

# PREservation FORMAts for culture information/e-archives

*Tallinn 11th of October 2017*

*Comformance checking and file format  
validation in PREFORMA*

*Börje Justrell*

# 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: Börje Justrell, Riksarkivet, [borje.justrell@riksarkivet.se](mailto:borje.justrell@riksarkivet.se)
  - Technical Coordinator: Antonella Fresa, Promoter Srl, [fresa@promoter.it](mailto:fresa@promoter.it)
  - Innovation Manager: Claudio Prandoni, Aedeka Srl, [prandoni@aedeka.com](mailto:prandoni@aedeka.com)



# Project Partners



- ❑ RIKSARKIVET, Sweden
- ❑ PROMOTER SRL, Italy
- ❑ AEDEKA SRL, Italy

**Project Coordinator and a memory institution**  
**Technical and Communication Coordinator**  
**Innovation and sustainability issues**

## ❑ **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 GHNIOMHAIREACHT BAINISTIOCHTA RIALTAIS AITIUIL, Ireland
- STIFTUNG PREUSSISCHER KULTURBESITZ, Germany
- AYUNTAMIENTO DE GIRONA, Spain
- Eesti Vabariigi Kultuuriministeerium, Estonia
- KUNGLIGA BIBLIOTEKET, Sweden

**PREFORMA General Presentation**



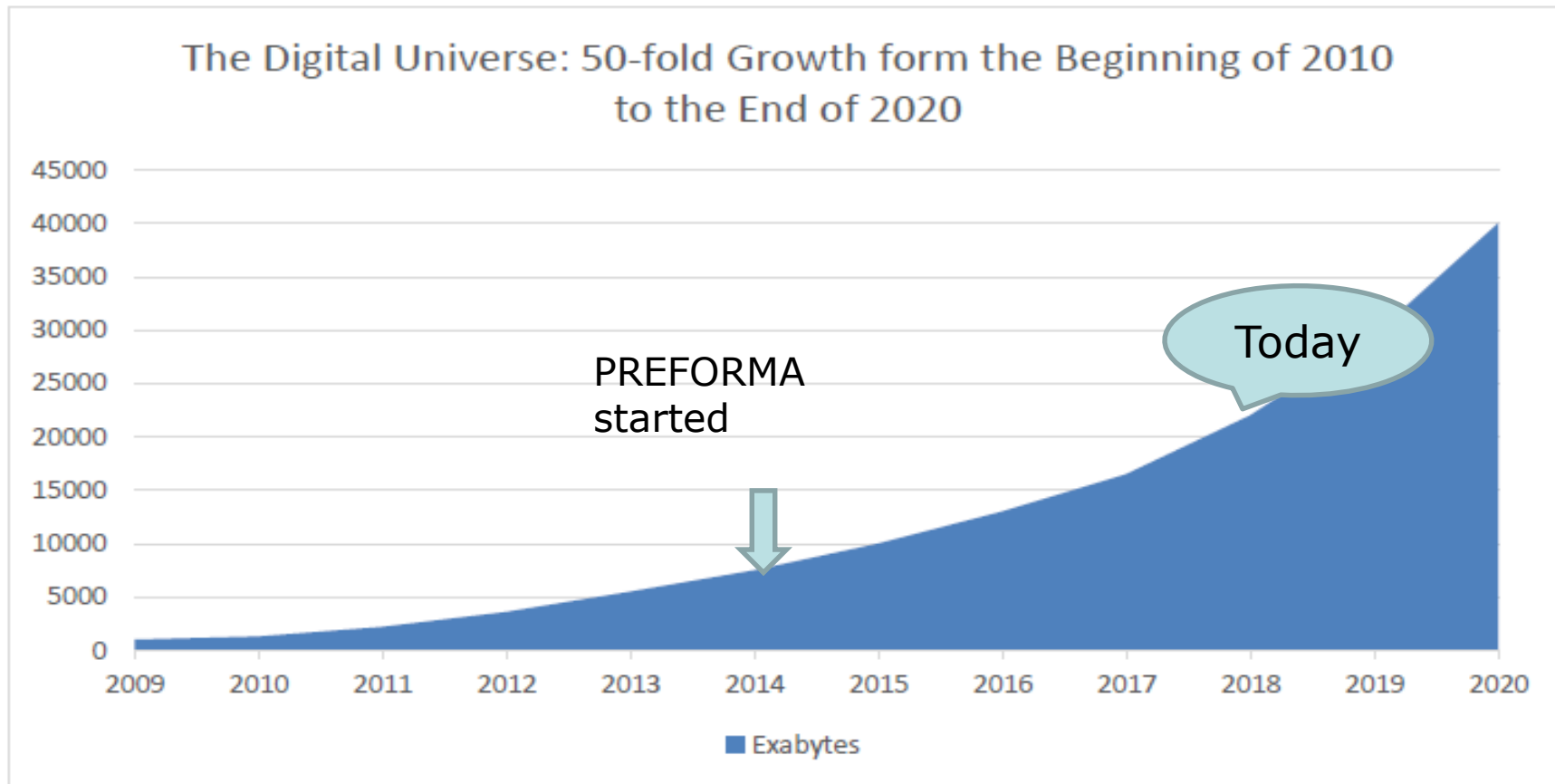
# Project Concept



- ❑ Memory institutions are facing **increasing transfers** of electronic documents and other digital 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 jeopardize the whole preservation exercise.



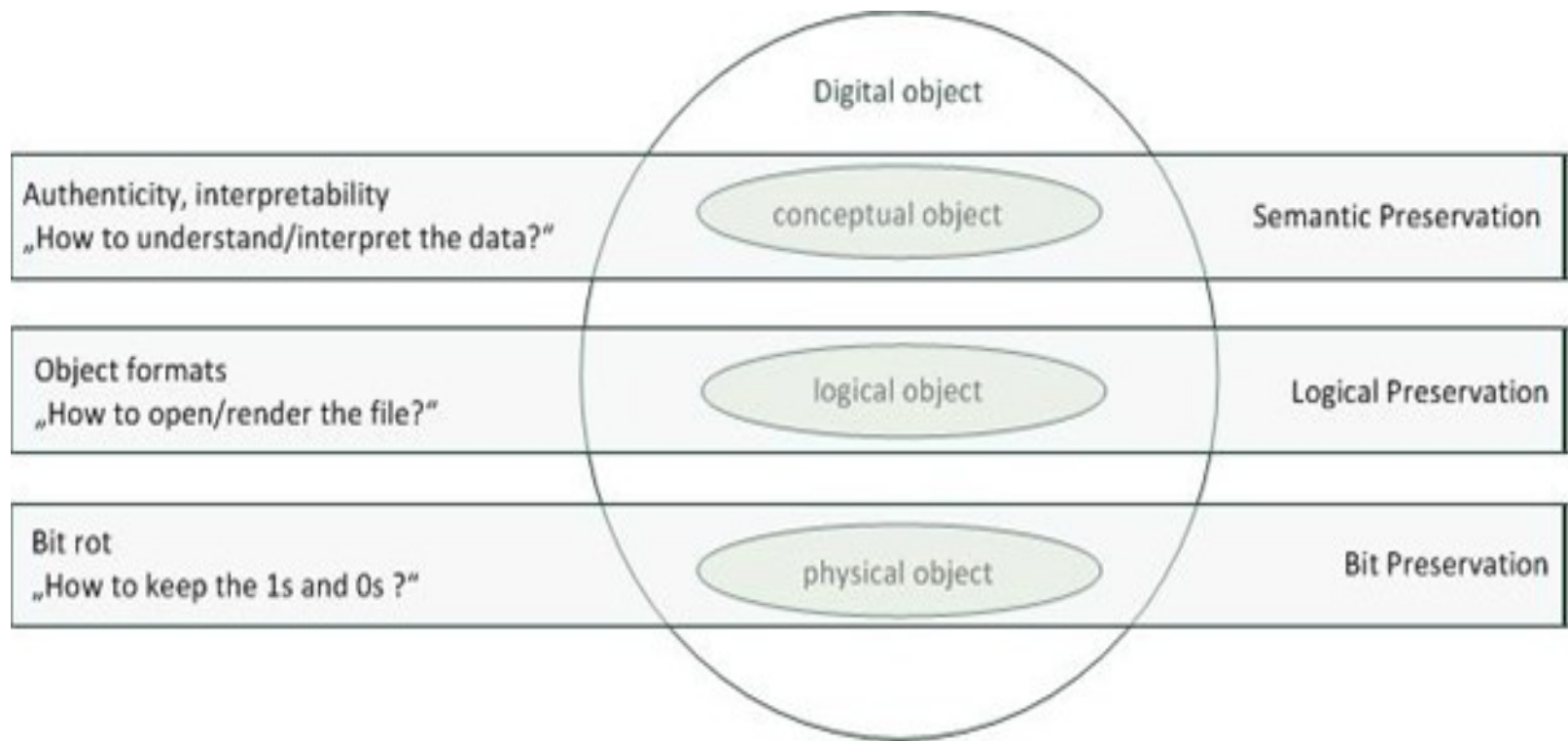
# Project Context



Source: IDC's Digital Universe Study, sponsored by EMC, December 2012  
PREFORMA General Presentation



# Project Context



# Project Aim and Objectives



- ❑ **The aim:** to implement good quality files in various standard formats for preserving 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 organisations 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.





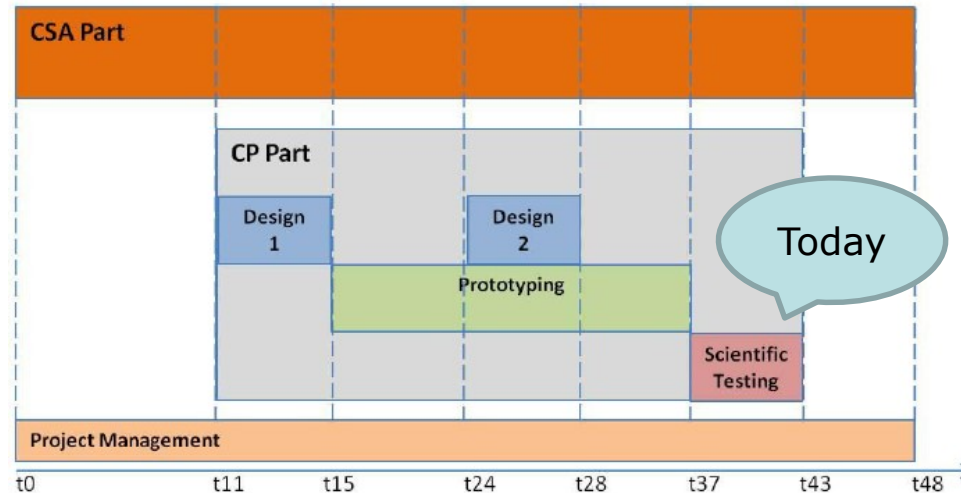
# Target users and stakeholders



- ❑ **Memory institutions** and cultural heritage organisations, involved in (or planning for) 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** (23 months): March 2015 – January 2017
  - First prototypes: March 2015 – October 2015
  - Re-design: November 2015 – February 2016
  - Second prototypes: March 2016 – January 2017
- ❑ **Testing phase** (6 months): February 2017 – July 2017



# PREFORMA Suppliers in the Design phase



1. veraPDF Consortium (led by Open Preservation Foundation and PDF Association) – The PDF/A conformance checker accepted industry-wide (PDF/A)
  2. Preservica - Digital Preservation Validation Framework (PDF/A)
  3. EasyInnova – Digital Preservation Formats Manager (TIFF)
  4. LIBIS - LIBIS/Aware checker for TIFF (TIFF)
  5. MediaArea – PREFORMA MediaConch - CONformance CHECKing for audiovisual files (MKV|FFV1|LPCM)
  6. UCL - OpenMediaCheck (MKV|JPEG2000|LPCM)
- In red those who have been selected for the Prototyping phase and the Testing phase



# Challenge Research and Development (R&D)



Empower memory institutions to gain full control over the technical properties of digital content intended for long-term preservation.



# Challenge R&D Strategy



- ❑ 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 (policy checking)
  - reports back to human and software agents
  - performs 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



- TEXT (strengthen the consensus)
  - 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 (improve the consensus)
  - ISO 12234-2:2001 (TIFF/EP)
  - ISO 12369:2004 (TIFF/IT)
- MOVING IMAGE (virgin path...)
  - MKV
  - FFV1
  - LPCM

# Open Source Conformance Checkers Deployment



The Conformance checkers are all available both for **local use** and as a **web-based application**

They allow for deployment in different infrastructures and environments

- as a **stand alone executable**
- as a **client-server application** to be deployed in network-based solutions
- as a **plug-in to be integrated in third party systems** (legacy or future ones) via API.



# Follow us!

## PREFORMA Website

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



A screenshot of the PREFORMA website homepage. The page is dated Thursday, 23 April 2015. The navigation menu includes WEBSITE, PROJECT, PARTNERS, TENDER, ACTIVITIES, OPEN SOURCE PORTAL, COMMUNITY, DOWNLOAD, and CONTACTS. The main content area features a "PRESENTATION OF THE PROJECT" section with a sub-header "PREFORMA, FUTURE MEMORY STANDARDS". Below this, there is a "MEDIA PARTNER" section for Digital Culture and a "RELEVANT NEWS" section with two articles: "Joint Conference on Digital Libraries 2015" and "International Conference on Digital Preservation (IPRES) 2015". The right sidebar lists the "COORDINATOR" (Riksarkivet), "TECHNICAL COORD." (PROMOTER), and "PARTNERS" (PACKED, Fraunhofer IDMT, University of Skovde, and University of Padova).

A screenshot of the PREFORMA Open Source Portal. The page features the PREFORMA logo, the European Union flag, and the Seventh Framework Programme logo. The navigation menu includes HOME, PROJECT, PARTNERS, TENDER, ACTIVITIES, OPEN SOURCE PORTAL, COMMUNITY, DOWNLOAD, and CONTACTS. The main content area has a large heading "VISIT THE OPEN SOURCE PORTAL" with a sub-heading "Give your contribution to the prototyping phase". To the right, there is a section for "UPCOMING EVENTS" titled "DAVID - FINAL USER-TEST-WORKSHOP" for April 28-29, 2015, with a photograph of a workshop. Below this, there is a section for "PREFORMA, FUTURE MEMORY STANDARDS" with a brief description of the project and its aim.

## PREFORMA Blog

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

