Information for All Programme

United Nations
Educational, Scientific and
Cultural Organization

IFAP Working Group on Information Preservation

Proposal 4 for the Preparation of the Draft Programme and Budget for 2012-2013 (36 C/5)

Promoting International Cooperation in Format Registries

The Issue

One of the key problems in digital preservation are the constantly changing file formats. Commonly used file formats become obsolete, new file formats evolve and new versions are no longer compatible to old versions of file formats. Often file formats are not fitting to preservation needs, they are not open and documented but developed for one single product without any possibility to be used more general. It is essential for the planning of a digital preservation strategy to deal with these issues. A database or registry of the thousands of existing file formats is needed to manage the information. As the file formats are used all over the world it makes sense to have a global approach for such a file format registry.

A file format registry should not only record the characteristics of file formats. It is also necessary to know what kind of software and hardware is needed to use the file formats. The structure of a file format registry should reflect the complex dependencies of system requirements for the usage of certain files. It should be possible to derive an effective preservation strategy from the data of the file format registry.

Finding a useful structure for the file format registry is important. But it is even more important to fill the registry with the detailed information of the file formats. This is a huge task because it needs data of many sources and has to be constantly updated. Often the information of widely used file formats is available from certain companies only. So there should be a co-operation with these companies and/or a legal obligation for them to contribute.

Best practice: PRONOM by the National Archives of the UK (http://www.nationalarchives.gov.uk/PRONOM) was the first practical approach to establish a file format registry. It is still very basic, but it is further developed in European projects like PLANETS. The Global Digital Format Registry (GDFR, http://www.gdfr.info/) was developed at the Harvard University Library. Its data model is much more complex than the one of PRONOM, but the actual implementation was never more than a proof of concept. In April 2009 GDFR and PRONOM started a joined development under the name Unified Digital Format Registry (UDFR, http://www.udfr.org/). The UDFR project is a co-operation of several international partners.

Recommendation:

The IFAP Information Preservation Working Group, therefore, recommends to the Director General

- to give digital preservation a high priority further on and promote the building up of format registries to support efficient digital preservation worldwide by including this topic into the next-term strategy,
- to support the global cooperation in maintaining such format registries by charging competent partners to work on this,
- but also to recommend the usage of formats fitting to preservation needs by drafting and spreading a paper which contains the necessary information about that issue.

All measures should be established in close intersectoral cooperation between Communication/Information, Culture, and Science. Within the CI sector measures should be implemented within framework of the Memory of the World Programme. Within the Science Sector this topic is important with respect to the preservation of scientific information and research data.

Possible partners in that topic are the National Libraries, Archives and Museums, national digital preservation agencies like **nestor** in Germany or the digital preservation coalition in the UK, standardisation agencies like DIN in Germany or ISO, and selected private companies like Adobe.

June 2010