**Persist Meeting, April 20-21, UNESO**

1. Opening Session

Indrajit Banerjee, Director of the Knowledge Societies Division of UNESCO's Communication & Information Sector welcomes the participants. He mentions the Vancouver Declaration and stresses that long-term preservation and accessibility of documentary heritage are key issues that need broad cooperation, including public and private partnerships. The new normative instrument that UNESCO hopes to adopt at the coming General Conference will bring a new impetus to the organisation's endeavours and ambitions in this direction. UNESCO appreciates the cooperation with IFLA, ICA and the Netherlands National Commission in the PERSIST project.

The President of the IFAP programme, Chafica Haddad gives a short introduction to the IFAP programme and the important work that has been accomplished in its many priorities areas.

Marco de Niet explains the structure and the 'philosophy' of the PERSIST project. Digital documentary heritage has quickly become the primary means of knowledge creation and expression, but there are no sustainable economic solutions for continuity of content. It is the shared responsibility of public and private parties to face this problem. PERSIST wants to find solutions via a global policy discussion between government, memory institutions and ICT-industry (the PERSIST triangle). Around these partners we can position the research community that should provide the policy discussion with the necessary empirical information about society at large, as digital continuity is a concern for all. PERSIST preferably uses existing channels of communication, and aims at an agile and pragmatic and open-minded cooperation.

1
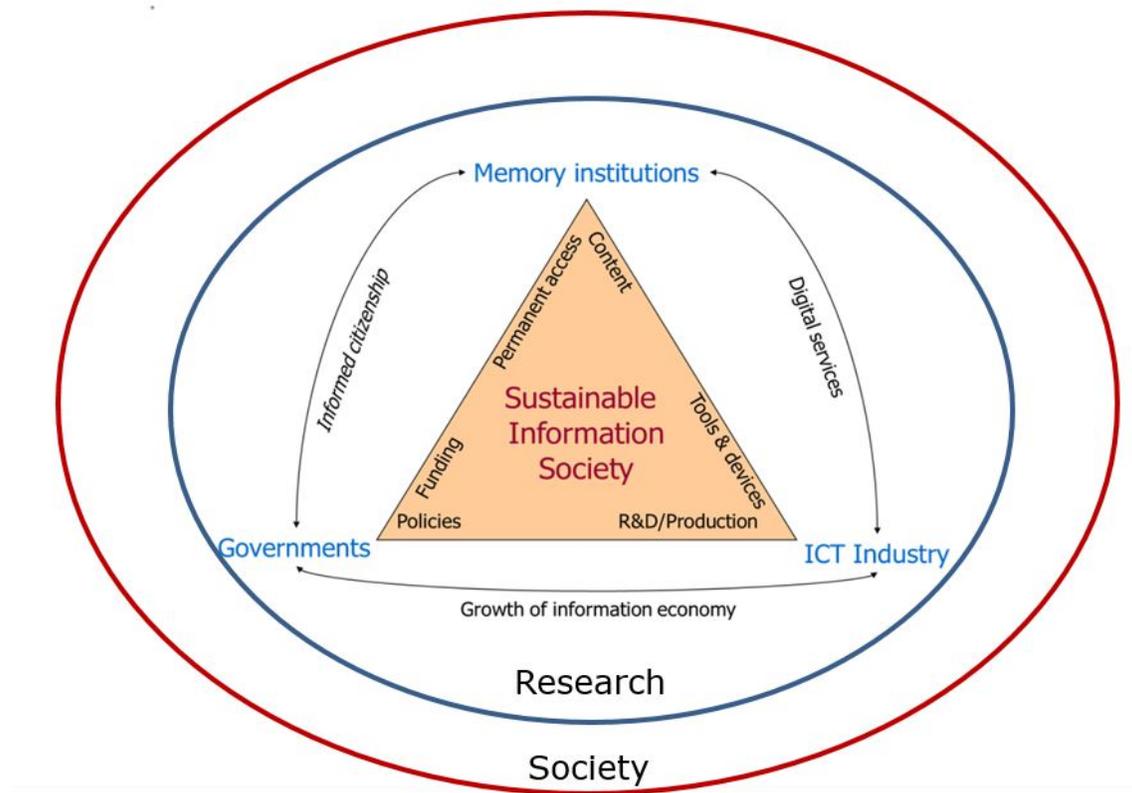
Selection for preservation should not only be driven by the opinion of experts, but also by *user needs*. A basic difference in the business models (in the widest meaning) that relate to documentary heritage is the following: for the ICT-industry, content follows users, but with heritage institutions and other public organisations it is the other way around: users follow content. Commercial partners are primarily interested in understanding behaviour and interest of information users, while for memory institutions most strategic decisions are derived from the heritage collections they manage. The 'global policy debate' which is PERSIST's ultimate aim should make the different stake holders aware of these differences. Lack of proper understanding of these different business perspectives can hinder cooperation. But cooperation is necessary, because digital amnesia is everybody's concern.

David Fricker and Ingrid Parent shortly introduce the work of the Technology Task Force and the Content Task Force respectively. Some of the main points of the ensuing discussion:

- The word 'ICT industry' does not cover this angle of the PERSIST triangle. It should also include civil society actors and universities. A better term might by the ICT-community of the ICT-ecosystem (Van den Oever).
- The PERSIST project is timely, because today it is still possible to determine who are the inheritors of the intellectual property of the products of parties like Microsoft and Google. Later, when these companies may no longer exist, this information will probably no longer be available (Mahadev).
- The PERSIST Steering Committee should decide to start a Business Task Force to carry out work for the register that falls beyond the orbit of the Technology Task Force (Milic-Frayling).
- Software and content are dissimilar; the protection for a text or a film given by classical copyright ends at a certain moment in time, and the content comes into the public domain. But old layers of software are constantly reused in new products, and so companies cannot give the licenses for it away. But they donate the executables of it to the Repository (Mahadev). Still, software is collected by heritage institutions, like the French National Library (Parent).
- PERSIST should not be too exclusively document-orientated and connect more with the museum world (Crofts). Likewise, the project should tap into the knowledge existing in computer museums (Delve) and gaming communities (Mahadev).
- There is a clear responsibility for the collection of the social media that accompany documents (Noha).

On the coalition with IT-community. How should it look like. Value proposition that we can offer. Natasa input.

Panel discussion on the policy issues around a Global Registry for Legacy Software

Participants; David Fricker (President ICA, chair), David Burrows (Microsoft), Fridrich Strba (LibreOffice), Jonas Palm (Swedish National Archives, memory of the World Sub-Committee on technology, Iskra Panevska (UNESCO), Jonathan Tilbury (Preservica)

At the ICA conference in Girona the idea of a Repository was born. Experts are invited to think about this in more detail and to formulate the value proposition that PERSIST could offer. The discussion centres around the following questions:

- Who could benefit from a Global Repository for Legacy Software?
- What suggestions major software tool providers may have about preserving old tools, considering that this is their legacy as well;
- What issues companies may have if they had to change licensing agreements or any other practices;
- What benefits they may gain, e.g., if they could 'deposit' software into the 'software bank' as soon as the demand for the software is in decline.
- What role has UNESCO, as the heritage institute of the United Nations, to play in making the repository a reality?

For the meeting very different types of companies and organisations have been invited, that might want to use the repository to donate legacy software, to use the collection for its work, or both. So the whole range of the ICT ecosystem was represented:

Preservica Tilbury explains that Preservica is a commercial vendor of preservation technology since 2003. Its clients are mainly smaller heritage institutions that lack the resources and know-how to this type of work for themselves. Repositories are not a new thing; Preservica works with PRONOM that is maintained by the UK National Archives (with a considerable investment of time and money - they would undoubtedly be very happy if their project could be transformed by UNESCO into an international undertaking). Preservica would probably be interested to use a repository for legacy software to donate its own software (migration tools and rendering tools); Preservica's corporate history is a part of the brand.

LibreOffice started its Document Liberation Project when it realised that the free en open source software it was developing could not handle a lot of materials from commercial vendors well, like Visio and CorelDRAW. The projects aim to understand the structure and details of proprietary, undocumented file-formats and encode the information in an open file format. LibreOffice's own software is copyleft, so uploading it in a repository would be no problem.

Microsoft: Microsoft, as one of the commercial companies in the ICT-ecosystem, has dramatically changed and is now very open to sharing and cooperating. Commercial vendors typically work with 5-7 years life cycles for software products. Windows XP had a long production cycle with its 13 years. In an ideal world Microsoft would like to keep its software updated longer. But the company is interested in cooperating with initiatives like PERSIST to tackle the problem of obsolescence. Microsoft would in principle be prepared to upload its legacy software, but the business model behind the Repository should be much more worked out then it currently is. But now is the good time to make this effort, as others have also stressed.

Mr Piacentino (Mozilla) expresses uneasiness with the term 'business model'. This word, however, should not be taken in the narrow sense of 'making money', but more broadly as 'creating value'. It is suggested that we should talk about 'sustainability models' instead of 'business models'.

Other companies that are in the centre of the ICT-ecosystem should be won over to join this coalition. There is a certain optimism that they will ('they would applaud to most of what PERSIST is aiming at' Killbride), but it would be good to have more insight in what uses that these companies could draw from the repository, so that we have more arguments than those of 'good citizenship'. De Niet stresses that there is also value creation in sustainable services. Fricker proposes to combine benefits for society and benefits for business – we should aim at nothing less than a change in the business culture of companies. Anderson warns that we will not succeed if we were forced to a fight with company lawyers. We must make our future partners clear that we are on the same side. Milic-Frayling points to Mozilla, a not-for-profit foundation, that nevertheless can survive in a very competitive environment.

Iskra Panevska stresses the need to involve national heritage institutions that have experiences with legal repository. She is convinced that UNESCO can make a strong case with companies if it stresses the cultural value of the materials that have been created with software.

Panel discussion of the Technology Task Force

The work with the companies for the Repository should ideally develop into a policy discussion on future-safe formats – provided, of course, that these don't hamper innovation. The Repository solves part of the special problems of digital continuity that has arisen in this first stage of the digital era. The solutions for new generations of software should be made 'higher up' in the production chain and is, for now, beyond the scope of the PERSIST project.
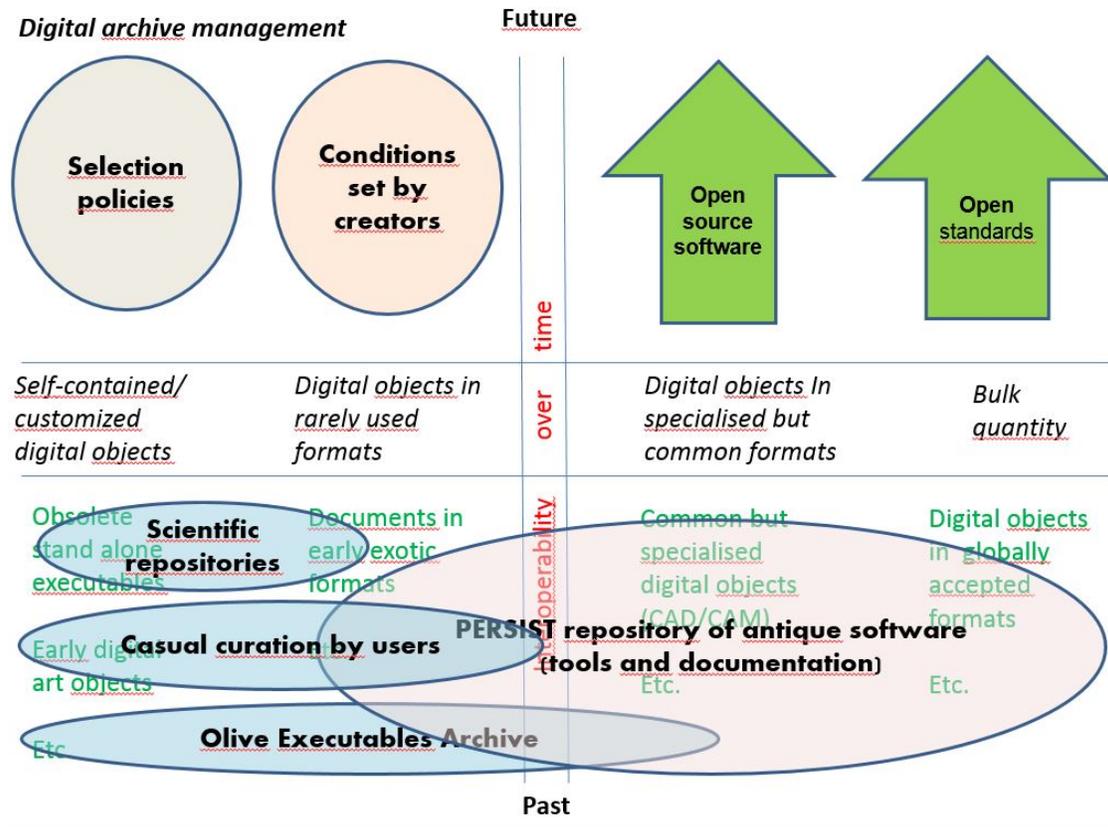
Van den Oever proposes a matrix of the transparency of software that could evolve into a rating of software. The register could help to bring such a rating system into being.

Strba stresses that working around the source code in order to test the behaviour of software is no rocket science: it has proven that it can be done. Only some documentation and simple metadata about the file forms in question are needed in order to figure out the algorithms.

Various experts speak about the special case of the preservation of games. For this type of heritage, the only preservation option is emulation. The need for good metadata is especially important. UNESCO could help to make meta-dating in this area a legal obligation (Anderson). Jason Scott has worked on a repository within the Internet Archive (https://archive.org/details/internetarcade). This work is very successful for games for '80 and '90s, but if you want to preserve the full range of software, you must go beyond Java. (Mahadev). Palm highlights the precarious situation with (private) pictures, and Tilbury mentions the many different formats in which scientific heritage is preserved.

Mahadev would applaud if PERSIST could help defining the notions of 'antique' and 'fair use' in the case of software.

Based on the discussions of the meeting of the technology task Force, De Niet proposes a fourfold division of digital objects and tentatively positions the PERSIST repository and other archives in the following way:

Digital archive management

Future

Selection policies

Conditions set by creators

Open source software

Open standards

time over

Self-contained/ customized digital objects

Digital objects in rarely used formats

Digital objects In specialised but common formats

Bulk quantity

interoperability

Obsolete stand alone executables

Scientific repositories

Documents in early exotic formats

Common but specialised digital objects (CAD/CAM)

Digital objects in globally accepted formats

Early digital art objects

Casual curation by users

PERSIST repository of antique software (tools and documentation)

Etc.

Etc.

Etc.

Olive Executables Archive

Past

2. Policy task Force - Good practices in national policies for digital sustainability

1. Marco de Niet: *A national infrastructure for digital preservation in The Netherlands*

The *Nederlandse Coalitie voor Digitale Duurzaamheid* (NCDD: *Netherlands Coalition for Digital Sustainability*) is network of national organizations with large digital collections initiated by the Ministry of Education, Culture and Science in 2008. It was inspired by the (DPC) in the United Kingdom. The Network presented a National Strategy for Digital Heritage in March 2015, comprising of three work programmes, centred around the words *visible* ('increase the visibility of collections, explore user demand, and promote the use and re-use'), *usable* (improve the possibilities for using collections by making them jointly accessible online, connecting and enriching data, and developing targeted services) and *preservable* (cross-sector sharing, utilisation, and scaling up of facilities for sustainable preservation and access, while devoting attention to cost management and the division of duties). The work programme *preservable* continues the original NCDD programme, the two other elements are new.

Dutch organisational culture is traditionally not inclined to top-down approaches and favours safe-regulation, but for this subject some an initiative form the Ministry was needed. The main NCDD principle is: share what is possible, keep separate what must be kept separate. The NCDD uses the

phrase 'Dutch collection', although no real centralised national digital collection exists. Knowledge and infrastructure are key interest. NCDD discourages small institutions to have their own e-depot.

Website: http://www.ncdd.nl/en/

2. Sudha Gopalakrishnan: *Digital Agenda - The Indian Scenario*

The production of digital materials is growing rapidly in India. The government stimulates access to the internet by investments in broadband and wifi in universities and schools via an initiative called 'Digital India'. The Information Technology Act has recently been changed in a more liberal direction. Large strata of Indian society are empowered by the digital transition.

The Department of Information Technology set up a National Digital Preservation Programme that aims at the long-term preservation of both born-digital and converted data. It is managed by the Centre for Development of Advanced Computing (C-DAC) in Pune (http://cdac.in/).

Various initiatives for digital heritage are implemented, both on the level of government and by NGOs. An interesting example is Sahapedia, an open, online encyclopaedia of Indian culture (http://sahapedia.org/).

Difficulties in implementing these various programmes stem from the size and internal diversity of the country: besides funding constraints and lack of standards, there is the sharp divide between city and rural areas and a lack of coherence between the various initiatives. Smaller collection need more attention than they have received up to now.

3. William Kilbride: *Digital Preservation Coalition (DPC) and Digital Preservation in the United Kingdom and Ireland*

The DPC works for four themes:

- Advocacy: for a political and institutional climate receptive to the need for digital preservation
- Training: competent and responsive workforces
- Technology and organisation: better tools, smarter processes and enhanced capacity
- Networking: closer and more productive collaboration

DPC has developed and continually updates a handbook for the Preservation Management of Digital Materials (www.dpconline.org/graphics/handbook/) and publishes Technology Watch Reports and other materials.

Recent trends that influence DPC's work are the growing recognition that digital sustainability is an important problem; the growing interaction with institutions that are not traditional heritage institutions, for example old people's homes; tool integration and data security & sensitivity – a vaster issue than obsolescence.

Kilbride stresses the importance of computer museums for digital preservation. Losing data is losing opportunity , but deletion of data, if it happens on purpose, is part of the solution. But the notion of 'data' itself is becoming more and more problematic. The borderline between content and software

is blurring. File formats is not a really problem. The real problems are capacity, workflows and money.

4. Robert Buckley: Digital initiatives in the United Arab Emigrates

The United Arab Emigrates are a young nation comprising Abu Dhabi, Dubai en five smaller Emirates. Most documents that record the early history of the region are outside the country (United Kingdom, Portugal, Netherlands). In 1968 the Center for Documentation & Research (CDR) was established, out of which the UAE National Archives grew (2008; http://www.na.ae/en/default.aspx).

A lot of the digital activities of the National Archives take place in the framework of the UAE National Charter 2021. This initiative, focused on the 50th birthday of the Emirates in 2012, aims to stimulate electronic governance in the country, and helps the strengthening of a national identity in the United Arab Emirates as well. Via the Watheq initiative, the National Archives increase awareness about heritage amongst the population by offering to scan personal records. People learn in this way that their family records are valuable at the national level, and that they can donate their archives. There are also programmes to stimulate the documentation of oral history.

The United Arab Emirates are an interesting case for UNESCO to study: its Archives are very young, at there is no national policy yet. There is a lot of training and awareness raising to be done, and a lot of the know how has to be imported. At the same time it is a technologically very advanced country.

From the ensuing discussion:

- Maréchal explains the plans of the International Association of Sound and Audiovisual Archives (IASA) for training and organising knowledge and invites PERSIST and UNESCO to partner in this programme. Annex A of this report describes these plans in greater detail.
- Programmes for access are in different levels 'inclusive' for people with disabilities. In the Netherlands there are certain legal obligations for government institutions, but not for museums. In Northern Ireland there are good practices. UNESCO organised a big conference on this theme in India in 2014 and currently works on guidelines for this theme.
- Kilbride mention the need for certification and benchmarks. De Niet agrees; the NCDD stimulates harmonization between the participating institutions, but strives not for not unification. There is more need for diversity then for unity. There should agreement on a minimum of standardization in order to participate in cooperation. If you ask too much compliance to standards, then you will lose the smaller institutions.
- When do digital sustainability or ICT for people with disabilities a viable business option for the industry? Awareness raising is necessary here: if you realize how much companies spend on litigation (220 billion dollars in the USA alone!), then record keeping can be appreciated as an important money saving investment.

Technology Task Force Parallel Meeting

Attendees:

David Fricker   Michiel Leenaars
Fred Van Kan   Fridrich Strba
Vincent Wintermans   Satyanarayanan Mahadev
Ryder Kouba   Rony Vissers
Jonas Palm   Marco de Niet
Guy Marechal   David Anderson
Jonas Palm   Hamad Al Mutari
Natasa Milic-Frayling   Iskra Panevska
Jos van den Oever   Robert Buckley
Noha Adly   Jean-Pierre Evain
Bram van der Werf

1. Welcome and Acceptance of Agenda

   • Meeting agenda accepted by attendees

2. Confirmation of the Minutes of the Technology Task Force Meeting of 2 March 2015

   • Minutes of 2 March 2015 confirmed

3. Presentation by Mr. Satyanarayanan Mahadev: Olive Executable Archive

   • Mr. Satyanarayanan discusses the challenge of archiving executable content (e.g. games, AutoCAD, word documents) due to many moving parts (hardware, operation system, etc.)

   • Possibility to preserve the entire environment, including the operating system, using a virtual machine (VM). Downside of VMs is huge file size.

   • Olive takes inspiration from YouTube and streaming software, rather than downloading. Access to files is not linear (i.e. like movies), but uses demand paging prefetching to quickly access files.

   • Mr. Satyanarayanan demonstrates Olive Sustainable Archive, showing the group Windows 3.1, the Great American History Machine educational software, and the NCSA Mosaic browser. These were all run online.

   • Options: pack up and carry the entire environment with you including OS.

   • Create a virtual machine, benefit of only preserving one piece of software (Olive) to access older software.

   • Online demonstration of Olive showed Windows 3.1, the Great American History Machine (educational software) and the NCSA browser (first browser, was able to load current web pages). The host environment is open-source, but the software is proprietary (e.g. Windows 3.1), so not accessible online to the general public.

   • The VM runs through VMNetX, which is open-source software allowing demand paging and prefetching. The VM runs a hardware emulator, which allows older operating systems and software to be used. The base layers need to be maintained, but the older software needs no constant updating.

- VMs can be run through the cloud, and the marginal cost of adding thousands of VMs goes down. This creates opportunities for collaboration in preserving software (e.g. nation X preserves Apple software, nation Y - Microsoft, etc.)

- Olive Executable Archive can be found here: https://olivearchive.org/

<u>Questions/Comments</u>

Van der Werf: Cost of the service is in people to maintain system, not in system itself. Libraries/archives don't have technical competence to run software. What is the business plan, and who will pay for it?

Satyanarayanan: Necessary to keep moving Olive forward to sustain archiving more recent material.

Maréchal commented that it would be great to find content, browse, select, navigate using Olive.

Satyanarayanan: crowdsourcing would be better than relying on librarians and archivists to create virtual machines; allow people to review quality of virtual machine and professionals to create host environment. Funding from archives/libraries to provide licensing fee for software.

4. <u>Presentation by Natasa Milic-Frayling: Technical Feasibility of Preservation</u>

- Digital obsolescence is an economic problem: digital preservation systems are digital and they need to be updated.

- The pragmatic approach is to virtualize entire desktop environments with their software; however, there are legal issues. Licenses need to be written to accommodate new uses (such as letting multiple users at once).

- The two main approaches to preservation are migration and emulation.

- For economic sustainability staff, infrastructure, and agile business development are necessary.

- Computer scientists are valuable partners, the Association for Computing Machinery (ACM) supports PERSIST as "an important step to ease effects of digital obsolescence and ensuring that we have effective means to reduce risk to digital assets." ACM can also help mobilize computer scientists to think about digital preservation.

- This is a good time to move on digital preservation; the Global Repository can be a sustainable solution to long-term access.

<u>Questions/Comments</u>

Van Kan: Archivists are always lagging behind industry and there needs to be a balance between innovation and preservation.

Milic-Frayling: VM/emulation allows viewing of old software on contemporary environment; necessary to continually maintain virtual machines on contemporary systems.

Strba remarks that there are multiple ways of digital preservation (both emulation and viewing files/their informational value).

Milic-Frayling agrees that both are valuable and that all options must be covered.

5. <u>Presentation by Michiel Leenaars: Reliable Persistance</u>

- Documents come from hybrid origins, and so no one single application can claim to reliably render every document.

- It is important to establish 'fair use' to use software that organizations have already paid for and to explore the possibility of re-licensing old software after a certain time period.
- Personal archiving and empowerment of people can lead to "casual curation" of digital materials.
- Creating and maintaining objects that aren't searchable is too expensive, since they lack functionality.

6. <u>Fred Van Kan: Role of Memory Institutions in Digital Preservation and Access</u>

- Archives have a long tradition of preserving heritage, therefore it is necessary to preserve and access digital materials.
- Van Kan gives an overview of history of digital records in archives as well as current challenges archivists are facing today.
- Two main ways to preserve and access digital materials: emulation versus migration.
- Emulation has the challenge of copyright restrictions, however it is easier to emulate rather than migrate files; migration is not a long-term solution for digital preservation.

7. <u>General Comments and Questions</u>

Fricker: How do you balance requirements of emulation with access?

Satyanarayanan: Cost can be shared and affordable for all if spread out over institutions; UNESCO can host events to raise awareness of opportunities and challenges, and financially coordinate funding for development of software.

Wintermans sees the role of UNESCO as a "global coordinator" who can point the stakeholders to tasks for digital preservation that have to be carried out.

De Niet argues that copyright is the elephant in the room. Is it viable for PERSIST to harmonize views on copyright issues?

Satyanarayanan: What does UNSECO want specifically? It is important to state this clearly. Leading vendors (e.g. Microsoft) can receive badge/recognition for signing up for making material accessible (value of cultural heritage).

8. <u>Seeking Consensus (the complete list of bullet points is Annex B)</u>

- Attendees discuss the definition, scope, and issues that need to be resolved for the Global Repository in an open forum. The definition decided on for the Global Repository was: "Repository should contain tools that enable heritage institutions to render and interact with digital heritage".
- Discussion focused on the content of the repository, and whether it would contain only tools to access digital content, or digital content as well. It was decided that the repository would focus on preserving software.
- Tools and documentation in the repository should be open source.
- It was recommended to begin with a pilot project, in order to work out the best practices, policies, and procedures which will set the foundation for future. The pilot can show value of PERSIST to potential partners and gradually accustom ICT-industries to the idea that participating in the repository is not harmful.
- The pilot program will have a limited number of licenses, which will direct the licenses obtained. The pilot will target "antique licenses" which are defined as older than ten years; companies could offer a limited license for the duration of the pilot.
- Discussion of how the Global Repository will achieve sustainability.

9. Tactical Implementation of the Global Repository (the complete list of bullet points is Annex B)

- Discussion from previous section continued, focusing on partnerships, sustainability, Repository governance plan, and the pilot project

- Potential partners include: national institutions (archives/libraries), open source community, and digital preservation groups around the world

- Discussion of revenue streams; limited options with UNSECO rules

- Crowdsourcing community is a possibility; individual institutions can create new virtual machines using open source software

- Ask Content Task Force for help in finding use cases for piloting

- Ask Policy Task Force for: potential licensing models, fair use, definitions of "antique", raising awareness with governments, requirements for preservation in the future

10. Next Meeting

- Video conferences, as well as plug into conferences for in-person meetings

- Have a virtual machine running with commitment from ICA, IFLA, and some institutions. Have a management and governance structure in place. Define brand for PERSIST and value proposition for potential partners

Content Task Force Parallel Meeting

Attendees (writers of the Guidelines in **bold**)

Ngian Lek Choh	Pio Pellizzari
Clement Oury	Wilbert Helmus
Katarzyna Ślaska	Robert Buckley
Sarah CC Choy	Tom De Smet
Robert Fisher	Ingrid Parent
Nicholas Crofts	Julia Brungs
Sudha Gopalakrishnan	

Presentation: Selection and collecting born digital heritage - Best practices and guidelines – Wilbert Helmus (presentation slides will be published at PERSIST's website: http://www.unesco.nl/digital-sustainability):

Information about the discussed Enumerate project can be found here: http://www.enumerate.eu/

Discussion following the presentation:

E-legal deposit:

This falls under legal deposit rules but there are selection procedures in place. The national library section at WLIC 2014 discussed this (http://conference.ifla.org/past-wlic/2014/ifla80/node/324.html). The National Library of France has a priority website list which is created by subject librarians during their day to day work. Furthermore the BnF archives according to specific topics (e.g. World War One).

Sound and Vision, the audio-visual archive of the Netherlands, has in its mandate the archiving of all public broadcast information. A selection process, by media coordinators, takes places when material is submitted.

The British Library includes web archiving in their legal deposit strategy.

General discussion on selection:

It was agreed that the documentation around the selection process also needs to be persevered.

Many institutions struggle with preserving multi-format materials.

Furthermore it was agreed that low usage of content does not diminish its importance. Low use might be due to access restrictions and availability.

Archivist approach selection of digital content differently, as they also regard personal documents not on the web.

The situation for museums is different again as most of the digital data is the information around the actual museum object. Often this information is spread over several systems, a unified policy across museums on how to archive this information would be recommendable.

It was agreed that selection is necessary even though the technology community believes that everything can be preserved.

Presentation: The Paradox of selection in the digital age – Bram van der Werf:

The presentation can be found here: http://library.ifla.org/1042/

Discussion following the presentation:

The subject of use equals a preservation priority came up again and it was stressed that preservation policies might drastically change over time and as a result erase history in the process.

The format of the content is not essential as long as it can be re-purposed. As for most other aspects of long term preservation, metadata management is essential to guarantee value of the content even after format change.

Institutions often see only the short term costs (staff) rather than the long term storage costs. Often commercial companies step in and take on some of the preservation efforts but these do not have a long term mandate and if/when they disappear, the content will be lost.

One long term solution might be the cooperation between small heritage institutions to tackle preservation and resulting financing problems.

The selection problem is not only confined to the digital world but often to the material world too as we cannot assess what part of today's waste will be tomorrow's heritage. However, the financial implications if everything is preserved will not be bearable especially if all content would be expected to be available online.

A further problem with preservation is duplication of content.

Privacy issues around content were discussed as these are prominent in a large amount of content preserved (e.g. personal letters, indigenous content).

A further preservation issue is the clash between a political agenda and the need to preserve a regions history (e.g. Hong Kong umbrella movement, Dutch Colonial History – Tropenmuseum).

Summary of morning's discussion/leading questions resulting from this:

- Do we have selection or do we go comprehensive on content?

Selection is both necessary and not. Human selection might not be enough to cover the full spectrum of what needs to be preserved and some random selection might need to be put in place.

Not everything can be preserved due to financing issues and the fact that amazing metadata would be needed to find content. Therefore selection is inevitable.

- What formats are included?

Are principles for all formats needed? This will be determined once the drafting of the guidelines has started.

A differentiation between online and offline materials might have to be made.

- Definitions and scope:

Terms needing definitions (not exhaustive):

Digital heritage

Collections

Content

Heritage

Objects

Metadata

Born digital (for reference see [2003 UNESCO charter](#))

Digitisation is often already a selection but not every digitisation is made for long term preservation but often for immediate access. Digitised material needs to be included in the Guidelines. However a priority will be on born digital material as there is no analogue back up/original.

Length?

The aim is to produce Guidelines with a length of 3-5 pages.

- Who is the audience for these guidelines?

The audience was identified as Member States of UNESCO, cultural institutions, research institutions, collection management institutions (including commercial ones and NGOs), and national level memory institutions. The secondary audience was identified as the information technology community.

- What are the roles of cultural institutions and can they be shared? Multi-institutional storage solutions?

It was agreed that collaboration on this issues is essential not only on a regional but also on a global level.

It was stressed that developed countries have a responsibility to help developing countries to preserve their heritage for the long term.

Legal responsibilities and restrictions occur and should be tackled in cooperation.

- Is legal deposit a selection?

Yes, there are selection criteria within an analogue and digital legal deposit.

However, not every country does have a legal deposit.

- Is sampling a legitimate kind of selection?

Questions arose on what the best way would be to preserve dynamic material. Sampling does not give an overview of what content will be preserved when the process starts, however it is a valid measure.

Cooperation and coordination between different institutions would guarantee a more effective and holistic sampling result.

Criteria of when sampling is valuable and when not needs to be included in the Guidelines.

- What is access for selected material?

Access is strongly and closely related to rights issues on content.

Open access does not mean that this is free.

Public access is part of the selection criteria for selection.

It needs to be made clear that content is worth preserving even if access is restricted for a long time (copyright restrictions) or ever (cultural/personal privacy restrictions).

- Different rules for web-pages and publications and other formats – format specific criteria

Tension between digital content and analogue content formats which is often due to resources.

There is also the problem of not being able to be comprehensive when it comes to archiving the web.

A further criterion is the availability of the content as the format might cease to exist and immediate preservation is necessary.

Generally it was agreed that different strategies for different resources might be necessary as some are more stable than others.

- Acquisition policies

Acquisition and evaluation policies are also selection criteria. Library and Archives Canada has 5 criteria to determine whether to preserve content.

- Criteria

Possible criteria:

Technical possibilities

Content quality

Value over time

Significance

Copyright restrictions

Quality is not always the most important issue, some low quality content needs to be preserved due to the uniqueness of this content. But content should be preserved in the highest possible quality available.

In a social media context, the accompanying metadata (e.g. comments section) is essential to understand the content and needs to be preserved. A rights issue on who owns this content might come up.

It was suggest working on decision trees for selection (6 steps like the Dutch model)

- Legal compliances/constrains

A focus of the Guidelines will also need to be legal compliances.

- Distributed preservation – replacements?

A risk management of copies needs to be in place. Content should not be preserved over and over again but some safe copies are needed. Institutions need to cooperate on this issue.

- Roles of designated communities as selectors?

Can we use the participator culture as selector?

Communities can be used to source content which otherwise would be lost. Furthermore communities and experts are also stakeholders of institutions and therefore in content selection and should be used as advisory groups.

- Preservation strategies – impact on format selection

Is migration and emulation a content or technology questions?

Institutions should be aware of how to preserve certain formats before acquiring them. However some content will need to be acquired regardless of formats.

As formats evolve, the ICT community will need to support the content selection community.

Format changes almost always results in loss of information, it needs to be insured that no vital part of the content is lost in this process and the significant properties are preserved.

- Format based selection

Should content in obsolete formats be preserved? Concrete information on the content (provenance etc.) needs to be available to prove that preservation is necessary.

If content is identified as essential, format takes a secondary role.

- Metadata

It was agreed that metadata is important as it describes the content. Both contextual and provenance metadata needs to be preserved.

The following questions/points were identified but not discussed:

- Value not content
- Who selects?
- Contextualisation

Discussion with writers group to define next steps for writing the Guidelines:

Structure of the Guidelines:

Preamble/introduction: Ingrid and Julia

Definitions of terms: Sarah

Sampling/ collection strategies: Clement

Role of institutions (esp. legal deposit): Ngian

Shared nature of long term preservation, redundancies: Katarzyna

Legal implications: Sarah

Metadata: Nick

Assessment value (risk assessment and costs), significance (current and future, including the decision tree): Rob

Social media/types of resources and strategies to capture these: ?

Timeline:

First draft in Google docs by June 1

Call with the group after the ICOM General Assembly: http://doodle.com/cf569xasc3gc24yb (please note, due to the geographical distribution of the writers, this might be late/early for some, we do hope that this would still be possible) Please fill in the doodle by **15 May 2015**

First draft by WLIC 2015 (mid-August 2015)

Working method:

A Google Document has been created to work within your sections: https://docs.google.com/document/d/1EbM7CbCBH3Hf5XC36C_sMsiYBZcbbmjZDi1ByDncpDU/edit?usp=sharing

Technology issues identified during the content discussion:

Relevant results from Wilbert's survey for technology task force to look at

File format conversions

De-duplication

Quality of content for preservation (e.g. youtube video has an original on a local computer – provenance)

Ability to preserve content (technical implications)

Metadata

Contextualisation

Random selection of content

Access to preserved content – not useable (soft and hardware obsolete)

Policy issues identified during the content discussion:

Information around museums objects can be found in multiple places, a unified policy on where to store museum object information is needed

Duplication policies, where and what is the master copy?!

Open access and open source is not free

Digital is global and laws are national – issues of exchange and access – jurisdiction – copyright restrictions

Access- can UNESCO speak on behalf of heritage institution to show governments that access should not be time restricted for keeping content

Cooperation – encourage developing country partnerships

Further actions:

CHIN guidelines – Rob to look into

BnF guidelines – Clement to elaborate

Annex A – Contribution of the IASA in PERSIST "Consultation meeting of experts" (see page 7 of the report)

Prepared by: Guy-Noël Maréchal Coordinator of the IASA-Organizing Knowledge Taskforce

The International Association for Sounds and Audiovisual archives [IASA] has a long tradition of collaboration with the UNESCO-MoW programme and its management, M. Abdelaziz Abid, Ms. Joie Springer and more recently Ms. Iskra Panevska.

The next IASA Conference will be held in Paris at the Bibliothèque nationale de France [BnF]. The keynote speech at that conference will be presented by the UNESCO.

At its last Conference, in Cape Town, IASA has decided to propose its collaboration to the UNESCO-MoW dynamics, as expressed in the UNESCO document [dated 2013-03-04] "Draft Action Plan for strengthening the Memory of the World" and "Preliminary Study of the Technical, Financial and legal aspects on the desirability of a Standard-setting instrument" and expressed in the associated document [dated 2014-09-04] "Draft Recommendation concerning the Preservation of, and Access to, Documentary Heritage in the Digital Age". IASA shares the objectives of that dynamics and has many relays in the UNESCO National representations. The details of these possible IASA contributions have been expressed in French and English in a 10 pages document sent end of 2014 to the attention of Mrs. Irina Bokova and Mrs. Iskra Panevska.

Today, IASA is represented by two persons: M. Pio Pellizzari one of the Vice-Presidents, in charge of the Education and Training Committee and me Guy-Noël Maréchal in charge of coordination the IASA-Organizing Knowledge taskforce.

The MoW 'Action Plan' and the associated 'Recommendations' on 'Documentary Heritage' have four main objectives involving many aspects

In particular, the PERSIST dynamics covers three of these aspects in the broader scope of 'Digital Heritage' (a scope larger that 'Documentary Heritage') in three taskforces:

1. The selection of Digital Born Contents

2. The technologies for the maintenance of their access

3. The associated policy management

The IASA proposal of contribution covers two other synergetic aspects, the "Training & Education" [T&E] and the "Organizing Knowledge" [OK].

The IASA proposal of contribution covers two other synergetic aspects, the "Training & Education" [T&E] and the "Organizing Knowledge" [OK].

- The 'Training & Education" is under the coordination of M. Pio Pellizzari. The IASA T&E covers a large range of modules. It covers recommendations and concrete hands-on and best practices for "Audio-visual digitization, recommendations on formats, Ethics, Cataloguing and Archiving" (six standards [some under ISO standardisation]) and one introductory module to "Conceptual Modelling". The modules are dedicated to 'training the trainers' or 'training the managers' or concrete education to the delicate operations. The IASA suggested to the

UNESCO MoW programme to create (or coordinate the creation) of a T&E platform in which the IASA modules could take a place, in parallel with contributions from other experts. Obviously, contributions from the PERSIST project should be integrated on selection of contents on the maintenance of old platforms, on emulation, on virtualisation, on migration and on policy management.

- In parallel with the T&E proposal, the IASA has made a proposal on the 'Organizing Knowledge' challenge. The Organizing Knowledge covers mainly the registration of documentary heritage, its modelling, its networking and contextualizing; the management of the rights, the interoperability (temporal, systemic, historical and cultural); the modelling of the processes associated with the life cycle of the Documentary Heritage Items (in particular the archival and maintenance processes and the involved contents: the Submission Information Packages [SIP]; the AIP and DIP of the OAIS standard.

  Created 5 years ago, the IASA-OK dynamics has now reached results. One of them is the adoption of an approach which has been successfully applied in diverse domains: the "Conceptualisation".
  - o The most well-known of these achievement is the CIDOC-CRM standard related to the 'Museum Objects". The key element of the approach is that the concepts can be mapped to 'Object Oriented' modelling or to pure 'Semantic' modelling or by 'Domain Ontology' modelling. I see here two key persons involved in CIDOC-CRM, M. Nicholas Crofts and M. Gerald Grunberg (and through him Patrick le Boeuf).
  - o The same power has been applied successfully in the audiovisual sector by the European Broadcasting Union. M. Jean-Pierre Evain is here and has been the designer of the EBU-Core and its plug-in for 'News' and for 'Sports'.
  - o The intention to model in a conceptual way the processes jointly with the modelling of the Works (in particular for the archival and maintenance 'the OAIS standard') started 8 years ago. In that period I was representing the non-profit-association TITAN in a European project called "MEMORIES" in which Mrs. Joie Springer was representing the UNESCO Memory of the World programme. During the project I have had contacts with M. Abdelaziz Habib, short contacts but so fruitful for the progresses. In 2006 the standards were not mature enough nor the computers powerful enough for practical implementation. Now, within the MediaMap project, the demonstration of the feasibility has been obtained and, for that, the project and the consortium have been granted of the "European Price Award" for the more promising contribution in the IT applicability by the Celtic Eureka. That conceptual model has been named "AXIS-CSRM". It is agnostic with regards of the way of modelling the 'Works' or of the modelling of the 'Processes' but has hooks for 'workflows' and for 'documentary heritage' items. In particular, the analysis have been made for hooking CIDOC-CRM, EBU-Core, GATE (an ontology for textual semantic representation) and Finite State Machines (local or in SaaS).

That is the reason why IASA has proposed to contribute in the OK domain with the assistance of the npa TITAN and the support of the EBU.

In practical, TITAN and IASA will organise a two days' workshop on the OK approach. We envisage of have it two days before the next IASA conference that will be in Paris at the BnF starting 28th September by a key note address by the UNESCO. The objectives of the workshop are on two folds:

- Opening a large consultation on AXIS-CSRM models and similar
- Open an analysis on the elaboration of a 'Proof of Concept' [PoC] project with typical DH content illustrating the approach.

A 25 pages summary of AXIS-CSRM is available in French and English. It will be sent to all the members of the Consultative meeting organised by PERSIST. The experts will also be invited to the OK workshop and possible AXIS-OK Proof of Concept. Tomorrow, at the parallel meeting of the Technology taskforce, I could elaborate more on the OK dynamics. In particular, the approach could be successfully applied for the migration of Documentary Heritage through Editorial Data called 'Semantic Alignment'.

Pio Pellizzari   Guy Maréchal

  Vice President IASA   IASA-OK leader

  IASA-Training & Education   TITAN-Research

  pellizzari@fonoteca.ch   guy.noel.marechal@gmail.com

  Office : + 41 919 61 64 00   Mobile : +32 496 21 14 99

Annex B – Notes taken by Ryder Kouba on overhead screen during "Seeking Consensus" and "Tactical Implementation" discussions of the PERSIST Technology Task Force (see page 10 of the report)

1. UNESCO can lead creation of repository, deposition of executables and informing community of risks

2. The definition adopted was: The Global Repository should contain tools that enable heritage institutions to render and interact with digital heritage.

3. A piloting effort to create one or more Virtual Machines in order to gain experiences with the organizational and legal challenges of obtaining legacy software.

4. A foundation (non-profit organization) provides interfaces to be used by third parties and delivers them based on the requirements of heritage institutions and a broader market.

5. Keep not just the tools, but the tool chains; focus only the technical dimension.

6. Through the piloting effort over a finite time period to develop the best practices and learnings needed to provide the foundations for the future.

7. Need to have a way to start collecting permissions, using various levels of licensing which enable usage.

8. Pilot will have a defined set of use cases, which will direct the licenses obtained.

9. Targeting executables of ten years or older (antique).

10. Scope goes beyond commercial executables, commission creation of tools; ie both the "creation" and "curation" of virtual machines.

11. The Repository may be implemented as a central place of documentation which points to where assets are; in distributed repositories which hold assets that are interoperable with each other.

12. How can we tell when software should be incorporated into the repository?

13. Monitoring system for file formats, relationship with PRONOM and other networks.

14. Secure and reliable and trustworthy repository.

15. Deposit has to include management layer to technically manage software .

16. Is the repository a back catalogue or an application programming interface (API)?

17. Use case, appealing to OAIS model, will give insights into what services the repository may deliver in the future.

18. Community support for ongoing sustainability.

Tactical Implementation of the Global Repository

Community partnerships: national institutions (archives/libraries) for pilot, open source community, accessibility, Digital Preservation Coalition, National Digital Stewardship Alliance (US), Digital Preservation Technical Registry (New Zealand), ACM.

For profit companies could offer a limited license for the duration of the pilot.

Repository should have management, authority and governance system .

Usage agreement for the life of the pilot with companies.

UNESCO can play a role in the bringing companies and institutions together for license agreements.

Pilots should be written in a few months to show the value of PERSIST to potential partners.

Technology Task Force needs to come up with resource plan and management plan in order to make pilot successful.

Sustainable operation requirements:

Member states/institutions are going to have to contribute, e.g. subscription model (hurts trust in community, can't work with UNESCO).

Revenue stream necessary to continue development.

Examples of subscription models already exist (e.g. Archive-It) for heritage institutions.

Trust is very important for crowdsourcing community as well as good communication (UNESCO website and focal point, simple server to bootstrap community, education).

Community can run open source software in their institution as well as creation of new virtual machines.

Zooniverse potential way of bringing people into the project.

To Content Task Force: help in finding use cases.

To Policy Task Force: potential licensing models, fair use, definition of "antique", raising awareness with governments, looking forward to preservation in the future.

Next meeting: video conference, plug into conferences for In-person meetings

Have a virtual machine running with commitment from ICA, IFLA, and some institutions. Have a management and governance structure in place. Define brand for PERSIST and value proposition for potential partners.

Annex 4 – List of Participants

Noha Adly	Ministry of Communication & IT, Egypt
Hamad Al Mutairi	National Archives United Arab Emirates
David Anderson	University of Portsmouth
Julia Brungs	IFLA
Robert Buckley	National Archives United Arab Emirates
David Burrows	Microsoft
Sarah Choy	Legislative Council of Hong Kong
Nicholas Crofts	ICOM
Janet Delve	University of Portsmouth
Getachew Engida	UNESCO
Jean- Pierre Evain	European Broadcasting Union
Robert Fisher	Library and Archives Canada
David Fricker	ICA
Sudha Gopalakrishnan	ED Sahapedia, India
Chafica Haddad	IFAP
Wilbert Helmus	Helmus Advies
Fred van kan	Gelders Archief
William Kilbride	DPC, UK
Ryder Kouba	American University in Cairo
Michiel Leenaars	NLnet
Maria Liouliou	UNESCO
Guy Maréchal	IASA/Titan
Natasa Milic-Frayling	Microsoft Research
Lek Choh Ngian	National Library Board, Singapore
Marco de Niet	DEN
Jos van den Oever	Ministry of Internal Affairs and Kingdom Relations
Clément Oury	ISSN
Jonas Palm	MoW Tech Subcie
Iskra Panevska	UNESCO
Ingrid Parent	IFLA
Pio Pellizzari	CCAAA/IASA
Jean Baptiste Piacento	Mozilla
Mahadev Satyanarayanan	Carnegie Mellon University
Katarzyna Ślaska	Poland National Library
Tom de Smet	Institute for Sound and Vision
Fridrich Strba	LibreOffice
Rony Vissers	Packed (Flanders)
Bram van der Werf	van der Werf Technologieadvies
Vincent Wintermans	UNESCO Netherlands National Committee