Moving Here Project

Digitisation Guidelines

Contents

Appendices

1. Guidelines Layout

These guidelines follow the layout described below.

1. Overview - A brief run through of the digitisation process. More detailed information regarding each process can be found in the appendices.

2. Moving Here Image Requirements

3. Moving Here Metadata Requirements

4. Appendices

This layout reflects the fact that the Moving Here project has many contributing partners each with a varying degree of knowledge of digitisation.

2. Overview

2.1 The Moving Here project is an ambitious undertaking involving a number of museums, archives and record offices up and down the country. Partner institutions will be aiming to digitally deliver roughly 200,000 images, audio and video clips, online. Clear communication to ensure consistent practice between these partner institutions will be the key to the success of Moving Here.

This document aims to provide the partners with a logical guide to digitising content and the standards at which that digitisation must be achieved. (See Appendix 1 for *nof-digitise Technical Standards and Guidelines, 2. Creation*)

2.2 The digitisation experience within the Moving Here consortium varies considerably from those institutions, which are at the forefront of delivering archival images online to those who have no knowledge of digitisation. Moving Here will therefore be a valuable opportunity for partners to share understanding and experience.

2.3 Given the disparity in experience a brief stage-by-stage overview of the digitisation process is helpful at this point to bring those institutions with less exposure to digitisation up to speed.

2.3.1 Digitisation is simply the creation of a computerised version of physical media. (See Appendix 13 for a *Digitisation Glossary*)

2.3.2 This physical medium varies hugely but will, for Moving Here at least, consist mostly of manuscripts and printed text. There will also be some three dimensional objects, audio and video clips.

2.3.3 In theory the best quality images will be obtained from creating the digital image directly from the original but this is not always possible or practical.

2.3.4 Before digitisation can commence there are four preparatory steps that should be made, bearing in mind that the more thorough the preparative work the smoother the digitisation process should be. The four steps are:

a) Selecting the items to be digitised. (See Appendix 2, *Content Identification*)
b) Establishing the physical makeup of the selected items. (See Appendix 3, *Establish Content Makeup*)
c) Grouping the items together according to their digtisation requirements. (See Appendix 4, *Content Grouping*)
d) Marking up the grouped items for digitisation. (See Appendix 5, *Content Markup*)

2.3.5 Establishing the physical makeup of the item will dictate the digitisation options available both in terms of specification: (See Appendix 6, *Imaging Issues*, and Appendix 7, *Imaging Options*)

a) Image type
b) Resolution
c) File type

and how to digitise the item (See Appendix 8, *Digitisation Hardware*):

a) Using a flatbed scanner

b) Using a digital camera

2.3.6 Together these factors, though governed by resource, will decide if the digitisation should be carried out:

a) In house (See Appendix 9, *Digitisation Options*)
or
b) Outsourced (See Appendix 9, *Digitisation Options*)

2.3.7 To ensure that once digitally captured the images retain their value, increase their availability, are searchable and are user-friendly corresponding descriptive and technical metadata needs to be created. (See Jone Garmendia's Metadata Standards and Cataloguing Guidelines)

2.3.8 A quality assurance procedure should be applied to both the images and the metadata. (See appendix 11, *Quality Assurance* for further details)

2.3.9 A preservation strategy for the digitised images also needs to be developed. (See Appendix 12, *Preservation & Storage*, for further details)

2.3.10 For Moving Here all the above points have to be carried out in accordance to the *nof-digitise Technical Standards and Guidelines*. (See Appendix 1 for *nof-digitise Technical Standards and Guidelines, 2. Creation*, or go online for an up to date version at http://www.ukoln.ac.uk/nof/support/)

2.3.11 As a final point it is worth noting that recommendations and standards for the creation of digital media together with the hardware and software used in the digitisation process are constantly changing. As with many continually developing processes cost also tends to fluctuate accordingly.

3. Moving Here Image Requirements

Image Requirements consist of three factors: Image Type, File Type and Resolution.

The guiding principle that should govern any decision regarding image type, file type or resolution is whether or not the digital rendering is legible online. Those that are not legible are unacceptable. This legibility must also be achieved within the constraints of file size and consequent download times.

The further options explored in Appendix 7 can be used in some part of the digitisation process if they prove more useful or easier to use as long as for delivery to the central team the minimum requirements outlined below are followed.

3.1 Image type

To ensure a degree of commonality between image types and the items they depict on the Moving Here Website the central team has proposed the following guidelines for image type according to the item being digitised. It is felt that this is necessary due to the number of partners involved and the possibility that if no image type standard was set the Moving Here Website would run the risk of becoming a group of disparate images where a map maybe digitised in true colour by one partner and bi-tonal by another. If these guidelines are followed the majority of items, which share the same or similar format, will be digitised using the same image type.

- **Bi-tonal should be used for paper media where manuscripts have been clearly written or printed and where there is good contrast. Bi-tonal can also be used to scan from the microfilm archive master**

- **Greyscale should be used for paper media where contrast is less well defined and for black and white photographs.**

- **24-bit True colour should be used for any document, map, or item where colour is important and provides extra meaning or where the aesthetic value of reproducing it in colour outweighs the resource factors such as cost, scan time, image file size and storage.**

- **Three-dimensional objects should also be digitised in colour.**

- **Video should be digitised as in whatever its original state.**

Bearing in mind the *nof-digitise Technical Standards and Guidelines* suggestion of ‘One scan for all’ it is up to the partners if they wish to scan images at a higher image type spec as long as when it is delivered to the central team for loading onto the website it conforms to the above guidelines. Partners should take into account the resource involved in manipulating and adjusting each image.

3.2 Resolution

- **The minimum resolution for the archival master image destined for the Moving Here project is 300dpi.**

Some partners may wish to choose a higher dpi for specific images but it is felt that this resolution offers good quality images (printers minimum) and file size is manageable.

- **Having been scanned at 300dpi, or above, images should be delivered to the central team on CD or via email at 150 dpi for delivery over the web.**

This reduction can be done either in house by the partner or externally by a scanning bureau, which can deliver two sets of the same images at two different resolutions, one archive master and one for web access. It should be noted that 150dpi images are often more legible than those of 300dpi and that 150dpi is still considerably greater than the 72dpi which is the default for most images on the web.

3.3 File type (updated 14 May 2002)

- **TIFF files should be used for the digital masters.**

- **Images delivered to the central team should be 'web ready' TIFF or JPEG files with a resolution of 150 dpi. (The central team will of course accept uncompressed TIFF files at 150dpi, as stated in the previous version of the guidelines, from those partners who have already embarked on the digitisation of content)**

- **Ideally where a metadata unit comprises of numerous images they should share a common file type, i.e. all JPEGs or all TIFFs.**

3.4 Image Referencing and File Structure (updated 14 May 2002)

When scanning the images should be named numerically and should be delivered with the following file reference structure:

Partner/
    Record/
        Subrecord/
            Digital Resource(s)

For example,

PRO/
    PRO1_1/
        0/
            0001.tif,0002.tif,0003.tif etc.

Notes:
- Three level hierarchy
- If a subrecord is not present, directory should be named 0
- If a directory name contains an illegal character ie.\,/,:,*,?,",<,>,| then this should be replaced by the _ character.

3.5 Video

The *nof-digitise Technical Standards and Guidelines* are very open when it comes to video standards, *"Video must be created and stored using the appropriate MPEG format or the proprietary formats MS AVI, ASF or Quicktime."*

Unless partners who are familiar with digitising video have specific in house standards it is recommended that video be digitised at the following rates.

- **Encoding to MPEG-2 @ 5Mb/sec as an archival master format**
- **Then transcoding to MPEG-4 @128Kb/sec. This will be made available online.**

MPEG-2 is considered the "de facto" standard for broadcast quality compressed video. The bit rate used depends on the quality of the input. Low resolution MPEG-4 downloads will create files that are just under 1Mb per minute in size.

3.6 Audio

As with video the *nof-digitise Technical Standards and Guidelines* are very open when it comes to audio standards: *"Sound must be created and stored using MP3, RealAudio, MS WAV or Sun AU formats."*

Similarly unless partners who are familiar with digitising audio have specific in house standards it is recommended that audio be digitised using the following file type.

- **MP3**

3.7 Audio and Video Delivery Online

There are two options for delivering video and audio online:

1. bite size downloads,
   or
2. streaming

The choice between bite size downloads and streaming can often be dictated by copyright. For example if the material is not 'owned' by those who put it online it will need to be streamed so that the source material remains protected.

Bite size downloads has the advantage of more manageable file sizes, though the duration of the video and audio will only be a few minutes at most. This does, however, mean that either:

1. a certain amount of editing, principally for video, may be required to deliver coherent playback

2. or that the 'bites' of playback are identified at the markup stage. i.e. editing takes place before digitisation. Editing prior to digitisation would require a 'markup' to be carried out identifying accurate in and out timecodes each side of every individual story. This can be a simple Excel document accompanying each tape.

Streaming has the advantage of providing longer video and audio playback but in order to do so it requires a large bandwidth to be maintained.

Capability for both streaming and bite size downloads will be built into the online delivery system.

3.8 File Size

Given the average download times below, it is important that file size is kept to a minimum to ensure that users get the most out of Moving Here. The table below provides a rough idea of download times according to different connection speeds.

File size	28.8k	56k	ISDN
60k	2 min	1 min	< 1 min
1-3 MB	20 min	10 min	5 min
4-11 MB	1-2 hours	40-60 min	15-30 min

As a general rule Moving Here will be aiming to keep file sizes to under 1mb. For the most part this should be achievable for images though it is likely that audio and video clips will exceed this.

4. Moving Here Metadata Requirements

4.1 Metadata is, in basic terms, data about data. The *nof-digitise Technical Standards and Guidelines* tells us that *'any digital object will benefit form the addition of metadata, as it will ensure that it can be more easily identified, understood and utilised'*. For Moving Here it can be divided into two distinct types; Descriptive and Technical.

4.2 Metadata Capture
Descriptive **(See the Metadata Standards and Cataloguing Guidelines available from Jone Garmendia,** jone.garmendia@pro.gov.uk**)**

4.3 Metadata Capture
Technical

Technical metadata includes information such as the file type, compression ratio, image orientation, resolution, scanner used, operators name, etc…

As there will be two sets of images created for Moving Here there will also be two corresponding sets of technical metadata, one for the archive master images and one for the online deliverables. The technical metadata for the online deliverables should be sent to the central team at the same time as the corresponding derivative image set.

Below is a simple excel input template for the capture of the technical metadata.

This data will be used for the image loading procedure and will also provide us with an audit trail should the image quality not be sufficient. Should it be necessary to replace an image the technical metadata also provides us with precise information as to how to recreate it.

- Technical metadata must be captured for each image & collections should be kept together.
- Each field can be populated at the time of scanning.
- If the digitisation is to be outsourced a technical metadata template should be completed by the bureau as specified by the partner in a scanning framework agreement.

Image_dimension_bytes: The number of bytes along the vertical and horizontal dimensions.

Image_dimension_pixel: The number of pixels along the vertical and horizontal dimensions.

Where a large image has been split into two smaller images, either 'y' or 'n'

Collection	Collection no.	Piece	Item	image no.	Technician	Scanning_ Start	Scanning_ End	Scanning_location	Scannner_used	Scanner_version	Software	image_resolution	image_dim_pxl	image_dim_byt	image_split
BT	26	1233	0	0001	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	1800 x 1000	4000 x 2200	n
BT	26	1233	0	0002	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	1800 x 1000	4000 x 2200	n
BT	26	1233	0	0003	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	1800 x 1000	4000 x 2200	n
BT	26	1233	0	0004	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	1800 x 1000	4000 x 2200	n
BT	26	1233	0	0005	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	1800 x 1000	4000 x 2200	n
BT	26	1233	0	0006	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	1800 x 1000	4000 x 2200	n
BT	26	1233	0	0007	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	1800 x 1000	4000 x 2200	n
BT	26	1233	0	0008	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	1800 x 1000	4000 x 2200	n
BT	26	1233	0	0009	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	2000 x 1500	4000 x 2200	n
BT	26	1234	0	0010	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	1800 x 1000	4000 x 2200	n
BT	26	1234	0	0011	C.Potter	01/01/02	01/01/02	RCD	big scanner	1	scanthis	150dpi	1800 x 1000	4000 x 2200	n
FO	372	3671	0	0001	P.Goodwin	01/03/02	02/03/02	eaccess	little scanner	16	scanthat	150dpi	1800 x 1000	4000 x 2200	n
FO	372	3671	0	0002	P.Goodwin	01/03/02	02/03/02	eaccess	little scanner	16	scanthat	150dpi	1800 x 1000	4000 x 2200	n
FO	372	3671	0	0003	P.Goodwin	01/03/02	02/03/02	eaccess	little scanner	16	scanthat	150dpi	1800 x 1000	4000 x 2200	n
FO	372	3671	0	0004	P.Goodwin	01/03/02	02/03/02	eaccess	little scanner	16	scanthat	150dpi	1800 x 1000	4000 x 2200	n
FO	372	3671	0	0005	P.Goodwin	01/03/02	02/03/02	eaccess	little scanner	16	scanthat	150dpi	1800 x 1000	4000 x 2200	n
FO	372	3671	0	0006	P.Goodwin	01/03/02	02/03/02	eaccess	little scanner	16	scanthat	150dpi	1800 x 1000	4000 x 2200	n
HO	54	58	0	0001	D.Davies	01/03/02	02/03/02	SuperScan, Wales	big scanner	1	scantheother	150dpi	1800 x 1000	4000 x 2200	n
HO	54	58	0	0002	D.Davies	01/03/02	02/03/02	SuperScan, Wales	big scanner	1	scantheother	150dpi	1800 x 1000	4000 x 2200	n
HO	54	58	0	0003	D.Davies	01/03/02	02/03/02	SuperScan, Wales	big scanner	1	scantheother	150dpi	1800 x 1000	4000 x 2200	n
HO	54	58	0	0004	D.Davies	01/03/02	02/03/02	SuperScan, Wales	big scanner	1	scantheother	150dpi	1800 x 1000	4000 x 2200	n

If the image is cropped, eg to remove or reduce a border area. Either 'y' or 'n'

Bit depth of each pixel, and whether multiple bits convey grey tones or colour. eg, 1bit bi-tonal, 8-bit greyscale, 24-bit colour

The colour space used for the image.

Positive or Negative

Portrait or Landscape

Addition of any comments eg to do with the legibility of the original or the details of image enhancing software if used etc.

Storage Media_id: A sequential number given to each cd.

Each CD should be titled in the following format. Institution Abbreviation_M(oving)H(ere)_Storage Media_id

image_crop	image_tonal_resolution	image_format	Compression	compression_ratio	Image_orientation	pos_neg	image_col_space	comments	SM_id	CD Title
n	8-bit greyscale	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	8-bit greyscale	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	8-bit greyscale	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	8-bit greyscale	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	8-bit greyscale	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	8-bit greyscale	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	8-bit greyscale	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	8-bit greyscale	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	24-bit colour	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	8-bit greyscale	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	8-bit greyscale	Tiff v6.0	uncompressed	null	portrait	positive	RGB		1	PRO_MH_1
n	bi-tonal	Tiff v6.0	uncompressed	null	portrait	negative	RGB		1	PRO_MH_1
n	bi-tonal	Tiff v6.0	uncompressed	null	portrait	negative	RGB		1	PRO_MH_1
n	bi-tonal	Tiff v6.0	uncompressed	null	portrait	negative	RGB		1	PRO_MH_1
n	bi-tonal	Tiff v6.0	uncompressed	null	portrait	negative	RGB		1	PRO_MH_1
n	bi-tonal	Tiff v6.0	uncompressed	null	portrait	negative	RGB		1	PRO_MH_1
n	bi-tonal	Tiff v6.0	uncompressed	null	portrait	negative	RGB		1	PRO_MH_1
n	bi-tonal	Tiff v6.0	uncompressed	null	Landscape	positive	RGB		2	PRO_MH_2
n	bi-tonal	Tiff v6.0	uncompressed	null	Landscape	positive	RGB		2	PRO_MH_2
n	bi-tonal	Tiff v6.0	uncompressed	null	Landscape	positive	RGB		2	PRO_MH_2
n	bi-tonal	Tiff v6.0	uncompressed	null	Landscape	positive	RGB		2	PRO_MH_2

4.4 Technical Metadata Capture for Audio and Video (updated 14 May 2002)

If audio and/or video is being digitised the following technical metadata (see examples below) should be captured in the simple excel template below.

A Audio
V Video
VA Video & Audio
AT Audio with Transcription

batch_id	all
technician	all
capture_start	all
capture_end	all
software	all
format_version	all
compression	all
Dimensions_pxl	V
resolution	A ,VA,AT
bit_rate	A,V,VA,AT
encapsulation	A,V,VA,AT
character_set	AT
associated_dtd	AT
structural_divisions	AT
dr_type	all

Batch 1 is video with audio metadata
Batch 2 is a video batch with no audio metadata
Batch 3 is an audio batch
Batch 4 is an audio batch with accompanying transcription

batch_id	technician	capture_start	capture_end	software	format_version	compression	Dimensions_pxl	resolution	bit_rate	encapsulation	character_set	associated_dtd	structural_divisions	dr_type
1	joe bloggs	01/01/2001 09:00	01/01/2001 11:00		quicktime 1.1	mpeg3	640x640		67 minutes 12 seconds	mpeg3				V
1	joe bloggs	01/01/2001 09:00	01/01/2001 11:00		aiff interleaved	mpeg3		44.1kHz	16 bit	real audio 2				VA
2	joe bloggs	01/01/2001 09:00	01/01/2001 11:00		quicktime 1.1	mpeg3	640x640		67 minutes 12 seconds	mpeg3				V
3	fred bloggs	11/01/2001 09:00	11/01/2001 10:00		wav			44.1kHz	8 bit	wav				A
4	fred bloggs	10/01/2001 09:00	10/01/2001 10:00		wav			44.1kHz	8 bit	wav				A
4	fred bloggs	12/01/2001 09:00	12/01/2001 10:00		MS word 97	zip file					ASCII	EAD	TEI element DIVn used	AT

5. Delivery

Once the images have been digitised, the images captured electronically and corresponding technical and descriptive metadata recorded they have to be delivered to the Moving Here central team at the PRO for loading onto the Moving Here website.

- The images (150dpi web deliverables) should ideally be burnt to CD and sent to the central team at the following address

  Peter Goodwin
  Moving Here Project
  Eaccess
  Public Record Office
  Kew
  Richmond
  Surrey
  TW9 4DU

  Alternatively the images can be emailed to the following email address,

  peter.goodwin@pro.gov.uk

  though to do so may prove impractical given the volume of images for initial deliveries. Individual images or very small groups, for example replacement images, can be emailed to the above address.

- The catalogue metadata contained either in the XML template or a word file should be delivered via email to the above address in advance of the images. (See Jone Garmendia's Metadata Standards and Cataloguing Guidelines)

- The technical metadata should be delivered via email to the above address or be copied to a floppy disc and accompany the CD containing the corresponding images.

- Collections should be delivered in their entirety. Though the partners can make as many or as few deliveries as they wish.

- Each delivery **must** be accompanied by a *Copyright Clearance Form (*see appendix 10 & See suggested *Procedures for Clearing Copyright for Moving Here Project* (available from Michael Gilliat, michael.gilliat@pro.gov.uk)) together with a schedule listing precisely what is being delivered.

APPENDIX 1

1. nof-digitise Technical Standards and Guidelines (2. Creation)

These guidelines are subject to change. Please consult the website at www.peoplesnetwork.gov.uk/nof/technicalstandards/creation.html for an up to date version.

---

2 Creation

This section covers the actual creation of digital resources by nof-digitise funded projects. Technical standards decisions made at this stage will have a fundamental impact on the manageability, accessibility and viability of the resources created.

2.1 Requirements

2.1.1 File Formats

Open standard formats should be used when creating digital records in order to maximise access to resources. (Note that file formats for the delivery of digital records to users are outlined in 5.1.1.) The use of open file formats will help with interoperability, ensuring that resources are reusable and can be created and modified by a variety of applications. It will also help to avoid dependency on a particular supplier.

However, in some cases there may be no relevant open standards or the relevant standards may be sufficiently new that conformant tools are not widely available. In some cases therefore, the use of proprietary standards **may** be acceptable. However, where proprietary standards are used, the project **must** explore a migration strategy that will enable a transition to open standards to be made in the future.

If open standards are not used, projects **must** justify their requirement for use of proprietary standards within their business plan, paying particular attention to issues of accessibility.

Text based content **must** normally be created and managed in a structured format suitable for delivery as HTML or XHTML. In most cases storing text-based content as HTML 4, XHTML 1.0 (or subsequent versions) or XML with another Document Type Definition (DTD), either in plain files or within a database of some kind, will be the most appropriate option. HTML and XML documents should be validated against a published DTD. In some instances, projects may choose to store text-based content using PDF. PDF (Portable Document Format) is a proprietary file format owned by Adobe that preserves the fonts, formatting, colours and graphics of the source document. PDF files are compact and can be viewed and printed with the freely available Adobe Acrobat Reader. However, as with any proprietary solution, there are dangers in its adoption and projects must explore a migration strategy that will enable a future transition to open standards to be made. They **must** also ensure that accessibility issues have been addressed.

Images, video, sound and music should initially be created and stored using appropriate open or proprietary standard formats as outlined below.

Images **must** be created using one of the following formats - TIFF, PNG, GIF or JPEG/SPIFF. In general photographic images **should** be created as TIFF images. However, in cases, for example when using cheaper digital cameras, it may be appropriate to use JPEG/SPIFF as an alternative. This will result in smaller, but lower quality images. Such images may be appropriate for displaying photographs of events etc. on a Web site but it is not suggested that such cameras are used for the large-scale digitisation of content. Line drawings, such as certain types of computer-generated imagery, **should** normally be created as PNG or GIF images.

Vector graphics **should** be created and stored using a rich open format such as the SVG (Scalable Vector Graphics) format. Use of the proprietary Macromedia Flash format may also be appropriate, however projects **must** explore a migration strategy so that they can move to more open formats once they become widely deployed.

Video **must** be created and stored using the appropriate MPEG format (MPEG-1, MPEG-2 or MPEG-4) or the proprietary formats Microsoft AVI, ASF or Quicktime.

Sound **must** be created and stored using MP3, RealAudio, Microsoft WAV or Sun AU formats.

2.1.2 Character Encoding

Projects **must** store and deliver text-based content in a way that is compatible with the Unicode UTF-8 encoding form. Unicode is a standard for representing characters, letters and numbers. Most modern office software products support UTF-8 by default. If appropriate, the UTF-16 encoding form **may** be used in preference.

2.1.3 Geographic Information Systems

Those projects seeking to employ Geographic Information Systems (GIS) **must** obtain appropriate permissions for use of any map data from third parties. Existing licences for use of data from Ordnance Survey and other companies may not extend to delivering services to the public over the Internet.

Projects **must** ensure that data sets combined for the purposes of delivering their service are of similar scale and resolution, and appropriate for being used together in this manner.

Commercial GIS products selected for use **should** comply with emerging industry standards from the Open GIS Consortium.

2.1.4 VRML/3D Content

Projects making use of three-dimensional virtual reality (VR) 'fly throughs' and models **must** consider the needs of users accessing their site using typical computers and modem connections.

These models are typically used in the reconstruction of vanished or ruined monuments, such as abbeys or castles, or even in simulating whole areas of an ancient landscape such as Salisbury Plain. Traditionally, models have been constructed and displayed using extremely powerful computer workstations, and this continues to be the case for the most detailed. For NOF projects, which are required to deliver the results of their work to a large audience via the Internet, such highly detailed models are almost certainly unhelpful. Nevertheless, there is scope for usefully incorporating less complex models into the Web sites made available to users.

In generating these models, projects **must** be aware that the majority of their users for the foreseeable future will continue to access the Internet using a 56k modem, rather than any higher bandwidth technology. Similarly, the specifications of the computers being used by typical visitors are likely to be significantly lower than those of the machines on which projects generate and test any such models. Projects **must** therefore consider the usability of their models in such conditions, and **should** test them using typical modem connections and home, school, or library computer systems.

Standards in this area continue to evolve, but projects **should** produce VR models compatible with the VRML 97 specification, and **should** seek to move to X3D once it is finalised, where such a move is viable.

Apple's QuickTime VR (QTVR) is not a true 3D image format, but does offer some useful functionality. Projects which do not require the full functionality of VRML 97 or X3D **may** wish to consider using QTVR instead.

2.2 Guidance

2.2.1 Data Capture

Standards for initial data capture are still developing and it is difficult to identify specific technical requirements at this time. In any case, different resource types may have quite different requirements. However, projects must demonstrate that they have considered the implications of the following three issues:

- the selection of materials for digitisation,
- the physical preparation of materials for digitisation,
- the digitisation process.

The JISC Image Digitisation Initiative (JIDI), the Arts and Humanities Data Service (AHDS) and the Technical Advisory Service for Images (TASI) all provide further guidance on this topic.

A variety of guidance regarding digitisation is also available in various publications. An important recent text is Anne R. Kenney and Oya Y. Rieger's, *Moving Theory into Practice: digital imaging for libraries and archives* (Research Libraries Group, 2000). Of importance also are the LG/NPO conference papers collected together in, *Guidelines for Digital Imaging* (National Preservation Office, 1998) - which are also available on the Web. In addition, the Digital Library Federation, the Council on Library and Information Resources and the Research Libraries Group have recently published on the Web some useful Guides to Quality in Visual resource Imaging.

2.2.2 Metadata

Metadata is structured data about data. Typically, the term refers to any data used to aid the identification, description, location or use of networked electronic resources. Any digital object will benefit from the addition of metadata, as it will ensure that it can be more easily identified, understood and utilised. Metadata will be needed through the whole life cycle of a digital resource and is consequently referred to throughout this document.

The types of metadata that need to be recorded during the digitisation process itself include, for example, information about the nature of the source material, the digitisation equipment used and its parameters (formats, compression types, etc.), administrative metadata about the agents responsible for the digitisation process itself, etc. In many cases, the only time that this information can be recorded is as part of the digitisation process itself and, in some cases, it may be possible to generate the output of this metadata from the digitisation software used.

There is, however, no single standard for this type of metadata. For images, a committee of the US National Information Standards Organization (NISO) has produced a draft data dictionary of technical metadata for digital still images (July 2000). A more compact set of sixteen metadata elements was published in 1998 by a Working Group on Preservation Issues of Metadata constituted by the Research Libraries Group (RLG).

2.2.3 Preservation

Preservation concerns apply both to the information object being digitised and to the surrogate digital object when it has been created. Those responsible for the project **must** weigh-up the risks of exposing original material to any digitisation process, especially where the items are unique, valuable or fragile. More guidance on this topic can be found in Guidelines for digital imaging conference papers.

Preservation issues **must** be considered an integral part of the digital creation process. Preservation will depend upon documenting all of the technological procedures that led to the creation of an object, and this - in many cases - can only be done at the point of creation.

Projects must consider the value in creating a fully documented high-quality 'digital master' from which all other versions (e.g. compressed versions for accessing via the Web) can be derived. This will help with the periodic migration of data and with the development of new products and resources.

APPENDIX 2

2. Content Identification

2.1 Before digitisation can begin the Moving Here content will need to be identified. Many of the partners have already progressed beyond this stage or are currently finalising their choices. For the benefit of those that have not yet made a selection or are trying to reduce content already selected the following guidelines should be taken into account. Factors that govern the identification of the content should include:

a) An item's relevance to the migration types and themes within them – are they important for the story element of the Moving Here project? - are they part of the educational content?
b) An item's availability for digitisation, eg if it is damaged or there are copyright issues. (If you have any queries on the copyright of a particular item the following website provides useful information on the subject including links to some detailed reference material: http://www.peoplesnetwork.gov.uk/nof/goodpractice.htm) See suggested *Procedures for Clearing Copyright for Moving Here Project* (available from Michael Gilliat, michael.gilliat@pro.gov.uk)
c) Beyond Moving Here if the partner institution wishes to reuse the data or incorporate it into a digital archive. Is there value in the item being digitally archived? Bear in mind future image use may demand high spec imaging.
d) Whether a digital image of a specific document already exists, though it should be noted that NOF will not pay for previously digitised material.
e) Similarly, are any of the items selected available in more than one format. eg original and microfilm. This may prove important in terms of resource availability and cost.
f) This leads onto another determining factor – are there resources available for digitisation? e.g. where there is a lot of similar relevant material it may be advisable to digitise representative samples rather than an entire collection. It should be noted that it is considered best practice to digitise collections in their complete state in order to maintain their intellectual integrity. However, resources may well curtail this 'best practice' and it should be remembered that the variety of items to be digitised will be of greater interest to the general user than a large collection of similar ones. This argument also relates to point c).
g) Is it technically possible to capture the information content of the original material?

The final point relates to one of the most important aspects of digitisation that has to occur before any scanning can commence. Establishing the physical makeup of the items to be digitised. This physical makeup will dictate much of the subsequent processes and is subservient only to resource.

APPENDIX 3

3. Establish Content makeup

3.1 Once the material for digitisation has been selected it is important that the makeup of the content is accurately established. The main elements to record are format, size, colour, quantity and condition. All of which will have a direct impact on the cost of digitisation and on timescales. The physical makeup of the content dictates the scanning options available.

3.2 Format: for example, original, facsimile, text documents, large maps, photographs, microfilm, audio, video, bound volumes, 3D objects etc… Content makeup should take into account whether any of the content identified is available for scanning in different formats, e.g. original and microfilm. If this is the case and it is unclear which format the scans should be taken from, a list of the pros and cons of using each should be drawn up. This should take into account such factors as quality of image, cost of digitisation, resources etc…

3.3 Size: clearly size will vary according to the format, but basically documents and other flat media should be divided into subA3 and greater than A3. Microfilm should be 35mm and the duration of audio and video clips should be noted.

3.4 Colour: The value of digitally reproducing an item in colour can be gauged by two factors, whether the colour is important to the items meaning and whether it has aesthetic value.

3.5 Quantity: Obviously the amount of pages in a collection, images on microfilm etc… must be recorded

3.6 Condition: A benchmark for condition of documents may be difficult to establish within a project such as Moving Here which involves many partners as one institutions 'good' condition maybe another's 'poor'. When attempting to establish the physical condition of the material itself it is important to take into account whether or not any adverse ‘condition’ may be corrected or at least reduced by digitisation. The following are examples of conditional aspects that you should be wary of, torn pages, sun bleached photographs, dirt, creasing of paper, yellowing, fading, quality of the Image, faint images, poor contrast, out of focus, challenging colours, video and audio interference.

Handling issues, fragility, uniqueness and transportation considerations should also be taken into account - many institutions will not allow original material to be removed from the building which will present further digitisation resource issues.

APPENDIX 4

4. Content grouping

4.1 Once the content makeup is firmly established you should group together, as far as possible, those items with the same or similar digitisation requirements. There will of course be items that are catalogued together that are of a mixed variety and would not otherwise be grouped together. Collections of mixed media should be kept together where possible. Some partners may feel it is more straightforward to group their content, in total or in part, by collection.

Partners should divide their content by format, size, colour content and condition, this will make the markup process more straightforward. It will also give clear indications as to the scanning specification required.

4.2 Format: In terms of format the divisions should be clearly defined. Originals, facsimiles, text documents, large maps, photographs, microfilm, audio, video, bound volumes, 3D objects etc… all will require a different scan set up and some will need to be scanned on different machines or photographed using a digital camera.

4.3 Size: The larger an item the longer the scan time. They may need to be scanned on special scanners or even have to be scanned in parts though obviously this should be avoided if possible.

4.4 Colour content: Where colour content is important to the item or has additional meaning it should be scanned in colour. All colour video and 3D objects should be digitised in colour.

4.5 Condition: Some items may either require conservation before they are digitised or extra care may need to be taken when the scanning is carried out.

4.6 The aim of establishing makeup is to narrow down the scanning options available according to the makeup of the specific content.

APPENDIX 5

5. Content Markup

Once the makeup of the content has been established and the collections divided into groups for scanning the markup can be carried out.

5.1 The aim of the markup is to make the scanning process as straightforward as possible and is simply a means of identifying exactly what is to be scanned and at what specification it is to be scanned at to the operator, be they in house or at a scanning bureau. The more information provided at the markup stage the smaller the room for error. The efficiency of the scanning process depends largely on the amount of preparation undertaken. If the scanning is to be outsourced much of the information recorded here will help when drawing up a scanning framework agreement for tender. (Contact peter.goodwin@pro.gov.uk for a general Scanning Framework Agreement if required.*)*

5.2 Of course the markup method will differ from partner to partner and from format to format so what follows are some suggestions about how this can be carried out given different formats. The importance of the markup is to convey the relevant scanning information to the operator. It does not follow that every item selected requires a markup, indeed now that the content has been grouped according to format they can be marked up as a group in terms of scanning spec but a list of what is contained within that group is a good idea especially if the scanning is being carried out externally. Doubtless each partner will settle on the method that will suit their needs.

5.2.1 A markup is particularly important if a selection of items within a collection has been identified for scanning. This may be due either to the selective relevance of those items eg a collection of marine passenger lists where only ships from specific ports are significant or alternatively a large collection of similar material from which a representative selection must be made in order to remain within the resource constraints. We would not wish the selected items to be separated from a large ordered collection.

5.2.2 Perhaps the best way to markup such a selection is to physically 'tag' those items to be scanned. For example if the format is paper media each selected page could be tagged with a white paper strip attached with a brass paper clip. Alternatively page identifiers could be recorded e.g. page nos. This would mean that the collection could remain together and that the operator would scan exactly the right documents.

5.2.3 For other non flat formats such as video, audio and three dimensional objects the markup is more straight forward as they should be more clearly defined. For audio and video duration should be noted as this will help determine cost and for three dimensional objects the markup should

record if there are any areas which the digitisation should focus on e.g. to avoid/highlight damage.

Once the markup is complete you will be in a position to know what is to be scanned and at what specification, the next step is to decide how it is to be scanned.

APPENDIX 6

6. Imaging Issues

As previously mentioned the physical makeup of the content dictates the scanning options available.

6.1 The *nof-digitise Technical Standards and Guidelines* for creating digitised content recommend that digital preservation issues be observed when producing digital content. Many of the partners may wish to bear this in mind when deciding on content - will these images be useful beyond Moving Here? *A good baseline to creating a digital file that will be long-lasting would be Scan Once for All Purposes – this means that all the complex and expensive preparation work will only need to be done once.*

The guidelines recommend that projects consider the value in creating a fully documented high-quality 'digital master' from which all other versions (e.g. compressed versions for accessing via the Web) can be derived. This 'digital master' file should be created at the highest suitable resolution and bit depth that is both affordable and practical. This master file then becomes the source for every other version of that item that the project will require, such as Web surrogates, versions for high quality printing and so on.

However, it should be remembered that the 'Scan Once for All Purposes' recommendation is just that. Some of the partners may feel that some of the images that they are aiming to have digitised as part of the Moving Here project would not be their first choice for being digitally archived or that the nature of the item, i.e. printed text would not benefit from being scanned at a high resolution in true colour. Clearly this depends on a variety of factors including condition, popularity, other surrogates available, resources for a digital archive, relevance for future projects etc… If this is the case it is possible to deliver the images for delivery over the web (<u>not</u> the digital masters) to the Central Team at the PRO at the minimum specification according to the medium.

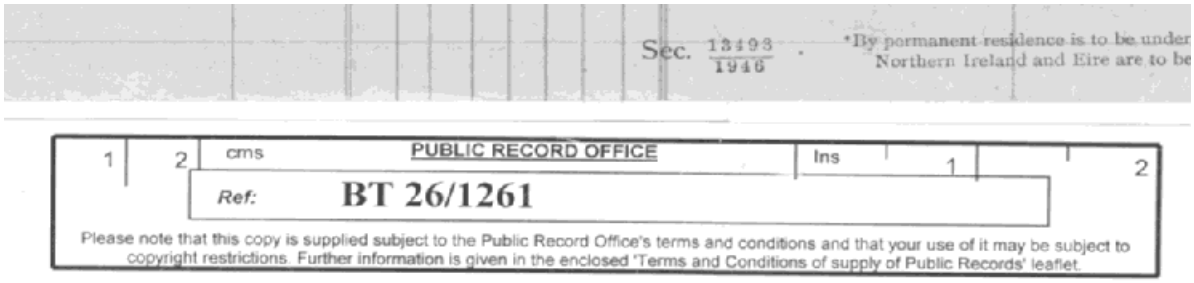
The advantages of this ‘one scan for all’ approach are clear; the disadvantages are primarily of cost, both in outlay and time. In theory the higher the scanning specification the higher the quality of the archive master though this also means the higher the specification the greater the scanning cost, scan time and also the larger the storage capacity required. For example scanning in true colour at 300dpi can result in a 70mb tiff file!

6.2 *The 'digital master' files will become an archive version of the data – it remains as pure a representation of the original as possible. Ideally more than one copy should be stored on more than one media type and in more than one geographical location, thus providing a degree of protection against data corruption, media failure and physical damage to equipment.* Each of the Partners will be responsible for preserving their own digital master images. This will be discussed further in the preservation and storage appendix.

6.3 Digital Watermarks: The benefits of a digital watermarking every image are threefold; it would act as a copyright notice, would identify the ownership of materials and would remind viewers of their limited copying rights. However for Moving Here digital watermarks have been deemed unnecessary as there will be copyright notices on the website (See suggested *Procedures for Clearing Copyright for Moving Here Project* (available from Michael Gilliat, michael.gilliat@pro.gov.uk)), users will be unable to download the master images and the images on the web will be at too low a resolution to allow quality printing. Digital watermarking video and audio is more complex. One way to watermark video is to have a watermark visible at the top right hand of the picture at all times. Streaming both audio and video will help to protect copyright. The *nof-digitise Technical Standards and Guidelines* advise nof funded projects to consider digital watermarking but is fully aware of the sometimes impractical nature of incorporating and maintaing them. It is up to individual partners to have their archival images watermarked should they desire to do so but the images delivered to the central team should not contain any.

6.4 A digital watermark should not be confused with a reference bar. Unlike digital watermarks, which are interwoven with the image, reference bars are placed at the foot of the image and contain information about the scale, collection origin of the image and may also contain a blanket copyright statement, as in the one below. They can either be added as a slip of paper at the point of scanning or they can be added after the scanning using image edit software like *Adobe Photoshop*.

The advantages of including a reference bar are clear, (it may indeed be standard practice for some of the partners), as it will act as an identifier for the document and as such will prove to be valuable beyond Moving Here. They can also be designed to contain other useful information should it be required.
Again it is up to the individual institution, if a reference bar is included. If they are to be included any reference to the Moving Here Project should be omitted to ensure value of the image beyond Moving Here.



APPENDIX 7

7. Imaging Options

For a clearer understanding of the options available and their impact the partners should be aware of the difference of Image type, resolution and bit-depth and more specifically what the *nof-digitise Technical Standards and Guidelines* advise. Although the Moving Here project will only adopt some of these options it is as well for the partners to be aware of all the variants so that they can understand why these decisions have been taken.

7.1 Image type

There are three main basic types of images: 1-bit black and white (bi-tonal), 8/16-bit greyscale and 24-bit colour.

7.1.1 A 1-bit bi-tonal image means that the bit can either be black or white which means that the image being scanned needs to be of good quality with good contrast and without faded detail. The primary drawbacks with this image type are the quality issues. However the latest scanning software means that good quality scans are achievable. For example the images of the census returns in the PROs census project have been scanned bi-tonal. A further advantage is that bi-tonal images do not often require compression as they usually have a small file size, e.g. a bi-tonal scan of an A3 page at 300dpi and saved as an uncompressed TIFF file should be no more than 300kbs. This image type maybe the best to use if scanning from microfilm.

7.1.2 16-bit greyscale. Greyscale images are better when the items to be scanned have faded text and the contrast is less good. While greyscale images are often considered more than adequate the consistent recommendation is that preservation or archival copies should be scanned as 24-bit colour.

7.1.3 24-bit colour is the best scanning choice in terms of quality but as the quality is improved the cost, scan/save time and file size all increase. The advantage is that once scanned the image can be manipulated and adjusted. Again however with this manipulation and adjustment there is a further factor of resource availability to be taken into account. For example if it is decided to scan a thousand images at a high spec, say 400dpi true colour, each image will need to be opened and changed to the required spec. The time it takes for either an individual or for a batch conversion program to open each of these files will be significant. The obvious advantage of the batch conversion is that it will require less human intervention.

Clearly given the nature of these differing image types a careful balance must be obtained which gives us the best quality image whilst maintaining a manageable file size.

7.2 Resolution

Resolution is measured in dots per inch (dpi). As with file types higher dpi results in increased image quality but it also means that with this increase in quality is an increase in file size. It is important to find a balance. As discussed the *nof-digitise Technical Standards and Guidelines* are in favour of a 'Scan Once for All' policy which means that a high dpi for best possible image quality should be sought. It should be noted however that sometimes better quality scans are obtained using a lower dpi depending on the state of the originals. It is important that for the Moving Here project that minimum standards are agreed to ensure both quality and a degree of commonality throughout the online service.

7.3 File type

The *nof-digitise Technical Standards and Guidelines* advises us that open standard formats should be used when creating digital records in order to maximise access to resources. However proprietary formats, such as Adobe's PDF (Portable Document Format), can be adopted provided that migration to a non-proprietary format are considered. It should be noted that the file type used to scan to is often not the one used for web delivery.

There are three main open standard image file types that the partners should be aware of, Tiff, Gifs & JPGs and each one has their own advantages and disadvantages.

7.3.1 TIFFs are the most widely accepted format of archival image creation and retention as master copy. TIFF files can be read by almost all platforms and it allows you to gather as much information as possible from the original and then saves this data. If the images are likely to be modified at some point in the future, eg for web delivery, then the images should be scanned as TIFFs. The one disadvantage of them is the large file size.

7.3.2 JPEG files are the strongest format for web viewing as they can be greatly compressed.

7.3.3 GIF files are limited to 256 colours without requiring much storage space and are viewed by many as a happy medium between the TIFFs and JPGs.

APPENDIX 8

www.ota.ahds.ac.uk/documents/creating/chap3.html

8. Digitisation Hardware

Whether or not the resources to purchase a scanner or a digital camera exist it is worth being aware of what is available and what equipment is appropriate for digitising specific content.

There are quite a few methods of image capture that are used within the humanities community. The equipment ranges from scanners (flatbed, sheetfed, drum, slide, microfilm) to high-end digital cameras. In terms of standards within the digitizing community, the results are less than satisfactory. Projects tend to choose the most available option, or the one that is affordable on limited grant funding. However, two of the most common and accessible image capture solutions are flatbed scanners and high-resolution digital cameras.

8.1 Flatbed scanners

Flatbed scanners have become the most commonplace method for capturing images or text. Their name comes from the fact that the scanner is literally a flat glass bed, quite similar to a copy machine, on which the image is placed face down and covered. The scanner then passes light-sensitive sensors over the illuminated page, breaking it into groups of pixel-sized boxes. It then represents each box with a zero or a one, depending on whether the pixel is filled or empty. The importance of this becomes more apparent with the discussion of image type below.

As a result of their lowering costs and widespread availability, the use of quality flatbeds ranges from the professional digital archiving projects to the living rooms of the home computer consumer. One benefit of this increased use and availability is that flatbed scanning technology is evolving continually. This has pushed the purchasing standards away from price and towards quality. In an attempt to promote the more expensive product, the marketplace tends to hype resolution and bit-depth, two aspects of scanning that are important to a project but are not the only concerns when purchasing hardware. While it is not necessarily the case that you need to purchase the most expensive scanner to get the best quality digital image, it is unlikely that the entry-level flatbeds (usually under 100 pounds) will provide the image quality that you need. However, while it used to be the case that to truly digitise well you needed to purchase the more high-end scanner, at a price prohibitive to most projects, the advancing digitising needs of users have pushed hardware developers to create mid-level scanners that reach the quality of the higher range.

As a consumer, you need to possess a holistic view of the scanner's capabilities. Not only should the scanner provide you with the ability to create archival quality images but it should also make the digitisation process easier. Many low-cost scanners do not have high-grade lenses, optics, or light sources, thereby creating images that are of a very poor quality. The creation of superior calibre images relates to the following hardware requirements:

- *the quality of the lens, mirrors, and other optics hardware;*
- *the mechanical stability of the optical system;*

- *the focal range and stability of the optical system;*
- *the quality of the scanning software and many other hardware and software features.*

Also, many of the better quality scanners contain tools that allow you to automate some of the procedures. This is extremely useful with such things as colour and contrast where, with the human eye, it is difficult to achieve the exact specification necessary for a high-quality image. Scanning hardware has the ability to provide this discernment for the user, so these intelligent automated features are a necessity to decrease task time.

8.2 Digital cameras

One of the disadvantages of a flatbed scanner is that to capture the entire image the document must lie completely flat on the scanning bed. With books this poses a problem because the only way to accomplish this is to bend the spine to the breaking point. It becomes even worse when dealing with texts with very fragile pages, as the inversion and pressure can cause the pages to flake away or rip. A solution to this problem, one taken up by many digital archives and special collections departments, is to digitise with a stand-alone digital camera.
Digital cameras are by far the most dependable means of capturing quality digital images. As Robinson explains,

> *They can digitize direct from the original, unlike the film-based methods of microfilm scanning or Photo CD. They can work with objects of any size or shape, under many different lights, unlike flatbed scanners. They can make images of very high resolution, unlike video cameras (Robinson 1993, 39).*

These benefits are most clearly seen in the digitisation of manuscripts and early printed books — objects that are difficult to capture on a flatbed because of their fragile composition. The ability to digitise with variant lighting is a significant benefit as it will not damage the make-up of the work, a precaution which cannot be guaranteed with flatbed scanners. The high resolution and heightened image quality allows for a level of detail you would expect only in the original.

APPENDIX 9

9. Digitisation Options

9.1 The digitisation options will, for the majority of the partners, boil down to two choices largely depending on the quantity of the content and resources available.

a) To scan the items in house (this may include the purchasing of a scanner/digital camera)
b) To outsource the scanning to a bureau

The likely outcome for many partners may be a combination of both.

This 'choice' may already have been decided by the selection of content as many partners will either have a policy restricting any original items being scanned off site or if there is no such policy they will almost certainly be reluctant to do so. It is in an instance such as this that it is important to be aware if any of the content selected is available in another format, like microfilm. Many of the partners will allow microfilm off site for digitisation though of course this change of format will in turn have an impact on the scanning specification.

Clearly large collections on microfilm are ideal for outsourcing as they are all similar and it keeps the scanning requirements simple.

Alternatively as it may not be possible to take the originals to the scanner it may be possible to bring the scanner to the originals. Some scanning bureau and other specialised digitisation services operate an option to lease equipment, and if necessary, staff to operate it, though often training can be provided as part of a package. If equipment is leased considerations such as cost, minimum lease period, available space etc… should be taken into account.

The following are scanning bureau which partners may like to contact for tender.

9.2 Digitisation Services:

Heritage Image Ltd - http://www.heritage-images.co.uk/
Higher Education Digitisation Service - http://heds.herts.ac.uk/
iBase - Contact: Peter Bridge peter.bridge@ibase.com or Davelle Watson davelle.Watson@ibase.comhttp://www.ibase.com/
Oce Digitisation Services - http://www.oce.co.uk/
Redcliffe Professional Colour Laboratories - http://www.redcliffe.co.uk/scan.htm
Trans Media - Contact: Mr G Pye (MD) enquiries@tmt.co.uk
Cendris - Contact: Andrew J Lowe Andrew.lowe@lason.co.uk
Microformat Systems - Contact: Dr J Baars (MD) 100315.315@compuserve.com
Max Communications - Contact: David Cordery davidc@maxcommunications.co.uk

If a partner decides to outsource some or all of their scanning they will need to draw up a scanning framework agreement upon which digitisation services will be invited to tender. The Central Team can provide a 'general' Scanning Framework Agreement for outsourcing digitisation should one be required. A covering letter should also accompany any scanning framework agreement providing a context for the scan requirements and specifying a production timescale.

Clearly the cost of outsourcing digitisation will vary according to the document makeup but below is a very rough idea of costings for the following scanning specs;

Colour 300dpi - £2.00/image
Greyscale 300dpi - £0.45/image
Bi-tonal 300dpi - £0.15/image

9.3 A further factor to take into account if digitisation is to be outsourced is quality control of the images as they are returned. This will ensure that the image quality is being maintained over a period of time and a variety of images. (See Appendix 12, *Quality Assurance,* for further details).

9.4 Many of the larger partners will have facilities to accomplish the digitisation of their content on site. Depending on resource availability it may even be possible for some of the smaller partners to 'outsource' their content in part or in total to one of the local larger partners. This would hopefully allay any fears concerning originals being digitised off site as they would be in the hands of experts. It would also assure consistent scanning and it should prove cost effective.

Similarly some of the partners will have expertise in digitising certain formats such as video and audio. Knowledge of which would be of great benefit to those partners who have such items in their content but who have not undertaken such a project before. The sharing of such knowledge between partners will be crucial.

Once content has been identified a production schedule will be agreed between each of the partners in turn and the central team with the specific aim of getting 40% of the consortium wide content digitised by the November launch.

APPENDIX 10

Copyright clearance form: *Moving Here* Project

To be attached to any material sent from the partners to the PRO in connection with the Moving Here Project

From: [Name of organisation]

To: *Moving Here* Project (Public Record Office)

Date:

Number and details of material submitted (see Schedule attached)

We can confirm that the status of material submitted is as follows:

- It is our own copyright and we grant permission for the material to be reproduced on the *Moving Here* website and by NOF and its Licensees in accordance with NOF's IPR Conditions[1], *or*

- We have obtained written permission from the copyright owner to reproduce the material on the *Moving Here* website and by NOF and its Licensees in accordance with NOF's IPR Conditions, *or*

- It is Crown Copyright and we have followed HMSO's guidelines, *or*

- It is out of copyright

Please note that for copyright owned by third parties the minimum clearance laid down by NOF's IPR Conditions is:

- Non-exclusive, royalty free, worldwide licence of the intellectual property rights (copyright) in the Material. For the purposes of private, non-commercial use. For the lifetime of the grant (grant payment period, plus three years), or 1 December 2007 whichever is earlier. For use on *Moving Here*, NOF and its Licensees, The People's Network, the National Grid for Learning or other similar public sector sites

We can confirm that the material submitted contains nothing that is libellous, defamatory or indecent

- This statement was included in NOF's IPR Conditions

Signed (on behalf of the organisation)……

- [1]New Opportunities Fund: Conditions of Grant covering Intellectual Property Rights, September 2001 version, and accompanying letters (August 14 and 29, September 21)

APPENDIX 11

11 Quality Assurance

There are two areas where a quality control check must be instituted to ensure;

a. Image Quality
b. Metadata Quality

11.1 Image Quality: This must be checked as thoroughly as possible and can be done as part of the scanning itself if the digitisation is taking place in house. If the digitisation is outsourced it is important that a quality benchmark is agreed between the partner and the digitisation service before digitisation commences. The scanned images from the digitisation service must be compared to the agreed benchmark upon delivery to ensure a consistent quality standard.

A further check between the images at the different resolutions is also a good idea just to make certain that the lower resolution does not affect the image quality too greatly.

Ideally 100% of the images should be checked for image quality, legibility, skew. However partners with large numbers of images to check can carry out the quality assurance based on a random sampling. A minimum of 10% of the images should be checked.

Methods for carrying out this QA will vary from partner to partner depending on where the digitisation has taken place. It is important to develop a simple QA system and record which images are checked, by whom, whether they pass or fail, reason for failure etc. so that an audit trail is established and accurate information about failed images can be returned to the scanner operator if necessary.

11.2 Metadata: Similarly the descriptive (catalogue) metadata should be thoroughly checked after it has been input into the XML template or word file as this will be the key to the digital images for the users and inaccuracies may result in failed searches.

A further QA will be carried out by the central team to ensure consistency throughout all of the Moving Here content. Any failures at this stage may result in images or catalogue information being returned to the partner from which it came.

APPENDIX 12

12. Preservation & Storage

12.1 The *nof-digitise Technical Standards and Guidelines* tell us that *'Preservation issues must be considered as an integral part of the digital creation process… Those responsible for the project must weigh-up the risks of exposing original material to any digitisation process, especially where the items are unique, valuable and fragile… Projects must consider the value in creating a fully documented high-quality 'digital master' from which all other versions eg web deliverables can be derived'.*

12.2 As outlined in the Moving Here business plan the proposed resource will consist of three elements - 'stories', digitised resources and a delivery mechanism.

12.3 The stories will essentially be HTML files with the full range of HTML elements - hypertext links, JPEG files, etc. These will be unique and will be preserved at the PRO by the central team.

12.4 The digitised resources will be created in the form of digital images, sound (audio), and film.

The digital images will exist in two forms - original archive masters and the derivative web deliverables. The original images will be in the form of uncompressed TIFF files. These will be retained by the museum, library or archive, which holds the original documents or objects. The individual partners will be responsible for ensuring the preservation of the archive masters for the lifetime of Moving Here. It should be noted that digital media is not recognised as an archival media as it can corrupt and degrade in a relatively short timeframe. To give the images the best possible chance of survival a system of refresh and migration must be adopted. This will entail storing the archive masters available on two different media eg CDs and DLT tapes and periodically refreshing them. The central team will be responsible for ensuring that the originating institutions have facilities to preserve the original images. As with the digitisation itself some of the larger partners may be able to offer some of the smaller ones advice and even space to hold their digital archive masters for the life of Moving Here. If necessary, the PRO will be willing to preserve these as part of its digital preservation system. Beyond the lifespan of Moving Here the images can be used and archived according to each partners requirements.

Similarly audio and video files will be retained by the museum, library or archive, which holds the original recordings (if applicable)

The web deliverable derivatives will be backed up and preserved for the Moving Here lifespan by the central team at the PRO.

12.5 The physical space required to store the archive masters on their respective media will vary according to the quantity of images and the scanning specification. A

high quality uncompressed TIFF image scanned in true colour at 300dpi will result in a file size of about 50mbs plus which means that you will not get many large images to CD. Although this matter may be inconsequential to some partners there are others were space will be at a premium.

APPENDIX 13

Annex 1: Digitisation Glossary (ibase.com)

Bit	The smallest element of computer data, derived from Binary Digit.
Bitmap	A generic name for an image file stored as a two dimensional array (map) of bits.
BMP	A Microsoft Windows proprietary image file format not widely used in a commercial context.
Byte	Collective term for 8 bits. 1 byte is the minimum amount of data used to define, for example an alphabetic or numeric character. In terms of digitisation each pixel scanned requires one byte of data for each colour channel. A full colour image therefore requires three bytes of data per pixel.
CMYK	From Cyan, Magenta, Yellow, Black - an image file format based upon the value of each of these colours. It is generally used to store images that are targeted at print, as opposed to monitors.
Colour depth	The extent to which the extremities of a colour are sub-divided into intermediate shades. For example, an 8 bit colour depth provides 256 shades between the extremes of the grey scale - black and white.
Dpi	Dots per inch - the number of individual elements per inch scanned or printed. For example, an image 450 dots high and 600 wide printed by at 150 dpi will produce an image 3" high and 4" wide.
File Type	There are several commonly used ways of creating a digital representation of an image, each is said to be of a particular format. Examples are BMP, JPEG, GIF and TIF.
File size	The size of a computer file is determined by the amount of information it contains. The more information stored about an image, the bigger the file will be.
Gigabyte	One thousand Megabytes. (Strictly 1,024 Mega bytes)
GIF	Graphical Interchange Format. An image file format frequently used on web sites.
Greyscale	A digitisation process which interprets all colour information as shades of grey between the extremes of black and white.
JPEG (or	Joint Photographic Expert Group. A compressed (reduced in size) image

JPG)	file format.
Kilobyte	One thousand bytes. (Strictly 1,024 bytes)
Megabyte	One million bytes. (Strictly 1,048,576 bytes)
Pixel	Equivalent to the dots in dpi. Pixel is the preferred term when referring to screen displays, whereas dot is most often used with reference to printing. A computer screen might for example be set to a resolution of 800 by 600 pixels, or 1,024 by 768 pixels.
Ppi	Pixels per inch - the number of individual elements per inch of a computer screen. For example, a screen 10" high and 13.3" wide set to 1,024 by 768 pixels resolution will have a ppi of 76.8. Thus an image 300 pixels high and 200 wide displayed on it will appear 3.9" high and 2.6" wide.
Resolution	The number of dots or pixels per inch scanned, printed or displayed on screen. The higher the resolution, the more detail available, the bigger the file.
RGB	Red, Green, Blue - an image file format based upon the value of each of these colours. It it generally used to store images that are targeted at monitors, as opposed to print.
Terabyte	One million Gigabytes. (Strictly 1,048,576 Gigabytes)
TIFF (or TIF)	Tagged Image File Format - the most commonly used image file format, capable of high compression with minimal loss of visual quality - lossless.
True colour	A somewhat vague term usually applied to a colour depth of 24 bits, providing the possibility of representing 16 million different colours (Strictly 16,777,216 colours).