Digital ArchiVing: guIdeline & aDvice 4

STANDARDS FOR FILEFORMATS

1. Preface

Importance

It is preferable to use as many standard file formats as possible to ensure the long-term readability of electronic records. Standard file formats owe their status to (official) initiatives for standardising or to their widespread use.

Advantages

Standard file formats have the advantage of being compatible with several softeware applications. Computerfiles do not need to be transformed immediately when one application isn't available anymore. Consequently, a number of conversions can be avoided. Many of the internal encoding of standard file formats are in the public domain so that in principle it is possible to develop new viewers at all times. It's also felt as a necessity to use file formats, with the status of standard, in the storaging of data to make the exchange of data possible.

Parameters and options

Most fileformats offer the possibility of adjusting several parameters.

Two parameters are important from an *archival* point of view: the ASC II storagemethod and storaging without compression.

ASC II: Files of which all the systemdata is stored through readable ASC II characters are more easily construed than through other applications.

No compression: Compression is not consistence with our goal to record as platform independent as possible. After all, to open a file one must use the matching decompressionkey.

Disadvantages

The consequence of the choice for the ASCII storagemethod and the storage without compression, results in a larger file size. Conversion to the standard file formats is sometimes accompanied with "freezing" or loss of functionality. It's for this reason that some files can no longer be used or worked on within their original application.

CONVERTING: WHEN?

The electronic records can be converted into another filingformat at different times.

1. Save electronic records as much as possible from the moment of their creation in the right standard folderformat. This is especially recommended for audio-visual records that you will be needing for years to come. This is not always possible for textual records.
   F.i. photos → TIFF; audio → WAV
2. Convert the files as soon as their immediate use has ended.
   WORD → PDF, XML
3. Convert the files while the original software is still running. Take care of the conversation before changing the software.
   F.i. autoCAD → DXF
4. The conversion has to take place at the latest before the transfer to the archival services.

2. Textual documents

FORMAT	VERSION	PARAMETERS and OPTIONS	USE
XML *eXtensible Markup Language*	1.0	– records of wich the layout is important: supply a XLS stylsheet – records of which the structure is important: compose a XML scheme (or DTD)	– e-mails – textdocuments – websites content – databases – metadata – spreadsheets
Unicode	3.0	- use clear-cut signs to separate fields - document the fields in the Unicodefile	- databases logfiles
PS *PostScript*	3.010	- do not use text- or image compressions - apply Level 3 - use ASCII encoding filters	- textdocuments with graphs and illustrations
PDF *Portable Document Format*	1.4	- use the ASCII storage method - do not use compression: switch the compression possibility off (monochrome, grey scale and full colour illustrations) - use tags: check the box option "Embed tags in PDF" - enclose only non-prevailing fonts (prevailing fonts: Arial, Courier, Helvetica, Tahoma, Times New Roman, Trebuchet, Verdana)	- textdocuments with graphs and illustrations -powerpoint-presentations - prints

TIFF *Tagged Image File Format*	6.0	– scans: controle whether everything is clearly readable. Much used resolution: 300dpi – store without compression – use only baseline and set TIFF-extensions: do not alter the TIFF- extension internally – RGB: image on screen/CYMK: prints – byte sequence (IBM or MacIntosh): free to choose – fill out filedata (metadata) in the file formats	- scanned documents
(X)HTML *(eXtensible) HyperText Markup Language*	HTML 4.01 (ISO-15445) of XHTML 1.0	– respect (X)HTML-syntaxrules – do not use rejected or non-standardised tags and attributes – register filerecords (metadata) explicitly in the HTML-header (see: *Digital ArchiVing: Guldeline & aDvice 5. Websites management for record-keeping.*)	– websites

3. Illustrations

3.1 Frames

FORMAT	VERSION	PARAMETERS and OPTIONS	USE
TIFF *Tagged Image File Format*	6.0	– same as textual documents	– full colour illustrations (photos) – scanned documents, – master copies
EPS *Encapsulated PostScript*	3.0	– apply ASCII-recordingmethode – indicate whether or not a thumbnail is being saved – variant DCS-1: CYMK-colour separation – variant DCS-2: supporting colours	– prints – 2 or 3 dimensional illustrations – text, graphs and illustration combination
JPEG *Joint Pictures Experts Group*	JPEG JPEG2000	– determine the parameters in function of the fileseize and the quality of the image	– photos on websites – workcopies
PNG *Portable Network Graphics*	1.0	– choose between 8-bit and 24-bit	– images on websites – images with *palette colours* (8 bit) or greyvalues (24 bit)
GIF	89	– limitations on the amount of colours	– cartoons (with

Graphics Interchange Format		= loss of information – **be carefull**: in principle, you need to have an Unisyslicentie to create GIF-images.	animations) – logos – black and white images – images with palette colours and large sections – images on websites

3.2 Vectorimages

FORMAT	VERSION	PARAMETERS and OPTIONS	USE
CGM *Computer Graphics Metafile*		– if possible, apply a profile – webCGM: exchange of dynamical CGM-images with hyperlinks	2 dimensional vector images
EPS *Encapsulated PostScript*	3.0	– same as with frames	2 or 3 dimensional images
DXF *Drawing eXchange Format*	16.1.01	– use the ASCII-recording method – save in the highest mode possible	2 en 3 dimensional CAD/CAM-derawings
SVG *Scalable Vector Graphics*	1.0		exchange of vectorial images over the web

4. Soundfiles

FORMAT	VERSION	PARAMETERS and OPTIONS	USE
PCM *Pulse Code Modulation*		– choose sample-rate and bitdepth in function of the quality of the analogue source – supply DAT-file in which the sample-rate and sample-resolution are explicitely recorded	digital master *copies*
WAV		– choose sample-rate en bitdepth in function of the quality of the analogue source – master copie: PCM-codec (compressionless) apply – exchange: MP3-codec (compression)	digital master copies or exchange
MP3	1.0	– determine the parameters frequency, bitrate and a number of channels in function of the quality and the fileseize that is wanted	exchange of soundfiles through networks. The use of streaming MP3 is possible.

5. Videofiles

FORMAT	VERSION	PARAMETERS and OPTIONS	USE
MPEG (*.MPG) *Moving Pictures Experts Group*	1.0	determine the number of frames (f.i. 30), the colourdepth (8, 16, 24) and the resolution (originall, 160 x 120, 320 x 240, 640 x 480) in function of the quality and the fileseize that is wanted	moving computerimages with/without sound
MPEG (*.m2v)	2.0	same as MPEG 1.0	digital television, multimedia

6. ? questions? Suggestions ?

Turn to DAVID for all your questions and suggestions:

e-mail: david@stad.antwerpen.be

telephonenumber: 00 32 (0)3 206 9415

More info: F. Boudrez, *Standards for digital records*, (http://www.antwerpen.be/david).