Proof of Concept in DCH-RP

Eva Toller,
National Archives,
Sweden

What is a Proof of Concept?

- A ‘proof of concept’ is
  - a mechanism to assess if
    - unknown (new) technology
      or
    - technology that I am not familiar with

  → has the potential to solve some of my problems

  → is sufficiently satisfying for my needs

What can a Proof of Concept test?

- Technology, e-infrastructure, and tools
  - ò**Technology**ò can be:
    - New hardware and software that (for example) has an impact on usage
    - *Current examples*: smart phones, tablets, social media, apps
    - *Earlier examples*: World Wide Web, SMS
    - *Future (?) examples:* wearable & implanted devices (ubiquitous computing, pervasive computing, everywareé)

cont. What can a Proof of Concept test?

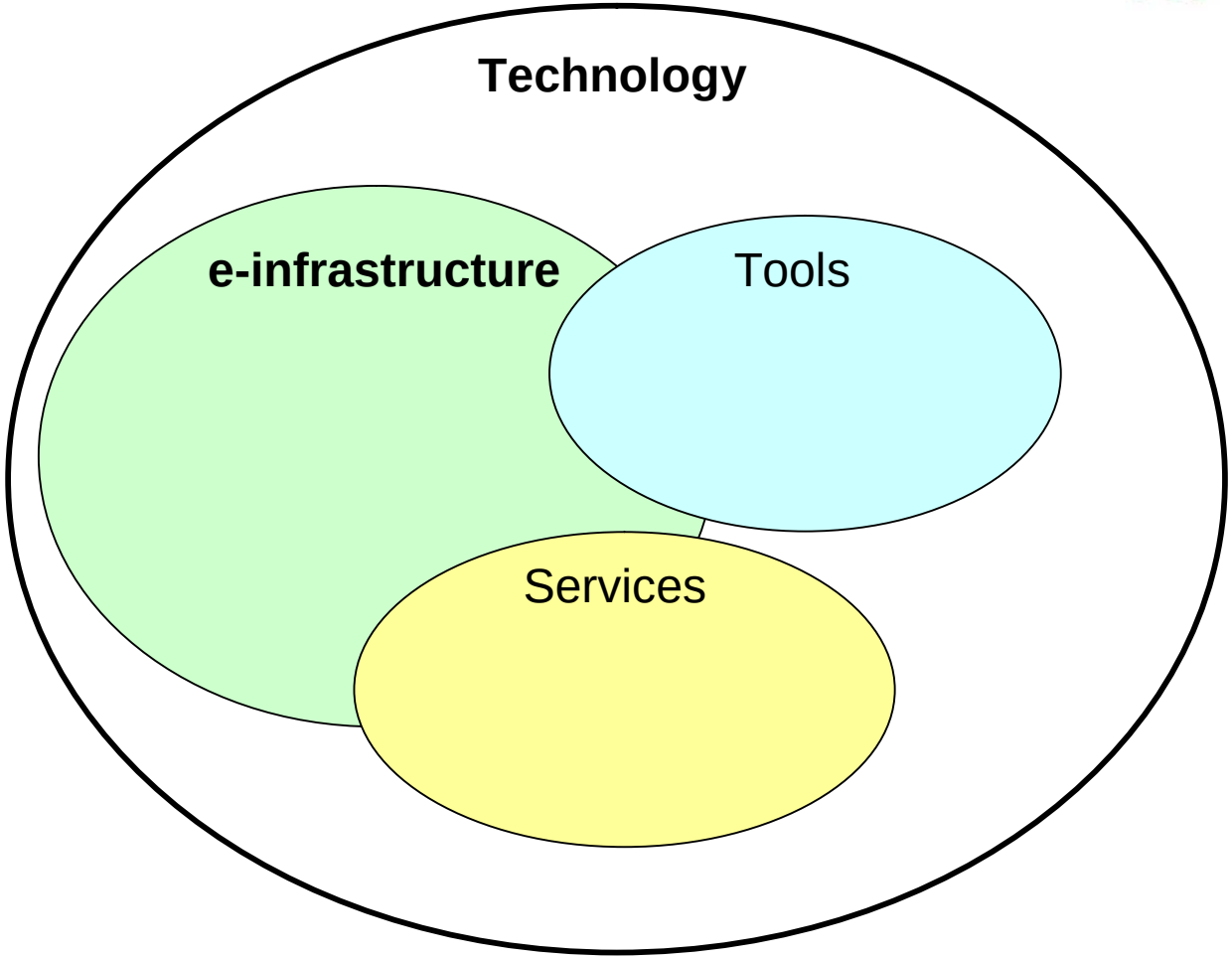
- Technology, **e-infrastructure**, and tools
  - è-Infrastructureò can be:
    - Regional, national, European, and world-wide grids, networks, and sites
    - *Examples of usage:* communities/communication, computing power, storage, dissemination, **preservation**

cont. What can a Proof of Concept test?

- Technology, e-infrastructure, and **tools**
  - õToolsò(and *services*) can be:
    - Tools for validation
    - Tools for mass conversion/mass migration
    - Tools for format identification
    - Guides and ðhelp desksò

cont. What can a Proof of Concept test?

IaaS
PaaS
SaaS
(PraaS)



When to use Proof of Concept

- When should òproof of conceptò be used (instead of some other evaluation mechanism)?
  - When resources are limited (time and man-power)
  - To answer if a specific tool is ñgood enoughò to solve my problems
  - To gain knowledge that can help me to decide if I should include a specific tool in my infrastructure

Proof of Concept in DCH-RP (2013)

- Overall objective: to validate assumptions and concepts expressed in the DCH roadmap for preservation **in concrete experiments**

- Fundamental concerns:
  - Operational challenges
  - End user concerns
  - New services and infrastructure integration

cont. Proof of Concept in DCH-RP (2013)

- Six Proofs of concepts (PoC) were conducted
- *Scenarios* form the basis of the PoCs
- The scenarios are short descriptions of hypothetical but realistic situations and/or problems
- Seven out of fourteen scenarios were covered
- All scenarios are listed in ὈD5.3 Report on first Proof of Conceptὸ: http://www.dch-rp.eu/index.php?en/61/deliverables

cont. Proof of Concept in DCH-RP (2013)

- Example of scenario for **Operational challenges**:

  Scenario 1.3 - Selecting a digital preservation solution in the case of an institution with only voluntary IT support

A little museum in Malta has a historical library and a digitised personal archive collection. The museum has staff of only 9 and only voluntary IT support. The director of the museum is aware of the need to organise digital preservation for the digitised documents, but is not sure how to do it. He receives periodically offers for long-term storage of digital content, but finds it difficult to select or to make a decision. He has practically no IT competence to rely on for decision-making, but is convinced that the decision should be forward-looking and accommodate the needs of the museum for the next 5 years.

cont. Proof of Concept in DCH-RP (2013)

- Example of scenario for **End user concerns** :

Scenario 2.1 - Researcher discovers a historical database

A researcher in history discovers a historical database resource presenting parish records. She would like to use the data, but she is also concerned to what extent these data could be trusted (authenticity, error rates introduced, errors caused by any transformations needed).

cont. Proof of Concept in DCH-RP (2013)

- Example of scenario for
  New services and infrastructure integration:

Scenario 3.3 - Integrating new services into existing infrastructure

The IT manager of a local art gallery is preserving the digital content using grid ñXò He attends a workshop on digital preservation where he hears about a new tool for checking the integrity of digital objects. He needs to implement it on the grid-based archiving solution.

Results of Proof of Concept in DCH-RP (2013)

- Scenarios that were actually used:
  - Scenario 1.1 - Using specialised research tools
  - Scenario 1.2 - Integrating a new tool into an existing institutional infrastructure
  - Scenario 1.3 - Selecting a digital preservation solution in the case of an institution with only voluntary IT support
  - Scenario 1.4 - Preservation from a consortium of collections on the cloud
  - Scenario 1.6 - Archived data retrieving
  - Scenario 2.2 - Research and select a tool serving a specific purpose
  - Scenario 2.4 - Gain access to archived websites

cont. Results of Proof of Concept in DCH-RP (2013)

For detailed results of the PoC tests, see:
http://www.dch-rp.eu/index.php?en/61/deliverables

- Scenario 1.1 - Using specialised research tools (KIK-IRPA, Belgium, and ICCU, Italy)
  - ***Original objective***: study of **Drambora** (auditing and risk assessment scheme).
  - Estimation: too much work to be included in PoC 2013!
  - Instead: test of **Scoremodel** (developed by DEN Foundation).
  - ***Recommendations***: The Scoremodel is a useful tool to test the integrity of a collection. It does not implement the full Drambora scheme but has the advantage that it is easily understood and usable by CH people. It can be used in the roadmap as an example for testing the integrity of data.

cont. Results of Proof of Concept in DCH-RP (2013)

- Scenario 1.2 ï Integrating a new tool into an existing institutional infrastructure (Belspo, Belgium, and ICCU, Italy)

- Scenario 1.4 ï Preservation from a consortium of collections on the cloud (Belspo, Belgium, and ICCU, Italy)

  - ***Objective***: preserving data and meta data on an external e-infrastructure.

  - Basic choice: the grid storage **e-Culture Science Gateway** (eCSG).

  - Not enough automation for large quantities of data
    (data had to be copied manually, meta data have to be filled in manually).

  - ***Recommendations***: The eCSG is currently limited to *small* collections, where the manual procedures are not too great an impediment. It is not yet adequate for large collections where a greater degree of automation is needed.

cont. Results of Proof of Concept in DCH-RP (2013)

- Scenario 1.3 – Selecting a digital preservation solution in the case of an institution with only voluntary IT support (Riksarkivet, Sweden)
  - ***Objective***: to test different tools for their simplicity of installation and usage
  - Two general tools: **Archivistôs Toolkit**, **DSpace**
  - Two more specialised tools: **Riksarkivet OpeN Data** (ROND), **XENA**
  - The two general tools were estimated to be too complicated for voluntary (non-professional) IT support.
  - ROND (de-identification) was considered useful but only in a very specific context.
  - XENA (format conversion) was considered useful (especially for batch conversion), although the quality of the result was not always acceptable.

cont. Results of Proof of Concept in DCH-RP (2013)

- Scenario 1.6 ï Archived data retrieving (Conservation Centre Kanut, Estonia)
  - ***Objective***: to get proof that backed-up archived information can be retrieved easily on a quarterly basis
  - Choice of tool: **IBM Tivoli Storage Manager.**
  - To set up the tool required advanced IT expert knowledge.
  - To use the tool required previous experience of command line interface programs.
  - The tool was sensitive to minor changes and modifications.
  - No recognition of format recognition, buté
  - é no limits on the file formats that can be used (neither on the file sizes).

cont. Results of Proof of Concept in DCH-RP (2013)

- Scenario 2.2 ï Research and select a tool serving a specific purpose (Riksarkivet, Sweden)
  - ***Objective***: to find a tool that can convert images in batch mode
  - Choice of tools: **AVS Document Converter 2.2**, **AVS Image Converter 3.0**, **Universal Document Converter**, **A-PDF DJvu to PDF**
  - AVS Document Converter 2.2 should only be used for small amounts of files (not reliable for batch conversion, and the quality is not always good).
  - AVS Image Converter 3.0 is not good for conversion to PDF, and cannot handle very large files. Worked for batch conversion between JPEG and PNG.
  - Universal Document Converter does not do batch conversion. It can handle very large files.
  - A-PDF DJvu to PDF was not tested since it did not run on 64 bit Windows 7.

cont. Results of Proof of Concept in DCH-RP (2013)

- Scenario 2.4 ï Gain access to archived web sites (Riksarkivet, Sweden)
  - ***Objective***: to check the ease of retrieving an archived web site (also to check the fidelity of the archived version versus the original site).
  - As a preparation, some tools were chosen to test the archiving itself: **HTTrack**, **SWAT**, **WARC Tools**, **Web Curator Tool**. Only HTTrack was both easy to install and easy to use (the other tools were not tested).
  - An end user was chosen to compare the fidelity of the archived version versus the original version. This test could not be done since the original version, http://www.linne2007.se/, was temporarily inaccessible.
  - Instead, the end user gave her impression of if the archived site could probably be mistaken for the original site. Apart from some dead links, the answer was ñyesò

Some Achievements & Lessons Learned

- Concrete metrics for tool assessment (ease of installation, management, and use). However, most of the assessments are subjective.

- Negative results: point out gaps in the roadmap (should *not* be considered as failure!)

- Input to the **Registry of Services**
  http://www.digitalmeetsculture.net/heritage-showcases/dch-rp/registry-of- services-and-tools/

Thank you!