Registry of File Formats

Scholars Portal Registry of File Formats

1.Policy Statement

SP requires immediate identification of the type of file format sent by the publisher in order to help mitigate risk posed by format obsolescence. To this end, SP employs the use of DROID, JHOVE and PRONOM through the FITS software package.

While Scholars Portal is not dependent on or restricted to any particular format or group of formats, it aims to use well-known, widely accepted formats that support long-term preservation. If a publisher wants to use a specific format not meeting these criteria, an agreement must be reached between the publisher and SP.

2. Implementation Examples

2.1. DROID

- Scholars Portal makes use of DROID for format identification during the ingestion process where a file format is associated with each file.

- DROID (digital record object identification) is a software tool developed by the National Archives to perform automated batch identification of file formats. It is a platform-independent Java tool, which is freely available to download under an open source license.1

2.2. JHOVE

- Scholars Portal employs JHOVE as a tool for further format-specific identification, validation and characterization of the file.

- JHOVE (JSTOR/Harvard Object Validation Environment) is an extensible framework for format validation created by a collaboration between JSTOR and Harvard University Library.

2.3. PRONOM

- During the process of DROID identification, a file format is associated with each file, and, where possible, the file is linked to the format's entry in PRONOM, the British National Archive's format registry.

- PRONOM is a resource providing impartial and definitive information about the file formats, software products and other technical components required to support long-term access to electronic records and other digital objects of cultural, historical or business value.3

Example characterization and reference to format registry:

```
<PREMIS:format>
   <PREMIS:formatDesignation>
      <PREMIS:formatName>Acrobat PDF 1.4 - Portable Document Format</PREMIS:formatName>
      <PREMIS:formatVersion>1.4</PREMIS:formatVersion>
   </PREMIS:formatDesignation>
   <PREMIS:formatRegistry>
      <PREMIS:formatRegistryName>http://www.nationalarchives.gov.uk/pronom</PREMIS:formatRegistryName>
      <PREMIS:formatRegistryKey>fmt/18</PREMIS:formatRegistryKey>
   </PREMIS:formatRegistry>
</PREMIS:format>
```

3. References

1. Sourceforge.net. (2009). DROID. Retrieved from http://droid.sourceforge.net /
2. Harvard University Library. (2009, February 25). JHOVE – JSTOR/Harvard Object Validation Environment. Retrieved from http://hul.harvard.edu /jhove
3. The National Archives. The Technical Registry PRONOM. Retrieved from http://www.nationalarchives.gov.uk/PRONOM/Default.aspx

4. Document History

Version	Date	Change	Author
0.1	09/27/11	Draft created	Aurianne Steinman
0.2	09/29/11	Formatted	Aurianne Steinman
0.3	10/28/11	Minor edits	Steve Marks

File	Modified
Microsoft Word 97 Document Scholars Portal Registry of File Formats.doc	Nov 03, 2011 by Aurianne Steinman

See also:

Environmental Monitoring of Preservation Formats