

# PREservation FORMAts for culture information/e-archives

## ***PREFORMA Future Memory Standards***

*Peter Pharow  
PREFORMA Design Phase Leader*

*Fraunhofer Institute for Digital Media Technology IDMT,  
Ilmenau, Germany*

*On behalf of the entire consortium*

# Project Identity Card



- ❑ PREFORMA is a **Pre-Commercial Procurement** project co-funded by European Commission under FP7-ICT Programme.
  
- ❑ **Start date:** 1 January 2014
- ❑ **Duration:** 48 month (end date: 31 December 2017)
- ❑ **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 (5)**
  - 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 (9)**
  - 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



# Pre-Commercial Procurement



- ❑ Pre-Commercial Procurement (PCP) is a competition-like procurement method
- ❑ Enables public sector bodies to engage with innovative businesses and interested parties in development projects
- ❑ Arrive at innovative solutions that address specific public sector challenges and needs.
- ❑ Solutions created through a phased procurement of development contracts to reduce risk.
- ❑ PCP more and more common within public sectors of EU.



# 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 **implementation of standards** cannot be guaranteed.
  - **Software** used for implementing standards 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 **aren't totally reliable**; different software for testing ends up in **different** results.
  
- ❑ This poses problems in **long-term preservation**. Data objects meant for preservation, passing through an **uncontrolled** generative process, can **jeopardize** the whole preservation exercise.



# Aim and Objectives



- ❑ **The aim:** to address the **challenge of implementing** various good quality standardized file formats for preserving data content in the 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 and deploy an open source software licensed **reference implementation for various file format standards**, aimed for any memory institution (or other organization with a preservation task) that wish to check conformance with a specific standard.



# Open Source approach



- ❑ PREFORMA is following an **open source approach**, with the aim of establishing a sustainable research and development community comprising of a diverse selection of contributors and users from different stakeholder groups.
- ❑ The open source nature **ensures long-term availability** of the software, independent of the memory institutions and suppliers involved in PREFORMA.
  
- ❑ Licenses
  - All **software** developed during the PREFORMA project will be provided under the two specific open source licenses: “GPLv3 or later” and “MPLv2 or later”.
  - All **digital assets** developed during the PREFORMA project will be provided under the open access license, i.e. Creative Commons CC-BY v4.0, and in open file formats (SOU 2009:86).



# Target users and stakeholders



- ❑ **Memory institutions** and cultural heritage organisations coordinating or representing them, that are involved in (or planning) digital culture initiatives.
- ❑ **Developers** contributing code for the PREFORMA tools as well as developers implementing the reference implementations in production software.
- ❑ **Research organisations** providing technical and domain expertise 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 culture, e-Infrastructures and policy arenas.



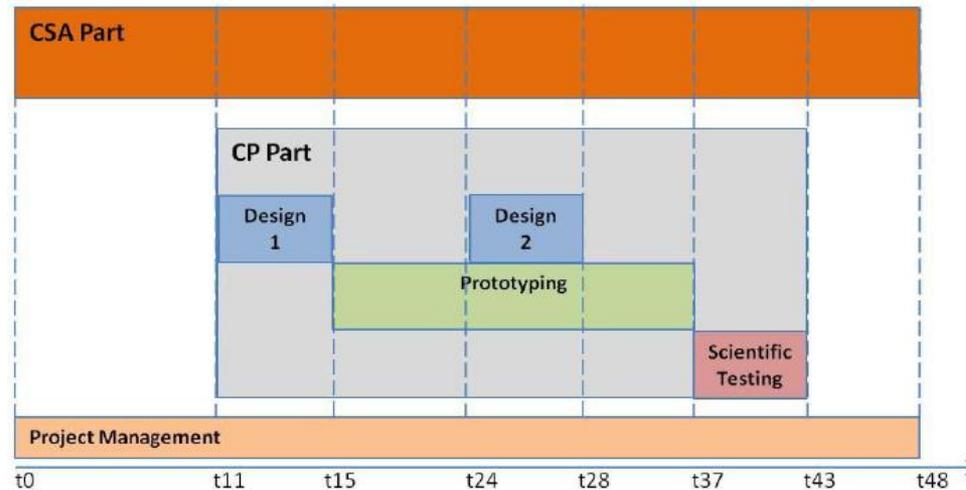
# Challenge Brief



- ❑ 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
  - performs simple fixes
  - reports back to human and software agents
  
- ❑ 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



# Project implementation schedule



- ❑ **Total budget for the procurement:** 2.805.000 EUR
- ❑ **Design phase 1** (4 months): November 2014 – February 2015
- ❑ **Prototyping phase 2** (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 3** (6 months): January 2017 – June 2017

# Preservation Formats



**MXF | MPEG | IMX | XDCAM HD422 |  
DPX | DCP | JPEG2000 | MOV |  
MPEG2 | AVI | MPEG4 | AVC | PDF  
1.4 | PDF/A1 | TIFF 6.0 | JPEG | RAW  
| AS-07 | MPEG|AF | PDF | MKV |  
FFV1 | OGG | Dirac | PNG | WebM |  
VP8 | OGG | Theora | PDF/A2 |  
PDF/A3 | LPCM**



# PREFORMA File Formats



Content type	Standard specifications
TEXT	ISO 32000-1:2008 (PDF 1.7) ISO 19005-1:2005 (PDF/A-1) ISO 19005-2:2011 (PDF/A-2) ISO 19005-3:2012 (PDF/A-3)
IMAGE	ISO 12234-2:2001 (TIFF/EP) ISO 12369:2004 (TIFF/IT)
MOVING IMAGE	OGG MKV  ISO 15444-1 (JPEG2000 core coding system) FFV1 Dirac  LPCM



# Lessons Learned: Iteration



PREFORMA

- ❑ 16 submissions received on time, 10 passed threshold process
  
- ❑ The procedure to evaluate the bids consisted of four different phases:
  - Phase 1: Individual evaluations (3 per submission, experts)
  - Phase 2: Consolidated reports (1 per submission)
  - Phase 3: Requests for clarification (in case of open questions)
  - Phase 4: Hearings (in order to clarify remaining open issues)
  
- ❑ The procedure was completed by a Consensus Meeting:
  - Final ranking of the proposals (in general and per file type)
  - Decisions taken and suppliers informed
  
- ❑ Negotiations started, completion expected today



# Lessons Learned: Data



## ❑ **Training data**

Comprises data set for extent and scope of digital objects acquired/ingested by the memory institution. Compiled and used only by suppliers for training software components but known to project partners.

## ❑ **Test data**

Comprises data set for extent and scope of digital objects acquired/ingested by the memory institution. Used by suppliers and project partners for testing the tools.

## ❑ **Demonstration data**

Comprises data set with a selection of files for digital objects acquired/ingested by the memory institution. Includes valid/conform files as well as corrupted files. Used by suppliers and project partners (and third parties) for demonstrating and endorsing.



# Lessons Learned: UCD

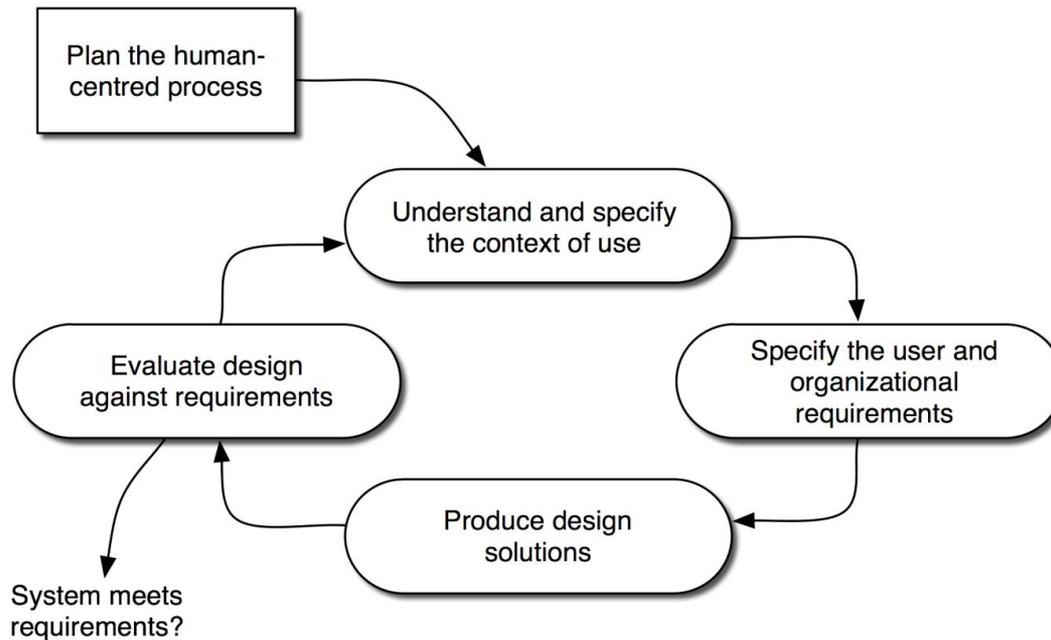
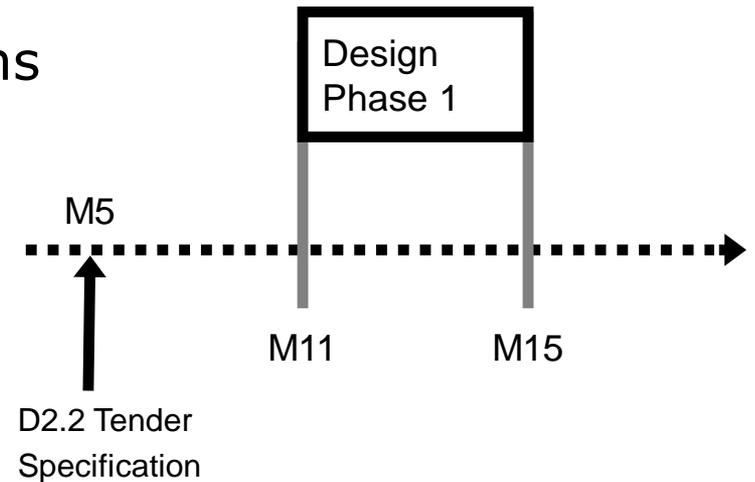


Figure: The human-centered design process (adopted from ISO 13407)

# Design Phase 1 (M11-M15)



- ❑ Six suppliers selected by the tender procedure are required to develop an appropriate design for the expected PREFORMA software
- ❑ Input = D2.2 Tender Specifications
  - M12: Functional requirements (function of the software and/or its components)
  - M14: Technical requirements (technologies, standards, specifications, etc. to be used)



- Requirements serve as basis for the suppliers' own in-depth req. analysis



# Design Phase 1 (M11-M15)



- Workshop is planned in M15 (March 6<sup>th</sup>, Brussels)
  - Presentation of the suppliers' functional + program specifications and software architecture documentations
  - 40 min presentation + 20 min discussion (challenge)
  - Central location for all suppliers to easily get access to
  
- Results serve as additional input for next phase
- Results serve as basis for selecting 3 winners
- Results serve as basis for new bid, negotiation



# Design Phase 2 (M23-M26)



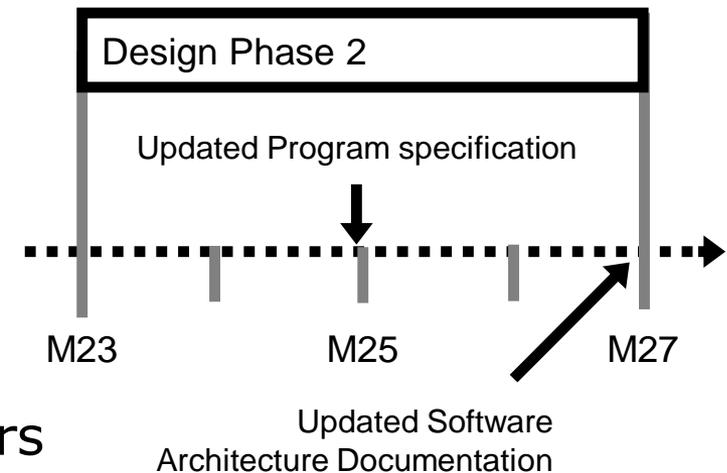
## □ Tasks to be done by the suppliers:

1. M23 – M26: Update of the initial design

➤ WP lead supports the suppliers in meeting the report notes, comments, recommendations, etc.

– Output = Updated program specification (end of M24)

– Output = Updated software architecture documentation (end of M26)



## □ Second workshop allowing suppliers presenting / defending their design

# Main public events



1. A **Training event for Open Source companies** that will take place in Stockholm in December 2015, in connection with the first Prototype Demonstration.
2. An **Experience Workshop** that will take place in Berlin in December 2016, in connection with the second Prototype Demonstration, where the PREFORMA partners will share with memory institutions their experiences of working with suppliers under R&D services agreements.
3. A **final conference** that will take place in Stockholm in December 2017 to present the results of the project.



# Follow us!

## PREFORMA Website

[www.preforma-project.eu](http://www.preforma-project.eu)



A screenshot of the PREFORMA website homepage. The page features a navigation menu with links for WEBSITE, PROJECT, PARTNERS, TENDER, ACTIVITIES, OPEN SOURCE PORTAL, COMMUNITY, DOWNLOAD, and CONTACTS. The main content area includes a "PRESENTATION OF THE PROJECT" section with a diagram showing a stack of documents being processed into a single document with a smiley face. Below this is a section titled "PREFORMA, FUTURE MEMORY STANDARDS" which describes the project's goal of addressing the challenge of implementing good quality standardised file formats for preserving data content in the long term. The page also lists coordinators (Riksarkivet), technical coordinators (Promoter), and partners (PACKED, Fraunhofer IDMT, University of Skövde, University of Jyväskylä). A sidebar on the left contains media partner information (Digital Culture) and relevant news from Digitalmeetsculture.

A screenshot of the PREFORMA website project page. The page features a navigation menu with links for HOME, PROJECT, PARTNERS, TENDER, ACTIVITIES, OPEN SOURCE PORTAL, COMMUNITY, DOWNLOAD, and CONTACTS. The main content area includes a section titled "PREFORMA, FUTURE MEMORY STANDARDS" which describes the project's goal of addressing the challenge of implementing good quality standardised file formats for preserving data content in the long term. The page also lists coordinators (Riksarkivet), technical coordinators (Promoter), and partners (PACKED, Fraunhofer IDMT, University of Skövde, University of Jyväskylä). A sidebar on the right contains information about upcoming events, including the "PREFORMA @ EUROMED INTERNATIONAL CONFERENCE 2014" in Limassol, 3-8 November 2014.

## PREFORMA Blog

[www.digitalmeetsculture.net/projects/preforma/](http://www.digitalmeetsculture.net/projects/preforma/)

EUROMED Conference – Limassol, 3 November 2014



# Thank you!

*Peter Pharow  
PREFORMA Design Phase Leader*

*Fraunhofer Institute for Digital Media  
Technology IDMT, Ilmenau, Germany*

*[phw@idmt.fraunhofer.de](mailto:phw@idmt.fraunhofer.de)*