Digital Preservation Time Capsule: A Showcase for Digital Preservation

Natascha Surnic, Andreas Rauber

Vienna University of Technology, Vienna, Austria
http://www.ifs.tuwien.ac.at/dp
{surnic,rauber}@ifs.tuwien.ac.at

Abstract. The importance of Digital Preservation is increasing steadily. The amount of digital data is growing fast, and maintaining the long term availability and accessibility of these data is turning into an increasing challenge. Creating an operational archival system requires to solve significant challenges in terms of analyzing and monitoring massive volumes of data while applying preservation actions. Still, simply demonstrating and grasping the issues surrounding Digital Preservation is non-trivial. It is very difficult to make digital preservation tangible, which is the main goal of the Digital Preservation Time Capsule. This paper presents the concepts of the Planets Digital Preservation Time Capsule, an appealing showcase demonstrating a range of activities in the context of Digital Preservation.

1 Introduction

The preservation of digital material has become very important. First of all the amount of digital data is growing fast and the material needs to be preserved. There are different ways to preserve digital material. It is possible to emulate the environment or to migrate data, for example. Another problem is that an appropriate storage media needs to be defined. To sum it up, we need to preserve the data in a long-term run. While numerous approaches and preservation actions have been devised, metadata schemas have been crafted, and storage solutions are on offer, really understanding the complexities of digital preservation, as well as the massive amounts of information required to be able to interpret digital objects, is a hard challenge. The Digital Preservation Time Capsule sets out to make these aspects easier to understand and to put them into context. The goal is to show the dependency of data objects required to turn them into information objects at all levels, starting from the basic primary objects via necessary repreentation information, software required to interpret the data, information rewuired to interprete the standards and other required documenation, the compilers for the viewer software and migration programs used, down to the actual operating system, and - eventually - the actual hardware architectures.

This article is structured as follows. An overview of related work is provided in Section 2. The objects to be placed into the time capsule are described in detail in Section 3. In Section 4 we discuss the expected results and give an outlook on future work to be done.

Fig. 1. Rosetta Stone [7]

2 Related Work

Digital Preservation has some very appealing and tangible showcases in the past event, if they may have been devised for different purposes. Examples are the Rosetta Stone or the Voyager Golden Record, which influenced the Digital Preservation Time Capsule.

2.1 Rosetta Stone

The Rosetta Stone ([7]), depicted in Fig. 1, is an Egyptian artifact which consists of Ptolemaic era steel with carved in text. The text is a single passage, a decree from Prolemy V, who described various taxes and instructions to erect statues in temples. The text is written in three different languages. Two are in Egyptian language, hieroglyphic and Demotic and one in classical Greek. The Rosetta Stone was created 196 BC and discovered by the French in 1799. Translations of the stone helped to decipher hieroglyphic writings. In OAIS [5] terminology, it provided vital descriptive information, albeit embedded, by offering several different representations of the same information. At its highest point the Rosetta Stone is about 114 centimeters high, approximately 72 centimeters wide and almost 28 centimeters thick. The approximated weight is about 760 kilograms. The Rosetta Stone is on public display at the British Museum since the year 1802.

2.2 Voyager Golden Record

The Voyager Golden Record [6], shown in Fig. 2, is a phonograph record, which was included in two Voyager spacecrafts. It was launched in the year 1977 and contains different sounds and images, which should provide an overview of the diversity of life and culture on Earth. It was developed for any intelligent extraterrestrial life form or far future humans that may find the record. The Voyager spacecrafts are not heading toward a specific destination but there are calculations where the spacecraft will be. The record is more seen as a Time Capsule,

than a communication media with any other species because if they ever will be found, it would only be in far future. The content of the record were selected by a committee of the NASA. About 115 images, which are encoded in analogue form, and a huge variety of different sounds are stored on the record. Examples for the sound files are, different animal sounds, the sounds of wind and thunder. Furthermore there are spoken greetings in 55 different languages on the record. Also music is represented on the record. In a selection of 90 minutes, a variety of music from different cultures represents the sounds of the Earth. The Voyager Golden Record stores an hour long recording of the brainwaves of Ann Druyan. The record is a construct of gold-plated copper with an ultra-pure sample of the isotope uranium-238, which is electroplated on the cover of the record. This isotope has a half-life of 4,51 billion years and was chosen because if any civilization will encounter the record, it will be possible for them to determine the age of the record from the remaining uranium. The sentence "To the makers of music - all worlds, all times" is handwritten on the record itself. Voyager 1 was launched in 1977 and has already passed the orbit of Pluto in the year 1990, left the solar system in 1994 and is now in empty space. Similar to the spirit of the Digital Preservation Time Capsule, the Voyager Golden Record contains selected information items represented in such a way as to be presumably intelligible by other / future species.

The cover of the disc, cf. Fig. 3, provides an appealing example of trying to achieve communication with unknown consumers of the information. Instructions on how to "read" read the disc are displayed graphically, with the rotation speed of the disc, for example, being provided in time units of 0,70 billionths of a second, the time period associated with a fundamental transition of the hydrogen atom. The originating location is depicted as the relative position of our solar system with respect to 14 pulsars[1].

Fig. 2. Voyager Golden Record [6]

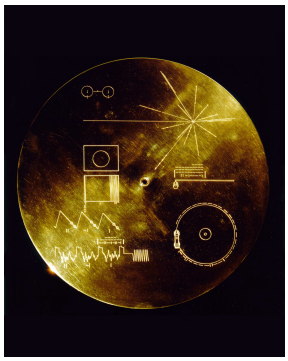

Fig. 3. Voyager Golden Record [6]

[1] http://voyager.jpl.nasa.gov/spacecraft/goldenrec1.html

2.3 LongNow foundation

The LongNow foundation [4] promotes awareness of the importance of long-term thinking and helps to understand the actual problem, its dimensions and to communicate these things beyond the core groups of experts, who are working on problem solving strategies. It created a series of projects [1] to demonstrate various aspects, such as the 10,000 Year Clock Project, the Long Server, or the Long Bets website.

3 Creating a Digital Preservation Time Capsule

In this section an overview of the activities of the Digital Preservation Time Capsule is given.

3.1 Selecting Primary Objects

A set of core objects, the Source Objects, will be selected. The selection of these objects will be small and should explain the core ideas within Planets. The Source Objects are:

Photograph in JPEG format which represents a popular format for images and is used by a vast majority of users of digital cameras. Photographs have been fascinating people since ages, that is the reason why we decided to use a photograph as a Source Object for the Time Capsule.

'Hello World' program in Java A simple and a few code lines similar to a 'Hello World' program in Java will represent the most elementary executable program in source code form.

Short movie on Digital Preservation showing the main aims of Digital Preservation in .MOV format which should act as a short training video for training events and deliver the key message of Digital Preservation to the public. Here are conformities in style with the Voyager Golden Record audio recordings.

Planets Project Homepage in HTML format acts as a representative of more complex object formats which do not only consist of a single file. The homepage focuses on simple, standard HTML-functionality, using formats, which are already within the scope of formats to be described.

Planets Brochure [3] in PDF format acts as a representative of one of the most wide-spread document formats. Furthermore it illustrates the key messages and goals of the Planets Projects.

3.2 Deriving Secondary Objects

The primary objects must be described by their technical and intellectual characteristics using a standard metadata scheme, like PREMIS [2].To guarantee a long term preservation, these objects must further be converted into a range of formats, which are more suitable for long-term preservation. For example, PDF

documents may be converted to PDF/A, or to an image format such as TIFF or JPEG2000. The Quicktime video may be converted to MPEG-4. Both, the tools used to perform the migration and the resulting objects need to be described using the metadata representation scheme, like the Primary Objects mentioned before. Furthermore, all file formats need to be documented by including the appropriate standards.

Viewer for the primary and secondary objects must be defined. The source and the documentation of the viewer must be documented too. In this case a simple PDF viewer, e.g. XPDF, a simple image viewer, e.g. GIMP, a simple video viewer, e.g. XINE, a minimalistic HTML browser, a minimal JAVA virtual machine and other compilers, all the way to a minimal operating system are needed. Further objects may be included. The list above represents only an excerpt of all the objects required to fully document the hierarchy of dependencies for any single digital file. After defining all the appropriate objects, these also need to be documented and described like the Primary Objects.

3.3 Storage

All the objects need to be stored. We choose a wide range of different storage media, like CDs, DVDs, USB flash storage, and HDDs. Conventional "analog" storage such as print-outs to paper, as well was hybrid storage on microfilm will be considered as well. In order to be useable, the storage encodings, the storage media and the reading technology need to be described using a standard metadata representation scheme as well. Descriptive information of the various data carriers, which are included in the Time Capsule, are given in this section. The data carriers should all carry, as far as possible, identical information. Exceptions are obviously planned for the print-to-paper and the microfilming. For some parts it is easier to represent the information in a digital way, instead of printing out a lot of pages to represent them. The achievement is, to include a rather large volume of representation information. Furthermore there will be more exceptions if we decide to include old data carriers, which have not been designed to carry such a large volume of data. To sum it up, the selected storage technologies encompass most types of storage principles, demonstrating storage capacity, size and the evolution of data storage during the years. The data carrier, the drive technology and the file systems must be described as well. This means, that the patent, the ISO Standard, and simple technical descriptions, as well as Wikipedia articles need to be included, adding to the list of secondary objects identified above. For example, the carrier of floppy discs, CDs, DVDs, Blu-Ray discs, HDDs, flash drives, selected tape type storage media and the drive technology description is needed. Furthermore, descriptions of the storage data carrier and their drive technology descriptions, microfilm description and the resolution settings for writing to microfilm need to be documented. An OLPC (One Laptop per Child) laptop is planned to be added to the Time Capsule, to represent a complete reading device. The laptop allows that all the objects can be rendered. The OLPC laptop serves as a host device for many other storage

technologies, especially, CD/DVD, external HDD, flash disk, as well as other devices, if they are connectible via USB. This is useful for exhibitions and for demo purposes of the Digital Preservation Time Capsule. All the objects, concerning storage, also need to be described and documented as the Primary objects.

3.4 Describing Metadata with PREMIS

All digital objects which are deposited in the Time Capsule will be described using a standardize meta-data container. At this moment, the plan is, to use PREMIS metadata, wrapped in a METS container.

3.5 Time Capsule as a showcase for Digital Preservation

The resulting material will provide a tangible showcase to the general public, offering itself for public exhibits at partner institutions, other libraries and archives, as well at museums of science and technology. The big achievement of the Planets Digital Preservation Time Capsule is to make Digital Preservation tangible. To achieve this goal, the Time Capsule will contain all the digital objects mentioned above, in multiple storage formats and encodings. The information will be stored on a wide range of different data carriers.

Objects contained in the Time Capsule The following list gives an overview of the planned physical items of the Time Capsule:

- several punch cards
- several floppy discs
- 1 or more archival grade CD
- 1 archival grade DVD
- 1 high-quality USB flash memory stick
- 1 external HDD drive
- 1 roll of microfilm
- paper print-outs
- 1 external CD/DVD/Blu-Ray reading device with USB connectivity
- 1 OLPC laptop

Deposit Locations for the Time Capsule The Time Capsule will be deposited in a prominent location to stress the key messages. Additional copies of the Planets Time Capsule will be deposit at partner institutions, memory institutions in general, libraries, archives and special museums, as long running test sets. The quality of storage carrier functionality and the quality of data reconstruction will be tested. The opening of the Time Capsules may happen in specific instances, like every 10, 20 etc. years. The primary deposit location of the Time Capsule should be an appealing, somehow exotic location, to offer people an emotional connotation with a longterm archive. There are two different types of deposit: On one hand, deposits that lock away a complete copy of the time capsule for a specific number of years; on the other hand a display deposit, which allows to show the concept of digital preservation at exhibitions with an open copy of the Time Capsule, to make Digital Preservation tangible.

4 Outlook

With different kinds of Media Activities, the Time Capsule should raise the awareness for Digital Preservation activities and challenges to a vast audience. Several activities, like press conferences and focusing on the goals and challenges in Digital Preservation, should motivate the general interest in Digital Preservation. Furthermore, exhibitions, at Planets partner institutions and/or science museums, can use additional copies of the Time Capsule for training purposes.

Acknowledgements

Part of this work was supported by the European Union in the 6th Framework Program, IST, through the PLANETS project, contract 033789.

References

1. Stewart Brand. *The Clock of the Long Now*. Basic Books, 2005.
2. PREMIS Editorial Committee. Premis data dictionary for preservation metadata (version 2.0). Technical report, PREMIS Editorial Committe, 2008.
3. The Planets Consortium. Preservation and long-term acess via networked services, June 2006. `http://www.planets-project.eu/docs/comms/Planets_Project_Brochure.pdf`.
4. The Long Now Foundation. The long now foundation, Aug 2009. `http://www.longnow.org/`.
5. ISO. Open archival information system - reference model (ISO 14721:2003). Technical report, International Standards Organization, 2003.
6. California Institute of Technology Jet Propulsion Laboratory. The Voyager Golden Record, Aug 2009. `http://voyager.jpl.nasa.gov/spacecraft/goldenrec.html`.
7. British Museum. The Rosetta Stone, Aug 2009. `http://www.britishmuseum.org/explore/highlights/highlight_objects/aes/t/the_rosetta_stone.aspx`.