DCH-RP and PREFORMA
Two case studies on the digital preservation of cultural heritage

Antonella Fresa

Technical Coordinator of DCH-RP & PREFORMA

Promoter Srl

DIPP 2015, Veliko Tarnovo, 29 September 2015

www.digitalmeetsculture.net

Introduction

- In the last decades, **cultural institutions** started to move their cultural content on the **digital world**. This has implied to develop new instruments, new rules of access, new standards to exhibit digital heritage, new ways of communication.

- The **amount** of digital cultural content is now so valuable that the issue of **preserving** the digital cultural heritage is becoming as much urgent as the preservation of tangible heritage.

- The digital cultural heritage sector has the challenge of the **complexity** of the information itself, due to the relationships that each cultural object has with its collection, with the memory institutions where it is held, with the other objects of the same nature and/or culturally interconnected, etc.

Digital Preservation: Organisational aspects

❑ Each digitisation programme is addressing the issue of preservation in a separate manner, while a **shared implementation** of common e-Infrastructure layers could be beneficial and cost effective to all.

❑ The **national policies** about cultural preservation (digital and tangible) needs to share and re-use best practices, procedures and workflows to avoid duplication of efforts and to reduce the cost.

➢ The use of the **e-infrastructures** is a pillar in this direction and the CH sector should progress towards its full integration in the new concept of open science.

➢ Implementation of the **preservation services** of such infrastructure is among the first priorities for the DCH sector.

Digital Preservation: Technological aspects

- ❑ **Migration** is the widest digital preservation strategy used by memory institutions. To avoid technical obsolescence, the digital objects are converted to new standardised file formats as technology changes. These conversions are expected to occur without information losses.

- ❑ 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 as 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, but instead fully controlled.

- ➢ It is of vital importance to fully **control the correct implementation of the standard**.

www.digitalmeetsculture.net

Two projects

Organisation

- **DCH-RP** was a Coordination and Support Action co-funded by the European Commission under FP7-ICT Programme.
- **Duration**: 24 month (1 October 2012 - 30 September 2014, SUCCESSFULLY COMPLETED)
- **Partners**
  - Central Institute for the Union Catalogue of Italian Libraries and for Bibliographic Information, Project Coordinator
  - Promoter Srl, Technical Coordinator
  - EDItEUR Ltd
  - National Archives of Sweden
  - Estonian Ministry of Culture
  - Collections Trust
  - National Institute of Nuclear and Particle Physics
  - Poznań Supercomputing and Networking Center
  - Belgian Science Policy Office
  - National Information Infrastructure Development Institute
  - TERENA National Research and Education Networks
  - EGI.eu European Grid Initiative
  - Michael Culture Association

Technology

- **PREFORMA** is a Pre-Commercial Procurement project co-funded by the European Commission under FP7-ICT Programme.
- **Duration**: 48 month (1 January 2014 - 31 December 2017, ONGOING)
- **Partners**
  - National Archives of Sweden, Project Coordinator
  - Promoter Srl, Technical Coordinator
  - Packed Vzw
  - Fraunhofer IDMT
  - University of Skovde
  - University of Padua
  - Netherlands Institute for Sound and Vision
  - Royal Institute for Cultural Heritage
  - Greek Film Centre
  - Local Government Management Agency
  - Prussian Cultural Heritage Foundation
  - Girona City Council
  - Estonian Ministry of Culture
  - National Library of Sweden

www.digitalmeetsculture.net

DCH-RP

digital cultural heritage – roadmap for preservation

a coordination action

www.digitalmeetsculture.net

DCH-RP objectives

❑ To develop a validated **Roadmap** for the implementation of a **preservation infrastructure for DCH**

❑ The vision for the next two decades: "*to implement a federated infrastructure dedicated to support the application of open science in the arts and the humanities*"

❑ This implies:

– Establishing conceptual and operational **coordination** among DCH and e-infrastructure organizations

– Defining a **Roadmap** and a set of **Practical Tools** which would help **monitoring** the progress of the implementation

– Make use of the shared implementation of **common e-Infrastructure layers** (identity federations, federated cloud, data infrastructure services, etc.)

CH and the e-infrastructures: requirements and benefits

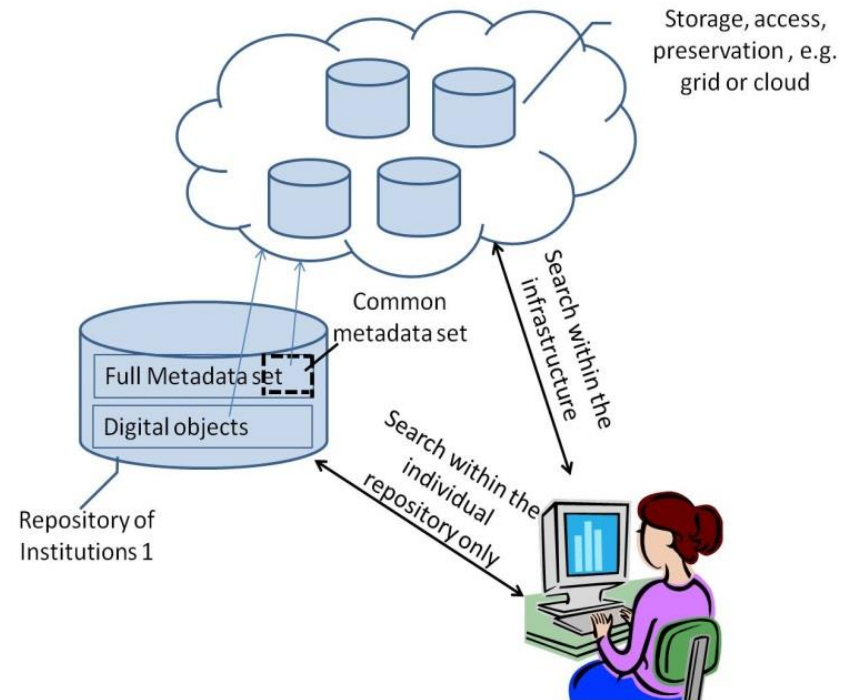| CH requirements	Benefit of the e-infrastructures
To reduce the **digital preservation costs**, avoiding separate solutions, implemented each time by each memory institution.	**Cost reduction** in digitization, cataloguing and metadata generation, substituting expensive human workforce with cheaper machine processes
To reduce the high investment due to the need of **human intervention** of experts.	**Permanent identification** of digital cultural objects and providers
To manage **different formats**: 2D images, 3D models, metadata, publications, digital exhibitions, virtual reconstructions, etc.	**Storage** and **preservation** facilities, ranging from short- medium- and long-term
	Search facilities to manage **semantic search** and **linked open data**
To support **contextual data**, which are very important for cultural research.	Computing resources to enhance **processing** and **visualisation** of **complex** cultural data (e.g. 3D modelling and VR representations)
Even if growing very rapidly, the actual **size** of DCH content is still rather **small** (if compared for example to the amount of data produced by other scientific domains).	Dynamic distributed virtual organisations, facilitating **collaboration** with **information** and **resource sharing** (e.g. virtual conferences, document sharing, blog and cooperation platforms, etc.)
The digitisation process is unique and **cannot be automatically replicated**	**Standardisation** in the data world, e.g. by developing a common reference model for the DCH sector

www.digitalmeetsculture.net

DCH-RP target users

- **Cultural institutions** that own digital cultural repositories (e.g. museums, libraries, archives)
- **Research** (e.g. researchers, scholars, teachers, students)
- **Teaching and learning** (e.g. schools, training centres, university courses)
- **Cultural and creative industry** interested to use and re-use DCH content
- **Private archives**, providing commercial access to content

Storage, access, preservation, e.g. grid or cloud

Common metadata set

Full Metadata set

Digital objects

Repository of Institutions 1

Search within the infrastructure

Search within the individual repository only

Characteristics of DCH-RP

❑ Focus on the **storage** phase, which means long, medium and short term actions

❑ Take a practical approach, based on **proofs of concept**
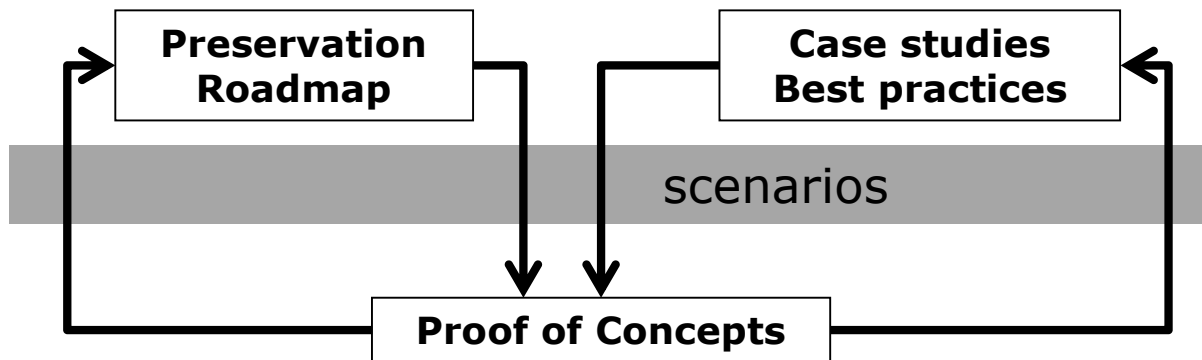
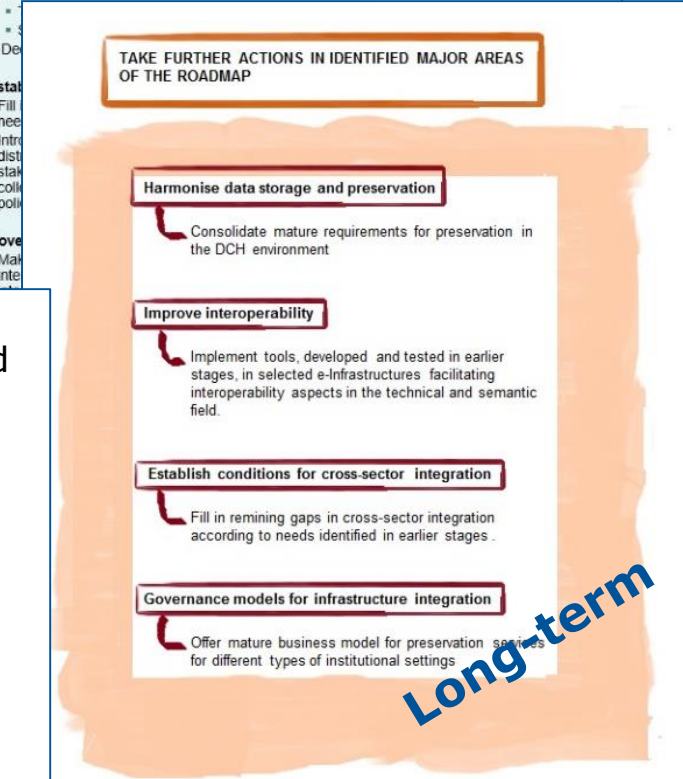❑ Coordinate with **other initiatives**, such as SCIDIP-ES, APARSEN, EUDAT, SCAPE, OPEN AIRE, CHAIN REDS, DARIAH, etc.

www.digitalmeetsculture.net

A Roadmap for long-term preservation of DCH

Short-term
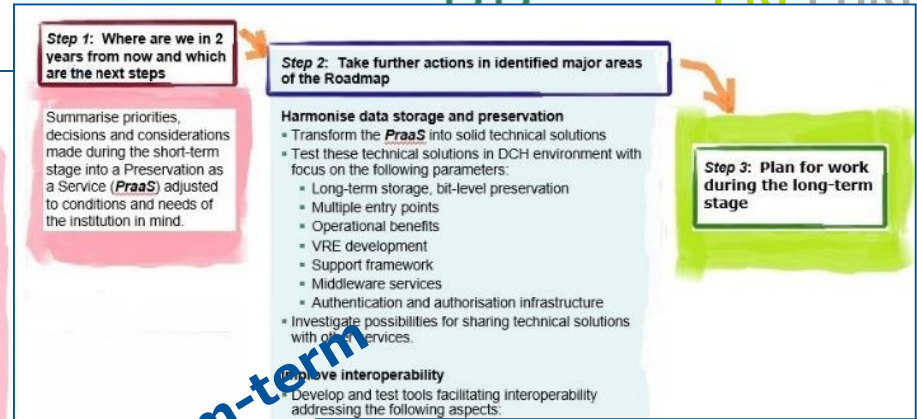
Medium-term

Long-term

❑ **Scope**
  – Help policy makers and programme owners to plan ahead
  – Assist managerial teams of cultural institutions in taking decisions related to digital preservation
  – Support cultural heritage institutions in defining practical action plan with a realistic time frame for its implementation

❑ **Areas**
  – Harmonisation of data storage and preservation
  – Improve interoperability
  – Cross-sector integration
  – Governance models

How to contribute

- In order to maintain the roadmap as a living document, the DCH-RP project created a **dedicated web-space** where it is possible to download the latest version of the roadmap as a handbook and to provide feedback and comments

- This web-space links also to other relevant material, information and services that are linked to the roadmap, such as the **Registry of Services and Tools**, a practical instrument to help decision makers, DCH communities, institutions and projects to plan the implementation of their digital preservation processes

www.digitalmeetsculture.net/heritage-showcases/dch-rp/dch-rp-roadmap-for-preservation

PREFORMA

preservation formats for culture information/e-archives

a pre-commercial procurement project

www.digitalmeetsculture.net

PREFORMA objectives

❑ Project's objectives
  – To implement **good quality files** in various standard formats for **preserving content** in a long term.
  – To give memory institutions **full control** of the process of conformity tests of files to be ingested into archives.

❑ Objectives of the Pre-Commercial Procurement
  – 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 organization with a preservation task.

www.digitalmeetsculture.net

The Open Source Challenge

- ❑ To 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

www.digitalmeetsculture.net

Conformance Checker

- ❑ Verify whether a file has been produced according to the **specifications of a standard file format**.
- ❑ Verify whether a file matches the **acceptance criteria for long-term preservation** by the memory institution.
- ❑ **Report in human and machine readable format** which properties deviate from the standard specification and acceptance criteria.
- ❑ Perform **automated fixes** for simple deviations in the metadata of the preservation file, leaving the original bitstream untouched and created a correct copy of the object to be preserved.

- ❑ At the time of: Creation, Transfer, Digitisation and Migration

Shell
Conf. Checker
Implementation Checker
Policy Checker
Reporter
Metadata Fixer
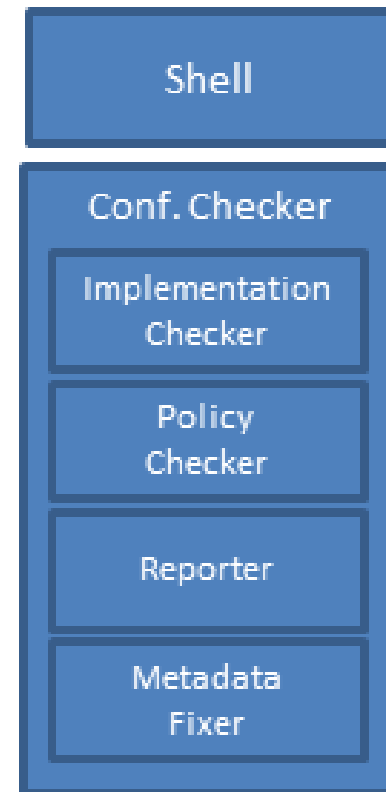
Standard Specifications

Content type	Standard specifications
TEXT	**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) **Ogg** (https://xiph.org/ogg/doc/) **Lossless JPEG2000** (ISO/IEC 15444-1:2004) **Lossless FFV1** http://www.ffmpeg.org/~michael/ffv1.html **Lossless Dirac v2.2.3** (http://diracvideo.org/download/specification/dirac-spec-latest.pdf) **Linear PCM** (IEC 60958-1 ed3.1 Consol. with am1: 2014)

www.digitalmeetsculture.net

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**, and it is open to all memory institutions and service providers, beyond the PREFORMA consortium.

❑ 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 (SOU 2009:86).

www.digitalmeetsculture.net

PREFORMA target users

- ❑ **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.

www.digitalmeetsculture.net

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

www.digitalmeetsculture.net

First results

□ Since July 2015 the **first releases of the Conformance Checkers** are available for download in the PREFORMA **Open Source Portal** http://www.preforma-project.eu/open-source-portal.html

□ An **Open Source Workshop** will take place on **7 April 2016 in Stockholm**, in connection with the first Prototype Demonstration.

How to contribute

❑ PREFORMA software development is carried out in a **collaborative environment** with memory institutions and experts.

❑ We are looking for memory institutions (or other organisations with a preservation task) willing to participate in the testing phase, by:

– **Providing datasets** to be used to test the prototypes that are being developed by the PREFORMA suppliers (including valid/conform files as well as corrupted files)

– **Testing the software** and sharing with us the results of such tests

– Helping us in **identify policies** and **workflows**, to refine and improve the conformance checkers

www.digitalmeetsculture.net

Follow us!

PREFORMA Website & Blog

www.preforma-project.eu

www.digitalmeetsculture.net/preforma
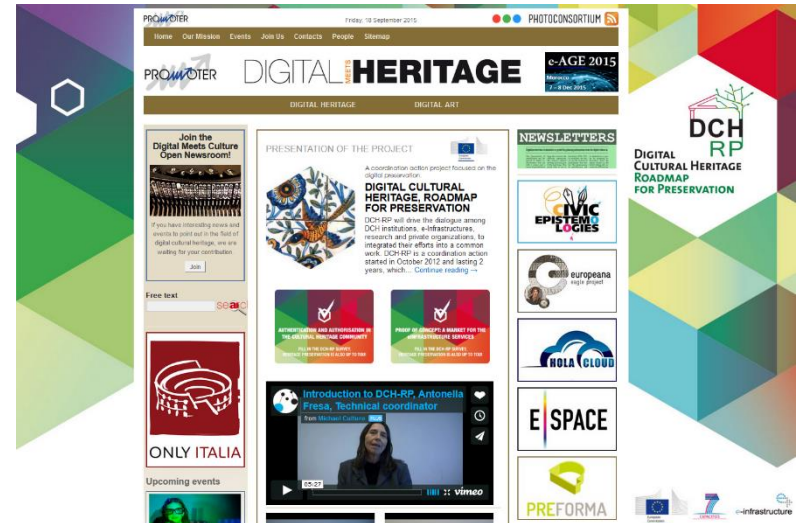
DCH-RP Showcase on Digital Meets Culture

www.digitalmeetsculture.net/dch-rp

www.digitalmeetsculture.net

Thank you!

Antonella Fresa

Promoter Srl

fresa@promoter.it

www.digitalmeetsculture.net

www.digitalmeetsculture.net