Quality Standards for Preserving Digital Cultural Heritage

Börje Justrell
Riksarkivet
Fyrverkarbacken 13 -17, Stockholm, Sweden
+46104767187
borje.justrell@riksarkivet.se

Antonella Fresa
Promoter Srl
Via Boccioni, 2, Peccioli (PI), Italy
+390587466881
fresa@promoter.it

Claudio Prandoni
Promoter Srl
Via Boccioni, 2, Peccioli (PI), Italy
+390587466881
prandoni@promoter.it

Peter Pharow
Fraunhofer IDMT
Ehrenbergstrasse 31, 98693 Ilmenau, Germany
+493677467363
phw@idmt.fraunhofer.de

Magnus Geber
Riksarkivet
Fyrverkarbacken 13 -17, Stockholm, Sweden
+46702561317
magnus.geber@riksarkivet.se

Erwin Verbruggen
Nederlands Instituut voor Beeld en Geluid
Media Parkboulevard 1, 1217 WE Hilversum, The Netherlands
+31624151991
everbruggen@beeldengeluid.nl

ABSTRACT

Memory institutions face increasing volumes of electronic documents and other media content for long term preservation. Data are normally stored in specific formats for documents, images, sound, video, etc., produced by software from different vendors. This software is controlled neither by the institution producing the files, nor by the institution that archives it. This obligates memory institutions to carry out conformance tests before accepting transfers of electronic collections, but again these are beyond the control of the institution and can be unreliable. This poses problems for long-term preservation. Digital content, meant for preservation, passing through an uncontrolled generative process can jeopardise the preservation process. The objective of PREFORMA (PREservation FORMAts for culture information/e-archives) – a Pre Commercial Procurement project co-funded by the European Commission under its FP7-ICT Programme – is to give memory institutions full control of the conformance testing of files created, migrated and ingested into archives. This is achieved through the development of open source tools which enable this process within an iteration that is under the full control of memory institutions. The project aims to establish a sustainable ecosystem involving interested stakeholders from a variety of backgrounds, including researchers, developers and memory institutions. The workshop will present the results of the project, including demonstration of the conformance checkers developed during the prototyping phase. This will inform a discussion with the digital preservation community – open source community, developers, standardization bodies and memory institutions – about the opportunities offered by PREFORMA and the challenges that are still to be addressed.

Keywords

standard file formats; digital cultural heritage; open source; digital preservation; conformance tests.

1. INTRODUCTION

PREFORMA (www.preforma-project.eu) is an EU-funded Pre-Commercial Procurement project working on one of the main challenges faced by memory institutions: the **long-term preservation of digital data**.

The main objective of the pre-commercial-procurement launched by PREFORMA is the development and deployment of three **open source tools**. These are developed for memory institutions, or any organisation with preservation requirements, wishing to check file-format conformance with a specific technical standard. The conformance checkers establish whether a file complies with a specific specification and facilitate memory institutions' acceptance criteria outside of file-format specifications. The software reports in human and machine-readable format which properties deviate from the specification and acceptance criteria, and can perform automated fixes for simple deviations in the file's metadata.

The process of conformity testing guarantees that the digital content to be ingested into the archives conforms to standards and, if necessary, the content can be re-processed for corrections.

The veraPDF consortium, Easy Innova and MediaArea are working on the implementation of the conformance checkers. The software development is carried out in a collaborative environment with memory institutions and subject-matter experts. The tools are being developed in an iterative process including a number of experiments with 'real' data sets from memory institutions and other organisations who have offered to participate in the testing phase, with the aim of demonstrating the software's effectiveness and refining the prototype releases.

The standards[1] covered in the project are:
- PDF/A for electronic documents
- uncompressed TIFF for images
- a combination of MKV, FFv1 and LPCm for AV files.

The first prototypes are available for download from the PREFORMA Open Source Portal at http://www.preforma-project.eu/open-source-portal.html.

Due to the need for sound, long-term sustainable solutions, the overall objective of the PREFORMA project is not "artefact centric", but instead aims to establish a long-term sustainable ecosystem around the developed tools, involving interested stakeholders from a variety of backgrounds.

2. TARGET AUDIENCE

The workshop is a step towards the establishment of this self-sustaining community and is aimed at anyone interested in digital preservation and cultural heritage:
- *Memory institutions* (museums, libraries, archives, etc.) and cultural heritage organisations coordinating or representing memory institutions that are involved in (or planning) digital culture initiatives.

---

[1] For further information on the standard file format specifications addressed in PREFORMA visit the project's website at http://www.preforma-project.eu/media-type-and-standards.html

- *Developers* who contribute, or are interested in contributing code to the PREFORMA tools as well as those developing commercial solutions and *enterprises* integrating the reference implementations into production software.
- *Research organisations* providing technical and expertise advice to cultural stakeholders.
- *Standardisation bodies* maintaining the technical specifications for the curation and preservation formats covered in PREFORMA.
- *Funding agencies*, such as Ministries of Culture and national/regional administrations, that own and manage digitisation programmes and may endorse the use of the PREFORMA tools in the digitisation, curation and preservation process.
- *Best practice networks* endorsing the use of open standards for creating and managing digital content.
- *Other projects* in the digital culture, e-Infrastructures and policy arenas that are considering the use of PCP.

The anticipated impact of PREFORMA is the reduction of curation and preservation costs, the improvement of curation and preservation capacity, increased competence in public organisations, including small archives, and enhanced independence from individual vendors.

3. PROGRAMME

The workshop will feature presentations by representatives from the PREFORMA project, live demonstrations of the conformance checkers and a round table discussion which will offer the audience the opportunity to share their views on possible improvements to the tools, possible integrations with other software / systems and what they perceive as the main challenges to be addressed in the future.

Draft Programme:
- 30 mins: presentation of PREFORMA, its main achievements and how to contribute
- 60 mins: working session with demonstrations of the conformance checkers
- 30 mins: break
- 60 mins: round table discussion: integration opportunities and future challenges
- 30 mins: lessons learnt and conclusions.