capture must take into consideration the attributes of the source documents: physical dimensions, physical presentation, level of detail, tonal range, and presence of color. These document attributes will determine the type of scanning approach and devices to be used in digital conversion. For instance, very fine lettering will require high resolution for full informational capture; oversize items (greater than 8.5” x 14”) can not be digitized on a flatbed scanner; fragile documents can not be scanned via an automatic sheet feeder.

The physical size and shape of the documents will determine the appropriate scanning device to be used. The first ‘sort’ is on document dimension; further subdivision is by physical presentation (bound, single-leaf), followed by document classification (e.g., printed text/line art); and then by tonal range and color attributes of the original source document (e.g., a black and white photograph). The final consideration will be the level of detail that must be rendered in the digital file. Capture requirements (resolution, bit depth) can be determined a number of ways. Once this information is gathered, digital conversion requirements and costs can be determined.



**TABLE 1: STEPS IN DETERMINING THE APPROPRIATE SCANNING METHOD**

**Physical Dimension:**

Documents may be classified by the physical size (height and width) into three categories:

1. *Regular size* ( $\leq 8.5'' \times 14''$ )
2. *Oversize* ( $> 8.5'' \times 14''$ )
3. *Microimages* (microfilm, microfiche, aperture cards, transparencies and negatives)<sup>1</sup>

**Physical Presentation:**

The physical presentation of documents will affect the scanning approach that can be used and subsequently the cost of conversion. The worksheet may be used for paper and film-based documents, not three dimensional objects. In general, documents may be classified as:

1. *Bound*
2. *Single-leaf or items that can be disbound prior to scanning*

**Document Classification:**

All paper and film-based documents fall into one of the following categories that will affect their digital recording:

1. *Typed or printed text/line art*: Distinct edge-based representation, with no tonal variation. Includes typed or printed texts, line drawings, woodcuts, maps, music scores, etc.
2. *Manuscript*: Hand produced text, line drawings, blue prints, maps, etc.
3. *Continuous tone or continuous tone-like reproductions*: Smoothly varying gradation of tones. Includes photographs and some original art work, such as charcoal sketches, watercolors as well as graphic art, such as aquatints, lithographs, collotypes, etc.
4. *Halftone or halftone-like*: Regularly spaced pattern of dots or lines, often placed at an angle. Includes graphic art that is created using a fine and highly regular pattern of lines or hatchmarks, e.g., line engravings, etchings, etc.
5. *Mixed*: Contains both text/line art *and* continuous tone or half-tone information. Includes newspapers, magazines, illustrated books.

**Tonal Range and Color Attributes:**

There are four scanning methods:

1. *Bitonal* (information is presented as either black or white; 1 bit per pixel)
2. *Grayscale* (shades of gray information can be presented; normally 8 bits per pixel)
3. *Color* (a full range of color hues and values may be rendered; normally 24 bits per pixel)
4. *Special treatment* (segmenting a page containing both illustrations and text and treating each segment optimally)

---

<sup>1</sup> We have excluded transparencies and negatives larger than 4 x 5 because they are not normally considered microimages (for which special scanning devices exist). Typically, larger size transparent media will be scanned with drum scanners, digital cameras, or some flatbed scanners.

The range of tonal difference between the lightest and darkest areas of an image will affect the method of scanning used. For instance, continuous tone images, such as photographs, must be rendered via grayscale or color scanning to ensure proper representation of the smoothly varying gradations of tones, and to preserve details in the highlights and shadows. A text/line art document can often be captured with black and white (bitonal) scanning. However, one that is badly faded, unevenly printed, smeared, or stained may exhibit a tonal range that cannot be captured adequately with black and white scanning and will require grayscale or color capture to ensure fidelity and legibility of the informational content. Full color documents or color monochromatic documents (e.g., cyanotypes) in which the color information is considered critical to the document's meaning will require color scanning.

**Detail:**

The level of significant detail in a document will affect the resolution or dots per inch (dpi) used in scanning. The higher the resolution, the greater the ability to record fine detail from the source document<sup>2</sup>.

**Example:**

Your institution is interested in converting brittle books. These books are regular size and can be disbound for scanning. The content is exclusively text/line art and there is a good contrast between text and background. The smallest significant character is 1mm in height. These documents can be scanned on a flatbed scanner with 1-bit scanning (black and white) at a resolution of 600 dpi to ensure full capture of small text.

To identify the recommended digital reformatting approach for your collection, first follow the steps to determine the appropriate scanning method (TABLE 1), then find your documents' category in the following document classification scheme:

<b>Legend for Scanning Devices</b>			
FL: flatbed scanner	DR: drum scanner	SH: sheetfeed scanner	DI: digital camera
SL: slide scanner	MI: microform scanner	RP: right angle, prism and overhead flatbed scanner	

***If your documents are regular size and bound, determine their appropriate categories and recommended scanning devices:***

- Typed or printed text/line art
  - ◆ black and white<sup>3</sup> →RP or DI; bitonal scanning
  - ◆ color<sup>4</sup> →RP or DI; color scanning

<sup>2</sup> For a detailed discussion of digital resolution requirements, refer to: Kenney, Anne R., Chapman, Steve. *Tutorial: Digital Resolution Requirements for Replacing Text Based Material: Methods for Benchmarking Image Quality*. Washington, DC: The Commission on Preservation and Access, 1995.

<sup>3</sup> Black media (e.g., ink) on light background or monochrome document in which color information is not significant.

<sup>4</sup> Colored media or support where color information is considered critical to documents' meaning.

**Legend for Scanning Devices**

FL: flatbed scanner	DR: drum scanner	SH: sheetfeed scanner	DI: digital camera
SL: slide scanner	MI: microform scanner	RP: right angle, prism and overhead flatbed scanner	

- Manuscript
  - ◆ black and white →RP or DI; bitonal or grayscale scanning
  - ◆ color →RP or DI; color scanning
- Continuous tone or continuous tone-like reproductions
  - ◆ black and white<sup>5</sup> →RP or DI; grayscale scanning
  - ◆ color →RP or DI; color scanning
  - ◆ color monochrome<sup>6</sup> →RP or DI; color scanning
- Halftone or halftone-like
  - ◆ black and white →RP or DI; bitonal with enhancements, grayscale
  - ◆ color →RP or DI; color scanning
- Mixed
  - ◆ black and white →RP or DI; bitonal with enhancements, grayscale, special treatment scanning
  - ◆ color →RP or DI; color scanning, special treatment scanning

***If your documents are regular size and disbound or single-leaf page, determine their appropriate categories and recommended scanning devices:***

- Typed/printed text/line art
  - ◆ black and white →FL, SH; bitonal scanning
  - ◆ color →FL, SH; color scanning
- Manuscript
  - ◆ black and white →FL, SH; bitonal or grayscale scanning
  - ◆ color →FL, SH; color scanning
- Continuous tone or continuous tone-like reproductions
  - ◆ black and white →FL, SH; grayscale scanning
  - ◆ color →FL, SH; color scanning
  - ◆ color monochrome →FL, SH; color scanning
- Halftone or halftone-like
  - ◆ black and white →FL, SH; bitonal with enhancements, grayscale
  - ◆ color →FL, SH; color scanning

<sup>5</sup> In continuous tone or continuous tone-like reproductions, use of black media produces gray scale information.

<sup>6</sup> Monochromic images where the color information should be retained. For example, the colors used in many 19<sup>th</sup> century photographic print processes, the purplish-brown colors of albumen prints, or the blue color of cyanotypes.

**Legend for Scanning Devices**

FL: flatbed scanner	DR: drum scanner	SH: sheetfeed scanner	DI: digital camera
SL: slide scanner	MI: microform scanner	RP: right angle, prism and overhead flatbed scanner	

- Mixed
  - ◆ black and white →FL, SH; bitonal with enhancements, grayscale, special treatment scanning
  - ◆ color →FL, SH; color, special treatment scanning

***If your documents are oversize and bound, determine their appropriate categories and recommended scanning devices:***

- Typed/printed text/line art
  - ◆ black and white →DI, RP; bitonal scanning
  - ◆ color →DI, RP; color scanning
- Manuscript
  - ◆ black and white →DI, RP; bitonal or grayscale scanning
  - ◆ color →DI, RP; color scanning
- Continuous tone or continuous tone-like reproductions
  - ◆ black and white →DI, RP; grayscale scanning
  - ◆ color →DI, RP; color scanning
  - ◆ color monochrome →DI, RP; color scanning
- Halftone or halftone-like
  - ◆ black and white →DI, RP; bitonal with enhancements, grayscale
  - ◆ color →DI, RP; color scanning
- Mixed
  - ◆ black and white →DI, RP; bitonal with enhancements, grayscale, special treatment scanning
  - ◆ color →DI, RP; color scanning, special treatment scanning

***If your documents are oversize and disbound or single-leaf pages, determine their appropriate categories and recommended scanning devices:***

- Typed/printed text/line art
  - ◆ black and white →DR, SH, DI; bitonal scanning
  - ◆ color →DR, SH, DI; color scanning
- Manuscript
  - ◆ black and white →DR, SH, DI; bitonal or grayscale scanning
  - ◆ color →DR, SH, DI; color scanning

**Legend for Scanning Devices**

FL: flatbed scanner	DR: drum scanner	SH: sheetfeed scanner	DI: digital camera
SL: slide scanner	MI: microform scanner	RP: right angle, prism and overhead flatbed scanner	

- Continuous tone or continuous tone-like reproductions
  - ◆ black and white →DR, SH, DI; *grayscale scanning*
  - ◆ color →DR, SH, DI; *color scanning*
  - ◆ color monochrome →DR, SH, DI; *color scanning*
  
- Halftone or halftone-like
  - ◆ black and white →DR, SH, DI; *bitonal with enhancements, grayscale*
  - ◆ color →DR, SH, DI; *color scanning*
  
- Mixed
  - ◆ black and white →DR, SH, DI; *bitonal with enhancements, grayscale, special treatment*
  - ◆ color →DR, SH, DI; *color scanning, special treatment*

***If your documents are microimages, determine their appropriate categories and recommended scanning devices:***

- 16 mm microfilm
  - ◆ high contrast →MI; *bitonal*
  - ◆ continuous tone →MI; *grayscale*
  - ◆ color →MI; *color*
  
- 35 mm microfilm
  - ◆ high contrast →MI; *bitonal*
  - ◆ continuous tone →MI; *grayscale*
  - ◆ color →MI; *color*
  
- Regular microfiche
  - ◆ high contrast →MI; *bitonal*
  - ◆ continuous tone →MI; *grayscale*
  - ◆ color →MI; *color*
  
- Full-frame microfiche
  - ◆ high contrast →MI, DR; *bitonal*
  - ◆ continuous tone →MI, DR; *grayscale*
  - ◆ color →MI, DR; *color*

*Legend for Scanning Devices*

FL: flatbed scanner      DR: drum scanner      SH: sheetfeed scanner      DI: digital camera  
SL: slide scanner      MI: microform scanner      RP: right angle, prism and overhead flatbed scanner

- Jacketed microfiche
  - ◆ high contrast      →*MI; bitonal*
  - ◆ continuous tone      →*MI; grayscale*
  - ◆ color      →*MI; color*
- Aperture cards
  - ◆ high contrast      →*some FL, DR, DI, MI; bitonal*
  - ◆ continuous tone      →*some FL, DR, DI, MI, grayscale*
  - ◆ color      →*some FL, DR, DI, MI; color*
- Transparencies and negatives (35 mm, 3 ¼ x 4 ¼ , 4 x 5)
  - ◆ high contrast      →*SL, DR, DI, RP; bitonal*
  - ◆ continuous tone      →*SL, DR, DI, RP, grayscale*
  - ◆ color      →*SL, DR, DI, RP; color*