Report from the PERSIST session at WLIC, Lyon, August 2014-10-03

Introduction

PERSIST (*Platform to Enhance the Sustainability of the Information Society Transglobally*) is a collaborative project of UNESCO, IFLA, ICA and other partners to address globally pressing questions on preservation strategies, technologies, selection, responsibility and division of labor. It assumes that on these issues a high-level global policy discussion is needed between heritage institutions, industry and government, and that UNESCO's Memory of the World Programme is a unique platform to conduct that discussion.

The idea for PERSIST was born at the Conference *The Memory of the World in the Digital Age: Digitisation and Preservation* in Vancouver (September 2012). The Declaration adopted by its participants states that:

'there is a pressing need to establish a roadmap proposing solutions, agreements and policies, that ensure long term access and trustworthy preservation. This roadmap should address issues like open government, open data, open access and electronic government. It should dovetail with national and international priorities and be in full agreement with human rights.'

PERSIST was launched as a project at the Conference *A Digital Roadmap for Long-Term Access to Digital Heritage* in The Hague in December 2013. In its initial stage the project is coordinated by the Netherlands National Commission for UNESCO and financed by the Netherlands Ministry for Education, Culture and Science.

The work for PERSIST is divided among three task forces: content, technical and policy. For the content task force, 'selection' has been chosen as the first subject of attention. Ingrid Parent (University of British Columbia, former President of IFLA) is responsible for this part of the project in the PERSIST Steering Committee.

Lyon World Library and Information Congress

The first activity of the content task force was organised in the framework of the IFLA World Library and Information Conference (WLIC) that took place from 17 to 23 August in Lyon, France. The UNESCO session in that conference was devoted to the problem of selection in the digital age. Under the chairmanship of Iskra Panevska from UNESCO, PERSIST was presented by two of its founders, Ingrid Parent and Martin Berendse (President of ICA). An introductory presentation of their essay 'The Paradox of Selection in the Digital Age' was made by Titia van der Werf and Bram van der Werf, followed by presentations of three national case studies by Michele Pickover (South Africa), Winston Roberts (New Zealand) and Clément Oury (France).

This report summarises the main points of the presentations and the discussion that the introductory essay provoked. It also makes some suggestion on the way forward for this part of the PERSIST project

The Paradox of Selection in the Digital Age:

Van der Werf and Van der Werf argue that the following trends in thinking about and 'doing' digital heritage have to be taken into account in order to draw up sensible selection policies:

- Digital heritage has been seen as a 'technology issue' since Jeff Rotenberg's seminal Scientific American article "Ensuring the Longevity of Digital Documents" from 1995. Yet it is becoming clear that it is much more a 'societal issue'.
- The growth of the web provokes the growth of what is potentially digital heritage:
    o Web-users have an unstoppable urge to self-publish.
    o Web-users create moreover a deluge of data by the traces of their digital behavior they leave behind them. These data are tapped by companies who want to sell and governments who want to spy.
    o Governments stimulate this growth by policies focused on transparency and on the stimulation of the digital economy.

    This unprecedented growth forces heritage institutions to set strict limits on what they ingest.

- Digital information seems 'free of charge', but this is an illusion. 'Free products' are bought at the expense of our privacy. The amount of electricity used worldwide to keep the digital universe alive is alarming.
- The really valuable parts of the web are not the bits which contain the 'content' (text, pictures, film..) but the bits that capture our on-line behavior. Businesses have succeeded in distinguishing between useless and useful information in the digital mass.

These trends warrant the following conclusions:

- Selection is an absolute necessity for heritage institutions. By concentrating on technology and on the digitisation of analogue materials, these institutions have been backing the wrong horses.
- The selection problem can be partially solved by internet users themselves: the removal of digital waste should be an important part of 'good digital citizenship'.
- Governments must devise and implement policies to stimulate such behavior. Policies should be implemented using positive incitements, rather than bans or fines.
- Since technology innovation is primarily driven by consumer behavior, it is in the minds of consumers that the foundations for digital preservation must be constructed.
- Heritage institutions cannot maintain their traditional selection policies in the digital era and should look for filters not on the supply side, but on the consumption side – good metadata are essential for this. Digital information that is not (widely) used should not count as heritage and should not be preserved by heritage institutions.
- Improving the interactions between the information industry, the public authorities and the cultural heritage institutions –the *trias hereditaria* – can potentially resolve the current digital preservation impasse.

The presentation gave rise to several objections and observations from other speakers and from the audience:

- After the invention of printing there was the fear that the printers would publish any kind of material after the small amount of important texts had been taken care of. Now we laugh at that glaring misunderstanding of the Gutenberg revolution. Will our fears for digital flooding be judged likewise in the future?
- Digitisation and collecting born-digital are not contradictory or mutually exclusive activities. Both are important.

- The problem is not 'selection criteria are too broad' but: 'the materials on the web are so vast that even with good selection criteria we ingest too much material'
- The question whether web-materials should be counted as 'publications' or not is not relevant, as libraries have always collected unpublished materials such as diaries, photographs, manuscripts etc.
- Participants discussed what possibilities there are to stem the avalanche of digital materials. There was skepticism as to the possibility that users could be educated to show restraint – what incentives could be given to them? Top-down approaches that do not open the door to censorship are also difficult to imagine.
- The image of libraries blindly ingesting anything that is digital because it is so easy to do so, is not correct. Selection does take place, but there is a lot to gain if we are more transparent and if we cooperate more closely.
- While we have to confront technological issues and resource issues which may limit our freedom to collect (as they always have), *what* we collect (or do not collect) is a philosophical and cultural question and we have to be careful that this discussion is based on what libraries actually do and need to do now in the digital context.

Three country presentations:

1. **Michele Pickover:** *Patrimony, Power and Politics: Selecting, Constructing and Preserving Digital Heritage Content in South Africa and Africa*

In South Africa, the archival sector is in disarray, lacks adequate skills and training (this is true for both the non-state and state archival arena), is under-resourced, and introspection and conscious self-reflection are largely absent. The archival practice of the Apartheid regime was conducive to the creation of a collective past that confirmed to the ideology of the rulers. But power relations continue to influence the way national stories are told today. The state and the ruling party lay claim to ownership and stewardship of South Africa's past and the 'liberation struggle' not under the guise of a discourse of inclusiveness or the construction of the homogenous 'rainbow nation' but through a carefully cultivated juggernaut which gives special advantage to a monolithic nostalgic legacy which has more value and is more valued.

In the South African context, ambitious digitisation projects are proving a highly ambiguous development and the representation of the "liberation struggle" through digitisation is intensifying contests over the redefinition of the archive. Many of these projects are fundamentally located in uneven power relations and perspectives which compromise national heritage; do not represent the views and interests of the developing nations; bolster inequities in globalisation; and exacerbate historic North/South imbalances. Increasingly the digitisation of South African heritage material for publication on the worldwide web is a space where the real challenges are not technological or technical but social and political. There should be a debate on the digitisation projects that should center on the following questions:

- How do we share knowledge without being exploited and deepening the digital divide?
- How do we enter into partnerships with countries in the North in ways that address and do not reinforce the digital divide?
- How do we ensure that such partnerships do not merely reformulate issues of heritage plundering and cultural asset stripping?

- How do we take into consideration issues of connectivity and context, use and power?
- How do we ensure that these digital projects do not serve to merely replace repatriation of actual heritage items?

Other African countries face problems that are similar to those in South Africa. Since 2009, the International Conference on African Digital Libraries and Archives (ICADLA) is a good platform to discuss shared problems.

To summarise: archives work in a highly political environment, and ethical and political issues have to be confronted in every digitisation project.

2. **Steve Knight & Winston Roberts:** *Digital Collecting at the National Library of New Zealand*

The National Library of New Zealand (NLNZ) is currently refreshing its Collections Policy which derives from the National Library of New Zealand Act which defines the purpose of the National Library and the functions and powers of the National Librarian.

At this moment, the library collects photos, websites, stories (digital stories and oral histories), government publications, monographs and serials, moving pictures, manuscripts (of books), manuscripts/personal papers, business and organisational records, blogs (captured or self-created), published CDs and DVDs, newspapers and newsletters embodied in emails.

It does not collect emails, databases, architectural plans / CAD, maps, auction catalogues, Twitter feeds, SMS / text messages, Facebook pages, online only music and video (yes and no), mobile apps, pre-press newspapers. The library might decide to start collecting puzzles, games, toys, podcasts, digital art, online training modules, grey literature.

There are dozens of different reasons why certain classes of digital objects are not collected. It can be due to copyright issues, a lack of policies or storage space, access complexity and technical problems (concerning both software and hardware), questions of quantity, lack of funding, staffing (capability and capacity), etc. It is often unclear what value must be attached to a digital collection and what prices are reasonable if there are no precedents (often in the case of digital art) or if vendors offer the same item to more than one library.

There are other issues that are currently being discussed at the NLNZ in the context of the Library's reflection on the renewal of its Digital Collection Policy:

- Does the distinction between original and published exist in the digital world? How do you distinguish an original from a copy?
- How to handle hybrid collections that include both paper and digital items? Do you prefer digital over paper, and if so, on what grounds?
- Collaborative collecting is problematic – how do we know who else is collecting, how these institutions handle access and what is there commitment to long term preservation?
- How do you organise digital interlibrary loans?
- What are the implications of licensing?
- What are the possibilities of demand-driven acquisition: rebalancing away from possible use toward immediate need.

- What are the boundaries of digital collecting? This not only touches on the jurisdiction of collecting institutions, but also the 'edges' of digital objects, i.e. the full experience/performance of an online object.

The NLNZ has also worked closely with its peer institutions in Australia (within NSLA – National and State Libraries of Australasia) to develop joint Principles of Digital Collecting.

3. *Clément Oury: A sustainable heritage in the digital age? - Preservation of born-digital heritage at the National Library of France*

Since the end of the nineties, heritage institutions have been involved in this "digital switch" by digitizing their collections to propose a worldwide access. At the same time, they started inventing ways of collecting, preserving and giving access to the new kind of heritage material produced by digital technologies themselves and especially by the web: the "born-digital" heritage. For the Bibliothèque nationale de France (BnF), this heritage is constituted of various kinds of documents: text and images, audiovisual content, interactive content like games and websites. Born-digital heritage enters the BnF on media and on-line, via legal deposit as well as via acquisition and donation.

Comprehensiveness, the traditional objective of legal deposit, is still the goal for digital publications on media. For on-line media, comprehensiveness is still strived for if for the heritage in question equivalent physical media exist, provided that publishing filters (scientific, artistic or commercial) do not hinder ingest and preservation and that the amount of content remains manageable. For other on line materials BnF strives for representativeness rather than comprehensiveness. The problems here are not only quantitative, but also connected with the tremendous change-rate of the materials and the disappearance of traditional filtering mechanisms. What can count as 'publications' on the web?

The harvesting practice of the BnF can be represented in the form of a triangle: at the base is the periodical harvesting of complete domains like .fr (France) or .nc (New Caledonia). This is largely done by automated processes and requires no cooperation with the publishers. Data are indexed, but not individual documents. In this way, some 4 million websites are harvested once a year.

Higher up in the triangle, websites selected by BnF librarians and partners are subject to frequent in-depth crawls. At the top of the triangle there is 'direct deposit' of carefully selected websites for which cooperation with publishers is established. In this layer comprehensiveness is again aimed for.

BnF cooperates closely with national and international partners. Selection is not decided in Paris only, but with librarians from across the country. Tools and services are shared with regional libraries via SPAR ("Système de Préservation et d'Archivage Réparti", "Distributed Archiving and Preservation System"). International cooperation takes place in the framework of the International Internet Preservation Consortium (IIPC).

(3) Future work

Ingrid Parent then proposed that the PERSIST content task force should continue its work on the problem of selection on the basis of the presentations and the discussions in the UNESCO session at WLIC, and to present at the meeting of the policy task force that is scheduled for early 2015 to write Guidelines for digital selection. For this goal, a position paper should be ready by the end of the year.

The Lyon meeting seems to point to three different approaches that could be pursued:

(a) Further elaboration of the questions raised and the suggestions put forward by the think piece of van der Werf and van der Werf, e.g.:
- how can heritage institutions in practice select on the basis of *use* (the bits about the bits) instead of on the basis of *content* (the bits)?
- How can governments influence users to change their on-line behavior in directions that are conducive to digital preservation?

(b) Stimulate the sharing of experiences and the division of labor between heritage institutions:
- Which lessons can be drawn from the experiences that have been made by web harvesting by institutions like the BnF and the NLNZ?
- What parts of the digital domain will remain uncollected if current practices are not changed, and how serious would be the losses?

(c) The ethical-political dimensions of archiving:
- How can UNESCO support countries and institutions with the development of good archival laws and – even more important – with the compliance to these laws?
- Is it possible to adopt ethical-political guidelines for international cooperation between archives, or for the selection of heritage in other countries? These could encompass both digitisation projects and the harvesting of born-digital heritage from the non-national domain.

References

Programme of the UNESCO Open Session: The PERSIST project – Digital preservation and content selection:

http://conference.ifla.org/ifla80/node/354

B. Van der Werf & T. van der Werf *The Paradox of Selection in the Digital Age*:

http://library.ifla.org/id/eprint/1042

M. Pickover, Michele *Patrimony, Power and Politics: Selecting, Constructing and Preserving Digital Heritage Content in South Africa and Africa*:

http://library.ifla.org/id/eprint/1023, http://library.ifla.org/id/eprint/959

C. Oury Preservation of born-digital heritage at the National Library of France

http://library.ifla.org/959/

Steve Knight & Winston Roberts *Digital Collecting at the National Library of New Zealand*

http://unesco.nl/sites/default/files/uploads/Comm_Info/nl_cims-535962-v1-nlnz_presentation_at_ifla_unesco_session_-_lyon_19_august_2014.pdf

Report of PERSIST Launching Conference : A Digital Roadmap for Long-Term Access to Digital Heritage, the Hague, 5-6 December 2013:

http://www.unesco.nl/sites/default/files/uploads/Comm_Info/digital_roadmap_-_report.pdf

Vancouver Declaration: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/unesco_ubc_vancouver_declaration_en.pdf

NSLA Principles of Digital
Collecting: http://www.nsla.org.au/sites/www.nsla.org.au/files/publications/NSLA.digital_collecting_principles_March2013.pdf