Grant Agreement 297384

Partage Plus

Report on standards to be used in the Partage Plus project

Deliverable number	*D1.1*
Dissemination level	*PU*
Delivery date	*May 2012*
Status	*Final*
Author(s)	*Chris Vastenhoud (KMKG) and Gordon McKenna (CT)*

Revision History

Revision	Date	Author	Organisation	Description
V0.1	2012-04	Chris Vastenhoud	KMKG	
V1.0	2012-05	Chris Vastenhoud; Gordon McKenna	KMKG	With revisions by CT

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.
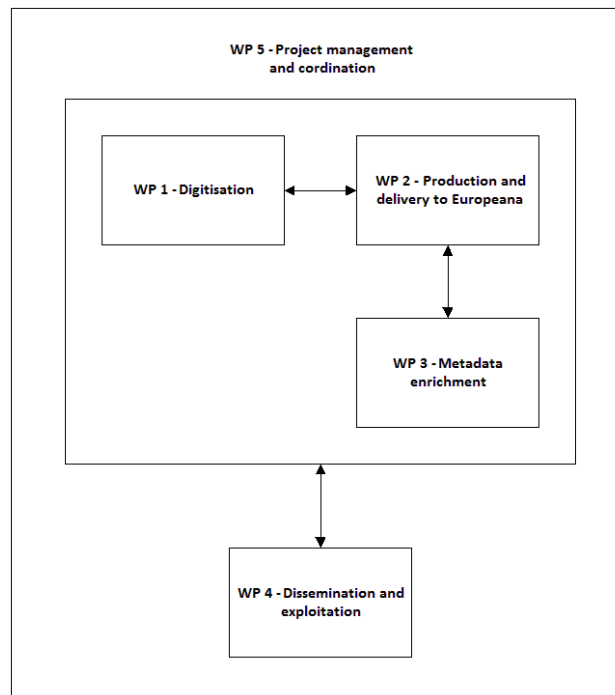
Contents

1   Introduction

This deliverable describes the digitisation standards to be applied in the Partage Plus project, which brings together Art Nouveau collections from across Europe. The project will digitise c75 thousand objects and other items including c2 thousand 3D models. The digitised collections will be submitted to the Europeana portal with a link to the partner's website.

It gives:

- An overview of the digitisation standards to be applied and file types to be used in the project;
- Directions in order to obtain and maintain the quality of the digitisation activities;
- Directions to ensure longevity and sustainability of the digital files.

The work package is closely related with WP 2 – *Production and delivery to Europeana*, and WP3 – *Metadata Enrichment*:

Partage Plus work package structure

2   Digitisation workflow

The generic digitisation workflow model which will be used in *Partage Plus* (WP 1) is based on that developed, in an academic context, by JISC Digital Media in the UK[1]. It has been used in many successful projects.

The workflow has four phases (the first is additional for *Partage Plus*):

1. ***Preparation*** – Issues to be considered and built into the procedures and timings (in the first three months of the project).

   - Choice of objects to be digitised;

   - Deciding on the uses of the content created;

   - Liaison with curators;

   - Movement of objects in and out of the digitisation facility;

   - Need for conservation work;

   - Provision of quality control for content produced ('sign off');

   - Availability of IT support for equipment;

   - Progress reporting for organisational and *Partage Plus* project management;

   - Creation of an organisational digitisation plan based on the workflow model.

2. ***Capture*** – The creation of an archive

   - Content (i.e. digital images, audio and video) should be captured at an agreed quality sufficient for all planned uses (see standards suggested below);

   - The original capture format will depend on the capture device. For example: a scanner, RGB TIFF; a digital camera, TIFF, Raw or JPEG. JPEG should be avoided;

   - Raw digital content to be stored in a backed-up archive (the *Master Archive*). This allows subsequent processes to be repeated or changed if necessary.

3. ***Optimisation*** – It will normally be necessary to undertake at least some skilled 'optimisation' work on the captured 'raw' content. The resultant files should be stored in an archive (the *Master Optimised Archive*).

   For images, the bulk of the content being provided by *Partage Plus*, this might include

   - Image size (cropping, orientation, perspective correction);

   - Image tonality (white balance, levels and curves);

   - Cleaning up (removing marks, spots, correcting other image problems);

   - Addition or editing of metadata to internal IPTC fields[2] within image header (including copyright tag);

   - Conversion from Raw or 'high-bit' TIFF format to 24bit RGB Baseline uncompressed TIFF file format;

   - Transformation to Adobe RGB 1998 colour space and profiled as such;

   - Limited sharpening (recorded in embedded technical metadata).

   Similar optimisation can be applied to other types of content.

---

[1] See: **http://www.jiscdigitalmedia.ac.uk/stillimages/advice/generic-image-digitisation-workflow**
[2] See: **http://www.iptc.org/cms/site/index.html?channel=CH0099**

4. **Surrogate creation** – Surrogates can now be created from the *Master Optimised Archive*. In general surrogates are created for various, service environment, purposes.

- For presentation on monitors (organisational websites and multimedia presentations). This purpose covers the provision of access to content via *Europeana*, including previews (thumbnails);

- For commercial printing (may be the identical to the *Master Optimised Archive* files);

- For in-house printing (may be the identical to the *Master Optimised Archive* files);

- For creation of smaller surrogates from online master files (usually smaller than *Master Optimised Archive* files).

3 Observations regarding digitisation

In order to ensure the sustainability, longevity and to maximise the repurposing of the digital information created during the project, Partage Plus will not limit the quality of digitisation to the specifications Europeana prescribes for access to the portal information.

For a sustainable service, and the best retrieval of information, it is essential to aim for the highest possible quality in capturing and storing records. There are a number of reasons to support high quality digitisation:

- Items should be digitised at the highest possible quality in order to aid use in applications;

- Advances in technology mean that data and image capture and storage, will be improving during the project and thus producing higher quality output. This is also true of web and display technologies;

- Repurposing – The higher the initial quality of the source, the more versatile it can be used for future applications;

- Although at a slower pace, Moore's Law[3] still governs. Hardware and storage capacity continue to double every 3 years at the same or lower cost, and therefore the cost of storage capacity should not be a consideration;

- Today, in general, capture technology uses higher resolutions than display technology, and therefore Europeana standards are met 'automatically'.

3.1 Back up and preservation demands

The digital products of the Partage Plus project must remain available and accessible for at least 10 years. The partners therefore must ensure that their front ends (website link) remain accessible and that master files and metadata are stored and preserved safely.

The project itself does not prescribe specific strategies for long term preservation, but will ensure that partners maintain a digital environment that is able to meet this demand.

All partners will produce a small report by August 2012 on their long term digital preservation strategy. KMKG and CT will advise where necessary.

3.2 Quality control

KMKG will ask to partners in May 2012, November 2012, May 2013, November 2013, and February 2014 to provide samples of the different file types they intend to use or are using. KMKG will assess whether the files meet the technical and quality standards of the files required and will offer advice where necessary.

3.3 Intellectual Property Rights (IPR)

All content that will be linked by to Europeana must be free of rights and claims for the purpose of display in Europeana, as previews, as under the stipulations of the Data Exchange Agreement (DEA)[4].

---

[3] **http://en.wikipedia.org/wiki/Moore's_law**
[4] **http://pro.europeana.eu/web/guest/data-exchange-agreement**

3.4 Display on partner websites

All partners delivering content to Europeana must provide digital files, metadata and a permanent link at object level to the respective objects on their website. This link ensures that Europeana visitors can navigate from the Europeana portal to the partner's website and explore the content in depth there. The specifications of this process will be addressed by WP 2.

4   File formats

For its activities in WP 1 – *Digitisation Partage Plus* will follow the most basic advice on technical standards when its partners carry out digitisation. This is to:

Use open standard formats when creating and delivering digital content.

Doing this will:

- Maximise access;
- Ensure that content is reusable. It can be created and changed by more than one piece of software;
- Avoid dependency on a single supplier with possible licensing restrictions.

Following this advice is easy for all of the common types of digital content. Therefore there must be a very good reason not to follow it. Even where a proprietary standard has been used it is good practice to have a plan to migrate to an open standard when one becomes available.

Detailed advice on which technical standards to use is based on the environment the material is being used in. In broad terms there are three '**use environments**':

Master
This is where the digital surrogate is created from an analogue original. Sometimes this is described as creating an archival master. This can be done by a number of techniques including: photography; scanning; sampling; OCR (optical character recognition), 3-D modelling, and so on. Born digital content will be archival by default.

The activity usually takes place at the collection holding organisation and in their systems.

Key concepts for digital content in this environment are:

- Maximum quality (e.g. no compression);
- Preservation;
- Open source.

Service
This is where users of the material are given **meaningful access** to a **single** piece of digital content. Delivery usually includes relevant metadata describing the significance of the material being accessed.

Keywords for digital content in this environment are:

- Usable quality (for service being offered);
- Reasonable speed of delivery;
- Rights protection (either by size restriction or other means).

Discovery
This is where users are given access to a set of thumbnails of digital content. The aim here is to review the results and move on to more detailed information. Delivery is usually part of the result set of a search and includes discovery metadata.

Keywords for digital content in this environment are:

- Maximum speed of delivery;
- Minimal size;
- Recognisability.

The service and discovery often appear together on the part of an organisation's website where they display their collections online. Portals do the same when they aggregate and deliver digital content.

Portals can also only implement the discovery environment. They point to the service environment on the collections holder's website.

Europeana is unusual in that it really only implements the discovery environment but uses a link, if available, to **appear** to give access to content in the service environment.

For full details on the standards recommended see the *Appendix – Standards Landscape* at the end.

The sections below give the **minimum recommendations** (file formats and quality) for the various media types, being created in *Partage Plus*, in the three use environments:

4.1 Image

4.1.1 Camera

Parameter	Use Environment		
	Master	*Service*	*Discovery*
File Format	RAW; DNG; TIFF	JPEG; PNG	JPEG; PNG
Colour Quality	8 bit greyscale 24 bit colour	8 bit greyscale 24 bit colour	8 bit greyscale 24 bit colour
Resolution	[maximum for equipment]	150-200	72
Maximum dimension (pixels)	[not applicable]	600	100-200

4.1.2 Photography guidelines

Here we do not prescribe detailed procedures for photographic recording. Partners can apply their own methods in order to maintain a level of continuity in their practices within their organisation. There are, however, a few subjects which might be of help, described below:

Backgrounds

Europeana and Partage Plus do not prescribe rules for backgrounds. However, it is advisable to consider the use of neutral grey as this has least impact on the entire image and in general ensures least conflict with the objects.

Post processing

In order to produce good quality images, post processing will be necessary. Post processing should be limited to colour correction and sharpening. Deformation, special effects adding or deleting parts of the object must be avoided. Supports, used to position an object can be removed digitally if this is preferred.

Lighting

Usually the objects are photographed in studio conditions. Partners are aware of light reflections in reflective materials and can position and adapt studio lights accordingly.

Position

Objects should be photographed from different positions, to produce different views of the object. Partners must be aware that Europeana presents only one image of the object, and this view should give an overall impression of the object. Detailed views or different positions can be presented on the partner website.

Ensembles

Often Art Nouveau objects form ensembles or are part of a group, for example ceramics, desktop sets or furniture. In such case the group should be photographed as a representative ensemble and separate items should then be photographed individually. Similar version or copies (like plates in a ceramic ensemble) are only photographed once. The metadata will link the individual items and the ensembles.

Colour control

Many partners with a permanent photographic studio will have their own procedures based on equipment and preferences. The application of colour spaces and normalisation are beyond Europeana's scope and control. For colour control a colour bar can be photographed with the object. On adjustment of the colours during post processing the bar should be cropped out of the picture.

4.1.3 *Scanning*

Parameter	Use Environment		
	Master	*Service*	*Discovery*
File Format	TIFF	JPEG; PNG	JPEG; PNG
Colour Quality	8 bit greyscale 24 bit colour	8 bit greyscale 24 bit colour	8 bit greyscale 24 bit colour
Resolution (dpi)	600 (photographs) 2400 (slides)	150-200	72
Maximum dimension (pixels)	[not applicable]	600	100-200

4.2 Audio

Parameter	Master Use Environment
File Format	Uncompressed [preferred]: WAV; AIFF Compressed [alternative]: MP3; WMA; RealAudio; AU
Creation quality	24-bit stereo and 48/96 KHz sample rate

Parameter	Service Use Environment
File Format	Compressed [preferred]: MP3; RealAudio; WMA Uncompressed [alternative]: WAV; AIFF; AU
Delivery quality	256 Kbps (near CD quality); 160 Kbps (good quality)

	Discovery Use Environment
Note	The discovery of audio content will be enhanced by a relevant image preview, e.g. a photograph illustrating the subject of the audio file.

4.3 Video

Parameter	Master Use Environment
File Format	Uncompressed [preferred]: RAW; AVI Compressed [alternative]: MPEG (MPEG-1, MPEG-2 or MPEG-4); WMV; ASF; Quicktime.
Creation quality	Frame size of 720x576 pixels; Frame size of (HDTV) 1280x720 pixels; Frame rate of 25 frames per second; 24-bit colour; PAL colour encoding

Parameter	Service Use Environment
File Format – for downloading	MPEG-1; AVI; WMV; Quicktime
File Format – for streaming	ASF; WMV; Quicktime

	Discovery Use Environment
Note	The discovery of video content will be enhanced by a relevant preview image (still) taken from the video file.

5 Conclusions

These digitisation guidelines and specifications represent the digitisation strategy for the Partage Plus project.

The approach described above may seem like day to day common practice, however the challenge is to maintain a preservation policy and individual digital repositories for the content produced over a considerable long time span, according to Europeana standards.

Appendix – Standards Landscape

Taking the idea of the use of standards to heart, we describe each standard in a Dublin Core (DC) derived format. 9 out of the 15 DC elements are used in the descriptions.

These elements are:

Title	The name (or names) under which the standard is known. In most cases both the abbreviated and the full name is listed.
Creator	The name of the organisation or individual who originally created the standard.
Publisher	The name of the organisation that makes the standard publicly available.
Date	The date on which the standard was *originally* published.
Identifier	A number or other identifier under which standard is published or a URL which points to the definition of the standard.
Rights	Whether rights restrictions, e.g. patents, apply to the standard.
Description	A textual description explaining the standard and its usage.
Subject	Keywords that identify the nature of the standard.
Relation	Other standards that this standard relates to, and associated websites.

The descriptions are aimed at a general reader. More technical details for the majority of standards discussed can be found in various places on the Web.

The purpose of this section is to allow the reader to have an easy reference to the range of relevant standards in one place.

Image

JPG

Title	JPG
	JPEG
Creator	Joint Photographic Expert Group
	Independent JPEG Group
Publisher	International Organization for Standardization (ISO)
Date	1990
Identifier	ISO/IEC 10918-1:1994
Rights	[Open Standard]
Description	JPEG is a still image compression algorithm based on the fact that the human eye cannot detect subtle differences in colour or contrast. JPEG is a lossy algorithm: the higher the compression factor the more information gets lost. An image that has been compressed using the JPEG algorithm cannot be completely reconstructed. The file format for JPEG compressed images is called JFIF. This file format is what people generally mean when they refer to "JPEG".
Subject	image format
	raster graphics
Relation	**http://www.iso.org** (ISO website)

PNG

Title	PNG
	Portable Network Graphics
Creator	World Wide Web Consortium (W3C)
Publisher	International Organization for Standardization (ISO)
Date	2003
Identifier	ISO/IEC 15948:2003 (E)
Rights	[Open Standard]
Description	PNG is an extensible format for the lossless, portable, well-compressed storage of raster images. PNG provides a patent-free replacement for GIF and can also replace many common uses of TIFF. Indexed-colour, grayscale, and Truecolor images are supported, plus an optional alpha channel for transparency. Sample depths range from 1 to 32 bits.
Subject	image format
	raster graphics
Relation	**http://www.w3.org/TR/PNG/**
	http://www.iso.org (ISO website)

TIFF

Title	TIFF
	Tagged Image File Format
Creator	Aldus Corp
Publisher	Adobe Systems Inc.
Date	1992
Identifier	**http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf** [Version 6.0]
Rights	[Open Standard]
Description	TIFF provides a general purpose data format and is compatible with a wide range of scanners and image-processing applications. It is device independent and is used in most operating environments. This non-proprietary industry standard for data communication has been implemented by most scanner manufacturers and desktop publishing applications.
Subject	image format
	raster graphics

Camera capture 'standards'

There are two 'standards' for master files when cameras are used. These are strictly not standards in the usual sense. However they do offer the storage of the maximum amount of information:

- *RAW* – This is the file type is most flexible as the camera only records and stores the original image recorded. Post processing is always necessary. A disadvantage is that it is proprietary for each manufacturer.

- *Digital Negative* (DNG) – a format which includes the RAW files, and which has become a more widely accepted format. DNG is proprietary to Adobe, but the company has made the format open, and is trying to have it become a standard.

Audio

AIFF

Title	AIFF
	Audio Interchange File Format
Creator	Apple Computer Incorporated
	Electronic Arts
Publisher	Apple Computer Incorporated
Date	1988-1989
Identifier	**http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-1.3.pdf** [Version 1.3]
Rights	Apple Computer Incorporated [?]
Description	A non-compressed audio format most widely found on Apple Macintosh computers. Lossless, it is commonly used with professional-level audio and video software and systems.
Subject	sound format

MP3

Title	MP3
	MPEG Layer 3
	Coding of Moving Pictures and Associated Audio for Digital Storage Media
Creator	ITU-T (International Telecommunication Union Telecommunication Standardization Sector)
	Moving Pictures Expert Group (ISO/IEC JTC1/SC29 WG11)
Publisher	International Organization for Standardization (ISO)
Date	1993 onwards
Identifier	ISO/IEC 11172:1993, Part 3: Audio
	ISO/IEC 13818, Part 3: Audio
	ISO/IEC 14496, Part 3: Audio (Amendement 1: Audio extensions)
Rights	[Open Standard]

Description	An audio compression format common on the Internet. Part of the MPEG standards, it can take larger audio recordings and shrink them down to a fraction of their size while losing little if any fidelity of the sound.
Subject	sound format
Relation	MPEG-1; MPEG-2; MPEG-3
	http://www.iso.org (ISO website)

WAV

Title	WAV
	RIFF
	Resource Interchange Format
Creator	IBM
	Microsoft Corporation
Publisher	Microsoft Corporation
Date	1991
Identifier	**http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf**
Rights	[?]
Description	Contain sampled audio. The sound information itself it stored in a container using the Resource Interchange File Format. The RIFF file stores data in chunks, including metadata. A WAV file can contain sound clips with different sample rates, number of channels etc.
Subject	sound format
Relation	AVI

WMA

Title	WMA
	Windows Media Audio
Creator	Microsoft Corporation
Publisher	Microsoft Corporation
Date	[?]
Identifier	[not available]
Rights	Copyright Microsoft Corporation
Description	Microsoft's proprietary competition to MP3. Optimised to deliver audio over the Web, particularly streaming, and using Microsoft products. Has integrated rights management, the advantages of a smaller file size and therefore transfer rates over other formats, including MP3. In addition the sound quality is said to be better.
Subject	sound format
Relation	WMV

RealAudio

Title	RealAudio
Creator	RealNetworks Incorporated
Publisher	RealNetworks Incorporated
Date	[?]
Identifier	[not available]
Rights	Copyright Real Networks
Description	Format optimised for delivery of audio over the Web.
Subject	sound format
Relation	RealVideo
Relation	RealMedia

AU

Title	AU
Creator	Sun Microsystems Incorporated
Publisher	Sun Microsystems Incorporated
Date	[?]
Identifier	[not available]
Rights	Copyright Sun Microsystems Incorporated
Description	A sound format for Unix systems. It is the 'standard' audio file format for Java.
Subject	sound format

Video

AVI

Title	AVI
	Audio Video Interleave
Creator	Microsoft Corporation (for Intel)
Publisher	Microsoft Corporation
Date	[?]
Identifier	[not available]
Rights	Copyright Microsoft Corporation
Description	The earliest video format for PCs. The size of image that can be displayed is dependent on the hardware being used. As compression and decompression functionality is part of Microsoft's *Video for Windows* package, there is support for this format in a range of hardware and software configurations. File sizes are high and therefore this format ill-suited for delivery over the Internet.
Subject	video format

MOV (Quicktime)

Title	QuickTime
Creator	Apple Computer Incorporated
Publisher	Apple Computer Incorporated
Date	1991 onwards
Identifier	[not available]
Rights	Copyright Apple Computer Incorporated
Description	Apple's proprietary video (and virtual reality) format and system. Built into the Mac's operating system, can be viewed on a PC by using a free to download player. Some of its technology was used in the development of MPEG-4
Subject	video format
Relation	Quicktime VR
	MPEG-4
	http://www.apple.com (Apple website)

MP4

Title	MPEG-4
	Very-low bitrate audio-visual coding
Creator	Moving Pictures Expert Group (ISO/IEC JTC1/SC29 WG11)
Publisher	International Organization for Standardization (ISO)
Date	1999 (version 1)
	2001 (version 2)
Identifier	ISO/IEC 14496 (Parts 1 to 10)
Rights	[Open Standard]
Description	A high compression version of MPEG-2. Version 2 has data protection and IPR.
Subject	video format
Relation	MPEG-2
	MP3
	QuickTime [used in development]
	http://www.iso.org (ISO website)

MPG

Title	MPEG-1
	Coding of Moving Pictures and Associated Audio for Digital Storage Media
Creator	Moving Pictures Expert Group (ISO/IEC JTC1/SC29 WG11)
Publisher	International Organization for Standardization (ISO)
Date	1993
Identifier	ISO/IEC 11172:1993 (Parts 1 to 5)
Rights	[Open Standard]
Description	Designed to be the equivalent of a video recorder format in the digital world, and to make use of the early model CD-ROMs as a delivery method. Standard television quality images, with a compression ratio of 50 to 1.
Subject	video format
Relation	MP3
	http://www.iso.org (ISO website)

Title	MPEG-2
	Coding of Moving Pictures and Associated Audio for Digital Storage Media
Creator	ITU-T (International Telecommunication Union Telecommunication Standardization Sector)
	Moving Pictures Expert Group (ISO/IEC JTC1/SC29 WG11)
Publisher	International Organization for Standardization (ISO)
Date	2000
Identifier	ISO/IEC 13818:2000 (Parts 1 to 11)
Rights	[Open Standard]
Description	An improvement to MPEG-1, with encoding techniques to allow for higher quality video and audio, and delivery from DVDs. High definition television quality images, with a compression ratio of 150 to 1.
Subject	video format
Relation	MPEG-1
	MP3
	http://www.iso.org (ISO website)

SWF (Flash Movie)

Title	SWF
	Small Web Format
	Flash Movie
Creator	Macromedia (now Adobe)
Publisher	Adobe Systems Incorporated
Date	1996 onwards
Identifier	**http://www.adobe.com/devnet/swf/pdf/swf_file_format_spec_v10.pdf** [version 10]
Rights	Copyright Adobe Systems Incorporated
Description	Originally just for the delivery of animated vector graphics it is now used for interactive audio and video. It can be viewed using a standalone player or via a web browser plug in.
Subject	animated vector graphics
	Interactivity

WMV (Windows Media Video)

Title	WMV
	Windows Media Video
Creator	Microsoft Corporation
Publisher	Microsoft Corporation
Date	[?]
Identifier	[not available]
Rights	Copyright Microsoft Corporation
Description	Microsoft's proprietary competition to MPEG-4. Optimised to deliver video over the Web, particularly streaming, and using Microsoft products.
Subject	video format
Relation	WMA

ASF

Title	ASF
	Advanced Streaming Format
Creator	Microsoft Corporation
	RealNetworks Incorporated
Publisher	Microsoft Corporation
Date	2004
Identifier	**http://www.microsoft.com/windows/windowsmedia/forpros/format/asfspec.aspx** [Revision 01.20.03]

Rights	Copyright Microsoft Corporation. All rights reserved.
Description	A proprietary format, designed to deliver, compressed, streaming video / audio content over the Internet.
Subject	video format