A Digital Roadmap for Long-Term Access to Digital Heritage

Conference organized by UNESCO, ICA and IFLA
in The Hague, The Netherlands, on 5 and 6 December 2013

Introduction

Keeping our digital information for current and future use is a cornerstone of the Information Society. Even though massive global information systems are in place, digital preservation is still a major challenge. The transition from research and pilots to large-scale, operational, long-term preservation is now accelerating but still has a long way to go. Finding sustainable economic solutions to safeguard the digital output of public and private sectors requires close collaboration between governments, industries, memory institutions and other stakeholders, including creators and consumers.

UNESCO's Charter on the Preservation of Digital Heritage already highlighted the need for broad co-operation a decade ago. The Vancouver Declaration, emanating from the 2012 UNESCO Conference 'The Memory of the World in the Digital Age: Digitization and Preservation', underlined this need more explicitly: "There is a pressing necessity to establish a roadmap proposing solutions, agreements and policies for implementation by all stakeholders and corresponding to national and international priorities which include the right to information, open government, open data and electronic government."

A consortium of institutions based in The Hague, the Netherlands, agreed to provide for a follow-up to this Declaration . These institutions are the Netherlands National Commission for UNESCO, the International Council on Archives (ICA), the International Federation of Library Associations and Institutions (IFLA), De Koninklijke Bibliotheek (the National Library of the Netherlands), LIBER and the Digital Heritage Netherlands Foundation. The Dutch Ministry of Education, Culture and Science has provided funding for the activities of the consortium in the period 2013-2015. The partners organized two events in 2013. The first was a preparatory workshop in Marseille, France, in October, entitled 'The Vancouver Roadmap – involving industry and government in problem-driven cooperation for digital sustainability'. The second, the main event, was the Digital Roadmap Conference held in The Hague, The Netherlands, on 5 and 6 December 2013. On the first day, around 80 participants were in attendance, representing governments, non-governmental and intergovernmental organizations, academia and the private sector. On the second day, discussions continued in a smaller assembly to take decisions on the next steps for the roadmap. This report summarizes the discussions and conclusions of the December conference as a whole.

Expert and panel discussion, 5 December

Miriam Nisbet, US National Archives and Records Administration, chaired the conference and delivered an opening address. Recalling the outcomes of the Memory of the World Conference in Vancouver in 2012, she reconfirmed that preserving digital information for current and future use is an urgent problem. "Large amounts of information in a dizzying array of formats is created every day", she said. "Ensuring access to this information into the future is critically important to prevent the loss of digital assets." Such loss, she emphasized, could have financial or legal consequences for any organization, and preventing digital amnesia and ensuring access are preconditions for the development of science, education, and the economy. "Digital preservation should be an integral part of modern information management, but we have not yet found a satisfactory solution to this problem", Nisbet stated. Any solution should at the very least include the following aspects: open

formats and standardization, and systems to maintain the integrity of the data over time. This requires cooperation between digital memory institutions, governments and industry, but their efforts are as yet fragmented. “They work in silence, and don’t share their findings with a long-term vision.” The current conference, as Nisbet concluded, would aim to explore solutions, identify all crucial stakeholders in this discussion, and set out an effective way forward.

Margaret Hedstrom, University of Michigan School of Information, argued that digital preservation is a social and economic challenge rather than a technical one. Commonly explored solutions, she felt, are generally too expensive, and do not scale to the vast quantity of data that needs to be preserved. Currently used software for digital preservation is not up to the task and requires too much human intervention on the part of the creators of digital content, the stewards, and the users. According to Hedstrom, a fruitful approach would be to care for and maintain media, to migrate data to software-independent formats, and to create adequate documentation.

This, however, gives rise to the question of who is responsible for digital preservation, and who will bear the cost. “Although memory institutions have long wished that creators would take most responsibility to create information that is easy to conserve and that is compliant with standards and requirements, this is not what happens in practice”, said Hedstrom. “Addressing this issue involves accepting the wisdom of the crowd in judgments on the value of information to be preserved and on priorities; it also requires memory institutions to shift from doing all of the preservation work to creating environments that allow digital preservation to happen from the outset. We should make digital preservation part of a process of continually improving the value of data, every time it is used. The more fun we make it, the more likely it is that people will engage.”

Over the years, much knowledge has been created about digital preservation, Hedstrom pointed out. “But we don’t know very much about the demand. That is the missing link in our knowledge. Who needs this digital information, over what timeframes, for what purposes, in what forms, and what would they be willing to pay to gain continued access?” Such exploration, as she concluded, could provide a basis for a division of responsibilities between governments and private sector with regard to how to satisfy that demand. “This should happen equitably”, she stressed, “so that individuals with a right to access are not disadvantaged by their ability to pay.”

Anne Thurston, International Records Management Trust, addressed the digital preservation challenges faced by developing countries. “Ultimately, it’s the governments and the donors that support them that will drive good governance and capacity for digital preservation”, she said. Developing countries’ increasing dependence on digital preservation makes them increasingly vulnerable, she continued, as digital information is generally not being structured. Metadata are not created, nor are digital repositories established. There is a lack of professional capacity to address this problem. At the same time, much documentation is still on paper. “The lack of preservation is a high and growing risk to citizens”, stated Thurston. “When regulatory frameworks and capacity are lacking, people are vulnerable to injustice.”

Thurston highlighted the general expectation that increased accessibility of digital information will allow people to question their governments’ actions, which will ultimately diminish the loopholes that currently allow for abuses such as fraud, environmental degradation and human rights violations. “But all of this requires trustworthy information”, she noted, “which is a notion not currently addressed on the post-2015 MDG agenda.”

The main question, according to Thurston, is how to fund the necessary transition. “We cannot expect it to be paid for by heritage institutions, nor by business until the demand pays for it”, she said. “Rather, we should aim to move the issue to government agendas by linking it to wider issues.” This can be done, she suggested, by framing digital preservation as a basis for social justice, and thus by linking it to development goals. She recommended forming partnerships with international donors and lenders and working closely with the World Bank. “This approach will roll out to development banks and governments across the world. The main challenge is how to articulate the real impact on people’s lives.”

Chalida Uabumrungjit, Thai Film Archive, pointed out that audiovisual archive issues are usually not considered a priority in discussions on digital preservation. She noted that in this context, ‘digital preservation’ is often confused with ‘digitization’, and stressed that digital preservation of audiovisual materials entails acquisition, inspection/identification, cleaning/repairing, duplication/making preservation copies, cataloging, making access copies available, and storing the original material in a climate control storage.

Uabumrungjit drew attention to the Open Archival Information System, the commonly accepted reference model for digital preservation systems, noting that in order to comply with international best practice for the long-term management of digital assets, any archival preservation system needs to fulfil the requirements set out in this standard. “But in the audiovisual context, this system creates some complications”, she said, “for instance relating to copyright and the mere quantity of the information.” An uncompressed feature film, she stressed, represents about 1.5 Tb of data, and film material is accumulating faster than it can be digitized. “Secondly, some films also represent a state of art, just like paintings do. Even a really good copy cannot replace original cinema. You have to preserve that experience for future generations.” Digital preservation is not just about objects, but also about the user experience of the objects.

Neil Grindley, digital preservation programme manager at Jisc (a British organisation that originated as the Joint Information Systems Committee), addressed the importance of estimating the cost of digital preservation. He said cost estimates are needed to: offer realistic and cost-effective curation services; support strategic planning; support tactical decision making; provide evidence of cost-effectiveness and value; enhance an organization’s credibility; and strategically align an organization. He named several areas in which cost estimates could be beneficial, including development, public-private partnerships, education and training, open data and open government, and risk management. However, he identified several challenges, including the poor usability of cost estimates and the lack of consensus on benefits and strategies. As potential solutions, he identified the provision of a standard vocabulary, a common cost-benefit model that can be adjusted to specific use cases, and clearly defined user guides and tools.

Regarding vocabulary, Grindley advocated the use of the term ‘digital assets’ rather than merely ‘digital information’. “The term digital asset sets up an expectation that the digital object in question has value and therefore requires curation”, he said. “But the analogy with financial investment doesn’t follow through, and we should not expect it to.” He doubted whether international collaboration would work beyond the operational level, suggesting that some competitiveness will always exist between organizations at the management level.

In the ensuing **question and answer session**, the speakers clarified and expanded on several issues:

- On the demand for digital information, Margaret Hedstrom explained that a substantial amount of work is already being done in the area of market research, but that the challenge lies in translating this information into useful products.
- Anne Thurston reaffirmed the connection between digital preservation and the post-2015 development agenda, with Margaret Hedstrom remarking that access to information and preservation should not be regarded as assets in themselves, but rather as important means to transparency, knowledge, education, cultural expression, etcetera. Neil Grindley noted that the value that is added along the chain of preservation and use should be made more explicit, and called for predictive analysis in this regard. Anne Thurston emphasized the need for increased communication between various development communities, but expressed confidence that "a tipping point has come".
- On incentives for researchers to embed access parameters into their research process, Margaret Hedstrom was not convinced that compelling researchers to share their data is the most successful mechanism to use. She stressed the need for a system that makes it easy, simple and transparent for scientists to comply, but said that "at the same time we need to understand and respect their reservations and constraints of sharing everything with everyone at once. But the better perspective we have on which data we need to preserve for the long term, the easier this will be."
- On the recommended 'radical redistribution of responsibilities', Margaret Hedstrom called for a move from a staged set of events where creators, information managers and users each have their separate tasks to a flow where data is continually improved. She also noted the need for a radical redistribution between human beings and intelligence systems. Neil Grindley agreed, noting that this calls for systems that are more intelligent. He added that a radical redistribution also implies changing behaviors, and making people realize that adding extra pieces of metadata will not only help other people, but also themselves.

Participants then engaged in a panel discussion.

Joe Macri, Microsoft Vice-President for Europe, Middle East and Africa, presented a business perspective on systems for sharing and preserving digital information. He regretted the fact that data storage practice is currently fragmented, and advocated using cloud computing solutions. Drawing attention to Microsoft's public cloud service called Windows Azure, he noted that an open computing environment reduces the risk of content becoming inaccessible. "Our aim is to build a body of companies that are committed to this shared objective", he said. "The cloud paradigm carries the promise to deliver, all the way down to the local level. We are confident that we can continue to innovate, and continue to ensure the long-term access to digital assets." All the same, Macri remained realistic: "Whatever we do, it has to be economically sustainable."

Martin Berendse, ICA President, acknowledged the fact that private parties need to rely on a business model in order to actively contribute to digital preservation solutions. He said that in order to allow them to adopt a long-term vision in this regard, other stakeholders need to establish and communicate their prioritization of needs, and help to find a financing model. "A redistribution of responsibilities may also mean that each party should focus on its strengths", he said. "Perhaps heritage institutions should not focus on technical solutions, but rather on societal solutions and creating linkages, for instance with industry."

Ingrid Parent, University of British Columbia and IFLA President from 2011-2013, said that libraries are not just there to provide access to information; rather, they are a force for change: 'strong libraries, strong societies'. She highlighted, among other things, collaboration with the World Intellectual Property Organization (WIPO) to ensure access to information, and IFLA's focus on being proactive in determining which archives and museums are at risk. Flagging IFLA's latest trend report (trends.ifla.org), she said one of the trends identified is that new technologies will actually expand as well as limit access to information, because of a lack of standards, financial resources and will to act. "In conclusion, we can only solve this through collaboration: private and public, government and non-government."

Javier Hernández-Ros, European Commission, said four aspects need to be paid attention to when dealing with the issue at hand: the organizational, legal, financial and technological perspectives. On the challenges of enabling access, he identified several questions: who owns the data, and who is on the receiving end? What happens when a project ends? "Some European member states have national strategies in place", he said, "but these have to be mainstreamed if big data and cloud computing are to serve as a driver for economic development. Any solution needs to be ambitious, but pragmatic. We need to collaborate towards common goals and measure effects and impact."

Joseph Alhadeff, chief privacy strategist at Oracle Corporation, and International Chamber of Commerce, stated that "preservation is not an end, but a means to an end". He stressed the need to identify which data needs to be preserved, noting that the vast majority of data is unstructured, and often produced as a byproduct. "Everyone may add value to information", he said, "but everyone may also destroy value by adding incorrect information. Crowdsourcing is not necessarily enhancing the information. It makes it more difficult for people to get credibility for what they're producing." A digital roadmap, he felt, therefore needs to focus on how to deal with this unstructured data.

Tjeerd de Boer, senior policy advisor, Cultural Heritage Department, Dutch Ministry of Education, Culture and Science, underlined the importance of governments taking part in these discussions and determining which public values need to be taken into account. "Use and preservation are two sides of the same coin", he said. "In order to maintain public support for heritage conservation, the information should be used." He highlighted the triangle of 'content, connectivity, and competence', noting that it helps us to look at different aspects of public value at the same time, and stressing that investments need to be made not just on the content side.

Joie Springer, programme specialist, UNESCO Communication and Information Sector and Memory of the World Programme, said we have to "make sure that preservation covers not just the bits, but also their meaning, and the information has to be contextualized". She stated that appropriate systems and strategic planning are needed to ensure long-term access to digital information, naming as examples systems that are forward-compatible, or can easily be migrated to new environments. She said the challenge lies in capturing global attention for these issues, and as societal values change, it is not easy to predict what the future needs will be. "Preservation is costly, but rebuilding even more so."

In the ensuing discussion, participants addressed the following question: **how can we strengthen collaboration and identify common goals?** Several ideas were offered, including:

- Relying on contracts between e.g. governments and user groups, rather than on individual contracts between a supplier and an individual consumer;
- Focusing on the beneficial aspects of competition, for instance, the fact that it drives innovation;
- Taking a policy-driven approach;
- Taking advantage of public-private partnerships (although the scalability remains a challenge); and
- Defining roles and responsibilities more clearly.

Addressing the question **which stakeholders should participate in this discussion**, and in which capacity, several participants remarked that coming up with appropriate business models is neither the responsibility of governments nor of memory institutions, but rather of interested companies. Others stated, however, that as soon as impact and value enter the discussion, it is indeed the memory institutions that have the necessary expertise with long-term care and delivering value to people. One participant felt that there is a role for the World Bank, notably its sector planning and government reform departments, and another stressed the importance of involving data creators as well as policy makers in the discussion. Opinions diverged on whether/how to reach out to the public at large. Some felt that it is more useful to narrow the scope and more clearly define what it is that needs to be preserved. "That may be the most difficult question of all. What might not be important to us today, may be hugely important to others tomorrow."

Participants also discussed how to create **incentives for industry to participate in the discussion and to collaborate to create new strategies and business models**. Collaboration, they concluded, may well be the most fruitful approach. Additional attention needs to be paid to legislation, and to funding schemes, for instance by reserving a set percentage of any research budget for preserving the resulting information.

At the end of the first day's discussions, participants formulated the following **conclusions and recommendations**:

About roadmap collaboration:

- We need more collaboration, but it takes creativity to find the right partners and representatives (and some compromising).
- We can't do everything together, so we need to define what the divisions are in the work needed for digital preservation.
- This dialogue can contribute to a better understanding (including metrics) of the needs for the use of data over time. That is where the business opportunities lie.
- We need a better contribution from content creators to the dialogue.

About domain perspectives:

- Memory institutions want to and should focus less on internal technological matters, and more on societal issues for establishing continuity of content.
- The focus of major ICT companies is on attracting more users. Content follows the users.

- Governments should look at information policies (including copyright) and sustainable funding, and contribute to the overall organization of the work (but should not interfere with business models).

About continuity of content:

- As this is a complicated topic, we need to be ambitious in our goals but pragmatic in our actions, and strive for more automation. It is not in the interest of any party to slow down innovation, even if it is disruptive.
- We should focus more on the continuity of data, and less on the continuity of media and software. We should also recognize that the majority of data that currently goes around is unstructured and not curated.
- We (i.e., society) need to define better what it is that we want to preserve (from a user perspective, both public and private) and identify how the roles in the information chains are changing.

Discussions in a smaller setting – 6 December

The conference resumed the following day in a smaller setting. Jānis Kārkliņš (Assistant Director-General for Communication and Information, UNESCO) attended the meeting by video conference.

Conference Chair Miriam Nisbet opened the session, highlighting the main conclusions and recommendations from the first day's deliberations. Summing up the challenges of information management, she said: "If you can't find it, you don't have it."

Participants debated **the definition of 'data'**, with some noting that since data implies 'raw data', it may be better to refer to 'digital information' instead. The latter term is broad enough to cover all relevant aspects, and also includes the meaning of data over time. Some felt, however, that in the context of unstructured data and research, the term 'data' may be preferable.

Addressing the issue of collaboration, participants discussed the need for a **platform where different stakeholder groups get together** to discuss digital preservation. Jānis Kārkliņš noted that there is an Internet Engineering Taskforce that sets internet standards, serviced by the internet society, volunteers and engineers, which addresses issues that need to be dealt with from a standardization point of view. The digital preservation community could consider setting up a similar organizational platform. Kārkliņš suggested that UNESCO could serve as a secretariat for such a platform. Others noted that such a platform may already exist, for instance the Research Data Alliance, which has working groups on issues such as metadata structuring, policy and preservation; and the Preservation and Archives Special Interest Group (PASIG), although the latter focuses mainly on technical issues. Margaret Hedstrom (University of Michigan) felt that "we need a strategy, not another organization". Joseph Alhadeff (Oracle Corporation and International Chamber of Commerce) supported forming a "malleable organization, with ad-hoc participation, and a nucleus that has sufficient representation". In any case, he urged the inclusion of 'market makers' in the discussion.

Following this last remark, participants acknowledged the fact that there are **different types of industry players**, as well as **different types of end users**, which calls for an analysis of their different

perspectives. “This will help us to understand their roles, where they are overlapping, how they are changing and how to create appropriate markets.”

There was strong support for increased effort to **identify the data that need to be preserved**. Javier Hernández-Ros (European Commission) stressed the broadness of the problem and the necessity to mainstream policies. Naming the example of the healthcare sector, he called for balance between privacy protection and data mining. Natasa Milic-Frayling (Microsoft) said the “fundamental, intrinsic problem is that the value of information depends on contemporary circumstances”. She stressed that business opportunities need to be made more explicit. Others added that there is a general need for readily understandable examples of the crucial importance of data preservation – with the understanding that ‘big data’ and ‘preservation of digital heritage’ are two different things.

Participants agreed on the need to **engage with policy-making groups** to establish what the preservation policy frameworks need to look like. There was a call for **interdisciplinary connections**, and for a vision on which issues can be addressed at the level of organizations or nations, and which ones require higher-level cooperation. Participants agreed on the need for **setting priorities and making choices**.

In conclusion, participants set up the following **framework for action**:

- **Goal:** enhancing the sustainability of the information society by establishing continuity of content (data, records, information);
- **Organisation:** a global platform with the convening power of UNESCO to bring together a policy arena, a content curation arena and a technical requirements arena, under the auspices of a Digital Preservation Board;
- **Proposition:** “What can I do for you and you for me?” (manifested in business models)
- **Method:** pragmatic, agile, quick-wins, open-minded, using existing channels of influence
- **Main questions to address:**
    - What are we going to preserve (balancing legal frameworks, institutional policies, and user-driven needs)?
    - Who is responsible for enabling the preservation?
    - How are we (the Platform) going to act on that responsibility (understanding that roles are changing)?
- **Time frame:** 1 year (with the next meeting in spring 2014)

In rounding up the discussion, Miriam Nisbet noted the need to “ensure the continuity of our own working group”, and suggested reconvening in conjunction with ICA and IFLA meetings in the next year. Margaret Hedstrom suggested adding two additional questions to the proposed framework: which are the business models and how is this effort going to be funded? Rapporteur Elco van Staveren presented a visual representation of the meeting’s proceedings, featuring the acronym that had been suggested as a name of the proposed platform: PERSIST (Platform to Enhance the Sustainability of the Information Society Transglobally).

Miriam Nisbet closed the meeting, congratulating participants on the “concrete steps that have been identified towards a Roadmap for Long-term Access to Digital Heritage”.