

https://dspace.org/

PRINCIPLES OF THE DIGITAL HERITAGE

Layton, Roger

2012-06-14
http://hdl.handle.net/10539/11517

Downloaded from a DSpace repository. https://dspace.org/

PRINCIPLES OF THE DIGITAL HERITAGE

Roger Layton

Roger Layton Associates (Pty) Ltd,
PO Box 7036, Westgate, 1734,
South Africa
Tel: +27 (82) 881-0380,
roger@rl.co.za

Presented at the 2nd International Conference on African Digital Libraries and Archives (ICADLA-2), University of Witwatersrand, Johannesburg, South Africa, 14th – 18th November, 2011

Abstract:
We were commissioned in late 2009 to create the National Policy on Digitization for the Department of Arts and Culture. This policy was subjected to public participation in early 2011, receiving positive comment and critique, and has now moved onto the next stage in its life cycle. As the project leader for this policy development I can now look back and reflect on the processes that led to the formulation of the recommendations and in particular how I envisioned the future of digital repositories and libraries. I made a specific point in the policy that heritage has the longest agenda of any human activity, which could be seen as eternal, with our largest group of stakeholders being future generations. Whereas libraries have been in existence for thousands of years, these have always involved the management of physical books and related documents, and the manner in which libraries are managed and administered has changed very little in this time. Computers have only a 60-year history, and initial computerized systems were primarily used for library administration and indexing. It was only with the introduction of the first digital networks and the World Wide Web that it was possible to share and cross-reference electronic documents using digital networks
.
The Digitization Policy was the core output of our work with the Department of Arts and Culture, and provided the key policy statement on how to enable a long-term digital heritage that was preserved and was accessible, and for which the issues of intellectual property were clarified. However, there were many other outputs that derived from this policy, including a collection of best practices associated with the digital heritage in all of its forms. I called this the Digital Heritage Body of Knowledge (DHBOK) and this was structured into the core areas of Principles, People, Processes, and Practices. In this paper I focus on the Principles which form the basis on which key decisions can be made when confronted with the plethora of choices that do confront organisations during various processes within the digital heritage. I could not prescribe a single set of choices or priorities that would cover every situation and every context, but I did manage to formulate a general set of principles to guide decision-making. The outcome was a set of 15 such Principles, drawn from extensive research of the available literature, of current practices and standards, and from interviews with key local and international stakeholders. These are the key Principles that define the notion of "success" within any digital heritage project, no matter whether this is a digital library, digital archive or virtual museum.
Keywords: eHeritage, digital heritage, oral history documentation, digitization policy, Linked Data, Web 3.0, Semantic Web, RDF, liberation struggle, victims of conflict, biographies

Introduction

The National Policy on Digitization was a large project, with the final outcome a Policy Statement which underwent considerable public consultation and review before being committed in March 2011. We were commissioned to produce this in late 2009, and it was intended to provide guidance on the increasing demand for digitization of the country's cultural heritage and how such requests should be handled. Our work was scoped to prescribe parameters on digital preservation, access and intellectual property. I was merely the "hired-hand" in creating this policy, and this is the property of the Department of Arts and Culture.

The Policy document was accompanied by a number of associated documents, provided as appendices to the Policy itself, of which one was the Digital Heritage Body of Knowledge (DHBOK). This DHBOK was to be the statement of collective best practices for the digital heritage community and was also intended to be a living document, encapsulating the changes in the practices of digital heritage, while the Policy represents the enduring and long-term position. The DHBOK is still under development and will be released for limited review shortly.

The development of the DHBOK proved to be a considerable challenge, and whereas the first versions focussed on the practices of digitization, it was a key requirement to include the entire spectrum of the digital heritage into a single practices document; a framework was established at the early stages of the project which then formed the basis for all other work. It was essential to the acceptance and relevance of the DHBOK that it was tightly connected to the Policy and the other documents, since these documents make specific reference to areas in which the DHBOK should be consulted for details of implementation. For example, whereas the elements of a digitization strategy are identified within the Policy, the details of how to create such a strategy can then be sought within the DHBOK.
It is not possible, within this short paper and presentation, to outline the full scope of the DHBOK as it currently stands, and I have thus focussed my attention on the vision that I used to create this, and the Principles which are incorporated as one integral element.

A Framework for the Digital Heritage

Our initial conception of the knowledge required within a full treatment of the digital heritage was largely limited to an analysis of world-wide best practices in the practical elements of conducting digital capturing of heritage. Much of this is already well-documented and available throughout the world, and our early work focused on collecting and collating this information. It became evident early in the project that a wider framework was required to encompass the totality of the digital heritage, and to explore how we can create digital solutions that can have a long, perhaps eternal, agenda – something that the entire ICT industry has failed to do.

The framework that I arrived at has been informed by a range of other best practice frameworks, guidelines and standards in a range of other disciplines. I sought an answer to the question, "What is the best practice framework to describe best practices?" This became

a meta-level problem before being then applied specifically to the digital heritage – in essence, to discover the best set of elements that would describe the digital heritage. The outcome of this was a framework consisting of five core areas, which I have referred to as the Five Ps of the Digital Heritage, and which form the integrated structure of the framework of the DHBOK.

The first area comprises the *Principles*, which provide the fundamental guidelines that prescribe, constrain and support decision-making throughout all digital heritage activities, including all of the Processes. It is these Principles which form the bulk of this paper.

The second area is the *People*, defining the roles played by organisations and individuals that are all stakeholders and interested parties to the digital heritage. These roles include Owner, Custodian, Repository, Funder, Producer and Consumer. They are conceptual roles which help to structure responsibilities so that there are clear distinctions for all levels of authority.

The third area comprises the *Processes.* One of my early structures was the development of an analytical framework for the study which led to the Policy, and one dimension of this analytical framework was a ten-stage process concerning the primary activities that cover the entire life-cycle of digitization and the digital heritage. From the initial framework, this ten-stage process has been refined, reworked and renamed in order to clarify the function and scope of each step. The steps are: Scoping, Strategizing, Planning, Preparing, Capturing, Describing, Loading, Storing, Accessing and Using. The framework is heavily biased towards action, and these steps provide the guidance on how these actions are to be conducted to meet best practice.

The actual Practices themselves -- "how to"-- are a specific set of guidelines, largely drawn from existing international best practices, covering a range of types of objects to be digitized, the challenges they present and the best practices associated with each: e.g., how to handle Dictabelt recordings and other near-obsolete technologies.

Finally, there is a range of Problem Areas which cut across the Processes, such as intellectual rights, legislation, and risk management.

Principles of the Digital Heritage

In my attempt to create a framework for the digital heritage I first developed a Process model, in which I examined the entire work of the digital heritage as a ten-stage process. This is a generic framework, and is applicable to projects, to programmes, to institutions as a whole and to the long-term and large-scale systemic efforts in digital preservation. This Process model informed the policy statements, but is not explicitly outlined within the Policy. In attempting to structure a best practices model, it became apparent that a focus solely on Processes and associated Practices was too limited and that a more fundamental level was required within this framework. This level is the fundamental level of the Principles that embody the underlying qualities and enduring values that define what we should be collectively aiming for in terms of best practice, in creating, managing, sustaining and

communicating the digital heritage. These Principles are intended to guide our work in the digital heritage at all stages.

We are continuously confronted with situations that demand decisions in implementing the digital heritage, and our response to each of these will have an impact long into the future. There are many choices and many variations for each decision we need to make and these Principles have been formulated to provide guidance to the decision-makers. While formulating these Principles it has been important for me to provide an explicit basis for these decisions, which can be then recorded as part of the provenance of the digital artefacts that emerge at the end of the process.

These Principles are not mutually exclusive and they overlap with one another. They are in a process of continual refinement as we learn more about why and how we make the right decisions. The Principles must be applied in a context-dependant manner, since there are many factors which influence which Principles are more important than others in a particular.

My conception of the best-practice Institutional Digitization Strategy includes a statement which explicitly states the basis of the decision-making process to be employed by individual institutions, identified by which of these Principles are to be used, and how they should be applied. I have learnt that there is rarely a right and wrong in terms of the decisions and have thus focussed on how the decision-makers will justify their decision-making processes in a way that they can then be held accountable for these decisions in the future.

These Principles are open to discourse and critique and will form a part of the DHBOK as its moves towards completion of the first public draft. The underlying goal of my work here is that there is such a set of Principles that we can turn to, and what follows is the first step on the path towards a principle-based digital heritage.

Principle of Principles

Any best practice structure is required to have an underlying set of principles that is widely accepted.

This principle is self-evident and positions the DHBOK as a principle-based set of guidelines and practices. It is this principle which requires strategies and decisions to be accounted for in terms of the application of specific principles. Each institution should make explicit which principles that they will use and why, and should also outline how they will be used.

Principle of Standards

Standards must be used where possible to enforce best practices and to enhance interoperability and these standards should be continually reviewed and refined.

Standards exist to ensure that we adopt consistent approaches which are widely agreed and which evolve. It is the absence of standards which has caused much of the problem of obsolescent technologies. Throughout the early stages of magnetic recording, photography,

and the digital world, each manufacturer created their own standards and this resulted in an almost total lack of interoperability of technologies.

This principle is applied by requiring each institution to select from an agreed list of acceptable standards the standards that they will be using, and then to justify why these are most appropriate to their needs. An example of this is in the selection of standards for digital images and digital audio, where there are many possible standards and each has separate benefits and risks.

Principle of Semantics

Digital heritage should not been seen as a collection of disconnected files and folders, but should contain powerful semantics that can enhance and embrace the context and meaning of the digital objects and render them capable of being connected in ways not anticipated at the time of their creation.

One problem in many early photographs and audio and video recordings is the absence of contextual information, which causes difficulty for historical research. At the time that these were created there would have been perfect contextual information available, but this was not recorded or made available. We find photographs in archives that cannot be connected to our world of history, but someone, sometime, knew why these were significant.

The current approach to implementing semantics is to create metadata for a digital object, using some metadata standard, and then to associate this with, or attach it to, the digital object. Modern digital standards allow for embedding of metadata within the file itself. This prevents the separation of the file content from the semantics that may occur if they are stored separately. Metadata is widely used for providing contextual information at the level of individual digital objects and new approaches to metadata are being developed which explore semantics within the content elements of digital objects (Layton, 2011). These approaches are also being used within the modern "Linked Data" movement, which is at the heart of the Semantic Web, also called Web 3.0 (Linked Data, 2011).

Heritage information, in all of its forms, is replete with meaning, and the management of heritage is one of the greatest challenges for information systems. The discipline of Artificial Intelligence has for a long time explored solutions for encapsulating meaning within data structures, in order to permit connection and reasoning from this data. This is the future to which we are headed, and we are merely at the start of the age of intelligent systems. When institutions consider the products of digitization it is essential that they consider how such semantics will be built into these digital objects so that maximum meaning is carried into the future. We should be providing not only the collections but the connectivity as well.

Principle of Collections

Heritage is best managed in collections, and this includes the digital heritage.

All custodians, including curators, librarians and archivists, create collections to enhance the management of physical items. Items may include museum objects, documents, books,

photographs, audio recordings, video recordings and a host of other types of items with enduring value. The digital heritage consists of both the physical and analogue items converted to digital form, as well as a completely new set of items that are "born digital", such as email records, digital photographs, digital audio recordings and a host of modern communications formats that use websites. The digital-age custodian is required both to continue the management of physical collections and to also manage digital collections; in both these functions there are respective requirements for long-term preservation practices. Digital technologies open up new avenues for considering the notion of a "collection" that is no longer bound by physical presence, creating opportunities for novel aggregations of content in the form of virtual museums, virtual archives, and virtual libraries, assembled for long or short-term purposes.

As part of a strategy, and in particular for the decisions on prioritisation, the custodial institution should clarify its position at the level of collections rather than at the level of individual objects, and should justify its position on selection at this level.

Principle of Universal Producers

The growth of digital content appears to have no upper bounds as everyone becomes a producer of such content, and as direct sensing and record systems increase in usage.

About ten years ago, at the end of the second millennium, digital cameras were emerging as a new technology, available at the top end of the market. Analogue photography was the primary method in use, with its corresponding requirement for film processing. These physical technologies had at that time been in place for a very long time, but were replaced almost overnight as digital cameras became available at low-cost. Today almost every mobile phone has a digital camera, digital audio recorder and digital video recorder, and their links to the Web allow such content to be sent to web servers and shared instantaneously.

Until the recent the past we used telephones primarily to speak to each other, and email facilities were only available to those with access to a computer. However, in today's world we have access to email, SMS, and a range of other Internet and Web-based services from our mobile phones. The net result is that everyone is now creating digital content, at an increasing rate, and this will become a difficult challenge for future historians.

On one hand it is good to have access to such content, reflecting the detailed thinking and actions of individuals during their daily activities, since very little of this was recorded in the past. On the other hand, the mass of data needs to be organized in order to be useful, reliable, accessible and authentic. Today's data and content is tomorrow's history, and if we do not store it we may lose it. Institutions that are responsible for recording history should be using today's content to assemble collections that can then be used in the future. For example, within the past few days, the ANC Youth League conducted a 50km march to the Union Buildings, and this may be seen in the future as a significant event for future historians. Today the information on this should be captured, from TV footage and professional photographers, as well as from the social media, Facebook and Twitter, and

from the mobile phone recordings from those participating, and any artefacts, such as posters, associated with this event should be collected. It is thus the responsibility of today's archivists to record today's activities for tomorrow's users. The question will be who will undertake this, since yesterday's news has already been discarded to make way for today's news in our throwaway world.

Principle of Respect for Rights

The protection and control of intellectual rights is a core economic principle of our modern world. This applies no matter whether there are commercial rights or open rights associated with digital content.

We live in a world in which there are legislated rights for owners of intellectual property. There is a range of rights, of which the most significant is copyright. This is a complex area of law; there is inadequate provision for the intricacies and variations of the digital heritage, and these issues have been identified as areas of concern (Nicholson 2008; South Africa. Department of Arts and Culture, 2010). Many online services are opting for an "open content" approach to distribution, rather than commercial archives, and this is challenging the continued role of commercial repositories. However, copyright clearance remains of concern even within the scope of open content systems.

Each institution must respect the rights of the owners of the copyright, as well as the moral rights (such as the right to be identified as the photographer of an image), and should take a position on how its own repositories are to be made available in terms of licensing of the content to the users. These are concerns that span the entire Process model for digitization, which I mentioned earlier, and they have particular relevance to the Using process, in which digital content is reused for other purposes, which may be commercial.

Principle of Trust

In the rapidly expanding world of digital content, there is an increasing need for access to valid, authentic, and reliable content. Provenance of digital content will become the distinguishing factor between trusted and untrustworthy repositories.

The majority of people will turn to Wikipedia, Google, Bing and Yahoo as their first point of attack when searching for information and the verb "to google" has become part of our daily vocabulary. However, there is no formal indication of the authenticity of the indexes, repositories and content resulting from such searches. It is important in the digital world to provide a distinction between trusted and untrustworthy repositories. In the non-digital world libraries, archives and museums have held originals whose authenticity can be established through provenance processes. The essential problem in the digital world is the ease with which digital resources can be copied, manipulated and changed; to be fully trusted, digital provenance is an essential but highly elusive goal.

I have concluded that trust is not a binary property of a digital repository but is a spectrum of positions, from the completely untrusted up to the totally authentic, and that a range of actions, with increasingly correspondent cost structures, can increase the trust level for a

digital repository. Whereas anyone can today create a repository and place this on the Web for public and commercial access, it is essential to ask how trusted this repository is and who says that it is trusted. Is it sufficient for each individual repository to state that it can be trusted or is there a need for some external agency to formally audit and declare this level of trust?

The Principle of Trust is that the more important a repository then the more trusted it should be, and that such trust should be maintained through actions within the total life-cycle of the digitization processes. The OCLC (2007) has produced a checklist for auditing and certification of repositories that wish to be called "trusted", and this includes a full organizational analysis of the repository management structure and processes. In the absence of the Principle of Trust, consumers will have no reliable method to determine whether and to what extent a digital repository can be trusted. The ideal for a trusted repository is that every item accessed is accompanied by irrefutable provenance information, possibly in the form of metadata, which provides citation back to all of the original sources and transformation used. This should also provide a universal citation linkage which is guaranteed to be persistent forever.

Any institution that wishes to create online repositories of digital content must explain how they intend to ensure authenticity of content. Given that this may be an expensive undertaking, beyond the means of smaller custodians, I have recommended that large-scale repository management is a specialist activity that should be conducted by a few competent and resource-rich institutions, rather than being expected from each individual custodian. These larger repositories can then provide outsourced repository management services to smaller custodians.

Principle of the Glass Bead Game

All records are essentially massively connected and it is the connections between the items which will become more important than the lists of inventories as we move into a semantic future.

I have long been fascinated by the future scenario portrayed by Herman Hesse in his novel The Glass Bead Game (1943/1990), which was cited in his award of the Nobel Prize for Literature in 1946. This work describes a futuristic society of the 23rd century that is dominated by knowledge and by the connections that are created within the various disciplines of knowledge, structured into a game that parallels the structure of the game of Go as it is played in Japan and other countries. Whereas Hesse does not outline the rules of his game, and wrote the novel as a future-historical, posthumous biography, it is the vision into a possible future world which grabbed my attention. I have long held the view that science fiction is not merely a set of stories of the future, but provides a sneak preview into futures towards which we are heading, driven by the imagination of the human mind. Fiction begets actuality.

One element of such science fiction literature is its treatment of future libraries and knowledge sources. In every story, no matter how far into the future the story is set, the

participants need to access information and to engage with some form of interface. Perhaps the most well-known is that of HAL in Arthur C. Clarke's 2001: A Space Odyssey, which was accessed by simply speaking to the spaceship. This is a dominant theme, in which computers of the future, with massive capacity for knowledge and information storage, will become increasingly humanlike, to the extent that we may not be able to tell the difference between machines and people, except that the machines can process billions times more information than we can and can arrive at conclusions far beyond what we can. This is the realm of Artificial Intelligence, my own research area, which provides a long-term promise for what we can expect in the future.

I argue that the vision of Hesse is not merely a fictional story of the 23rd century but is achievable within the next 10-20 years in the way that it treats knowledge as highly integrated. I am currently exploring models for large-scale repositories that can support the dynamic creation of new stories from existing sources, allowing us to reconstruct history on dimensions and themes not considered at the time that this history was being recorded. I had considered calling this the Principle of Story-Telling or the Principle of Massive Connectivity, in which our data stores are not seen as independent, disconnected lists but rather are highly connected, and for which new search structures are needed. This is outlined in a recent paper I have given to the National Oral History Conference, that outlines my current work in researching and developing novel approaches to address how such information is described, stored and accessed (Layton 2011).

Each custodian that is considering building up a large-scale digital repository is required to ask and answer questions on how the connections will be embedded into their repositories, and how these will then support improved forms of access to information, where such information is not seen merely as lists but as dynamic complex connections of related information. This is the next step in the evolution of intelligent repositories.

Principle of Ethics

Ethical considerations arise frequently in decisions on the digital heritage and it is important to have an explicit stance on how such decisions are treated.

Most libraries, archives, and museums hold artefacts that challenge our conceptions of right and wrong. However, such value systems are relative and institutions and observers hold differing positions on both ethical and moral issues. A photograph of naked slaves or of bare-breasted Zulu maidens may be offensive to some in spite of their historical value. Certain words used in the past have become socially unacceptable, and their usage may constitute hate speech.

One recent example is Julius Malema's "c**lie" slur (IOL News 2011) in using a derogatory reference to people of Indian descent, of which he appeared to be unaware, but which was quickly picked up by the media and by action groups as offensive. The online news service that posted this story specifically omitted the full word and rather disguised the word using * placeholders, which would have been done within the bounds of its own ethical stance. Such questions are not new, and remain the same in the digital world as they are within the

physical world of libraries, museums and newspapers. A similar position is taken by TV stations in providing content within boundaries of their viewers, and allowing self-censorship based upon ratings of the programme content.

What this Principle considers important is that each institution has an explicit statement of its position, and that this is included within the strategy document.

Principle of Eternal Preservation
In the absence of anything to the contrary we should expect that what we consider today to have enduring value will be valid for future generations.

How long should we preserve the contents within our memory institutions? Is there a time-limit after which information may no longer be relevant? Or are we expected to ensure that digital objects will remain accessible forever? We are often so concerned with our current generation's needs that we fail to plan for future generations. We are limited in our ability to predict future technologies, but physical museums, libraries and archives do not have a sell-by date and are expected to continue forever, and to be relevant as long as humans of the future will find them important.

It is an impossible task to build a digital repository today that will last forever, since there are too many challenges and risks that cannot be identified and planned for. However, we can build a repository that will have sufficient change control procedures to accommodate continual reformatting and conversion, in order to adapt to unknown future changes. This is possible but it comes at a cost. When an institution is planning strategy it will need to consider the lifetime of its repositories, and how it will ensure continuity into an unknown future in which digital content that has enduring value is preserved for future access no matter how far into the future this may be.

Principle of Constant Change
Everything is always changing in the digital world. We cannot stop this, but we can adapt.

There are two fundamental theories that apply when we consider the impact of change on our digital repositories. The first is the theory of evolution, as formulated by Darwin, in which all species will adapt to their environment, and those that adapt better will increase their chances of survival. This theory equally applies to digital content and repositories in which those that adapt better to changes in the environment will survive whereas those that do not adapt will become less relevant and will die out. There are many examples of important databases and data stores which are no longer accessible because the technologies on which they operate are not available.

The second theory is called the Second Law of Thermodynamics, which states that all systems are in a constant state of increasing entropy, so that they become more chaotic and less orderly over time. While this law of physics is applied to heat it is also relevant to other situations by way of analogy, since the way to reverse this process is by decreasing the heat, or lowering the temperature, which will then create order.

All digital content, and all digital repositories, are in an environment of constant technological change, and each institution must understand its environment sufficiently to plan how it will adapt to these changes, and should explicitly indicate these within its strategy. These form a part of a large risk-management function, although risks normally are concerned with uncertainties, yet the changes in technology are inevitable and are not uncertain. What we do know, for certain, is that staying the same will result in increased entropy, greater chaos, and less relevance. What we do not know is how to adapt. This is something we must learn.

Principle of Community

There is a strong association between the community and heritage since communities are owners, producers and consumers of heritage, and in the case of the oral history are also the long-term custodians of this heritage.

This Principle considers custodial institutions to be located within a system of communities who have produced heritage, who are the owners of this heritage, and who are also consumers of this same heritage. A community is a grouping of people with a common purpose, and each institution may have a number of such groups that are stakeholders and interested parties to the success of the institution. These communities may include the original tribes whose artefacts and records are being held, as well as their descendants. These communities also include the researchers who need access to these, and government institutions that have made their records available for long-term digital preservations. This Principle applies in that each institution should be aware of the communities who have an interest in the institution, and should also have well-defined interactions with these communities.

The Open Archival Information System (OAIS) standard (Consultative Committee for Space Data Systems 2009) defines the "designated community" as the people and organisations with an interest in a digital repository. My Principle of Community extends this to divide this designated community into constituent subsets which may have differing needs and relationships to the repository.

Principle of Sharing

The ultimate reason that we create digital heritage is to allow this to be shared with others, both current and future.

We keep our heritage for long periods of time, purely for the purpose of being able to share this with others in the future. There are situations in which information is embargoed until some future time, as is the case with sensitive military information. For many years the important records of the Second World War code-breakers at Bletchley Park in England was kept secret, with the result that most of the people involved, through whose efforts the war ended two to four years earlier than it might have done, could not be honoured until after their deaths. In most situations, however, archival information is available without such restrictions, and should be shared as much as possible to increase its relevance.

I argue that the largest future community of users are the history students in high schools throughout the country, who will tap into a range of trusted digital repositories as part of their daily research. This will not only provide for a new approach to history as a living subject, rather than the mere memorisation of facts, but will also allow for interactions between these scholars and experts who will be online for discourse and engagement on new perspectives and theories. I received a welcome surprise at the recent 2011 8th National Oral History Conference (OHASA) in which scholars from the ages of 12-17 made excellent presentations based upon their own research. I then imagined how much more powerful we can expect these learners to be in a future in which digital resources are available to all. Custodians must determine how best to share their digital content, and how to ensure that this is accessible to those who can benefit from it. As school-level learners increasingly use the Web to find sources, let them in turn be directed to authentic sites of content rather than to use the first entries that appear on a Google search.

Principle of Identity and Diversity

Digital repositories run the risk of normalising content and creating uniformity. A conscious effort is required to encapsulate diversity and differences.

One of my personal highlights at the OHASA conference in October 2011 was a presentation by two learners, both in Grade 10, and both 15 years old. They were Setswana-speaking and were from a Mafikeng school. Their paper concerned the impact of technology on their culture, and how they, with access to TV, mobile phones, and the Internet, were losing the culture of their parents and grandparents. They posed the question of whether this culture would then disappear forever and what could be done to preserve it. We run the same risk when creating our digital content and when storing these digital objects within large-scale digital repositories. This is the risk of the loss of the cultures that we are attempting to preserve, since repositories are mostly only available in English and position these historical records as a completed past rather than as a living current.

There is a potential for new technologies not only to bridge the digital divide but to play an important role in restoring cultural diversity, and it is a core responsibility of custodial institutions to support this. Whereas the Web has been available primarily to computer users who will always have a good grasp of English, the availability of the Mobile Web on mobile phones now provides Internet access to many rural communities where English is mostly unknown, and it is these rural areas which are the sole remaining heartland of these diverse cultures.

In one generation all of this cultural diversity may be gone, as new generations of the youth are being taught in English for all of their secondary school years and being exposed only to English as the basis for their reading, writing, and electronic communications. I recently asked a group of young people what language they used when talking to on SMS and MXiT, and they indicated that they only use English, for the reason that the SMS abbreviations are not developed or understood in their home languages.

Every institution should indicate clearly how it will provide for cultural diversity through the opportunities provided with the digital heritage.

Principle of Future History

The earlier that we can record current activities as history, the better prepared we will be when we want to access and use these records.

At every moment we are all creating history in everything we do, and we are recording increasingly more current activities as we write, use computers and digital cameras. Much of the history of the past has had to be retrieved from old, forgotten and lost records, since it was not possible at the time to know what would be considered important and of enduring value in the future.

The principle of future history is that the selection of materials for long-term preservation should commence at the point of production, and should not wait until future historians need this to help answer various questions that they pose. One key recommendation arising from this principle is that all government records should be structured into digital archives and should then move to national and provincial archives immediately rather than waiting until later. Whereas we cannot predict the information needs of future historians, we can provide the best possible digital archives structure to support their future studies by delaying selection decisions in favour of mass storage.

Discussion

It is important to have a principle-based approach to the digital heritage, and this requires that we have an explicit set of principles on which to base our work. This paper has provided an initial set of common principles that can be used, and I hope that these will form part of a large discourse in which they can be reviewed, modified and added to as we learn more about the essential elements that should inform our decisions and actions.

The digital world is very young compared with libraries, archives and museums and we will be learning more about the critical success factors over time. However, we are at a unique point in history, in which we are shifting from physical custodianship to digital custodianship, and this shift will occur only once in the whole of our human history. The shift is occurring now, within our generation, and it is our responsibility, and only ours, to make sure that this is done in the best possible way for the benefit of the future generations of users that we serve. It is likely that the digital records we are creating today may be the only history that future generations will have access to, and thus the work of our current generation is of exceptional significance and carries with it unique responsibility.

I hope that my work here will help contribute to the growing body of knowledge of the digital heritage, and that we can create increased openness and sharing through the adoption of a shared conception of best practices and of a drive towards open and free content.

References

Consultative Committee for Space Data Systems (CCSDS). 2009. Reference Model for an Open Archival Information System (OAIS). Draft Recommended Standard. CCSDS 650.0-P-1.1 (Pink Book) Issue 1.1. August 2009. Available: http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf. Last accessed 28 February 2010

Hesse, H.1943, 1990. *The Glass Bead Game.* Henry Holt & co. 1990. Original in German 1943.

IOL News. 2011. Outrage over _c**lie' slur. IOL News. 20 Oct 2011, 08:04am. Retrieved from http://www.iol.co.za/news/politics/outrage-over-malema-c-lie-slur-1.1160919?pageNumber=1 (31 October 2011).

Layton, R. A. 2011. "Linked Data" approach to biographical information: a case study of unsung struggle heroes. 8th National Oral History Conference. Mafikeng. 10-14 October 2011.

Linked Data. 2011. World Wide Web Consortium. Retrieved from http://www.w3.org/standards/semanticweb/data, (2 October 2011).

Nicholson, D. 2008. Digitization: considering the 3 C's: copyright, contracts, Creative Commons. 1st African Digital Curation Conference. CSIR, Pretoria, South Africa, 12-13 February 2008.

OCLC.2007. *Trustworthy repositories audit & certification: criteria and checklist*. Version 10. February 2007.

Oral History Association of South Africa [(OHASA]. www.ohasa.org.za.

South Africa. Department of Arts and Culture.2010. Draft National Policy on Digitization. V8. August 2010. Available from www.dac.gov.za.