

DIGITAL PRESERVATION CHALLENGE 2008

2nd edition

Mac Kobus
Urbanstrasse 49
D-73728 Esslingen
Germany

mackobus@gmail.com

07/31/2008

Introduction

There is never enough time, so I decided to enter this challenge on 30th of July 2008 with less than 24 hours to go.

I will not deliver explicit solutions, but my ideas what can be done. Please excuse the incompleteness and the draft-like writing.

A handwritten signature in black ink, appearing to read 'Mac Kobus', written in a cursive style.

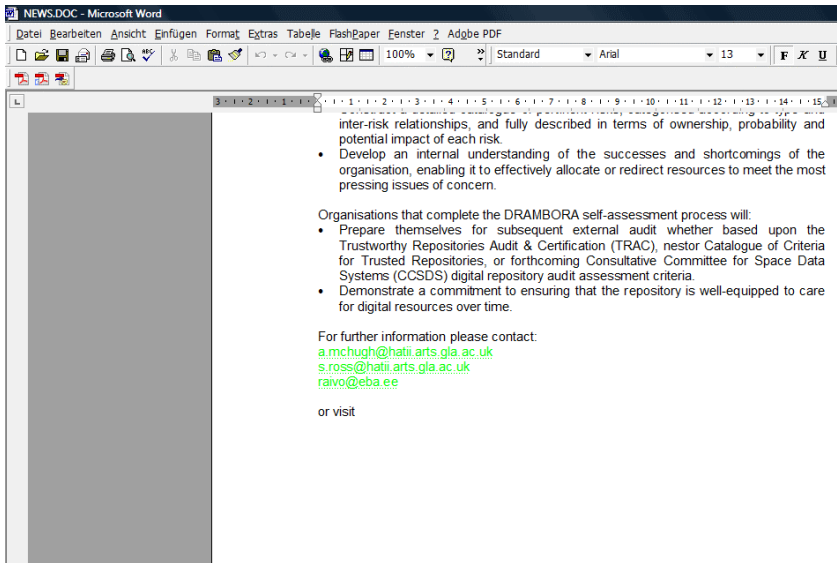
Mac Kobus (07/31/2008)

Task 1: Legacy Application File

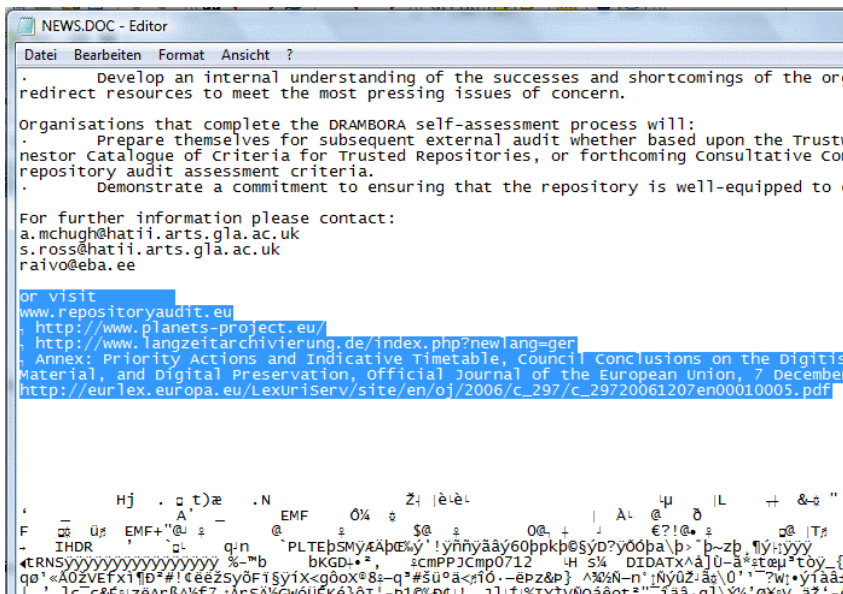
Fortunately the archivist still had the right hardware alive and running to read the old tapes. The 4 folders contained the following files:

Graphic: logo Pur.ltp

Text: NEWS.DOC should be a Microsoft Word-file. The file opened with my version of Word¹, but an abrupt ending with "... or visit" didn't seem right. Opening the file with another simple editor² brought up more text.



NEWS.DOC opened with MS Word 2000 – something is missing



NEWS.DOC opened with simple editor – there it is

¹ Microsoft Word 2000

² notepad.exe

It still has to be verified, that it is an MS Word-file. If it is, there has to be found out with which version of MS Word the file was processed.

Unknown 1: NEWS is very likely a Microsoft Office Document Image Writer-file³, which is an office-document converted to an image. It contains a spreadsheet.

Unknown 2 contains Lotus 1-2-3 spreadsheets.

LONDON: INCOME STATEMENT 1989: Goodwin's Sports Supply (London)
OSSOURCE: INCOME STATEMENT Year Ending 12/31/88: Bright Associates
QSTARGET: FINANCIAL RATIOS Year Ending 12/31/88: Bright Associates

First I viewed the files with the simple editor. There were hints, that these might be spreadsheets. They opened with no problems in MS Excel. I saved them as .csv and .wk1 files. With .csv some connections between the files are lost, but the sheets are then very easy accessible and readable without special software.

To preserve the spreadsheets over a long time, they should be converted to an open, well documented format (e.g. open office or CSV). A proprietary format, like the one EXCEL provides could also be an option, because it is very well established and wide-spread. Conversion and the software used should always be documented. I would recommend to choose more than one format.

If the spreadsheets should be easy accessible at all time, they should be compatible with the actual system used in the institution.

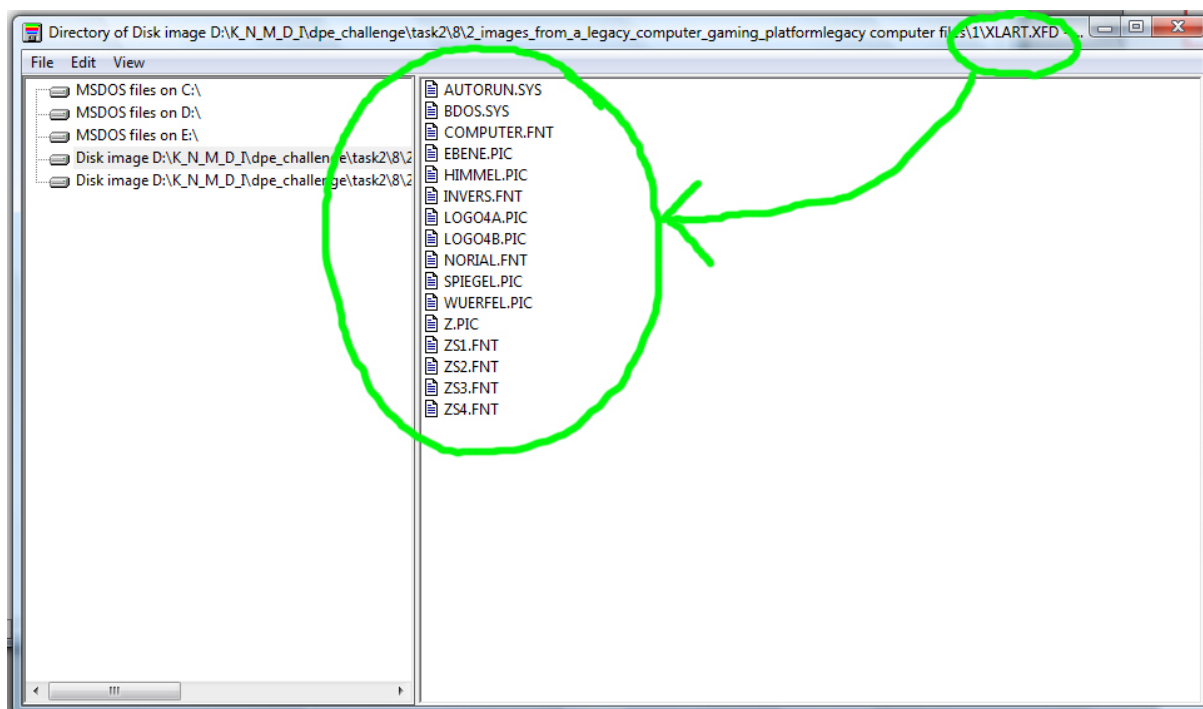
³ Version 4

Task 2: Images from a Legacy Computer Gaming Platform

The file with the extension .XFD⁴ seems to be an disc-image of an Atari-8-bit-system, probably an Atari 800XL. The disc contains a system-file BDOS.SYS (Basic Disk Operating System)⁵, several fonts (.FNT) and graphic elements (.PIC).

Looking at the .XFD file with a HEX-editor⁶ showed up some greeting notes and “cracked.by” and “proudly.presents”, so there might be some legal issues with that.

The directory of the disc-image can be browsed with GEMULATOR EXPLORER 2.03⁷.



Screenshot of GEMULATOR EXPLORER 2.03 showing the XLART.XFD-file

The easiest way to access and view the files will be to emulate an Atari 800XL-system on a new(er) computer, capable of running the emulation-software⁸.

Once the graphic files are extracted from the image-file, they should be converted to a well documented, lossless format like TIFF or JPG2000. The new files should be kept together with the original files and a documentation of the steps taken. It would be even better to include the used software, if possible.

⁴ “XFD: Xformer Disk Image. This image is similar to the ATR but lacks the ID and format header, it is essentially a large unmarked blob of data. Compatible only with the PC-Xformer emulator.”

<http://www.atarimax.com/ape/docs/DiskImageFAQ/> [07/31/2008]

“XFD-Image format invented by Emulators Inc, for their ST Xformer emulator. Identical to .ATR except without the 16 byte header.” <http://dotwhat.net/xfd/5300/> [07/31/2008]

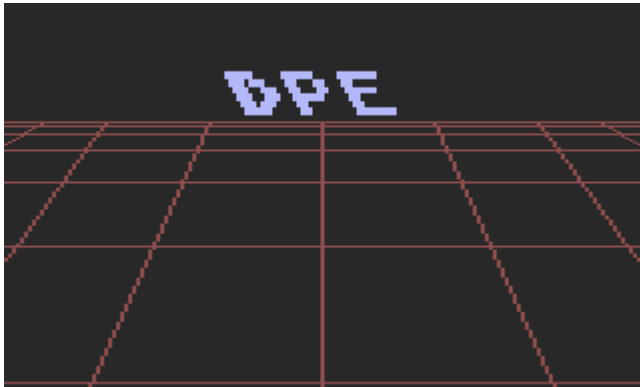
⁵ <http://de.wikipedia.org/wiki/CP/M> [07/31/2008]

⁶ Hex Workshop 5.1 <http://www.hexworkshop.com/> [07/31/2008]

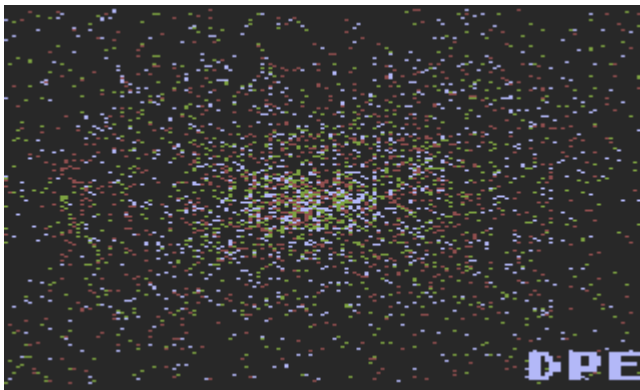
⁷ <http://www.emulators.com/explorer.htm> [07/31/2008]

⁸ e.g. A8E-emulator <http://www.zerstoerung.de/download.html> [07/31/2008]

The 2 images in folder 1 (.PIC files) can be viewed with the freeware picture/text converter PIXLator⁹. They look like this:

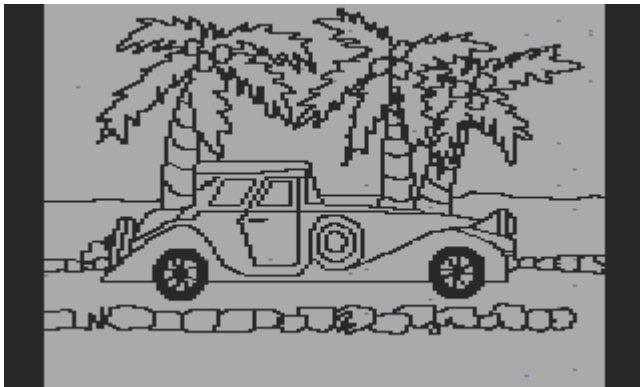


GRID.PIC



SKY.PIC

Assuming the files in folder 2 relate to the other ones, I simply added the file extension .PIC to show them.



CAR.PIC

⁹ PIXLator Version 2.1.0 http://www.mipro-online.de/waseo/ATARI_XL_XE/atari_xl_xe.html
[07/31/2008]



EINSTEIN.PIC

Task 3: Obsolete Database

- X -

Task 4: Electronic Art

What are the significant properties of the art-works to preserve?

If possible interviews with the artists and the viewers should be done to document the whole art experience.

The right system environment has to be established (or has to be emulated) to run the EXE-files properly. I guess it will be something like Windows95/98.

I have much more thoughts to this, but no time left.

Task 5: Web Archiving

It is only ½ hour until the deadline (4 p.m. GMT), so I have to hurry up. I harvested the Site with the HTTrack Website Copier quickly, without reading the guidelines and the proper adjustments. Sorry for that. And I was not able to establish a Linux-System to try other Web-Harvesting-Solutions like Heritrix.

There is the question: What should be captured? Is it the information or the appearance of the site? Is it the interaction with other websources? Can functions provided by servers be captured, too? At least they have to be documented. If we know what we want to preserve, we can start building up a solution for it.

The links in the harvested site should be modified, so the site can still be navigated without "slipping" out of the captured site while browsing it. This must be very well documented. Other details, that have to be documented are: Time of Harvesting, Missing files (and why they are missing), excluded files or databases and much more.