

Digital Cultural Heritage Roadmap Project: Roadmap Concertation Workshop

Tallinn Technical University, Tallinn 23-25th April 2014

About the event

The Digital Cultural Heritage Roadmap Project (DCH-RP) invited WK to participate in a 2 day workshop to review their 'Interim roadmap for digital preservation', an early draft of the main deliverable of a 2 year EC-funded project which is due to be complete later in 2014. The full text of this document is available online at: <http://www.dch-rp.eu/getFile.php?id=221>

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 identify 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 (Holger Brocks of FTK / APARSEN, Raivo Ruusalep of the National Library of Estonia / 4C, Kuldar Aas of the National Archives of Estonia / E-Ark, Marco De Niet of DEN/NCDD and Tim Devenport of EDItEUR).

Introduction

Borje Justrell (RA Sweden) – Introduction to the DCH Roadmap

Background is familiar – large amounts of data, heterogeneous institutions, rapid development of tools and services but not well integrated into practical workflows of cultural heritage institutions. DCH RP is a logical follow up to INDICATE and DC-Net projects. Data is becoming more complex and interoperability in workflows will increase open access, and 'hard sciences' are already demonstrating impact. So the project's basic assumption is that existing e-infrastructure for research and academia could be used for digital preservation and digital cultural heritage. Underlying this is an assumption that it will be possible to establish or enhance common procedures and policies that support interoperability.

DCH-RP will produce a roadmap for digital preservation services for digital cultural heritage – a short term, practical plan which is supported by appropriate demonstrators. Deliverable 3.1 is a plan for a roadmap; Deliverable 3.2 is a study of standards and interoperability best practice; Deliverable 3.3 is a basic registry service for digital preservation. These components were used as the basis for an

intermediate roadmap (Deliverable 3.4) which is in discussion today. This is a work in progress. It describes a working model for the implementation of distributed digital preservation services for the digital cultural heritage community, based then on concrete steps and services to address them. This workshop is an important point in to refine the roadmap which will be completed by September.

Some immediate lessons: digital preservation is about perpetual change; key concepts are distinct functional and technical requirements, solid models for handling business issues, authenticity, usability; the prospects of using e-infrastructure for digital preservation in DCH institutions is promising. Main challenge is whether to use OAIS as the underlying model and map e-infrastructure services here, or to use the service architecture models provided by e-infrastructure and embed preservation services into them. NOTE TO SELF – distinguish information model and functional model of OAIS. Even if we chose to adapt the functional model we may still need to address the information model?) The maturity of the sector is an issue – how do we ensure that the DCH sector is prepared to exploit the opportunities? (NOTE TO SELF: surely this points to need for training?)

Discussions about sustainability for any recommendations.

Session 1:

Marie-Veronique Leroi (Patrimoine Numerique, France)

Commentary from French partners on the interim roadmap. The focus is on functional and technical aspects of preservation; for a document aimed at policy makers there is surprisingly little about the legal or political implications; no effective definition of a digital object – is it about digitised content, or metadata or born digital, or research data; preservation approach depends on the type of digitised content (ie digitised content is different from born-digital); reference to Paris Declaration on co-operation on sharing standards) and UNESCO charter;

Maurizio Messina (Biblioteca Nazionale Marciana, Italy)

Primary concern from this partner is fitting DP into digitisation workflows. Italy has good experience of library networking and integration between museums and archives, there is common use of resource discovery metadata and a growing integration between memory institutions and e-infrastructure. But there has been a significant financial problem for the sector, relatively slow staff turnover, poor awareness of DP and limited commercial involvement in DP in Italian cultural heritage institutions. The libraries – the whole sector – is waiting for preservation that can be implemented easily. Trust models don't work for distributed digital preservation.

Hilke Arijs (Royal Institute for Cultural Heritage, Belgium)

The focus on digitisation in cultural heritage means that digital preservation is both pressing need but currently lacking. It's more than storage and backup. But the standards and tools like OAIS and Drambora are not easy to use. Deployment of multiple micro-components is very attractive because the skills are not in the community: but it creates multiple micro-dependencies which is a risk we've not explored. Need for advocacy and need for training.

Eva Stenskold (Ministry of Education and Research Sweden)

E-infrastructure is heavily populated by acronyms. It is off-putting but it is actually a sign of strong collaboration which is both needed and taken for granted. Threat to implementation of the road map is the gap between the cultural heritage sector and the science sector which are funded from different sources and have different expectations. In Sweden there is a determined effort to bridge this gap, based on a major re-organisation of things like research networks to ensure that cultural heritage agencies are engaged.

Kuldar Aas (National Archives of Estonia)

Start on the assumption that the cloud is good! But there are different cloud models – IaaS, PaaS and SaaS – and it's not necessarily clear which approach is best for each institution. There may not be a one-size fits all model. National agencies may be the people who run the infrastructure for the others, and at least may need to maintain a very different level of expertise (ie chose IaaS) based on their national and statutory roles than smaller ones which might need simply to access the service (ie PaaS) or access occasionally individual tools or applications (SaaS) in their workflows.

Marco De Niet (DEN, Netherlands)

Netherlands takes a decentralised approach to cultural heritage sector. Only one third of cultural heritage agencies have a preservation policy and only one quarter is connected to digital infrastructure which could offer preservation. The majority rely on in-house collections management systems. More than half report some loss of data, and many were unwilling even to say answer the question. There is a cultural coalition on digital preservation which has its own working plan that has 4 phases of work which actually align pretty closely with the DCH-RP. There is no strong sector leader in the cultural sector to drive digital preservation and although there is need for such it needs to be based on a collaborative and representative approach. The sense of urgency varies between agencies which depends on the maturity of information policy. That means we need to get as close as possible to existing workflows. Object driven preservation dominates and the need for process focussed preservation: so we need a different way of thinking about repositories. And we need to link preservation and access.

Vania Virgili (DARIAH – CNRI, Italy)

DARIAH is an international research infrastructure for arts and humanities and there's a linkage to DCH-RP.

Remigiusz Lis (Silesian Digital Library, Poland)

Regional model of digitisation in digital libraries in Poland. Digitisation is often a marginal activity of institutions and mostly project based, and the emphasis is on descriptive metadata and presentation: archiving is therefore a gap in what is already a marginal and fragmented activity. That's the context that the roadmap would have to deal with in Poland.

Sarunas Bagdonas (Lithuanian Art Museum)

Integrated museums information systems for Lithuania is provided nationally as infrastructure serving diverse audiences. 63 different institutions involved. No new digitisation efforts and

therefore the trick is to use what has already been digitised. It's easier in some sense to preserve physical objects in the museums than to preserve the digital ones. Use or abandon! IT serves a purpose: it doesn't exist on its own. Systems, software, infrastructure etc have to be a service in the broadest sense: but the terms software, infrastructure are not interchangeable. Who serves whom?

William Kilbride (DPC)

(Notes from presentation as appendix 1).

Leif Laaksonen (CSC/RDA/e-IRG, Finland)

Roadmap, in practical terms is a plan that matches short term and long term goals for the development of a specific technology (or policy or organisation). It's a process of negotiation and perhaps compromise between different possibilities and opportunities. Questions to be considered by the partners ... who will pay for infrastructure: there is no free resource. Is the infrastructure actually available for use by DCH? How might standards and processes and agreements that can be already exist be deployed to our advantage: like CODATA, RDA, WDS? How will resources will be findable and usable? The key one is about money.

Jacques Dubucs (BNF, France)

Roadmap is ambiguous between digitised cultural heritage and born digital objects. This is a problem because these are very different. For example we can put boundaries around a book and authenticate in a way that we can't do with a born digital collection.

Open discussion on E-infrastructure perspective on DCH and DP

- Thanks noted to reviewers and recognition of need to get outside of 'project bubble'
- Discussion between DCH/DP people and E-Infrastructure people could be termed service-providers versus service users
- Is the technology being driven by technology itself or is it being driven by actual needs: a solution in search of a problem or a demand pull
- Have we properly articulated the needs of the DCH community? For example process (function) seems to be well understood but packaging and information modelling is not.
- Who pays and who administers?
- What is the difference between born digital and digitised material. How might this difference be managed and what about the shades of grey between these two poles?
- Oceans of insignificance – how do we deal with the large volumes of not very interesting material?
- Physical location of the digital objects matters within information governance? How to ensure proper disaster management while keeping data within constrained geographic locations. Data is a 1st class citizen but sometimes is under a travel ban.
- Is a preservation infrastructure just for digital cultural heritage possible or is digital preservation always a generic service offering.
- What happens when we move from Capex to opex functions?

- Cultural heritage agencies that need digital preservation most are the small and fragile ones that can afford it least
- Is there another validation step needed before the conclusion of the project?
- Can preservation be embedded into VREs or other desktop tools so that preservation is not a post-production task?
-

Other sessions / presentations not noted: Laila Valdovska, Mate Toth, Jurgen Kepler, Michel Deschler (EGI), Haardi Teder (EENET) and Holger Brocks (FTK/APARSEN)

About this document

Version 1	Written at workshop	23-24/04/2014	WK
Version 2	Distributed	24/04/14	DPC members



DCH RP

Infrastructure Concertation Meeting 23/04/2014

If you have a map then you need a scale?

- What's the big picture?

The things that maps won't tell you?

- Effort (ie cost)

What's going to stop us today:

- (Money?)
- Technology?
- People?

Shameless advert

Digital preservation typically makes bleak reading ...



When asked about how long their digitised resources would be available for, JISC-funded projects said ...

‘In perpetuity’

‘Indefinitely’

‘50 years’

‘10 years then elsewhere’

‘until 2014’

‘forever or for three years’

DPC/Portico/ULCC 2010



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

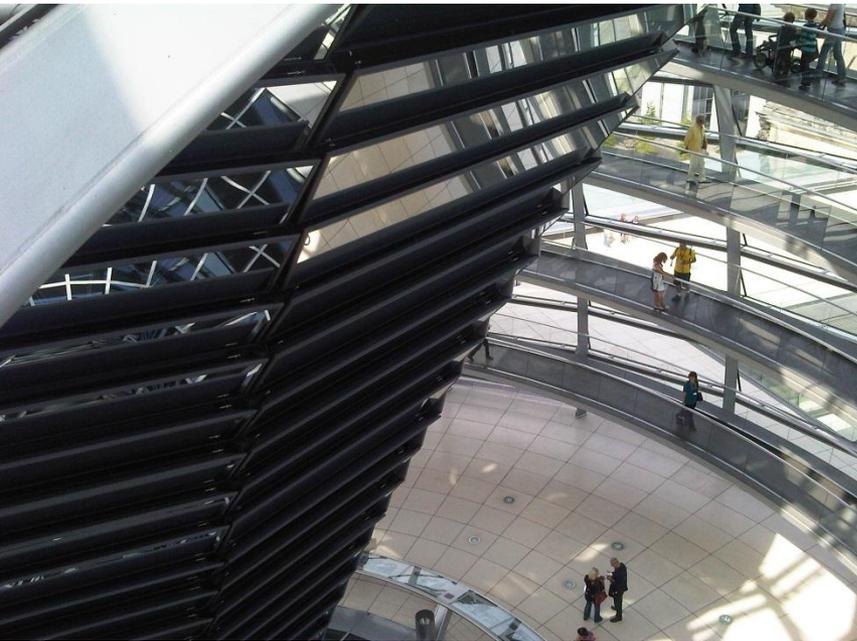
...therefore...

- Technology can create barriers to reuse

- So, managing data in the long term protects and creates opportunities

- Not something we do for its own sake

We do preservation because we want to be:



1. Transparent

e.g. Data Protection, Freedom of Information ... childcare, human tissue

2. Safer

e.g. preparedness, detection, disaster, recovery, audit

3. Smarter

e.g. scientific value, access to heritage, value of social knowledge

4. Wealthier

e.g. efficient business, management of IP, employment, planning, creative

5. Healthier

e.g. managed life history, research and safe innovation

6. Greener

e.g. evidence-based policy development, efficient data retention

And because of



1. Legal Compliance

e.g. Sarbanes-Oxley, Data Protection

2. Regulatory Compliance

e.g. power generation, aviation, banking

3. Legal protection

e.g. patents, mis-selling, detection, audit

4. Unanticipated exploitation

e.g. petro-chemical, music,
pharmaceuticals

5. Business Continuity and improvement

e.g. product recall, disaster recovery

6. Business Value

e.g. getting the right information to the
right people at the right time in a format
they can use



Digital **Preservation** Coalition

Digital preservation is not just about 'data':

Digital preservation is not just about 'access':

Digital preservation is not just about 'tools':

it's about

people and

opportunity

Start

Here!



Safer

Smarter

Harmonisation of data storage and preservation?

Improved interoperability?

Fostering the conditions for cross-sector integration?

Governance models for infrastructure integration?

Healthier

But there are not destinations:

They are means to an end

Wealthier

Greener

Fairer



Why this matters?

Money turns out to be the major problem facing the future of our digital heritage' (Rosenthal 2012)

Data volumes increasing 60% pa

Data storage increasing 25% pa

Data budgets increasing 2% pa



CHEAPER?



LOWER BARRIERS?

Does DCH-RP help this?

Maybe!

‘existing e-infrastructures are **efficient** channels for the delivery of advanced services’

‘it’s possible to establish **common** policies and processes’



CHEAPER?

LOWER BARRIERS?

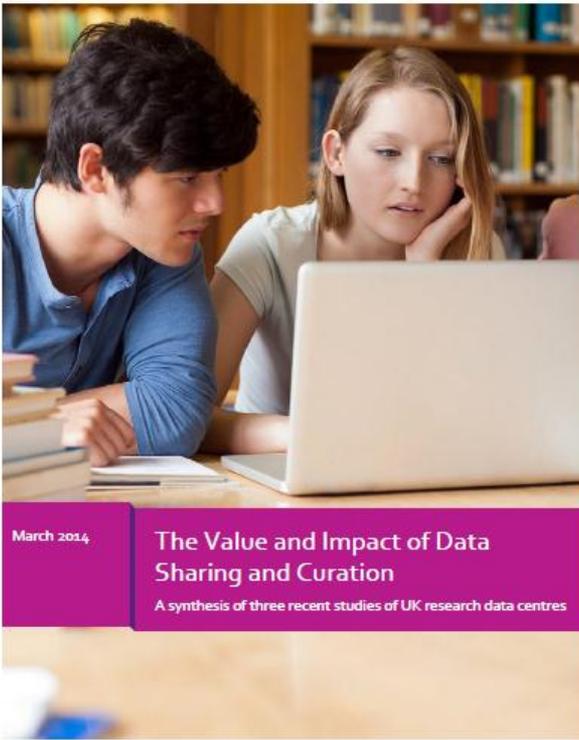
Does DCH-RP help this?

PROVE IT!

Are these e-infrastructures actually more efficient? What about other market models? Why not commercial? National? Sectoral?

Is it actually cheaper to get people to work together? What is DCH? Do they have much in common with each other?

Jisc



Does DCH-RP help?

The roadmap needs to consider how implementation of its four basic components deliver financial advantage ...

Four things to examine

Return on Investment from Data Centres

Collaboration to clarify the costs of curation

Commercial providers

Workforce developmet



We can tell a good story about costs!

quantitative analysis indicates that: ...

- The value to users exceeds the investment made in data sharing and curation via the centres in all three cases – with the benefits from 2.2 to 2.7 times the costs;
- Very significant increases in work efficiency are realised by users as a result of their use of the data centres – with efficiency gains from 2 to 20 times the costs; and
- By facilitating additional use, the data centres significantly increase the returns on investment in the creation/collection of the data hosted – with increases in returns from 2 to 12 times the costs.

<http://blog.beagrie.com/>



Digital**Preservation**Coalition

What's going to stop us today?

- Money
- Technology
- People



<http://blog.beagrie.com/>



The technology? Some observations that impact on the roadmap

1. No longer clear what 'data' is
2. What are the implications for cybersecurity and acceptance testing if you need to store and execute an archived technology stack?
3. Sensitivity review (redaction) and compliance?
4. Balancing the need and practicality of metadata gathering and management?
5. Scaling up while managing (reducing ?) costs?
6. Integrate with existing infrastructure





Specific implications for the roadmap arise in sections 5.2.1 and 5.2.2

5.2.1 Harmonise data storage and preservation

5.2.4 Enhance interoperability



5.2.1 Harmonise data storage and preservation

(Is this roadmap actually a roadmap about harmonising data storage and preservation?)

Digitisation workflows should be a lot simpler to harmonise than collection of born digital collections?

e.g. metadata capture

Need to get best (any!) DP tools embedded in mainstream software

Preservation from the outset



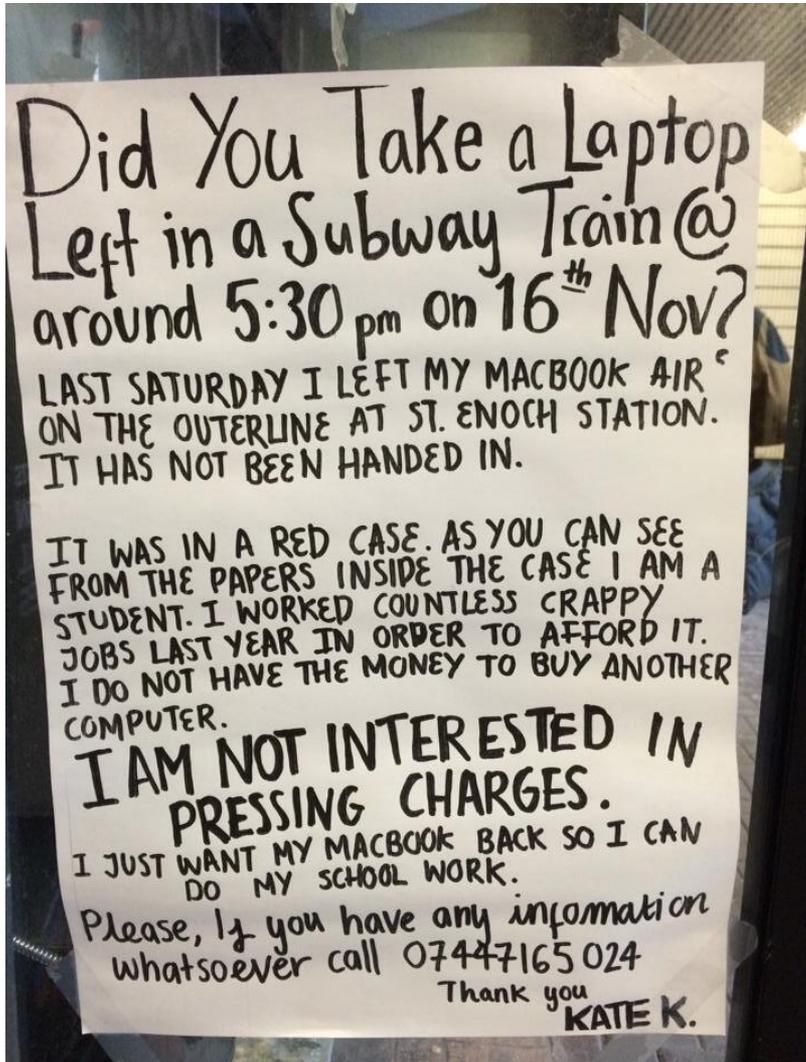
5.2.4 Enhance interoperability

Is this really just about digitisation?

Interoperability of workflows – is this realistic given diverse mission of DCH institutions?

What *are* DCH institutions? Libraries? Museums? Galleries? Publishers? Archives?

Naming things is very controversial in cultural institutions: vocabulary ontology building is tough



The people? Some observations which can mostly be substantiated

1. Community building is harder than technology building
2. There is significant distrust of (opposition to) cloud providers in the cultural heritage sector and it's not just about the technology
3. The majority of DCH institutions are small, making them hard to reach
4. There is a significant skills gap
5. Staff involved in preservation are not numerous, have numerous other responsibilities and are scattered
6. The digital preservation community is highly fragmented



Digital **Preservation** Coalition

Specific implications for sections 5.2.3 and 5.2.4

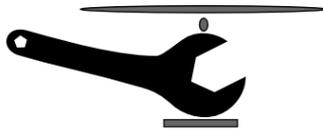
5.2.3 Establish conditions for cross sector integration

5.2.4 Establish a governance model for infrastructure integration

Connecting digital preservation expertise

[Home](#) [Blogs](#) [Paul Wheatley's blog](#)

COPTR tools registry beta launch



Almost a year ago, I presented a proposal to the Aligning National Approaches to Digital Preservation (ANADP) group to create a community tool registry. I was frustrated by the profusion of tool registries and the lack of coordination between them. Pooling the knowledge in one place would result in a far better resource. It would be easier to discover new tools, to share experience in using them and to help avoid the tool



development duplication we've seen so much of in the past. As ANADP II kicks off today in Barcelona, I'm very pleased to announce the beta launch of COPTR: the Community Owned digital Preservation Tool Registry.

We've been working to collate, combine, de-duplicate and align the contents of 5 existing tool registries from: The Open Planets Foundation (OPF), The National Digital Stewardship Alliance (NDSA), The Digital Curation Centre (DCC), The Digital Curation Exchange (DCE) and the Digital POWRR Project. There were of course quite a few duplicates to weed out but the scope and depth of COPTR now supercedes anything out there that I've seen previously, albeit with an inconsistency of depth between resulting tool registry entries. Each source registry had its own differing characteristics. At one end of the scale the DCC registry had really strong detail but coverage of well under a hundred tools. The DCE registry included over three hundred tools but with each tool described in far less detail.

5.2.3 Establish conditions for cross sector integration

- Assumption that this is not already happening
- Model existing good practice
- Derive maximum benefit from existing community endeavours:
 - Registry services
 - Workforce development
 - Policy co-ordination



5.2.4 Establish a governance model for infrastructure integration

- Assumes that there are not already organisations that could provide such governance ... RDA, OPF, DPC, nestor, NCDD, APARSEN, Archivemática, Digital Repository of Ireland
- Etc.
- How do I get my data out of the Cloud?
- Focus on 'Trusted Digital Repository' seems redundant (and I wish you would say this explicitly!)
- IaaS changes the market place





And is there a converse side to this whole discussion

- Ask not if DCH institutions can use the cloud for preservation ...
- But ask if DCH institutions can collect and preserve the cloud?
- (Might all preservation ultimately look a bit like a web archiving workflow?)





Digital**Preservation**Coalition



5 Awards

- Research and Innovation
- Training and communications
- Student work
- Safeguarding digital legacy
- Industry

<http://www.dpconline.org/advocacy/awards/digital-preservation-awards-2014>



DCH RP

Infrastructure Concertation Meeting 23/04/2014

If you have a map then you need a scale?

- What's the big picture?

The things that maps won't tell you?

- Effort (ie cost)

What's going to stop us today:

- (Money?)
- Technology?
- People?

Shameless advert