DIGITAL PRESERVATION USING OPEN SOURCE XENA SOFTWARE

Richa Sharma[1], A.J. Singh[2]

[1]*Dept. of Computer Science H.P. University Shimla*

[2] *Associate Professor, Dept. of Computer Science H.P. University Shimla*

Er.richa37@gmail.com, aj_singh69@yahoo.co.uk

Abstract: *Digital preservation is a technique used to store error free digital data for long term. While we are still able to read our written heritage from several years ago but digital information created merely a decade ago is in serious danger of being lost creating digital dark age. By a series of open source software by National Archieve of Australia this paper propose a Xena based digital preservation platform for long term preservation of digital data by converting file formats to preserve one.*

Keywords: *Digital preservation, Xena*

I.   Introduction

Digital preservation refers to the various methods Of keeping digital material alive in to the future [1]. Many of the digital files or formats are dependent on particular environment. Any change to the environment leads to the obsolescence of that digital format by creating digital obsolescence."Digital preservation raises challenges of a fundamentally different nature which are added to the problems of preserving traditional format material" [10]. Digital data should be preserved in a way so that data Could be easily retrieved by future users in the same way as today's user. The issue with many organizations recognizing the need to preserve digital

data before they fall victim to digital obsolescence. The question we need to address therefore is "What does this long term means in context of digital preservation." Long term may extend indefinitely" [2].   Here long enough to be concerned with the impact of changing technology including support for new media and data formats or with changing user community [3].For this it should match user requirement

1) The end user should be able to access the preserved digital document.
2) The content should be executable, i.e. the host machine should be able to render the document in its original environment. The end user should be able to interpret and understand the content of the digital document [5].This paper discusses the working of xena software which will help end user for preservation of data long term.

II. Other Syrategies to Preserver Digital Data

Most of the current literature on preservation focuses on archiving and preservation technique. The rapid development of information technology makes the use of hardware and software used to read digital data becomes obsolescent soon. But the information in digital format is so valuable and is urgent to preserve these resources for long term utilization. The concept of digital was generated in 1980s [14].Some of the related preserving techniques are migration, emulation and archiving. According to cedars and CAMileon projects" Migration is the preservation approach which has been most widely practiced to date."At its simplest it is defined as copying or conversion of digital object from one technology to another. Whilst preserving their significant properties. In this area it also had certain problems some of the data or attribute of the digital object may be lost during migration the authenticity of the record may be then compromised, which is major issue for digital archivists [4]. The other factor such as ensuring the authenticity and usability of preserved digital document are always considered. An exception to this is Glandey's work on preserving trustworthy digital objects [9, 11] however Glandey's viewpoint on authenticity relates to trusting the content of preserved digital document. As such he suggests the end users requirement with regard to how they can view and use the preserved document are not considered.

III. Open Source Digital Preservation

Open source software is software for which source code is freely available to anyone. To develop the suite of digital preservation software the National Archives adopted open source development methodology. The open source methodology covers:

Software licensing

External contribution

Version control

Supporting file formats [5].

At present main source digital preservation systems include: Xena (National Archieve library of Australia) [15],

Dspace (MIT and HP company) [6].

Greenstone [8] (University of Waikato).

IV. About XENA

Xena digital preservation software has been developed by the National Archieve of Australia to aid in the long term digital preservation[13].The latest version of xena Version[5.0.0] was released Dec 9, 2009.The main function of xena is to recognize the file formats of digital data and convert them in appropriate preservation file format based on open standards. Xena architecture composed of [12]:
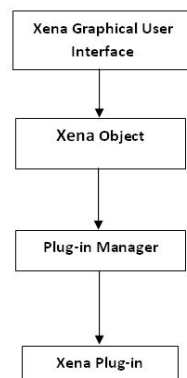
Xena object: The xena object allows to do basic functionality in the system including plug-in:

Loading plugins.

Guessing file types.

Normalizing files.

Exporting normalized files.

Plug-in Manager: It loads the plug-in and distributes the input source file to each of the plug-in to determine the file format.

Xena Plug-in: It consist of one or more components, each having specific role in the conversion process (such as file format detection, file conversion and creation of the xena .xml file) [12].

Xena software aids digital preservation by performing two imp tasks:

Detecting the file formats of digital objects.

Converting digital objects in to open formats for preservation.

V. How Xena Perform file Format Identification

Xena software can be downloaded through National Archieve of Australia [13].Xena uses different plug-in to deal with various file types. Some plug-in components are:

[12] Plug-in structure:

Xena Type: Corresponds to a supported file format.

Guessor: Determines the type of a given xena input file .Normaliser: It takes the xena input file and transform in to an xml file.

View: It displays Xena files.

For the purpose of file format identification there are most important plug-in of Xena they are Guesser and Normaliser. The relation between the plug-in   normaliser and guesser is described by example of image plug-in and image.
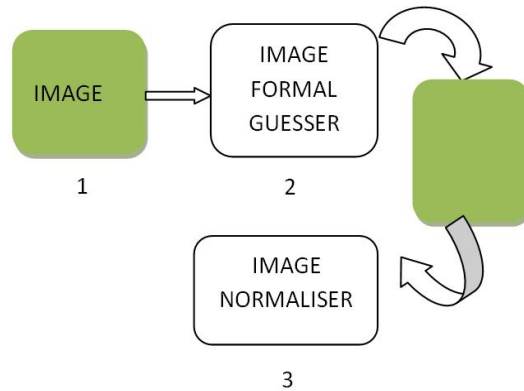
1. Plug-in reads unknown image format
2. Image guesser identifies the image format.
3. Image plug-in normalizes image with image normalizer.

Normalization involves migrating a digital object from the original software in o an open source. Normalization aids to preserving electronic records. Normalization is opt as a rule for open source no normalization format meet all the requirements of any organization therefore combination of approaches may be necessary [7].

Xena stands for xml 'electronic normalizing for archieve and is also known as normalizing software which convert the format in to xml so that format can then be view by xena viewer'.

VI. Objective of Study

The objective of the present study is to look in the tool and software in favor of open source and customize according to the need to preserve long term digital data so that it can be easily accessible by future user.

Research Methdology

As it is discussed earlier Xena is an Open source tool used to converting file formats and then preserve them as a digital preserved object. Mainly these two properties are satisfied by this open source tool. This tool is downloaded from the open source site Source forge net and then installed it

in windows vista. After successful installation of the software next step done is to configure the preferences where the output will be stored and then next step is to configure the plugins such as for image conversion image magic convert is configured by setting path. Another plugin is open office.org for supporting numerous file format. In this paper some of the file formats are included which were normalize by normalizer and hence comes out with certain outputs. The file formats included such as: JPEG,TIFF, PDF files. They put in xena interface window by selecting the directory and then start with the normalization when given the command After the normalization starts next   window will appear and shows those files which were in the xena interface window and give the nomalization results of the given file formats.After the normalization of considered files done then the conversion of these files in to .xena can be seen through xena viewer at Xena output which was configured earlier.

Results:

source	Guessed Type	Normaliser	Success	Zoom factor	Destination
1.file/users/abc	JPEG	JPEG	✓	63.33%	2011026-675 raster graphic-fish-40x46 squaresex.jpg.xena
2file/.users/abc	TIFF	Image TIFF Normalized	✓	49.48%	2011059-blue hills.tif-9a967fb1.xena
3.file/users/abc	JPEG	JPEG	✓	55.23%	2011096-life......jpg-9f2c6fle.xena
4.file/users/abc	TIFF	Image TIFF Normalized	✓	37.11%	2011096-moon.tiff-6111a5eb.xena
5.file/users/abc	TIFF	Image TIFF Normalized	✓	100.00%	2011096-tiff-106.tiff-20f9a747.xena
6.file/users/abc	JPEG	JPEG	✓	63.33%	2011068-winter.jpg-40-cfbd29.xena
7.file/users/abc	PDF	PDF	✓	100%	201094-1433-6780-2-ab%20(1).pdf.xena

TABLE1: Normalization results and conversion of file formats in .xena

The file formats of data were normalized by Normalizer and hence converted in to .xena and saved in destination directory as a preserved Digital object. If any type of image cannot able to be opened in xena viewer window due to corrupted Files then its coding part can be easily through Raw XML view .Xena is a very useful tool to preserve digital Data. It helps user to preserve even very obsolete file formats for long term.

VII. Conclusion

In this study researcher analysed that xena is a useful tool for long term preservation of digital data. It is helpful in house digital preservation programs. It is type of tool which helps any organization to preserve digital records in open formats so that important information can be accessed from a wide range of application or a wide range of computing platform. Xena software now includes some file format to be normalized including Images, audio, documents and hence plug-in architecture enables the software to support many other file fomats.

References

[1]   Convay, Paul (1996): Preservation in the digital word. Washington DC: commission on preservation and access Available at http://www.clir.org/pubs/reports/ /Conway 2/index.html

[2]   Consultative committee for space data systems (2002): Reference model for an Open Archival Information system (OAIS) Washington, DC: CC.SDS secretariat   P1-1.

[3]   Consultative committee for space data system (2002). Reference model for an open archival information system (OAIS) Washington DC: CC SDS secretariat P1-1.

[4]   Cedars Project, cedars project website URL: http://www.leeds.ac.uk/cedars

[5]   Dissecting the digital Preservation software platform version 1.0 RKS; 2009/4026 the archives URL:<http://www.naa.gov.au>

[6]   Dspace: http://www.dspace.org[2008] 4-14

[7]   Electronic records. http://Xena.sourceforget.net/.preserving

[8]   Greenstone: http://www.greenstone.org [2008 -3-13]

[9]   H.M. Glandey, Principles for Digital Preservation: Focusing on End User needs rather than those of Archiving institutions. In communication of the ACM, vol49, Pg no.2

[10]   Hedstrom M (1997): Digital preservation: a time bomb for digital libraries. Retrieved December 4th 2007 from http://www.uky.edu/kierman/DL/hedstro m.html

[11] H. M. Glandey; (2004): Trustworthy 100-Year Digital Objects; Evidence after Every Witness is dead, In ACM Transactions on information system, Vol. 22, No.3, Pg No. 406-436.

[12] Iconsource: http://oxy-gnome.org/

[13] http://Xena.Sourceforge.net/

[14] Song Xiadrao (2005) (in Chinese): The long term preservation of digital information: master dissertation [D], Dept. of common Management, Sichuam Univ.

[15] Xena. http://www. Xena.org [2008-3-13]