‘If a tree falls in the forest’: recording and sharing digital preservation knowledge on formats, software and dependencies at the NLA.

David Pearson
National Library of Australia - CurateGear
6 January 2012 (UNC)
**Mission:** The primary objective of preservation activities within the NLA is to maintain the ability to meaningfully access digital collection content over time.

- **Contextual Information** – About Content
- **Dependency Information** – About Formats etc.
- **Systems to Ingest, Manage, Report and take Actions**
- **Systems to Access – Master or Derivative**

A. ‘Logical on Physical Stuff’

B. ‘Logical on Physical Stuff’

‘Stuffed?’
Required preservation processes

The Library must be able to:

• Understand what it holds in its collections;
• Understand what its preservation intentions are for every digital object and what it is entitled to do to realise its intentions;
• Understand what is required to provide access, existing inhibitors to access, and the current level of support the Library is able to provide;
• Evaluate and monitor the degree of risk arising from collection composition, preservation intentions and available level of support within the Library for digital collection content, and monitor for risk conditions arising during general Digital Library operations;
• Anticipate the effects of changes in support;
• Recognise planning triggers, and plan and take appropriate action on a scale appropriate to the size of the target; and
• Audit the effectiveness of its preservation arrangements and modify the arrangements if necessary.
Given this we designed this data model for expressing some of these concepts:
Perhaps we will use this model instead?
The Library expects to take into account indicators of ‘preservation intent’, ‘significance’, and ‘level of support’ within monitoring and reporting activities, and in evaluations of risk and prioritisation for preservation planning and action.
Preservation intent – indicates the expectations for preservation for content:

- whether content is to be preserved;
- who is responsible for preservation of the content;
- the period over which content must be preserved;
- the required level of support for access to the content over time; for example, that the Library intends to actively maintain the ability to both present and modify content, or only to present content, or does not intend to actively maintain access to content beyond its expected useful life.

Preservation intent may also extend to include more specific characteristics to be supported, based on curatorial input or constraints imposed by rights policies or agreements with rights holders.
Significance – indicates the relative priority required for taking preservation action to maintain access to content, as determined by collection curators; for example, content rated as highly significant would be prioritised for preservation planning and action before content of lower significance.

Level of support – indicates how well a digital collection object is supported within the Library, based on a combination of how much is known about the object and its components (including their file formats), and the degree to which supporting software or hardware environments are available.
Reference knowledgebases

Enable staff to create, update and maintain reference information knowledgebases on:

a. File formats and versions

b. Software and hardware components that support access to file formats and versions, for maintaining access to managed content;

c. The level of support available for particular file formats and versions:
   i. sets of software or hardware components available to support access to formats;
   ii. functions supported, both for providing access to content and for use in preservation action – for example, presentation, modification, batch processing;
   iii. fidelity of support – how well functions are supported; and
   iv. known risks, including potential inhibitors to preservation, associated with formats or supporting software or hardware;

d. Preservation intent descriptions and parameters for sets of content.
Other systems are also required to interrelate in this ecosystem such as:

• Preservation monitoring, reporting and prioritisation

• Preservation options and preservation action planning

• Preservation action evaluation
Digital Preservation knowledge base

The following pages contain the prototype digital preservation knowledgebases.

To browse different knowledgebases click on the tabs to navigate between different knowledgebases or search using the search box provided. For more information on the rational and background for this project see background information page.
File formats

We need to have more information about the file formats used in our institution so we can effectively deal with the issue of file formats. These pages will provide you with general information about some of the formats used on our sites as well as links to other resources such as specific lists of formats or other institutions, formats which are being used internally at the site, and formats that are generally used in digital preservation. The criteria outlined in the Format Standards page will be used as the basis for selecting, preserving, and maintaining these formats.

Together with the preservation needs, the information gathered and presented in this monograph book is intended to assist in making qualitative judgements as to which formats we should use for our own creations.

These pages include a Trail that is intended to give you an idea of the format history of a particular file format. If you need to see the data of a file, you should probably be a more programmatic approach for interpreting the data of a file, and to help you have a clearer understanding of:

- What formats are being used by different digital presentation oriented software applications.
- The advantages and disadvantages of these formats.
- Why are some formats preferred over others?

The list of file formats in this monograph book:

All file formats

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIF</td>
<td>image</td>
<td>is an animation of the GIF format.</td>
</tr>
<tr>
<td>JPEG</td>
<td>image</td>
<td>is a compressed format.</td>
</tr>
<tr>
<td>PDF</td>
<td>document</td>
<td>is a portable document format.</td>
</tr>
<tr>
<td>TIFF</td>
<td>image</td>
<td>is a tagged image file format.</td>
</tr>
<tr>
<td>MOV</td>
<td>video</td>
<td>is a video file format.</td>
</tr>
<tr>
<td>AVI</td>
<td>video</td>
<td>is a video file format.</td>
</tr>
<tr>
<td>WAV</td>
<td>audio</td>
<td>is a audio file format.</td>
</tr>
<tr>
<td>MP3</td>
<td>audio</td>
<td>is a audio file format.</td>
</tr>
<tr>
<td>FLV</td>
<td>video</td>
<td>is a video file format.</td>
</tr>
<tr>
<td>MP4</td>
<td>video</td>
<td>is a video file format.</td>
</tr>
<tr>
<td>MPEG</td>
<td>video</td>
<td>is a video file format.</td>
</tr>
<tr>
<td>WMV</td>
<td>video</td>
<td>is a video file format.</td>
</tr>
<tr>
<td>SWF</td>
<td>application</td>
<td></td>
</tr>
<tr>
<td>HTML5</td>
<td>text</td>
<td>is a hypertext markup language.</td>
</tr>
<tr>
<td>XML</td>
<td>text</td>
<td>is a text document.</td>
</tr>
<tr>
<td>RDF</td>
<td>text</td>
<td>is a text document.</td>
</tr>
<tr>
<td>JSON</td>
<td>text</td>
<td>is a text document.</td>
</tr>
<tr>
<td>RSS</td>
<td>text</td>
<td>is a text document.</td>
</tr>
<tr>
<td>Atom</td>
<td>text</td>
<td>is a text document.</td>
</tr>
</tbody>
</table>

Software

This monograph book of digital preservation standards contains a list of useful tools for managing and maintaining digital content.
**BMP Format**

**General Information:**
BMP is a bitmap image format. It was developed by Microsoft for use with Windows operating systems. BMP is a straightforward, lossless format that stores pixels in a linear sequence. It is useful for storing images that will be printed or displayed on screen, as it can support both 8-bit and 24-bit color images.

**Format type:** Image

**Vendor:** Microsoft

**Extensions:** .bmp

**Technical capacity:** BMP supports bitmaps with up to 16.7 million colors (24-bit) and supports transparency, making it suitable for a wide range of uses.

**Conditions of use:**
- The BMP format is widely supported by both Windows and other operating systems, ensuring widespread compatibility.

**Adoption:**
BMP is the native image format for Windows and is used by many applications, including Microsoft Paint.

**Strengths:**
- BMP is simple and straightforward, making it easy to use.
- It supports transparency, allowing for the seamless integration of images into other documents.

**Limitations:**
- BMP is not as space-efficient as some other formats, such as PNG or JPEG, especially for images with a large amount of color information.

**Stability:**
BMP is a simple, stable format, which means that it is less likely to be affected by changes in the Windows operating system.

**Additional information:**
- Free software programs are available for editing BMP images.

**File format versions for the BMP Family**

<table>
<thead>
<tr>
<th>Version</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP 1</td>
<td>Released in 1985</td>
</tr>
<tr>
<td>BMP 2</td>
<td>FileType (1987)</td>
</tr>
<tr>
<td>BMP 3</td>
<td>FileType (1989)</td>
</tr>
<tr>
<td>BMP 4</td>
<td>FileType (1991)</td>
</tr>
</tbody>
</table>

**Software:**
- Paint
- PaintShop Pro
- Photoshop

**Hardware:**
- Any Windows operating system

**Physical quantity:**
- The BMP format is lightweight in terms of file size compared to other formats like JPEG or GIF.
Software

Digital content is dependent on software in order to be meaningfully interpreted. These applications vary greatly in functionality and complexity. This page contains information about software that can be used to manipulate, display, or access the content of the Digita Library preserved in this service library. By using this software, a user can use the information to assess media, view paths to maintain access to our digital content. For the sake of both users and providers, more detailed information such as detailed dependency listings are not included. If more detailed information is required at a later date, it will be added as required.

Where possible, information about how to obtain the software, or if there is a copy in the Digital Preservation software library, has been included.

<table>
<thead>
<tr>
<th>Name</th>
<th>Operating System</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LibreOffice</td>
<td>Linux, Mac, Windows</td>
<td>Open source productivity suite developed by LibreOffice Foundation.</td>
</tr>
<tr>
<td>Adobe Reader</td>
<td>Windows, Linux, macOS</td>
<td>A PDF viewer and editor.</td>
</tr>
<tr>
<td>Mozilla Firefox</td>
<td>Windows, Linux, macOS</td>
<td>An open-source browser and web development platform.</td>
</tr>
<tr>
<td>Microsoft Word</td>
<td>Windows</td>
<td>A word processing program developed by Microsoft.</td>
</tr>
</tbody>
</table>

Search the Knowledge Base

Valid entries include:
- LibreOffice
- Adobe Reader
- Mozilla Firefox
- Microsoft Word
Photoshop CSS

General information:
Photoshop is a very popular image manipulation and graphic program developed by Adobe. The program itself is very powerful and provides support for a very broad range of functions and uses. For example, it can be used as a tool to retouch photos, or to create new images entirely from scratch. It has a very broad user base, particularly for those who work in the graphic or photographic industries. The default saving format for Photoshop is PSD, a proprietary industry standard, and can be opened in a variety of programs.

Software type:
Application

Current version:
CS5

Highest level of internal support:
SDE

Requirements:
Microsoft Windows XP with Service Pack 3, Windows Vista, Windows 7, Windows 8, or later. On Macintosh, Photoshop CS6 requires OS X 10.6.8 or later. Additional free space required during installation: 2GB hard drive space required during installation. If installed on removable media, a hard drive or USB thumb drive is recommended. A hard drive is recommended for faster performance. System requirements: 1.6GHz Intel Core 2 Duo processor, 1GB RAM, and 1024MB of video RAM for video editing. Additional free space required during installation: 2GB hard drive space required during installation. If installed on removable media, a hard drive or USB thumb drive is recommended. Additional free space required during installation: 1GB hard drive space required during installation.

Vendor availability:

<table>
<thead>
<tr>
<th>Name</th>
<th>Vendor support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adobe Creative Cloud</td>
<td>Current supported</td>
</tr>
<tr>
<td>Adobe Creative Suite</td>
<td>Current supported</td>
</tr>
</tbody>
</table>

Optional support:

<table>
<thead>
<tr>
<th>Name</th>
<th>Optional support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adobe Creative Suite</td>
<td>Standard</td>
</tr>
</tbody>
</table>

Additional information:

- Adobe
- Help
# Hardware

The various hardware configurations that are considered to be relevant to accessing elements of the collection should be known and their dependencies understood.

<table>
<thead>
<tr>
<th>All file formats</th>
<th>Sort by</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>Zip Drive</td>
<td>Zip Drives support Zip Disks, which are relatively low...</td>
<td></td>
</tr>
</tbody>
</table>

**Home**

- File formats
- Software
- Hardware
- Storage
- Physical carrier

**Home**

- File formats
- Software
- Hardware
- Storage
- Physical carrier

Search the knowledge base
relations test view
Zip Drive

General Information:
Zip Drives support Zip Disk, which are relatively low capacity (less than 1GB) storage devices that were introduced into the market by Iomega in the mid-1990s. Although they gained some popularity, they were soon superseded by the advent of lower cost storage mediums, such as CD-R, and later by removable USB flash memory. The market for Zip drives is now very low. Zip Drives come in both internal and external form factors, using a variety of connectors (USB, SCSI, Parallel, etc). External models use an external power pack.

Additional Information:
Wikipedia Zip Drive page
Iomega web site

File format versions for the Zip Drive family

<table>
<thead>
<tr>
<th>Version</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZIP 100</td>
<td>ZIP 100 version of drive.</td>
</tr>
<tr>
<td>ZIP 200</td>
<td>ZIP 200 version of drive.</td>
</tr>
<tr>
<td>ZIP 750</td>
<td>ZIP 750 version of drive.</td>
</tr>
</tbody>
</table>
ZIP 100

General Information:
ZIP 100 version of drive.

External Identifiers:
So what?

Thank you

Everything, for Everyone Forever

http://www.flickr.com/photos/ricksmit/15671245/