



# A Virtual Research Community for the Preservation of Digital Cultural Heritage – Welcome and Introduction

*Antonella Fresa*

*Technical Coordinator of DCH-RP*

*Promoter Srl*

# Agenda



- ❑ 16:50 – 17:00 Welcome and introduction (Antonella Fresa, Promoter Srl)
- ❑ 17:00 – 17:05 EUDAT: last achievements and next steps (Norbert Meyer, Poznan Supercomputing and Networking Center)
- ❑ 17:05 – 17:10 SCAPE and SUCCEED: last achievements and next steps (Hildelies Balk, National Library of the Netherlands)
- ❑ 17:10 – 17:15 APARSEN and SCIDIP-ES: last achievements and next steps (David Giaretta, Alliance for Permanent Access)
- ❑ 17:15 – 17:20 CHAIN REDS: last achievements and next steps (Federico Ruggeri, INFN / GARR, Italy)
- ❑ 17:20 – 17:35 Discussion. Towards a Virtual Research Community for the Preservation of Digital Cultural Heritage: topics and ways of cooperation (use cases, MoUs, etc.), target users, features and services to be provided, interoperability issues

# Objectives of the Session

- ❑ A **network of projects and initiatives** working worldwide in the domain of DCH, digital arts, digital performances, digital humanities, digital preservation and e-infrastructures was born during the [EGI Community Forum in Manchester](#)
- ❑ This session intends to continue the discussion which was carried out in the last months through a series of joint events when we tried to identify **common goals, strategies and synergies** between the projects
- ❑ Now, we would like to progress in setting up a common plan for establishing a **Virtual Research Community** for the preservation of DCH, to be acknowledged at European level by the world of e-infrastructure and to target the opportunities offered by Horizon2020



DCH-RP Workshop the creation of a Virtual Research Community dedicated to the cultural heritage and the humanities

### EGI Community Forum 2013 in Manchester

Between 8th and 12th of April 2013, EGI.eu and UK NGI, a partnership between GridPP and the National e-Infrastructure Service (NES), hosted the Community Forum 2013. The event, which was held in Manchester in conjunction with the 3rd EMI Technical Conference, was very successful and provided a great opportunity to reflect upon the growing diversification in usage of the European Grid Infrastructure from across all research disciplines as well as the widening international collaboration that is taking place. [Continue reading →](#)



Panel discussion chaired by Wim Jansen (DG Connect, EC) at the International Conference on Theory and Practice of Digital Libraries TPDL 2013

### eInfrastructures and Digital Libraries... the Future @ TPDL 2013

Towards an e-infrastructure for the Digital Cultural Heritage sector and the humanities research: which are the needs and the requirements of the DCH? How can this sector benefit from the new and innovative solutions offered by the e-Infrastructures? Experts try to answer these questions. [Continue reading →](#)



DCH-RP workshop organised in the frame of the 2nd EUDAT International Conference

### Digital Preservation of cultural data @ EUDAT, Rome

The need for novel more efficient and affordable solutions for digital preservation is increasing in the Social Science and Humanities. Aim of the workshop is to share and present the advancements in the state of the art, to understand which could be the benefits offered, in particular to the Digital Cultural Heritage sector, by the new technologies and e-infrastructures. [Continue reading →](#)



Networking session at the ICT2013 Conference in Vilnius

### A Virtual Research Community for the Preservation of Digital Cultural Heritage @ ICT 2013, Vilnius

Objective of the session, organised by DCH-RP, SCIDIPES, APARSEN, EUDAT, CHAIN REDS, DARIAH, DASISH and CLARIN, is to bring together projects and initiatives working world-wide in the domain of DCH, digital arts, digital performances, digital humanities and digital preservation to find synergies and discuss opportunities for cooperation, in particular around the theme of using e-infrastructures for the preservation of DCH. [Continue reading →](#)

# CH and the e-infrastructures



CH needs and requirements	Benefit of the e-infrastructures
High investment in the production of DCH data due to the need of <b>human intervention</b> of experts.	To allow for <b>cost reduction</b> in digitisation, cataloguing and metadata generation by substituting expensive human workforce with cheaper machine processes
High <b>costs of digital preservation</b> , due also to the use of separate solutions, mostly implemented each time by each memory institution.	To support the <b>permanent identification</b> of digital cultural objects and providers
Several different kind of information, <b>different formats</b> : 2D images and 3D models, metadata, publications, digital exhibitions, virtual reconstructions, etc.	To facilitate <b>storage</b> and <b>preservation</b> , ranging from short- medium- and long-term
	To improve <b>search facilities</b> to manage <b>semantic search</b> and <b>linked open data</b>
<p><b>Contextual data</b> are complex and very important for cultural research.</p>	To enhance <b>processing</b> and <b>visualisation</b> of <b>complex</b> cultural data (e.g. 3D modelling and VR representations) through the computing resources offered by research e-infrastructures (grid, cloud)
Even if growing very rapidly, the actual <b>size</b> of this content is still rather <b>small</b> (if compared for example to the amount of data produced by other scientific domains).	To enable dynamic distributed virtual organisations, facilitating <b>collaboration</b> with <b>information</b> and <b>resource sharing</b> (e.g. virtual conferences, document sharing, blog and cooperation platforms, etc.)
The digitisation process is unique <b>cannot be replicated</b> unless the whole work is done from scratch.	To contribute to <b>standardisation</b> in the data world, e.g. by developing a common reference model for the DCH sector

# An infrastructure dedicated to the cultural heritage



- ❑ The longer term plan is to implement a dedicated CH infrastructure able to provide **distributed safe storage, seamless powerful access**, effective **preservation services** to a “continuum” of data coming from the largest number of small and big cultural institutions
- ❑ This infrastructure should be able to **aggregate and interoperate** existing features as well as services that are under development, often (unfortunately) following separate ways
- ❑ **Target users:**
  - **Cultural institutions** that own digital cultural repositories (e.g. museums, libraries, archives)
  - **Individuals** (e.g. researchers, scholars, teachers, students)
  - **Teaching and learning** bodies (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



# DCH-RP: first outcomes



- ❑ A **registry of Services** was implemented and published online
- ❑ The first **Proofs of Concept** is just been completed: cultural institutions experimented the actual use of distributed computing and storage infrastructures (such as the **e-Culture Science Gateway**: <http://ecsg.dch-rp.eu/>) to store and manage cultural digital resources, starting from real scenarios and use cases
- ❑ The first version of the **Roadmap** will be released next month

**PROMOTER DIGITAL CULTURE** Tuesday, 05 November 2013, 11:25 UTC

Home Our Mission Events Register / Edit Profile Contacts People Sitemap

**PROMOTER DIGITAL HERITAGE MEETS DIGITAL ART PROMOTER**

DIGITAL HERITAGE DIGITAL ART

**DCH RP** DIGITAL CULTURAL HERITAGE RESEARCH PROJECT FOR PRESERVATION

This registry is maintained by [DCH-RP project](#)

**Category**

- 2 CMS/webpage
- 75 file processing tool
- 26 framework/distributed system/client-server
- 4 other

**Content type**

- 4 audio
- 2 database
- 1 databases, documents
- 3 disk images
- 2 disk images,

**Usage**

- 1 access control
- 1 acquisition
- 1 acquisition, integrity check, conversion, storage, replication, view
- 10 acquisition

**Status**

- 10 in progress
- 101 maintained
- 6 not maintained
- 22 unknown

**Licence type**

- 7 free
- 111 open source
- 17 paid
- 4 unknown

**139** [DETAILS](#) • [SUMMARY](#)

Name	Functionality
AccessToSiard	A collection of scripts to automatically convert MS Access files to the SIARD format.
ACE (Audit Control Environment)	Set of tools to help archives monitor the integrity of collections. It provides a mechanism to allow a 3rd party to independently verify a collections integrity.
Acrobat XI Pro	Create, modify, convert from/to PDF documents. Convert a Web page or an entire Web site to a single PDF. offline viewing.
AFF	Tools for the creation of disk images, used in conjunction with the AFF open and extensible file format to store disk images and associated metadata
Apache Commons Imaging	library that reads and writes a variety of image formats, including fast parsing of image info (size, color space, ICC profile, etc.) and metadata
Apache Lucene	text search engine library
Apache POI	Java API for Microsoft Documents
Apache Tika	toolkit detects and extracts metadata and structured text, using existing parser libraries
Archive-It	Web archiving service to capture, build, and manage collections of web content.
Archivematica	Digital preservation system that is designed to maintain standards-based (OAIS, DC, METS, PREMIS), long-term access to collections of digital objects. Micro-services design pattern to provide an integrated suite of software tools. Users monitor and control the micro-services via a web-based dashboard.
Archivists Toolkit	Archival data management system. Supports accessioning and describing archival materials; establishing names and subjects associated with archival materials, including the names of donors; managing locations for the materials; and exporting EAD finding aids, MARCXML, METS, MODS, Dublin Core records.
Archon	Platform for archival description and access. Record information about collections and digital objects and view, search, browse that information in a fully-functional public web site. Export MARC and EAD records.
AudioScout	Distributed audio content indexing system. It can index a large collection of audio content for the purpose of later recognition of unknown signals. Robust to noise, different encodings and other types of distortion, based on pHash
AVS Document Converter	Transfer regular text formats to e-Pub format and create e-books. Open and convert such e-book formats as DjVu and FB2 to all key formats supported by AVS Document Converter.
AVS Image Converter	Convert images between JPEG, PDF, RAW, TIFF, GIF, PNG, RAS, PSD, PCX, CR2, DNG, APNG, etc. Resize, rotate, apply effects, watermark pictures.
b2x Translator	Tools for batch conversion from binary to XML MS Office formats.
BagIt Library and Tools	Creation, manipulation and validation of data files according to BagIt specification
BitBlocker	scan file sharing sites and BitTorrent trackers for copyrighted material, based on pHash

**OAIS stage**

**ingest**

- 23 no
- 116 yes

**data management**

- 108 no
- 31 yes

**archival storage**

- 109 no
- 30 yes

**access**

- 113 no
- 26 yes

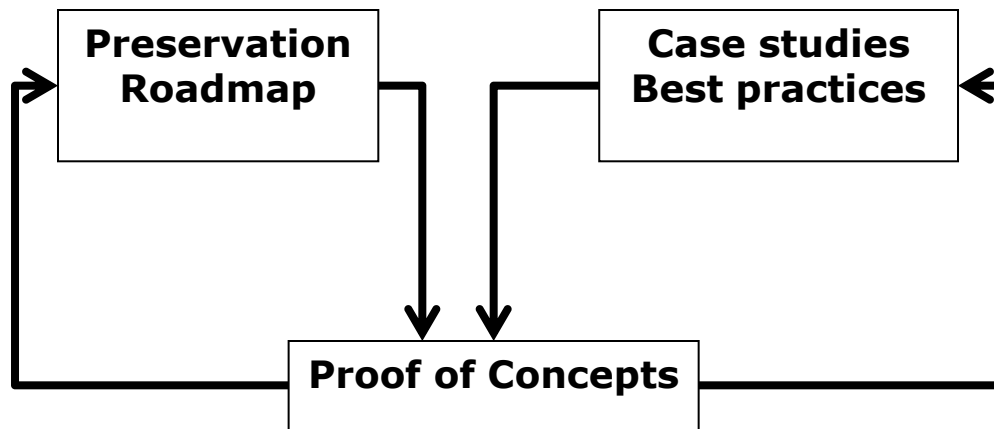
**preservation planning**

- 131 no
- 8 yes

**administration**

- 113 no
- 26 yes

# DCH-RP: first outcomes



- ❑ Cultural data are curated by many different persons: data **management** and **administration** + **user access control** are very important.

- ❑ **Security** of the data is very important for cultural institutions: **trust building** is a key factor when it is not determined where data are stored
- ❑ Functionalities and services offered by e-infrastructures should not **impact** on the outgoing traffic of the institution
- ❑ Access to the e-infrastructure services should be **simple** without requiring IT specialist knowledge



# DCH-RP: what's next?



- For the **second Proof of Concept** DCH-RP is planning to experiment with services offered by other projects (e.g. EUDAT, APARSEN, SCIDIP-ES, SCAPE)



- We are looking for other **new services** to be used with data made available by the cultural institutions participating the second Proof of Concept



# A new challenge: controlling the file formats



- ❑ **Migration** of digital objects into new formats is the most used long-term digital preservation strategy adopted by the memory institutions. In order not to lose information during the conversions, it is of vital importance to fully **control the file formats**.
- ❑ **Metadata** is normally stored in XML and specified in different schemas controlled by the community of professional curators through international organisations.
- ❑ **Data content** are normally stored in specific file formats for documents, images, sound, video etc. which are usually produced by software from different vendors.
- ❑ Even if files are in standard formats, **the implementation of standards cannot be guaranteed** because the software implementing standards for the production of the electronic files is not in control neither by the institutions that produces it nor by the memory institutions

# PREFORMA: Future Memory Standards



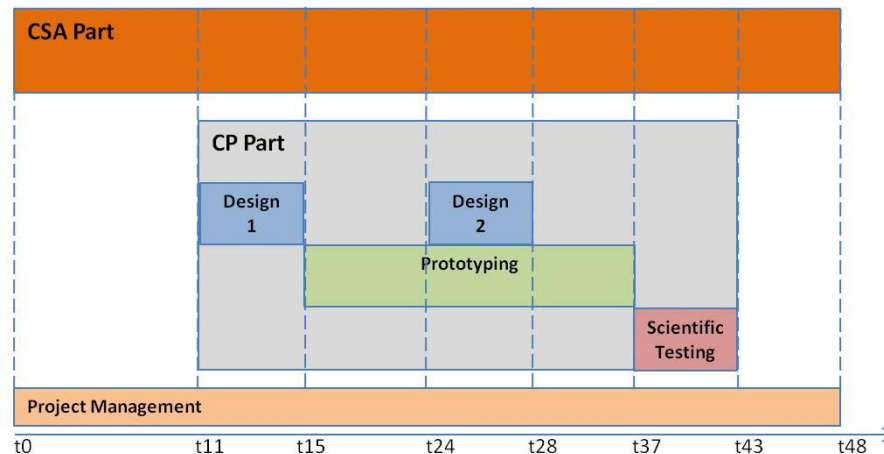
- ❑ **PREFORMA** (under negotiation) will start next year to launch a PCP for the provision of a set of **open source tools** to check the **conformance** of digital archives implemented in the most common **file formats for documents, images and AV content**.
- ❑ The PREFORMA consortium is composed by:
  - **Swedish National Archives** – Project Coordinator
  - **Promoter** – Dissemination and Communication Leader
  - **University of Padua, Fraunhofer, University of Skovde and Packed** – responsible for the technical specifications and the assessment of the tender
  - 9 memory institutions providing requirements and evaluating the results: **Sound&Vision, KIK-IRPA, Greek Film Center, Local Government Management Agency, SPK, Ayuntamiento de Girona, Estonian Ministry of Culture, National Library of Sweden** + Swedish National Archives
- ❑ **R&D suppliers** that will be selected by open tender procedure and they will act as subcontractors of the memory institutions

# Call for tender



PREFORMA

- ❑ **March 2014:** Pre-announcement of the Call
  - ❑ **April 2014:** Info day
  - ❑ **June 2014:** Deadline for submission of the offers
  - ❑ **December 2014:** Selection of the winner suppliers and start of the work
- 
- ❑ **Budget to be spent in the tender: 2.880.000 euro**



# Stay tuned on [www.digitalmeetsculture.net](http://www.digitalmeetsculture.net)



Staring at Culture,  
from the hi-tech point of view

DIGITAL MEETS CULTURE .net

A cutting edge **online magazine**



Shot the QR Code with your smart phone/tablet, get connected with [www.digitalmeetsculture.net](http://www.digitalmeetsculture.net)!

**Digital Meets Culture** is an interactive online magazine where digital technology and culture collide. It provides, on a truly global scale:

- the widest information about **international projects** and initiatives for the digitization, preservation, and enjoyment of digital cultural heritage;
- tales and talks, as well as **amazing multimedia** contents, for a broad perspective of the latest in digital art;
- Up to date **showcases** of the ongoing most relevant acts in the field.

Digital Meets Culture counts on a **steadily growing visitors number**, actually over 10,000 per month, from all over the world and high-profiled on heritage digital preservation, digital art, humanities computing, cloud computing. That's why **it is critical to be showcased** on Digital Meets Culture, either for business or research dissemination purposes.

Digital Meets Culture is proudly edited and published by an international experts team at **Promoter SRL**, since the 90's **European leader** in the ICT project management and business promotion.



[www.digitalmeetsculture.net](http://www.digitalmeetsculture.net) | [digitalmeetsculture@promoter.it](mailto:digitalmeetsculture@promoter.it) | +39 0587 466881  
PROMOTER SRL, ICT Research & Innovation | via. Bocconi 2, I-56037 Peccioli, Pisa | [www.promoter.it](http://www.promoter.it)





# Thank you!

*Antonella Fresa*

*Promoter Srl*

[fresa@promoter.it](mailto:fresa@promoter.it)

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