

# PREservation FORMAts for culture information/e-archives

*PREFORMA  
and the Open Source Community  
Borje Justrell,  
Swedish National Archives  
Coordinator*

# Project Identity Card



- ❑ 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](http://www.preforma-project.eu)
- ❑ **Contacts**
  - Project Coordinator: Borje Justrell, Riksarkivet, [borje.justrell@riksarkivet.se](mailto:borje.justrell@riksarkivet.se)
  - Technical Coordinator: Antonella Fresa, Promoter Srl, [fresa@promoter.it](mailto:fresa@promoter.it)
  - Communication Coordinator: Claudio Prandoni, Promoter Srl, [prandoni@promoter.it](mailto:prandoni@promoter.it)



# Project Partners



- ❑ RIKSARKIVET, Sweden **Project Coordinator and memory institution**
- ❑ PROMOTER SRL, Italy **Technical and Communication Coordinator**
  
- ❑ **Technical partners**
  - PACKED EXPERTISECENTRUM DIGITAAL ERFGOED VZW, Belgium
  - FRAUNHOFER-GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V, 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 GHNIOMHAIREACTH BAINISTIOCHTA RIALTAIS AITIUIL, Ireland
  - STIFTUNG PREUSSISCHER KULTURBESITZ, Germany
  - AYUNTAMIENTO DE GIRONA, Spain
  - Eesti Vabariigi Kultuuriministeerium, Estonia
  - KUNGLIGA BIBLIOTEKET, Sweden



# Project Concept



- ❑ 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, the correct **implementation of standards** cannot be guaranteed:
  - The software used for the production of the electronic files is not in control neither by the institutions that produces them nor by the memory institutions;
  - Conformance tests of transfers are done by memory institutions, but are not totally reliable; different software for testing could end up in different results.
- ❑ This poses problems in **long-term preservation**. Data objects meant for preservation, passing through an uncontrolled generative process, can jeopardise the whole preservation exercise.



# Project Aim and Objectives



- ❑ **The aim:** to implement good quality files in various standard formats for preserving digital content long-term.
- ❑ **The main objective:** to give memory institutions full control of the process of conformity tests of files to be ingested into archives.
- ❑ **The main objective of the PCP launched by PREFORMA:** to develop an *open source software* for the management of the whole conformance test process, supporting a range of standards, addressing the needs of any memory institution or other organisation with a preservation task.



# Open Source approach



- ❑ PREFORMA is following an **open source approach**, with the aim of establishing a sustainable research and development community comprising a wide range of contributors and users from different stakeholder groups.
- ❑ The open source nature **ensures long-term availability of the software**, beyond the memory institutions and suppliers involved in PREFORMA.
  
- ❑ Licenses
  - All **software** developed during the PREFORMA project will be provided under two specific open source licenses: “GPLv3 or later” and “MPLv2 or later”.
  - All **digital assets** developed during the PREFORMA project will be provided under Creative Commons CC-BY v4.0, and in open file formats.



# 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
  
- ❑ Establish an ecosystem around an **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



# Target users and stakeholders

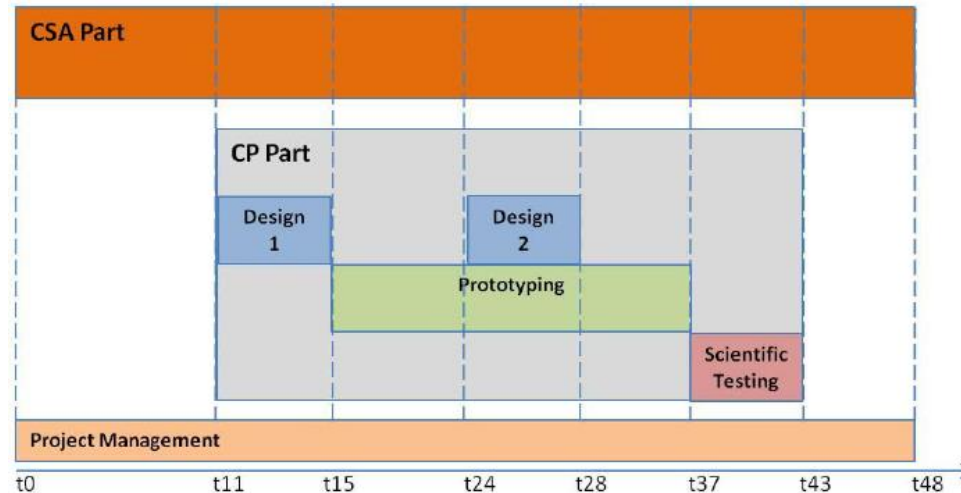


- ❑ **Memory institutions** and cultural heritage organisations, involved in (or planning) digital culture initiatives.
- ❑ **Developers** contributing code for the PREFORMA open source tools.
- ❑ **Research organisations** providing technical advice to cultural stakeholders.
- ❑ **Standardisation bodies** 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.
- ❑ **Any other organisation** planning for long-term preservation of digital content.





# 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



# PREFORMA Suppliers in the Prototyping Phase



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)



# Requirements for Long-term Sustainability of Open Source Communities



To achieve long-term sustainable open source communities of relevance for memory institutions and other stakeholder groups, contracted organisations need to successfully manage and conduct a number of fundamental activities:

- **effective communication** of long-term vision of goals and plans for how the project will evolve over time;
- use of an **open collaboration platform** (such as GitHub) and use of open source tools with associated work practices;
- provision of **source code** under **clear licensing and IPR conditions**;



# Requirements for Long-term Sustainability of Open Source Communities



- **provision of executables** - there shall always be executables for several different deployment platforms;
- **provision of build environment and its source code** - the specific tool chain used for creation of a running instance of the open source code (the build environment) must be provided under an open source license, i.e. a license approved by the Open Source Initiative ([www.opensource.org](http://www.opensource.org)).
- **long-term sustainability of a technical specification of a file format** - promoted through implementation in open source software for which there is transparent information on how the specification of the file format has been interpreted.



# PREFORMA Open Source Projects Provided via Three Channels



1. Development releases provided via an open collaboration platform (GitHub)
  - all code maintained and openly available
2. Stable (monthly) releases of source code, executables, and build environment available at the Open Source Portal for easy download and use under the PREFORMA licenses.
  - all code maintained long-term (after being provided by the suppliers)
3. Releases of developed software provided via each supplier's own channels



**Today: Open Source Preservation Workshop** in Stockholm, in connection with the first Prototype Demonstration.

## Forthcoming:

- **Experience Workshop** that will take place in Berlin in November 2016, in connection with the second Prototype Demonstration
- **Final Conference** that will take place in Stockholm in the end of 2017 to present the results of the project.