

Preserving Digital Cultural Heritage: the PREFORMA Project

Antonella Fresa

Promoter Srl

PREFORMA Technical Coordinator

Problem statement



- ❑ Memory institutions are facing **increasing transfers** of electronic documents and other media content for long term preservation.
- ❑ Data content are normally stored in specific **file formats** for documents, images, sound, video etc., and these files are usually produced by software from different vendors.
- ❑ Even if the transferred files are in standard formats, standards implementation and particularly **conformance test** process is locked-in, under the control of the vendors.
- ❑ This poses problems in **long-term preservation**. Data objects meant for preservation, passing through an uncontrolled generative process, can jeopardies the whole preservation exercise to be put in place by the memory institutions.

Proposed solution



- ❑ To bring the conformance test phase, under the control of the memory institutions, by providing an **open source** suite of software tools
- ❑ To develop a set of tools able to address the problem from a **multimedia** dimension: documents, still images and audiovisual content
- ❑ To provide **interoperability** mechanisms to allow the integration of the conformance test tools into the legacy systems of the memory institutions and the extension of the tools to new formats

PREFORMA Project



- ❑ PREFORMA is a **Pre-Commercial Procurement** project co-funded by the European Commission under FP7-ICT Programme.
- ❑ **Start date:** 1 January 2014
- ❑ **Duration:** 48 month (end date: 31 December 2017)
- ❑ **Total budget for the procurement:** 2.805.000 EUR
- ❑ **Website:** www.preforma-project.eu

Project Partners



- ❑ RIKSARKIVET, Sweden **Project Coordinator and memory institution**
- ❑ PROMOTER SRL, Italy **Technical Coordination and Communication Manager**

- ❑ **Technical partners**
 - PACKED EXPERTISECENTRUM DIGITAAL ERFGOED VZW, Belgium
 - FRAUNHOFER IDMT, Ilmenau, Germany
 - HOGSKOLAN I SKOVDE (University of Skovde), Sweden
 - UNIVERSITA DEGLI STUDI DI PADOVA, Italy

- ❑ **Memory institutions**
 - STICHTING NEDERLANDS INSTITUUT VOOR BEELD EN GELUID, Netherlands
 - Koninklijk Instituut voor het Kunstpatrimonium, Belgium
 - GREEK FILM CENTRE AE, Greece
 - LOCAL GOVERNMENT MANAGEMENT AGENCY-AN GHNIOMHAIREACHT BAINISTIOCHTA RIALTAIS AITIUIL, Ireland
 - STIFTUNG PREUSSISCHER KULTURBESITZ, Germany
 - AYUNTAMIENTO DE GIRONA, Spain
 - Eesti Vabariigi Kultuuriministeerium (Ministry of culture), Estonia
 - KUNGLIGA BIBLIOTEKET, Sweden

1. veraPDF Consortium (led by Open Preservation Foundation and PDF Association)
 - The PDF/A conformance checker accepted industry-wide (PDF/A)
2. EasyInnova
 - Digital Preservation Formats Manager (TIFF)
3. MediaArea
 - PREFORMA MediaConch - CONformance CHecking for audiovisual files (MKV|FFV1|LPCM)

Overall R&D Objective (The PREFORMA Challenge)



□ Develop an **open source conformance checker** that:

- checks if a file complies with standard specifications
- checks if a file complies with the acceptance criteria of the memory institution
- reports back to human and software agents
- perform simple fixes

□ Use cases:

- Conformance Checking at **Creation Time**
- Conformance Checking at **Transfer time**
- Conformance Checking at **Digitization time**
- Conformance Checking at **Migration time**

File formats standard specifications



Content type	Standard specifications
ELECTRONIC DOCUMENT	PDF 1.7 (ISO 32000-1:2008) PDF/A-1 (ISO 19005-1:2005) PDF/A-2 (ISO 19005-2:2011) PDF/A-3 (ISO 19005-3:2012)
IMAGE	TIFF/EP (ISO 12234-2:2001) TIFF/IT (ISO 12369:2004)
AUDIOVISUAL	MKV (http://www.matroska.org/technical/index.html) Lossless FFV1 http://www.ffmpeg.org/~michael/ffv1.html Linear PCM (IEC 60958-1 ed3.1 Consol. with am1: 2014)

Pre-Commercial Procurement (PCP)



- ❑ Pre-Commercial Procurement (PCP) is a competition-based procurement method, which enables public sector bodies to engage with innovative businesses and other interested parties in R&D projects to arrive at innovative solutions that address specific public sector challenges and needs.
- ❑ These innovative solutions are created through a phased procurement of development contracts to reduce risk.
- ❑ Pre-Commercial Procurement is becoming more and more common within the public sectors of the European Union.

Open Source approach



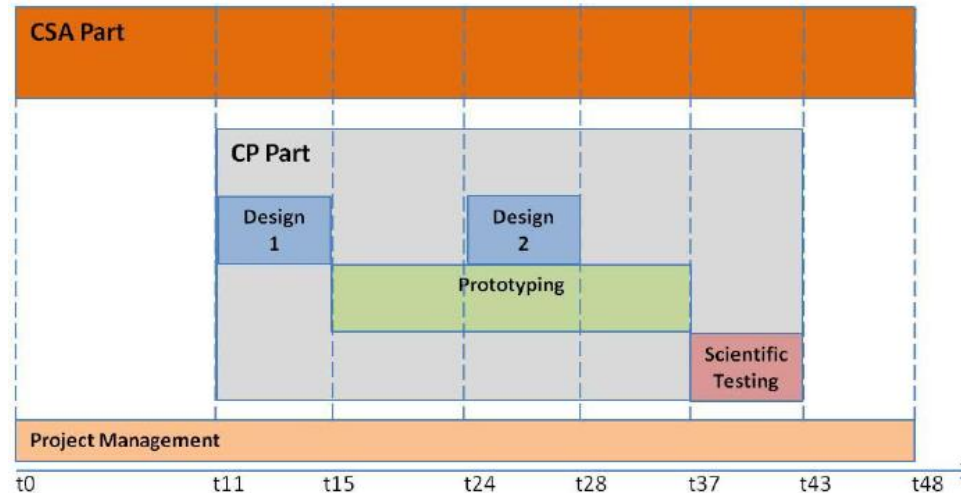
- ❑ Establish an ecosystem around the **open source reference implementation** that:
 - generates useful feedback for those who control software
 - advances improvement of the standard specification
 - advances development of new business cases for managing preservation files
- ❑ PREFORMA is following an **open source approach**, with the aim of establishing a sustainable research and development community comprising developers and users, from different stakeholder groups.
- ❑ The open source nature ensures **long-term availability** of the software, beyond the group of memory institutions and suppliers involved in PREFORMA.

Target users and stakeholders



- ❑ **Memory institutions** and cultural heritage organisations, which are involved in (or planning) digital culture initiatives.
- ❑ **Developers** contributing code for the PREFORMA open source tools.
- ❑ **Research organisations** providing technical advice to cultural stakeholders.
- ❑ **Standardisation bodies** maintaining the technical specifications of the 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 process.
- ❑ Other **projects** in the digital cultural heritage domain.

Project implementation schedule



- **Design phase** (4 months): November 2014 – February 2015
- **Prototyping phase** (22 months): March 2015 – December 2016
 - First prototypes: March 2015 – October 2015
 - Re-design: November 2015 – February 2016
 - Second prototype: March 2016 – December 2016
- **Testing phase** (6 months): January 2017 – June 2017

Training data

To be used by the suppliers for training the software components during the prototyping phase

Test data

To be used by the suppliers and by the PREFORMA partners for testing the software components during the prototyping and testing phases

Demonstration data

To be used by suppliers, by the PREFORMA partners and by third parties for demonstrating and endorsing the PREFORMA solution

How content providers contribute



- ❑ Cultural institutions and other content provider outside the PREFORMA Consortium participated in the **requirement** analysis
- ❑ They contribute by providing **datasets** used to test the prototypes , including both valid/conform files and corrupted files
- ❑ They participate in the workshops and public events organized by the project, follow the project online, provide feedbacks and comments on the project evolution, joining a long-lasting **community on common interest**

Open Source Portal

www.preforma-project.eu/open-source-portal.html



PREFORMA

- ❑ First prototypes of the conformance checkers are available for download

The screenshot shows the PREFORMA Open Source Portal website. At the top, there is a navigation menu with links: HOME, PROJECT, PARTNERS, TENDER, ACTIVITIES, OPEN SOURCE PORTAL, COMMUNITY, DOWNLOAD, and CONTACTS. The main content area is titled "OPEN SOURCE PORTAL" and contains a description: "This section provides an overview and references to each open source project that is currently working in the prototyping phase. It acts as an entry point for all interested suppliers and memory institutions allowing easy navigation to all externally hosted resources." Below this, there are three project entries:

- PROJECT N. 1. VeraPDF: THE PDF/A CONFORMANCE CHECKER ACCEPTED INDUSTRY-WIDE**
by *Open Preservation Foundation, PDF Association, Digital Preservation Coalition, Dual Lab, KEEP SOLUTIONS*
A unique collaboration, the VeraPDF Consortium brings together an end user community and a software industry rooted in the principle of interoperability based on ISO standardized technology... [access project page >>](#)
- PROJECT N. 2. DPF MANAGER: DIGITAL PRESERVATION FORMATS MANAGER**
by *Easy Innova*
DPF Manager is an open source modular TIFF conformance checker that is extremely easy to use, to integrate with existing and new projects, and to deploy in a multitude of different scenarios... [access project page >>](#)
- PROJECT N. 3. MEDIACONCH - CONFORMANCE CHECKING FOR AUDIOVISUAL FILES**
by *MediaArea.net*
MediaConch is an extensible, open source software project consisting of an implementation checker, policy checker, reporter and fixer that targets preservation-level audiovisual files for use in memory institutions... [access project page >>](#)

On the right side of the page, there is a section titled "PREFORMA OPEN SOURCE PROJECTS" with links to "PDF/A CONFORMANCE CHECKER", "DPF MANAGER", and "MEDIACONCH". Below this, there is a section titled "OTHER EXISTING TOOLS" with links to "ARCHIVEMATICA", "JPYLYZER", and "MEDIA FILE CHECKER". At the top right of the page, there is a text box stating: "This project has received funding from the European Union's Seventh Framework Programme under grant agreement no 618566".

Cooperation with other projects



❑ **BenchmarkDP**

- Use shared methodologies and approaches to establish an objective frame of reference for the evaluation of the conformance checkers

❑ **Europeana Space**

- Integrate the conformance checkers in the Technical Space, a web based application for the development of applications and services based on digital cultural content

❑ **E-ARK**

- Use the PREFORMA tools in E-ARK pilot archival services

❑ **AppHub**

- Additional channel to distribute the conformance checkers
- Evaluate and incorporate the code quality and OSS risk management best practices developed by the AppHub community

Next appointments



- ❑ **Open Source Workshop,**
Stockholm, 7 April 2016
- ❑ **Experience Workshop for memory institutions,**
Berlin, December 2016

Follow us!



PREFORMA Website

www.preforma-project.eu

A screenshot of the PREFORMA website homepage. The header includes the PREFORMA logo, the European Union flag, and the Seventh Framework Programme logo. A navigation menu lists: HOME, PROJECT, PARTNERS, TENDER, EVENTS, OPEN SOURCE PORTAL, COMMUNITY, DOWNLOAD, CONTACTS. The main content area features a large call-to-action: 'VISIT THE OPEN SOURCE PORTAL' with a subtext 'Give your contribution to the prototyping phase'. Below this is a section titled 'UPCOMING EVENTS' with a 'VIEW ALL' link. The first event listed is 'BRUSSELS 26 JANUARY 2016 APPHUB SQUAT FEST'. A brief description follows: 'The aim of the AppHub project is to support the market outreach strategies of EU-supported open source by launching AppHub, the European open source market place. AppHub is a service platform that will help the market to seamlessly identify, position and implement the software outcomes of your projects. The main objective of the SQUAT Fest is to assist European open source project to bring their software into the AppHub European Open Source Marketplace.' A 'READ MORE' link is provided. Below the event list is a 'NEWS FROM THE BLOG' section with a 'VIEW ALL' link. Two news items are visible: 'VERAPDF 0.8 NOW AVAILABLE' and 'COOPERATION BETWEEN PREFORMA AND BENCHMARKDP'. At the bottom, logos for 'veraPDF' and 'Benchmark DP' are displayed.

A screenshot of the PREFORMA Blog page. The header shows the date 'Friday, 18 December 2015' and a navigation menu: WEBSITE, PROJECT, PARTNERS, TENDER, ACTIVITIES, OPEN SOURCE PORTAL, COMMUNITY, DOWNLOAD, CONTACTS. The main content area features a 'PRESENTATION OF THE PROJECT' section with a 'RESERVED AREA' label. Below this is a 'MEDIA PARTNER' section for 'DIGITAL CULTURE'. The 'RELEVANT NEWS FROM Digitalmeetsculture' section includes a post from 'LoCloud' titled 'Last chance to submit your ideas in the shreds My LoCloud Services' and another titled 'LoCloud international competition - deadline 1 November!'. An 'IN FOCUS' section highlights a 'PREFORMA OPEN SOURCE PRESERVATION WORKSHOP' held on '7 APRIL 2016' at 'Kungliga Biblioteket, Stockholm, 7 April 2016'. A 'CONTACTS' sidebar lists: Project Coordinator: Borje Jurett; Technical Coordinator: Antonella Fresa; Communication Coordinator: Claudio Prandini; Scientific Coordinator: Nicola Fero; Project website: www.preforma-project.eu. The sidebar also lists 'COORDINATOR' (Riksbank), 'TECHNICAL COORD.' (PROMOTER), and 'PARTNERS' (PACKED, Fraunhofer IPT).

PREFORMA Blog

www.digitalmeetsculture.net/preforma

PROMOTER
Information technology, research and innovation

ILIDE 2016
Jasná, Slovakia, 16 March 2016



Contacts

Project Coordinator:
Borje Justrell, Riksarkivet
borje.justrell@riksarkivet.se

Technical Coordinator:
Antonella Fresa, Promoter Srl
fresa@promoter.it

Communication Manager:
Claudio Prandoni, Promoter Srl
prandoni@promoter.it

Thank you!

Antonella Fresa
PREFORMA Technical Coordinator

Promoter Srl

fresa@promoter.it

QUESTIONS

?