Document Distribution Note
Release to Members: Immediate
Release to Public: 22/03/2015

Digital Cultural Heritage Roadmap Project: Roadmap Conference

Biblioteca Nazionale Centrale, Rome 22nd September 2014

About the event

The Digital Cultural Heritage Roadmap Project (DCH-RP) invited WK to speak at a conference to review and introduce their 'Interim roadmap for digital preservation'. WK had previously attended a review workshop which discussed and examined an early draft of the main deliverable of this 2 year EC-funded co-ordination action which is due to be complete later in 2014. The full text of the draft roadmap is available online at: http://www.dch-rp.eu/getFile.php?id=221 and the final version of the roadmap will shortly be published online.

DCH-RP is a coordination action supported by EC FP7 e-Infrastructures Programme, launched to look at best practice for preservation standards in use. The project aim to harmonize data storage and preservation policies in the digital cultural heritage sector; to progress a dialogue and integration among institutions, e-Infrastructures, research and private organisations; to identy models for the governance, maintenance and sustainability of the integrated infrastructure for digital preservation of cultural content. It involves 13 partners from EU countries and will move to external partners from Europe and other countries.

These notes are intended to provide an informal briefing for members of the DPC not able to attend in person. They only represent the sessions that WK was able to attend was able to make notes. For an authoritative and comprehensive report, readers are encouraged to contact the organisers or speakers directly. No other DPC members were present though a number of close partners were present (including Tim Devenport of EDItEUR).

Introduction

Andrea de Pasquale (Biblioteca Nazionale Centrale di Roma) – Welcome and Introductions

Welcome to the National Library

Rossana Rummo (Istituti Culturali ed il Diritto d'Autore)

Welcome to the national library and brief words about digital strategy in the Italian presidency.

William Kilbride (DPC) Strategies for preserving the cultural heritage

See attached notes

Wim Jansen (European Commission) The EC's activities for digital infrastructure, state of the art and future prospects

Project officer for a chain of FP7 projects to mobilise a community around the better use of infrastructure. Making progress slowly but definitively. New call coming soon (closing Jan) which might be interesting for the DCH community. E-infrastructure as a component of developing

research capacity. Society is complex and therefore infrastructure need to cut across disciplines and institutions if it is to be effective. It need to enable innovation rather than stifle it, and this means also reaching out to industry for new markets and new economies. Drivers for change – more computing power, global connectivity of researchers with large and complex data sets (big data), openness in research is better for everyone. Operational continuity (sustainability) based on the appropriate allocation of EU and national budgets. E-infrastructure exists to integrate resources and services – networking, computing, data, software and user interfaces. It's a big part of the programme and comes in two parts: fostering innovation potential of research infrastructures and their human participants; and fostering commercial and industrial exploitation and integration of research outputs. Virtual research environments are going to be the next component (call 3). Virtual Research Environments are groups of researchers typically widely distributed what are working together through ubiquitous and trusted and easy access to services for scientific data, computing and networking in a collaborative virtual environment. Typically VREs address the needs of specific communities, combine academics and industries, involve bottom-up and user-oriented services and are based on e-infrastructures. A call for proposals is now out for 'Capacity building in inter-disciplinary research through community led development and deployment of service-driven digital environments for large scale cross-disciplinary research collaboration and data interoperability.' Looking for more-effective collaboration between researchers and increased take-up of collaborative research by new disciplines; easier discovery of data tools and partnerships. (42 million in total, 2-8meuro per project) call 3 e-infrastructures.

Another open call (call 4, topic 4) is designed to bridge the gap between science and ICT. New professions and skills for E-Infrastructure. Professional recognition and wider opportunities for training along with educational programmes for data scientists and data librarians. Create a community of practitioners from one sector or another in order to provide better support for student and professors in computing and networking. These people exist already but are often poorly recognised. Highly necessary but poorly paid and often lost at the end of the project. There is no scope for them to publish or be recognised for their work. This is crazy because the community needs to be bigger if the e-infrastructures are to develop. So define / update university curricula for e-infrastructure and mobilise and motivate them better. Smaller amounts of money 2.5 M euros overall and deadline for proposals is 14/01. It would presumably provide a roadmap for a later programme which would be much better funded and much more detailed in requirements.

Rosella Caffo (Co-ordinator of DCH-RP project and Director of ICCU) Summary of the project and the roadmap

Introduces the project but skip from the overview to pick up some more specific aspects of the project. Memory institutions create and collect huge quantities of digital content and need to consider preservation. DCH-RP is the latest iteration of a series of projects (DC Net and Indicate) and moves from analysis of the space to provide a practical roadmap for the future. Main aim is to ensure that memory institutions and e-infrastructure providers should talk together. It was a small project and small budget (810Ke) 2012-14 but the benefits are highly strategic and potentially very important. Roadmap is important and will be covered more later. The point is that all the components already exist and the question is really about concertation and configuration rather

than building anything new. For the DCH institutions there are specific challenges about access to E-infrastructure so efforts on authorisation services so that institutions can participate in a trusted environment, and put their valuable digital resources there. E-culture science Gateway has examined the transparent access to digital cultural heritage. Also engaged a range of EC projects including APARSEN, ARIADNE and others. Lesson learned – need to raise awareness, continue experimenting with services, intensify collaboration between stakeholders, market of preservation services is embryonic, need to develop clear guidance on what to preserve and different strategies of preserving different digital products. Issues coming out the other side of the project: sustainability (cost and community) remains an ongoing issue.

Italian EU presidency has made a feature of digital technologies: the development of Europeana and seeks an opportunity to foster creative and collaborative uses of digital objects. 2 other conferences coming up soon: Rome on 2nd October and then again 13-4th November. (rosa.caffo@benicultuali.it)

Federico Ruggieri (GARR consortium) The role of e-infrastructure for preservation of cultural heritage

Infrastructures are a very important facility and we must reference not just the scientific components but also the technical points. GARR is an organisation of research entities, managing a research network. The advantage is that we can do a great deal with shared resources. Massive academic network covers the whole of Italy – fibre reaching 500 sites and is also connected to Geant wth additional interfaces to Malta and others. CH community is well represented in GARR. A new extension to Southern Italy will connect even more institutions with greater bandwidth. New addition of data storage facilities and computing infrastructure – 10PB and 8K CPUs distributed around southern Italy. Resource sharing model – everyone shares and gets access to resources: hard to build a sustainable business model so modelled a little bit as a Cloud. Allied services also provided: identity management, web-conferencing, storage and PaaS. Solutions are not just technical though. Four services for cultural heritage agencies – Storage, SaaS (authentication, web conferencing, video-streaming), PaaS (Gateways) and IaaS. GARR is a state of the art research infrastructure with connections to the world. CH institutions can build on it. Sustainability is guaranteed by the national network. International cooperation is essential. It's a win-win situation for GARR and the DCH agencies involved.

Börje Justrell (Royal Archives, Sweden) Digital Cultural Heritage Roadmap for Preservation

DP doesn't yet have workflows and tools which are universally applicable. Current solutions also need a lot of adaptation to be deployed in other sectors or institutions. DCH is producing a lot of data which is also becoming more complicated through time. DP is done mainly in-house or in-sector. Cost could be reduced and interoperability enhanced by shared common workflows. Existing scientific infrastructure could be efficiently deployed for the cultural heritage sector. So we need to understand the needs of the DCH sector and what tools / services the E-infrastructure can make available. Distributed preservation for digital cultural heritage is an attractive promise but we also know that preservation is not simply a post-production task. It's inevitable that there are variations of practice in DCH sectors and institutions so need to understand and respond to these differences, such as by encouraging harmonisation. Case studies and examples of good practice

have helped with this. An intermediate roadmap was reviewed by partners with an enormous amount of feedback. The market is in its infancy so a lot of work to be done. This conference allows for more feedback and will be completed by the end of the year. It will include short term, medium term and long term recommendations and actions. There is a lot of work in the first two phases which loop round the same four topics - Improve interoperability; Harmonise storage and preservation; Establish cross-sector integration; Establish governance model. The final phase is still a 'blank' because there are so many unknown features. It will be presented as a handbook which will help to steer thinking. But it remains a living document and needs to be updated as we progress. Roadmap is a way to arrive at a certain destination. Allied to this is Preforma which is doing some simple commercial tendering for the validation of digital archives. Also a new proposal (DP-Infra) submitted to the EC on Deployment of Roadmap services and will offer services from the PreForma project.

Roundtable discussion on E-infrastructure perspective on DCH and DP

Giovanni Bergamin (natonal Library of Italy / Florence), Enzo Valenti (GARR), Marie Veronique Leroi (French Ministry of Culture), Fulgo Marelli (ESA), Franco Niccolucci (Ariadne), M Mayer (EUDAT), Mariella Guercio (Sapienza / CINI), Tim Devenport (Editeur), Michel Deschler (EGI), Bore Justrell (Author of roadmap) Wim Jansen (EC – chair of discussion), Mirella Serlorenzi (Soperintendenza per Archaeologici di Roma), Luigi Birguglio (ScidepES).

- Who are the target groups? Not one group but at least three – Policy makers, DCH Institutions and E-Infrastructure Providers. E-infrastructure needs to be paid for so there needs to be a return on investment.
- What if the funding runs out? How to follow up if there is no new funding. Fall back to national infrastructure providers. Partly it could be funded nationally or via the EC. No agreement yet.
- Agencies most likely to benefit are small DCH institutions: also the agencies least able to engage. How to support them better?
- Are there good examples of collaboration? ESA already commits to preservation of its data sets so there is infrastructure there. A lot of collaboration already in this field in earth sciences. Environment always resistant to change so any new perspectives (such as the re-planning of the archival process) is faced with inertia.
- CASPAR project involved collaboration between culture, art and science communities. SCIDEPes has picked up the CASPAR tools and applied them to Earth Sciences. It shows that DCH can bring benefits for the infrastructure providers too. Layer of community development is also needed.
- EC has spent a huge amount of money on DP. Are we able to collaborate when the EU stops funding? At a certain stage the community needs to come to an agreement about the value of the tools. It has to exist beyond projects. Publishers are a model of this – PORTICO, LOCKSS etc are now service providers in this space.
- PUBLISHERS ARE ABLE TO ARTICULATE THE VALUE OF THE COLLECTIONS.

- Some doubt that there is any hope that we can maintain this collaboration without core funding from the EC. The focus on international projects is the Achilles heel of projects and development. Failure to take into account local and other sources of funding is critical.
- F Niccolucci - Sustainability gives me a rash. Advocacy is critical. Good things are sustainable, bad things are not. The only way to make progress is to guarantee and show how the community of use are asking for this: success stories, demonstrations of value, bottom-up approach. These are the only way to succeed.
- Think more creatively about the solutions and business cases.
- ESA has a mandate to preserve collections: constantly trying to illustrate what you can do with the retained data and its importance, eg 30 years' worth of satellite images of Arctic Ice showing the decline which has happened in that period. Similarly a weather satellite which happened to be passing over the epicentre of the Fukushima Earth quake, quite by chance acted as a seismometer as the atmosphere was disturbed by the shockwave from the quake. Suddenly there was a tremendous interest in data from such satellites from people with no background in air density perturbation.
- Cultural heritage sector is so large and diverse that the assumption that it can provide a single point of view is naïve. Infrastructures need to speak to someone but not clear that they are speaking to the right people.

About this document

Version 1	Written at workshop	22/09/2014	WK
Version 2	Distributed	24/04/14	DPC members

DigitalPreservationCoalition

Digital Cultural Heritage Roadmap Project

Preservation strategies for digital cultural heritage

Part One: Cultural Heritage and Digital Preservation
Part Two: Challenges for the future
Part Three: The DCHRP Roadmap

Part One: Cultural Heritage and Digital Preservation

Digital resources are extraordinarily flexible and surprisingly fragile…

… what need we do to ensure that our generation's digital creativity becomes a meaningful digital legacy for the next?

DigitalPreservationCoalition

Digital Preservation make bleak reading:

a game we can all play

<Enter your details here>

……

……

……

DigitalPreservationCoalition

What's the problem?

Obsolete media
Obsolete software
Obsolete wrappers (file formats)
Bit rot
De-synchronisation
Poor representation information
Lack of funding
Loss of rights
Encryption and security
Physical loss (fire, theft, flood …)
Virtual damage (malware, …)

etc …

DigitalPreservationCoalition

Digitization and Digital Preservation

Digital Cultural Heritage

Access and engagement
Surrogacy
Crowdsourcing
Comparison
Etc …

But …

Sustainability
Syncrhonization
Technical obsolescence
Maintenance
Etc …

DigitalPreservationCoalition

Preservation policies in digitisation

18 questions…

Who →
- are your users long and short term?
- maintains the intellectual integrity of the content?
- maintains the technical integrity of the content?

What →
- is the content and where did it come from?
- formats and metadata do you use?
- is the size of the collection?

Where →
- is the master copy of the metadata?
- is the master copy of the principle content?
- are other copies held?

When →
- How long do you expect content to be available for?
- (if) things go wrong what are the consequences?

How →
- will the collection be created?
- maintained?
- migrated or emulated?
- will the collection be updated ?
- Will the metadata be updated?
- do you track who did what?
- do you keep data synchronised?

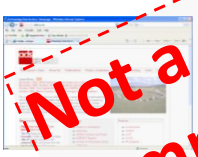
Policy needs to be funded

How much will this cost?

How much does preservation cost?

Lifecycle costs of digital objects
vs
Lifecycle costs of books
vs
Lifecycle costs of museum objects
vs
Lifecycle costs of archives
vs
Lifecycles costs of historic environment

How much does a repository cost
Here's two I prepared earlier ...



Setup:
Tens of thousands

Setup:
Tens of millions?

Not a direct comparison

Assertion based on my experience:

We are able to understand and assert the value of museums, libraries, archives, heritage ...

But

We are poor at understanding and asserting the value of digital.

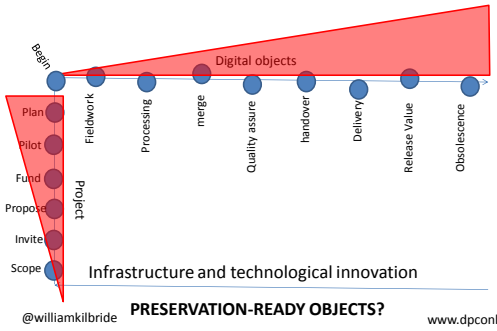
(Is data the new oil?)

Digital preservation expensive ..?

Not necessarily: it's an unfunded mandate

Don't throw money at it:
Get the mandate properly incorporated

When should preservation start?



PRESERVATION-READY OBJECTS?

Digitization and Digital Preservation

Digital Cultural Heritage

Sustainability not just technical
Sustainability not just money
Sustainability not just planning

Sustainability (also) about community

DigitalPreservationCoalition

What's the problem?

- Digital data (images, documents etc) have value
- They create opportunities

...but...

- Access depends on software hardware and people
- Technology and people change

...therefore...

- Technology can create barriers to reuse
- So, opportunities are lost

DigitalPreservationCoalition

asking the right questions

Digital preservation is not just about 'data':
Digital preservation is not just about 'technology':

it's about

people and opportunity

Part Two: Some thoughts in the future

The file is not necessarily the atomic unit of data. Preservation or records management approaches which fetishize files are never likely to be sufficient.

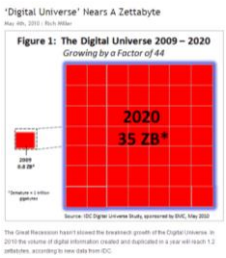
DigitalPreservationCoalition

Six (or Seven) Observations

1. **Data is growing,** budgets are not
2. **Big data complex data** as a metaphor for our future problems: does the cloud / GRID help?
3. Does the cloud / GRID make it **easier to engage** in digital preservation
4. Why is digital preservation **so hard**?
5. What is **data** anyway?
6. What is **trust** going to be like in the distributed world?

Does '**preservation as a service**' help?

DigitalPreservationCoalition



Data is growing on three axes

Scale
Complexity
Expectation

Capacities are not

Data volumes: 60% pa
Storage capacities: 25% pa
Data centre budgets: 2% pa

DigitalPreservationCoalition

... successful, practical digital preservation is going to be about **workflow and capacity** as much as about **obsolescence or authenticity.**

DigitalPreservationCoalition

Digital Preservation Approaches 1990-2014

Migration
- Intervention at data layer to ensure information objects
- Based on significant properties of content and performance
- Quick start, low cost, ready quality assurance, focus on data/access
- loss of authenticity, poor with complex objects

Emulation
- Intervention at software / OS layer to ensure operation of software
- Based on significant properties of the environment and its behaviours
- Slow start, high technical threshold, access less transparent
- retains authenticity, geared towards complex objects

Migration has done all the running in the last 10 years (20 years)

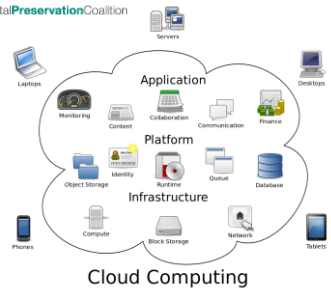
DigitalPreservationCoalition

Big data / complex data

Web archives
Sound and vision
Digitised content
Email

Complex, vast, valuable, heterogeneous
Difficult to move
Difficult to access
Greater than the sum of its parts

DigitalPreservationCoalition



…. the delivery of computing as a service rather than a product

DigitalPreservationCoalition

Infrastructure as a service GRID Computing

Scalable and Elastic
- Services scale on demand to add or remove resources as needed

Service based
- The service could be considered "ready to use" or "off the shelf"
- Offers IaaS, PaaS, and SaaS

Economical
- Services share a pool of resources to build economies of scale
- Metered by Use : Pay-as- you- go

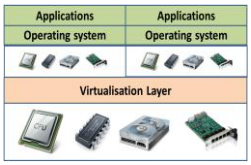
Evolvability
- Supports for migration and upgrades.
- Services are configurable

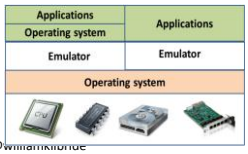
DigitalPreservationCoalition

Virtualization Vs Emulation

Virtualization

- Virtualisation puts a layer between physical hardware and controls access to that machine.
- Each guest machine (VM) that is built on top of the abstraction layer (hypervisor) is then provided access to the physical host's resources without modification.
- The hypervisor act as a traffic cop by allowing certain amount of the physical resources to be used by the guests, as well as manages resource sharing when more than on guest system try to access the resources.





- Emulation
  - Duplication of functionality of systems, be it software, hardware parts, or legacy computer system as a whole, needed to display, access, or modify a certain contents.

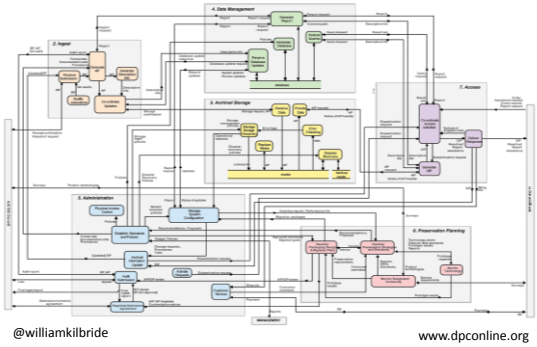
DigitalPreservationCoalition

Does virtualisation make DP easier?

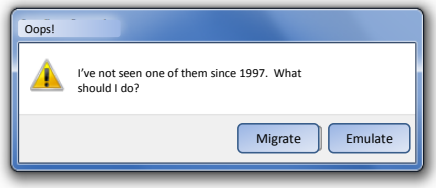
- Elastic
- Bypass (generic) corporate IT
- Industrial scale preservation?
- Highly dependent on services
- Pret a porter?
- Highly dependent on location and configuration of services
- Easier deployment of complex solutions – virtualisation

DigitalPreservationCoalition

We've made this all very hard.



DigitalPreservationCoalition



DigitalPreservationCoalition

What is data any way??

❄□◆♋❒♎◆ ♋
✞■♓❖♏❒◆♋●
♎♏♍●♋❒♋◆♓□■ □♐
♓■◆♏❒♎♏□♏■♎♏■♍♏

or

Towards a Universal declaration of interdependence.

DigitalPreservationCoalition

From Trusted Digital Repository to untrusted digital commons

It's not the repository we trust – it never was.

the **people**, the **tools**, the **services**, the **policies**, the **business plan**, the **mission**, the **organisation**, the **context,** the **user-focus.**

How is this different?

DigitalPreservationCoalition

Can infrastructural approaches help core DP issues?

- •Storage – yes!
- •Compute – yes!
- •Costs – maybe (maybe not)
- •Skills – to some extent
- •Making emulation (virtualisation) realistic?

Part three: The DCH Roadmap

Common procedures and workflows would reduce the cost in terms of time and money and would contribute to the interoperability and open-ness of data.

Existing e-infrastructures could be efficient channels for the delivery of advanced services that can be used by the digital cultural heritage sector for digital preservation.

Core assumptions

- Digital cultural heritage collections need preservation
- Expensive and tricky to accomplish individually
- Significant economies of scale and scope are possible
- E- infrastructures (esp Grid) have proven their value in the hard sciences
- E-infrastructures are flexible and have capacity to assist

Core challenges

- Considerable diversity in workflows in cultural heritage sector
- Skills gap in cultural heritage agencie
- E- infrastructures (esp Grid) have limited preservation experience
- How to develop trust in distributed preservation infrastructure
- How to ensure sustainability
- How to respond to other emerging technologies

Practical actions

- Harmonizing storage and preservation practices in DCH institutions
- Facilitate dialogue between DCH institutions and E-infrastructure providers
- Understand and create the conditions in which these two sectors can integrate their efforts
- Examine and agree models for the governance, maintenance and sustainability of infrastructure

Timescales

- Short Term (2014-5) Define topic
  - Improve interoperability
  - Harmonise storage and preservation
  - Establish cross-sector integration
  - Establish governance model
- Mid Term (2016-7) next steps
  - Enhanced data storage and preservation
  - Enhanced interoperability
  - Programme of integration
  - Redesign and implement governance
- Long term (2018 onwards)
  - Review, update, consolidate

Audiences

Who is paying attention?

Policy makers
- Create a legal and institutional setting which makes progress possible

Digital Cultural Heritage
- Harmonise (simplify?) practice and identify priorities (eg training)

E-Infrastructure providers
- Provide access to infrastructure in trusted but simple ways

In conclusion: Some Questions

What this means ….

Some questions?

What will success look like?

Why chose E-infrastructures over Cloud?

How will we train our staff?

How realistic is it to harmonise workflows?

How will we track and respond to user needs?

What other sorts of partnership are needed?

Where will the money come from?

Digital Cultural Heritage Roadmap Project

Preservation strategies for digital cultural heritage

Part One: Cultural Heritage and Digital Preservation
Part Two: Challenges for the future
Part Three: The DCHRP Roadmap