DRAFT REPORT

Explaining Pres Actions

(A Working Document)

David Pearson (dapearso@nla.gov.au)
Nick del Pozo (ndelpozo@nla.gov.au)

National Library of Australia, Digital Preservation
4th August 2009

Version	Description	Date
0.1	initial draft	30.6.2009
0.2	formatting changes and minor content updates.	1.7.2009
0.3	got rid of preservation 'difficulty' levels, added more diagrams.	4.8.2009
0.3.1	input from Maxine Davis – consistency changes, adjusted migration diagrams	6.8.2009
0.3.2	input from Andrew Long.	18.8.2009
0.3.4	Input from Colin Webb.	21.9.2009
0.3.5	changes after review	2.10.2009

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Table of Contents

Table of Contents 1

Table of Figures 3

Abstract 4

Introduction 4

References 6

Primary Preservation Actions 8

1.1 Migration 9

1.2 Take No Action 20

Secondary Preservation Actions 21

1.3 Emulation: 22

1.4 Renderers 29

1.5 Technological Museum 37

Aknowledgements 42

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Table of Figures

Figure 1: Migration 9
Figure 2: Stack Migration 11
Figure 3: Linear Migration 13
Figure 4: Risk Based Migration 17
Figure 5: Window of Opportunity Based Migration 16
Figure 6: On Demand Migration 19
Figure 7: On Receipt Migration 15
Figure 8: Take No Action 20
Figure 9: Emulation 22
Figure 10: Matroyshka Method 24
Figure 11: Migrate Emulated Environment 25
Figure 12: Emulated Environment for Each File Format 26
Figure 13: Emulated Environment for Consolidated File Formats 27
Figure 14: Generic Emulated Environment 28
Figure 15: Renderer 29
Figure 16: Rewrite Renderer 31
Figure 17: Emulate Renderer 32
Figure 18: Find New Renderer 33
Figure 19: Renderer for Each Format 34
Figure 20: Renderer for Each Consolidated Format 35
Figure 21: Generic Renderer 36
Figure 22: Techology Museum 37
Figure 23: Access Path for Each Format 39
Figure 24: Access Path for Each Consolidated Format 40
Figure 25: Generic Access Paths 41

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Abstract

This document attempts to present the different preservation methodologies that are currently available to the library, as well as provide an indication how these methodologies could be used to preserve digital materials. It is the goal of this document to outline the current thinking of the digital preservation branch in such a way that it is possible to decide, in close collaboration with each collecting area, which digital preservation strategies are most appropriate for each collecting area. It is also intended that this should provide a means for identifying where a collection area may require additional resources, as well as where new technologies should be acquired, or developed.

Introduction

This document identifies a number of preservation methodologies that have been identified as building blocks to help build potential strategies for approaching the preservation of digital materials. The methodologies which are examined are:

- Migration;
- Emulation;
- Application based rendering;
- Collecting and maintaining a 'Technology Museum'; and
- Taking no action.

For each preservation methodology the following will be presented:

- What the methodology is, or what it purports to do;
- How it works;
- Its perceived advantages and disadvantages; and
- Different strategies for approaching or maintaining the methodology.

All these preservation actions presume access to the bit-stream. That is to say, that it is possible to access the physically stored data without being technologically inhibited (e.g., if the digital object is stored on a CD-ROM, it is still possible to read the CD-ROM, and the data stored on the CD-ROM is still intact). Ensuring digital materials are not stored on obsolescent carriers is a significant preservation issue, but is not the subject of this paper. For more thinking on this topic, see Clifton and Langley (2007), Elford *et al.* (2008), or del Pozo *et al.* (2009).

This document makes references ideas and terminology which is better defined in the Rethinking Repository Requirements, and Preserving Digital Objects Within the National Library of Australia, two forthcoming documents which, together with this document, make up a suite of technical and theoretical papers addressing various ideas and problems in digital preservation. As is alluded to in these other papers, in order to make a more informed decision about which preservation actions will most adequately preserve a digital object over time, relative to the needs of the preserving organisation, a certain degree of knowledge is

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

required. Ideally, as much as is possible, the following knowledge areas should be known, or there should be some clarity in each of these areas:

- The intention of the preserving institution for the object, and an understanding of which aspects of the digital object should be maintained, or which elements of the original digital object are most important to preserve.
- The perceived intention of the original creator of the digital object, for how others should experience the material.
- Knowledge of how the digital object will be accessed both before and after any preservation action.
- Knowledge of the file part of the digital object (where a digital object contains a file part), including the structure of the file format, and the content of the file.

It is proposed in the Preserving Digital Objects Within the NLA document, that a digital object can be thought of as having various aspects, each of which may require a different strategy for long term preservation. These aspects are:

- The physical arrangement of the data
- The binary sequence of the file derived from the physical arrangement
- The information that the binary sequence can be decoded to convey
- The interpretation that a user may derive from the information

Each preservation action outlined below can be seen as potentially preserving one or more of these aspects, sometimes at the expense of others. Although it is difficult to generalise in this area, where it has been seen as appropriate, it has been indicated to what degree a given preservation action will require knowledge in these different aspects (data, action and experience), and to what degree it can be generally expected that a preservation action will preserve the different aspects of a digital object.

Although in some cases a given preservation methodology may *lessen* the degree of knowledge required in a given area, no preservation action will ever entirely eliminate the need to have at least some understanding in all these areas. For example, a preservation strategy that incorporates an emulation layer may not necessarily require a great deal of knowledge about the file structure, but may require more knowledge of the expected experience.

This document attempts to present each of the included preservation actions as valid ways of preserving digital materials, but also to help identify in which circumstances they may be more or less viable. In some instances, methodologies contain variations which may be more or less appropriate for different institution or individual circumstances, and for different types of collection materials. These have been identified in some instances.

For the purposes of this document, and in order to help facilitate the building of preservation strategies, the methodologies below are presented as facilitating either Primary Preservation Actions, or Secondary Preservation Actions. A Primary Preservation Action here indicates an action that directly *changes* the digital material/data. A Secondary Preservation Action here indicates an action that changes the way in which the digital materials/data is *accessed* (either

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

to reproduce an experience or otherwise). For example, migration is seen as a Primary Action, as it directly changes the file itself in order to facilitate access. Alternatively, emulating an access environment for a file is seen as a Secondary Action, as it changes the way in which that file is accessed.

These actions have been presented from the perspective that many different methodologies may be employed to preserve a digital object over time, and in general more than a single preservation methodology will be required to adequately preserve any digital object. Indeed, it is proposed that it may generally be required to employ more than a single preservation action at once.

Likewise, it is assumed that the preservation actions that it is possible to carry out on a digital object will vary and change over time, as will the most appropriate action. Initially, the appropriateness of preservation actions will depend upon the type of material and the intention of the individual carrying out the action. Further to this, however, the right actions to pursue may additionally be affected by the age or obsolescence of the material, which may initially rule out certain actions entirely. It may also change as knowledge about an object increases, or the preservation intent towards the object changes over time. In some cases, certain preservation actions may make it easier or more difficult to recover from these possibilities, and so this should be taken into account. Similarly, certain preservation actions may be required as pre-requisites for other actions, or in some instances, as vital parts of an institution or individual's overall preservation strategy.

For these reasons, the actions described below are not presented as singular preservation paths, but as different elements that could contribute to the overall goal decided upon by the individual or organisation. In all cases, this document attempts only to provide an indication of our current thinking in these areas, and of the consequences or benefits of any given action, and should be taken as such.

This document was created to address a need at the NLA to try to understand and articulate to a broader audience the appropriateness of certain preservation methodologies. It is also intended for this document to be useful for facilitating a better informed conversation between business areas across the Library.

References

Clifton, G. and Langley, S. 2007. 'New forms, new techniques: challenges of preserving digital materials', in 'Contemporary Collections', preprints from the AICCM National Conference, 17-19 October 2007, Brisbane, pp.37-41.

del Pozo, N., Elford, D. and Pearson, D. 2009. 'Prometheus: Managing the Ingest of Media Carriers', in *Proceedings of DigCCurr 2009, Digital Curation Practice, Promise and Prospects*, University of North Carolina at Chapel Hill, North Carolina (2009), pp.73-75.

del Pozo, N., Long, A., Pearson, D. 'Preserving Digital Objects Within the NLA', forthcoming.

del Pozo, N., Pearson, D. 'Rethinking Repository Requirements', forthcoming

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Elford, D., del Pozo, N., Mihajlovic, S., Pearson, D., Clifton, G. and Webb, C. 2008. 'Media Matters: developing processes for preserving digital objects on physical carriers at the National Library of Australia', *IFLA World Library and Information Congress*, Quebec City, Canada (2008). At http://www.ifla.org/IV/ifla74/papers/084-Webb-en.pdf

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Primary Preservation Actions

Primary Preservation Actions are action which directly impact on the digital material/data being preserved. The methodologies presented below are ways in which this can be achieved.

1.1 Migration 9
1.2 Take No Action 20

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

1.1 Migration

What it is:

Migration is the process of converting a piece of digital content from its original file format into a new format that can more easily be accessed without having to maintain contemporary software and hardware. The basic premise is that the file format needs to be changed in order to facilitate accessing the file in the simplest possible way at any given time. Therefore, migration favours access over immutability.

There is no specific requirement that a target file be migrated to a single destination file. It might be preferable to store the properties that have been identified as significant across multiple files, or using multiple storage mechanisms (e.g., a file and a database). There are two basic ways in which a file can be migrated to a new format, namely stack and linear migration, which are examined in greater detail below.

While there are many potential approaches to migration, there is no specific requirement that an individual or institution select to only employ a single migration schedule or strategy. For example, some types of files in a collection might be better suited to migration on receipt, while some might better suited to risk based migration (both explained in greater detail below).

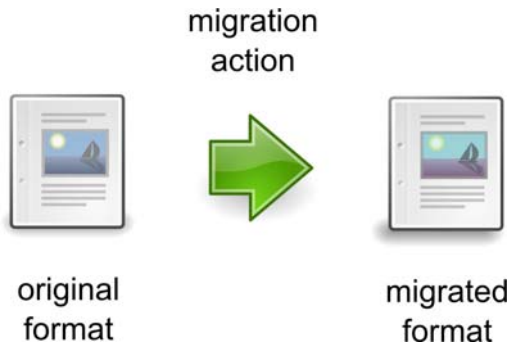


Figure 1: Migration

How it Works:

1. Original file format is acquired; and
2. File Format is changed to another format.

Pros:

- The digital object will be stored in such a way that the experience can be recreated with an acceptable degree of change over time, without maintaining the environment contemporary to the original file.
- Potentially, digital objects could be stored in a file format which is better suited for long term preservation and access, which in some cases may simplify the maintenance of large sets of file types.
- If the source and target file formats are well enough understood, the level of loss could be controlled, documented, and acceptable relative to the specific collecting area.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

- If the file formats migrated into are well understood, the change to the experience could be controlled, documented, and deemed acceptable by the collecting area.
- If the file formats migrated into are well understood, it may reduce the amount of preservation action needed to preserve the experience over time. This may also reduce the overall rate of change over time.

Cons:

- Requires a clear articulation of which properties or elements of a file are most important to preserve between migrations, and a way of measuring how effectively these properties have been transferred to a new format. These may vary according to the intended use of the file.
- Requires constant preservation planning, and assumes a great deal of understanding of the systems being migrated.
- If the change cannot be understood, controlled and documented then it may not be an acceptable action to take on a file.
- If a large amount of material needs to be migrated at one time, and depending on the mechanism for migration, it may be difficult to ensure an acceptable level of loss in each digital object.
- Preservation planning needs to occur for each file format, and any permutation of that file format. For example, an application that adds proprietary information to a file format may significantly change the nature of the format without changing the way the format is identified.
- Significant properties of the file may be lost, if there is not sufficient understanding of both the source and target file formats, and the migration process being used.
- Complex digital materials (formed from the relationships between many files) are inherently difficult to migrate, and migration may not adequately preserve their dependencies, unless there is a significant level of knowledge about the digital object in question.
- If a bulk of materials are migrated into a file format for which we subsequently lose all access, access to all this content is lost.
- Given that the knowledge about, or intent towards a file may change over time, the parameters for what constitutes the 'most significant properties' of a file may change over time, and at a time where they may already have been lost (and recovery from the original is no longer possible).

Further Reading:

Brown, A. 2006. Archiving Websites: a practical guide for information management professionals. Facet Publishing, London. pp.92-99.

Harvey, R., 2005. Preserving Digital Materials. K G Saur, München. pp.147-153

PADI, Migration, list of references. Viewed 1st July 2009 <http://www.nla.gov.au/padi/topics/21.html>

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Migration Strategies

This paper identifies two different primary methodologies for migrating a digital object from one format into another. Each has its own specific draw backs and advantages, and may be appropriate in different situations. It is suggested that both strategies will have to be used in concert, at different times in the life of a digital object. These particular strategies are presented from the perspective of migration over time, but are also valid in terms of digital objects that may only be migration a single time in their lifetime.

- **Stack Migration**

In this method, the original file (preservation master) is always used as the basis for the source of the migration over time. Therefore, any derivative copy is a direct derivative of the original. Over time, the effectiveness of each migration is relative to the technology and knowledge available at the time. Eventually, it is assumed that access to the original will be lost, and this will no longer be a viable preservation action.

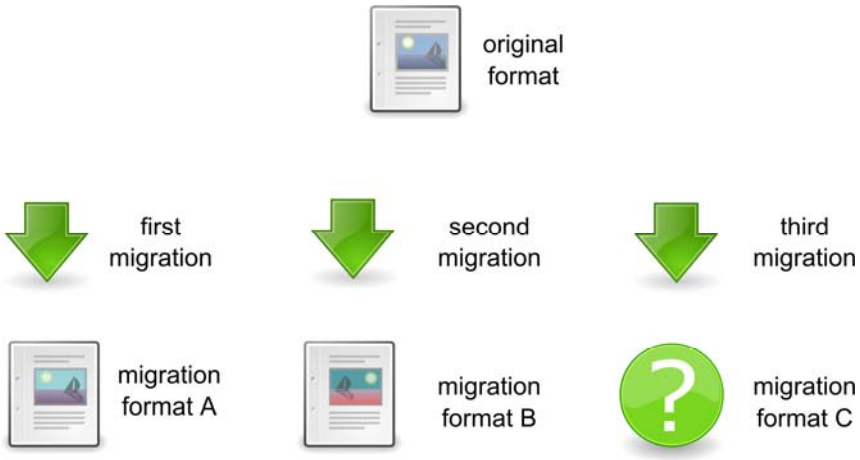


Figure 2: Stack Migration

How it Works:

1. Original file format is acquired;
2. File Format is migrated to another, new format. The original is also used for any subsequent migrations;

Pros:

- Does not create cumulative loss over time.
- If one migration did not convey significant properties from the original, so long as access to the original is still possible, these properties can be collected in a subsequent migration.

Cons:

- Over time access to the original may be lost.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

- Knowledge of the original file must be maintained over time.
- The effectiveness of each migration is relative to many factors, and there is no assurance that each migration will be better than the last.
- If access to the original is still available, there may be no benefit from migrating to a new format.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

- **Linear Migration**

In this method, the most recently migrated derivative of the file is used for the source of the next migration.

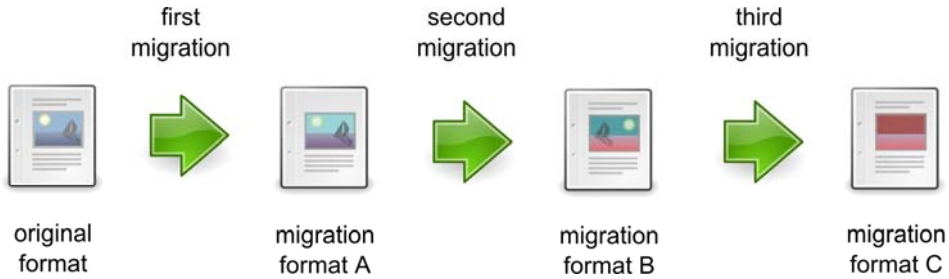


Figure 3: Linear Migration

How it Works:

1. Original file format is acquired;
2. File Format is migrated to another, new format. Subsequent migrations will use the previous (the most recent) format as the source.

Pros:

- If the formats that are migrated from and to are well understood, the loss of significant properties may become more controllable over time.
- Solves the problem of the original file becoming inaccessible.

Cons:

- If the formats that are migrated from and to are not completely understood, the cumulative loss to a file may become unacceptable over many migrations
- Although change may be acceptable in a single migration, this may be compounded, and eventually become unacceptable over time
- By the time change in one element of the file has become unacceptable, it may no longer be possible to access a previous version for which that change is not unacceptable, and it may not be possible to recover.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Migration Catalysts and Schedules:

Irrespective of which type of migration is used, there could be various catalysts or schedules which dictate when a file format should be migrated over its life span. A number of likely catalysts have been identified below.

- **On receipt**, which is as soon as the file comes into custody.
- **Window of opportunity**, which is at some point that an opportunity presents itself to take action
- **Risk based**, which is when an external or internal risk is identified that requires action be taken to avoid losing access to the file
- **On demand**, which is at the time the file is requested by an external or internal party.

What constitutes the best catalyst for a file is depending on the type of file, and the preserving institution's preservation intent for that file. Additionally, the timings for these catalysts may overlap in many instances. For example, a risk based migration may for some files be an ingest migration. It is expected that a preserving institution would use a variety of timings to migrate their files, rather than adhering to a single strategy over time. The different timings are outlined below.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

On Receipt Migration

The file is migrated to a new format as soon as it is ingested.

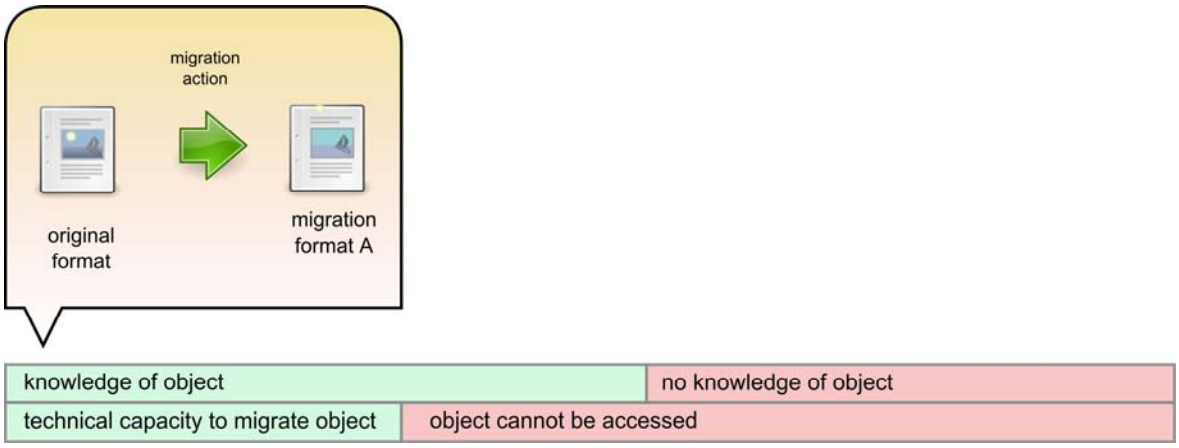


Figure 4: On Receipt Migration

Pros:

- Potentially perform migration action on digital object when tools and systems are available that are best capable of doing so; and
- Provides the maximum possible time frame for taking future preservation actions;

Cons:

- Creates immediate overhead both on technical systems and human resources at time of ingest;
- Assumes immediate knowledge of digital materials; and
- May not be possible to migrate files on ingest. May need to a different migration schedule at a later stage.
- May not always immediately have enough information (either about the format itself, or which properties are believed to be most significant) to migrate to the best possible format.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Window of Opportunity Migration

The file is migrated when it is possible to preserve the most amount of properties that have been identified as significant. This is distinct from risk based migration as while the same variables that would affect any migration still need to be tracked, external parameters that would affect the obsolescence of a file format do not need to be tracked. This reduces the overall number of variables which need to be accounted for, as action is taken as soon as it is possible to do so.

In order for this process to be effective, multiple migrations may take place until all the properties that have been identified as significant have been preserved. This means that at any time the original is stored concurrently with a derivative format that has the greatest amount of transferable significant properties. Alternatively, a number of formats, or different storage mechanisms (e.g., database) may be used to preserve these properties.

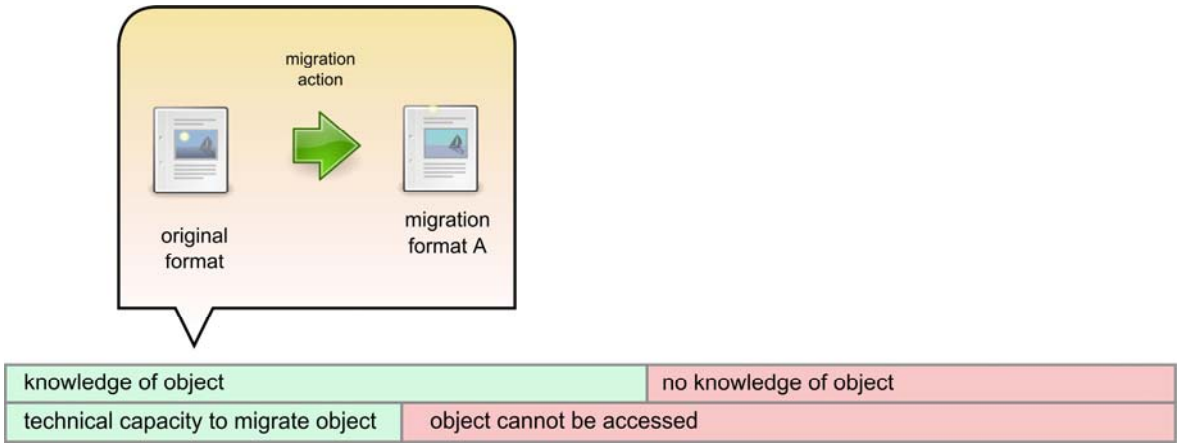


Figure 5: Window of Opportunity Based Migration

Pros:

- Migration occurs at best possible time to preserve as many significant properties for any given file;
- Comparative to risk based migration, is dependant upon a smaller and more controllable set of variables; and
- Will always have an accessible copy which contains the most possible significant properties possible.

Cons:

- Digital objects may still be susceptible to obsolescence if there is no Risk based assessment mechanism.
- May require greater overall load on systems over time, to maintain a digital object in various transformative states.
- Requires constant surveillance of file formats, to understand when a good migration

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Risk Based Migration

The file is migrated when it is deemed 'at risk' (but before it becomes obsolete). Ideally, this occurs some time before we lose our ability to access the file.

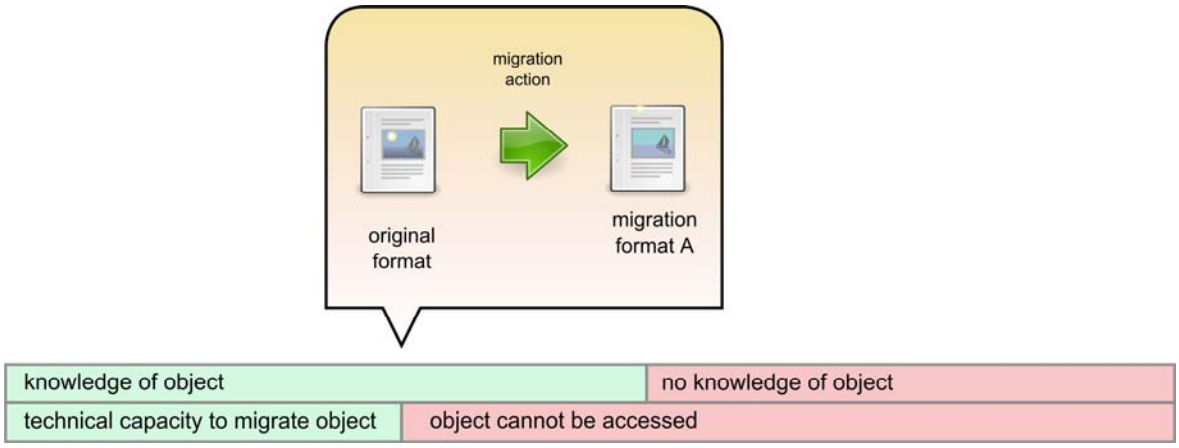


Figure 6: Risk Based Migration

Pros:

- Migration actions are only invoked on files that are at deemed 'at risk'. This is less time consuming for the user who preserves the digital materials, than migrating all content upon ingest;
- Arguably, in some instances, when a file is deemed 'at risk', this may be the most appropriate time to migrate the file given that there may be the greatest number of tools, and most knowledge available for migrating the file.

Cons:

- Information about risk which forms the basis for decisions may not be reliable or applicable.
- Risk is subjective and dependant upon many possible variables. Some of which are potentially unascertainable. As a result this makes it incredibly difficult to track risk for many file formats. This may have implications for staffing.
- There must be a reliable mechanism for identifying that a file format is at risks. If risk cannot be identified then action cannot be taken.
- Requires constant surveillance of file formats, to know when a format will be at risk.
- Makes the assumption that appropriate tools will be available to migrate content at the time that content is at risk.
- May not always provide enough headroom to take action.
- Risk based migration is a reactive approach to digital object maintenance. That is to say, the individual or institution will be guided to a greater extent in their timing by external forces (relative to other migration strategies). This may not always be a convenient or desirable situation for an organization to find themselves in.

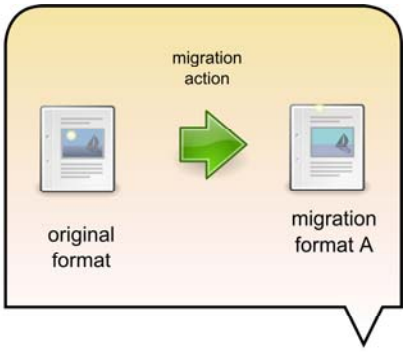
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

- solution presents itself.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

On Demand Migration

The file format is migrated when it is requested by an interested party. This means that potentially some files may never be migrated.



knowledge of object	no knowledge of object
technical capacity to migrate object	object cannot be accessed

Figure 7: On Demand Migration

Pros:

- Over time, potentially least amount of load on system of all migration strategies; and
- Content that is consistently useful over time will more likely be migrated.

Cons:

- Makes assumption that appropriate tools will be available to migrate content, at the time that content requested;
- Content may become inaccessible long before it is requested;
- All knowledge of the object may be gone long before it is requested; and
- Some content might only be deemed useful after it can no longer be accessed.
- Overall load on systems is unpredictable.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

1.2 Take No Action

What it is:

This approach does nothing. This could be done either because there is currently no reason to perform a preservation action, there is an expectation that a preservation problem will be addressed externally, or the institution is making a conscious decision to not preserve a given digital object. This methodology is classified as a primary action as it still focuses on the digital object, and may potentially effect the digital object over time (e.g., object may change through 'bit-rot' over time if no action is taken).

original format
(not changed)

Figure 8: Take No Action

Pros:

- Does not require any effort;
- Does not require any special skills; and
- Does not require preservation planning.

Cons:

- If this is the only action taken, loss of access to some, eventually all, digital materials is guaranteed;
- Because no action has taken place, it may not be evident that access to digital materials has been lost until access is attempted;
- If a change in strategy occurs some time in the future, it may be very difficult to understand undocumented older environments and formats in order to take some new form of action;
- If everybody takes this action, no solutions will ever present themselves; and
- Places a great deal of trust in variables outside the control of the organisation.

Futher Reading:

Havey, R. 2005. Preserving Digital Materials. K G Saur, München. pp.118-120.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Secondary Preservation Actions

Secondary preservation actions do not effect the material/data itself, but change the way in which that material is accessed, and subsequently how access to that material is preserved over time. The following methodologies and their various permutations represent options for changing or deciding the most appropriate way of maintaining access to digital objects over time.

1.3 Emulation: 22
1.4 Renderers 29
1.5 Technological Museum 37

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

1.3 Emulation

What it is:

Emulation is the process of creating a 'virtual' version of the original environment that was used to access a given file. The virtualised environment is accessed via an emulation application on modern hardware and software. This allows access to the original content to be maintained (without changing this content), through the emulated computer. Emulation retains the experience, and the original form of the data, and to a degree the performance, but does not necessarily retain the original form or performance of the hardware. This may have implications depending on the preservation intent being articulated.

It is noted that an emulated hardware configuration and operating system on its own may not be enough to adequately access digital materials beyond their original arrangement. It will in most instances be necessary to pursue this methodology in conjunction with specific renderers (outlined below).

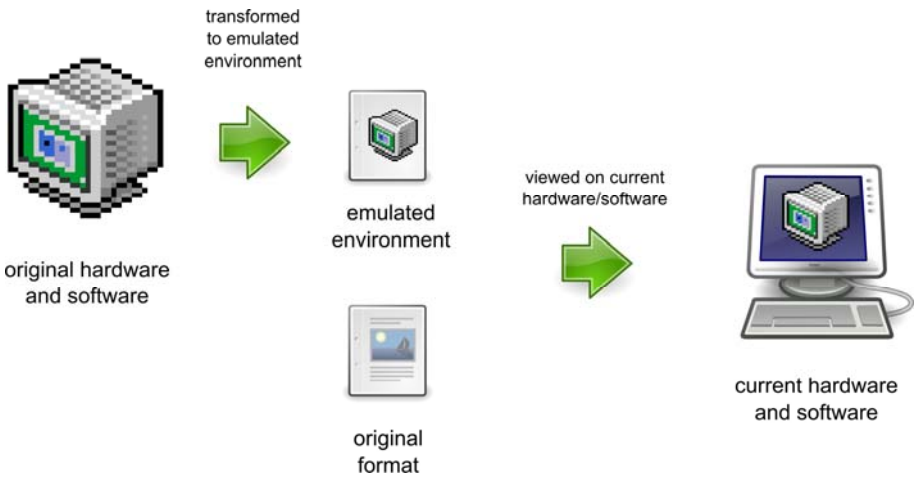


Figure 9: Emulation

How it Works:

1. A contemporary access environment for a digital object is encapsulated into an emulated environment;
2. The emulated environment is accessed using a current hardware and software platform; and
3. By using the current hardware and software platform to access the emulated environment, the emulated environment is used to access the target file.

Pros:

- Does not change the file format, so as long as access is maintained, there is no loss to content;

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

- Maintains an environment which is contemporary with the digital object, which itself may be considered an important contextual part of the digital material being preserved; and
- May be the only practical way of preserving access to some digital objects.

Cons:

- Requires constant preservation planning, and assumes a great deal of understanding of the systems being emulated;
- Requires an articulation of what constitutes an acceptable reproduction of the environment being emulated.
- It may be difficult to integrate this methodology with a content or preservation management system.
- Emulation only provides a 'surface level' reproduction of the original access environment: There are certain tactile or input elements that may not be accurately replicated;
- The ability to accurately reproduce materials is limited to the ability of the emulator: Although it is currently possible to reproduce a wide range of machines, the types and combinations of software and hardware that can accurately be emulated is by no means exhaustive. for example, some software may contain copy protection or activation protocols that may limit or even prevent their functionality on an emulated system;
- Not all access environments can be emulated or reliably emulated. For example, our ability to emulate older machines may be limited;
- Emulated machines can themselves be technically classified as 'file formats', and as such are susceptible to all the same issues as other digital content;
- Emulated environments represent complex chains of dependency, and are therefore more difficult to manage than just the digital material itself;
- Overtime, it may not be easily ascertainable if the file has been accurately rendered, resulting in an unpredictable experience for the end-user; and
- Arguably necessitates a simplistic view of what constitutes hardware and software and their interdependencies.

Further Reading:

Brown, A. 2006. Archiving Websites: a practical guide for information management professionals. Facet Publishing, London. pp.87-92.

Suchodoletz, D. and van der Hoeven, J. 2008. 'Emulation: From Digital Artefact to Remotely Rendered Environments', in Proceedings of the Fifth International Conference on Preservation of Digital Objects (iPRES2008), The British Library, London 29-30 September 2008, pp.92-98.

PADI, Emulation, list of references. Viewed 1st July 2009 <http://www.nla.gov.au/padi/topics/19.html>

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Emulated Environment Preservation Strategies:

As indicated above, an emulated environment can also be seen as a digital object itself. As such, actions need to be taken in order to maintain access to these environments over time. A number of potential preservation strategies are outlined below.

Matroyshka Method

The emulated environment, together with the environment it is currently being run on, are both encapsulated into a new virtual environment. The new virtual environment is now run on current software and hardware. Potentially, overtime there could be many layers of emulation which are needed to access the original target file.

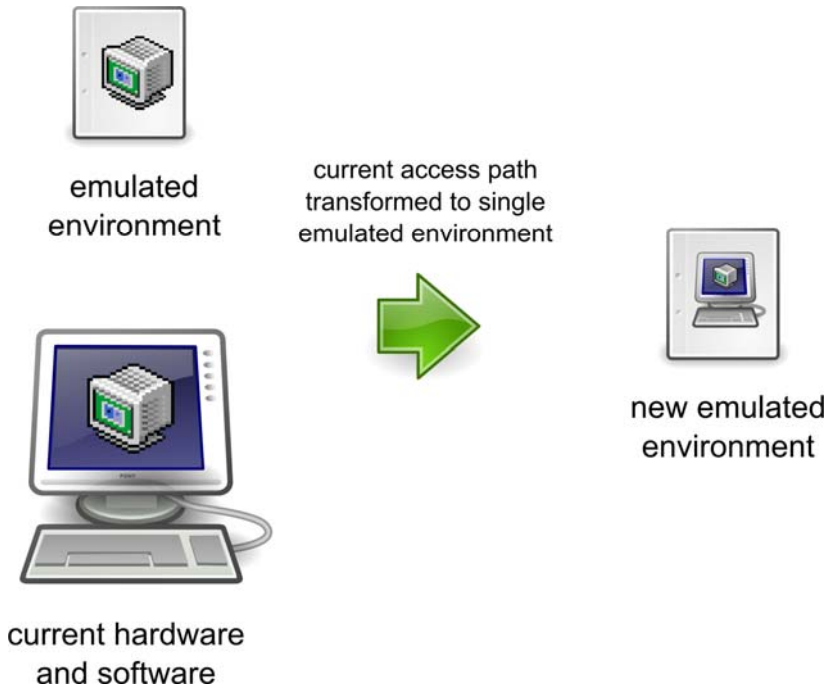


Figure 10: Matroyshka Method

Pros:

- Potentially, this method will most accurately preserve the original environment.

Cons:

- This method is evidently a convoluted way of preserving access to digital materials. With every additional layer of emulation, access to the original file becomes more complex and thus potentially less sustainable; and
- It will become harder to access the original digital materials each time a new set of environments is encapsulated.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Migration

If we consider that an emulated environment is a complex digital object, it is possible to implement the same preservation strategies that would be used on other digital objects (including emulation!) to preserve access to the emulated environment. The same pros and cons apply. This specific scenario uses risk-based linear migration.

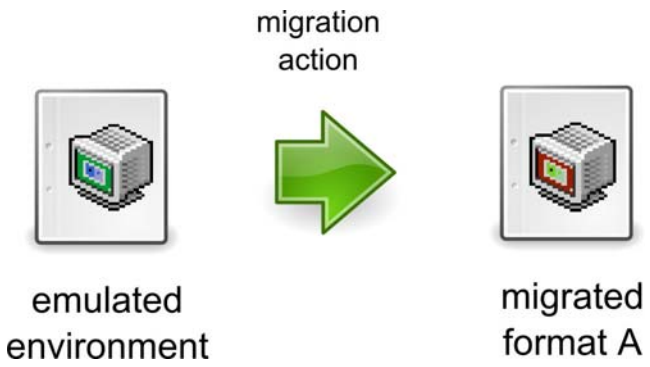


Figure 11: Migrate Emulated Environment

Pros:

- Comparative to the Matroyshka Method, this approach maintains a degree of simplicity, which may make it easier to maintain over time.

Cons:

- Changes at the level of the emulated environment may result in changes to how the original material is accessed that may not be predictable; and

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Emulation Environment Access Strategies

Given a basic emulation approach, there are various ways in which the emulation of digital materials can be approached. This section outlines a number of scenarios for creating and maintaining emulated environments.

Emulated Environments for Individual Formats

In this case, for each given file format that needs to be accessed there is a corresponding emulated environment which provides access to that file format. The file is copied to the virtual machine, which is then used to access the file format.

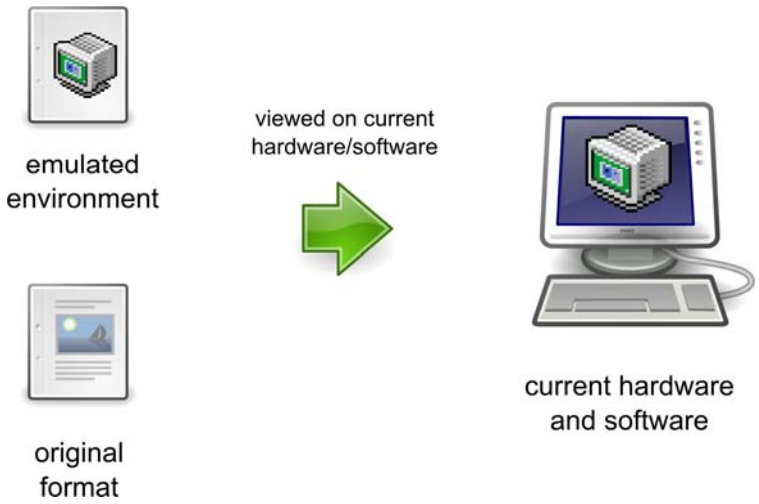


Figure 12: Emulated Environment for Each File Format

Pros:

- Potentially access the digital objects with the least amount of change in the experience.

Cons:

- If there are sufficient changes to a file format, it should be considered a new format. As such, there may be many emulated environments for each file format, or some emulated environments for only a single file; and
- In a worst case scenario, it would be necessary to generate a new emulated machine for each file that is to be preserved. Asides from permutations in file formats, this could become necessary if the access context of each file were individual enough to warrant a specific set of components to be emulated in order to most faithfully view the document, as it was originally intended to be viewed by its creator.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Emulated Environments for Consolidated Formats

In this method, various file formats that share similar characteristics are migrated into a single file format. While there is still a single emulated environment for each format, the total number of emulated environments is drastically reduced.

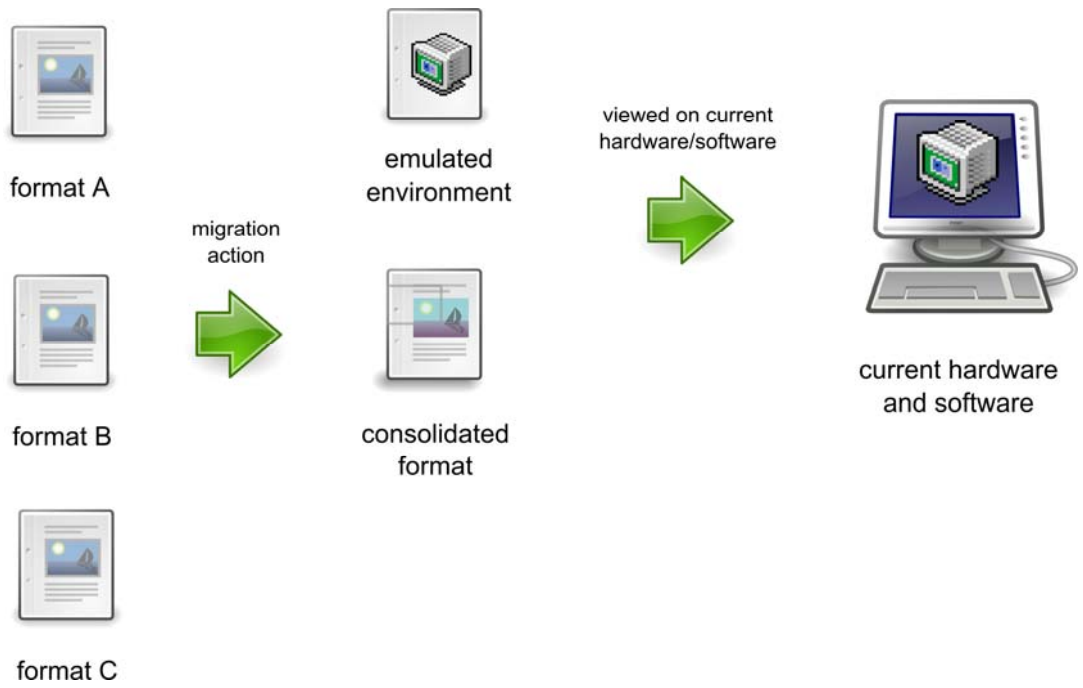


Figure 13: Emulated Environment for Consolidated File Formats

Pros:

- This would make it more practical to maintain virtual environments

Cons:

- If this solution undertaken, there would be little additional gains from viewing the consolidated format via an emulator, as in any case a consolidated format which is currently viewable using current hardware and software could be selected as the target format.
- Acceptable level of loss must now be articulated for both the emulated environment, and the formats which are being migrated into a consolidated format. All the cons for general migration are applicable.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Generic Emulated Environments

Emulated environments will be created that provide access to the greatest number of file formats possible. This could be done in conjunction with a Consolidated Formats approach, or at the level of Individual Formats.

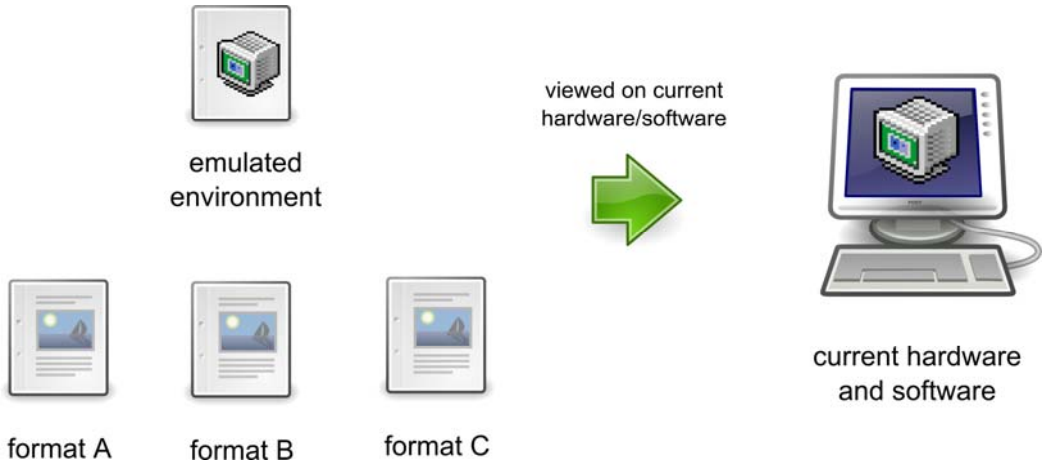


Figure 14: Generic Emulated Environment

Pros:

- Reduces the overall number of emulators; and
- Does not require that files be migrated to new formats.

Cons:

- The more formats an emulated environment can access, the more dependencies are present in the emulator, and the harder it is to carry out preservation actions on the emulated environment.
- The additional dependencies also mean a greater level of knowledge is required to predict the consequences of any preservation action.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

1.4 Renderers

What it is:

An application that runs with current hardware and software is used to access the digital object. The software itself could either be written internally, or procured from another party. It could either be a first party application, if it is written by the same organisation responsible for creating the file format, or a third party application in all other cases. A renderer can also be categorised as the application which created a specific digital object, or an application which can access the file format of a digital object without necessarily being the creating application.

Like in the case of emulation, a renderer can itself be considered a digital object, and so steps must be taken to preserve the renderer, or access to the digital objects it services will be lost.

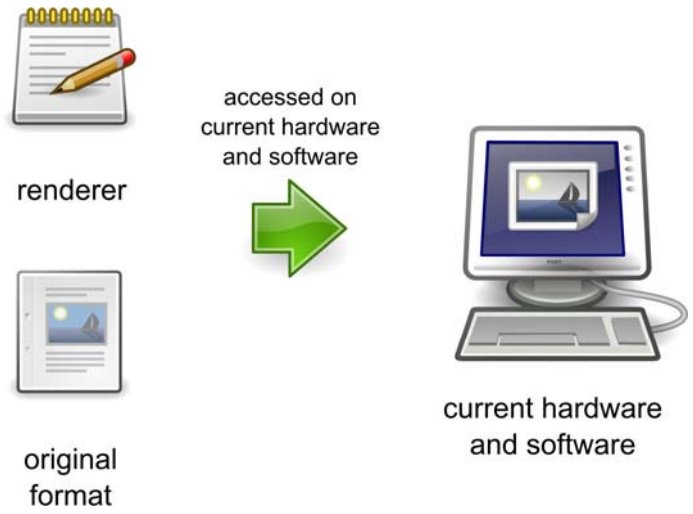


Figure 15: Renderer

How it Works:

1. Original file format is acquired; and
2. It is accessed using an application that runs on modern hardware and software.

Pros:

- Allows original format to be viewed on current hardware and software;
- Allows a faithful rendering environment for the file; and
- It may be possible to tailor the renderer according to the properties that have been identified as significant, in some cases providing a more useful experience.

Cons:

- A renderer may eventually have to be emulated or re-written to work on current hardware and software.
- The author's creating environment may be significantly different from an institution's access environment (e.g., some applications may have different plug-ins), which may impact on how appropriate a given renderer is for any given file.
- If the renderer is written internally, requires a significant amount of knowledge about the file format being accessed.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

- Requires a significant amount of knowledge about the file being accessed in order to choose the most appropriate renderer.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Renderer Preservation Strategies:

As indicated above, a renderer can also be seen as a complex digital object itself. As such, actions need to be taken in order to maintain access to the renderer over time. A number of potential preservation strategies are outlined below.

Rewrite Renderer

The renderer is rewritten to work with newer hardware and software.

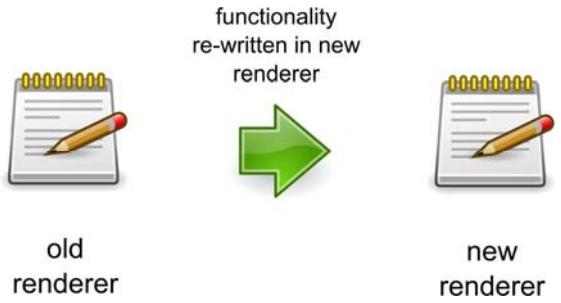


Figure 16: Rewrite Renderer

Pros:

- Maintains access to digital object without complicating the access mechanism.

Cons:

- Requires a great deal of knowledge to be retained for both the file format being accessed, and the process of the renderer being rewritten. This knowledge may be very difficult to retain.
- May not be possible to rewrite renderer, if the renderer is proprietary software. Open source renderers may be easier to rewrite, but may lack the documentation required to make this a practical exercise.
- Requires a greatest effort over time to maintain all renderers.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Emulation

An emulated environment is created that can run the renderer.

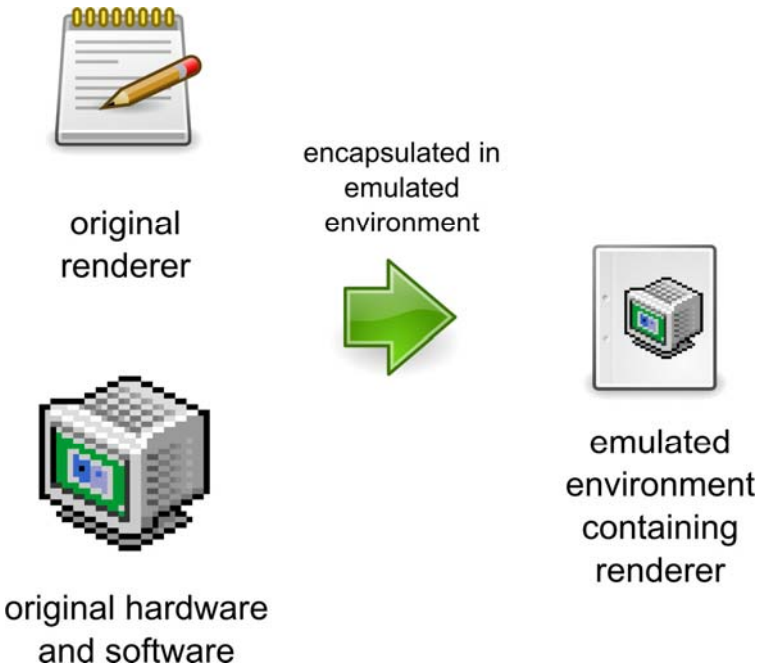


Figure 17: Emulate Renderer

Pros:

- Does not require the renderer to be rewritten in order to maintain access.

Cons:

- Complicates the access to the digital materials initially provided by the renderer.
- Makes the renderer susceptible to the same preservation concerns as emulation.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Find a New Renderer

The current renderer is abandoned, and a new renderer is found.

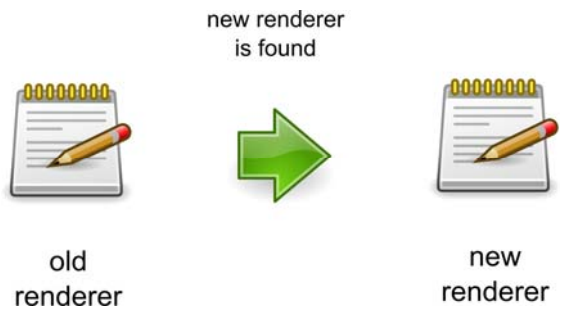


Figure 18: Find New Renderer

Pros:

- If a new renderer can be found, is the most efficient solution.

Cons:

- The new renderer may access digital objects in a way which is not as adequate (relative to the needs of the preserving institution) as the old renderer.
- If no suitable replacement renderer is available, the institution may have to migrate their materials, or rewrite the renderer. By this time both these options may no longer be practical.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Renderer Access Strategies:

This section outlines a number of strategies for employing renderers as an access mechanism for digital materials.

Renderers for Individual Formats

A new renderer is created for each file format type.

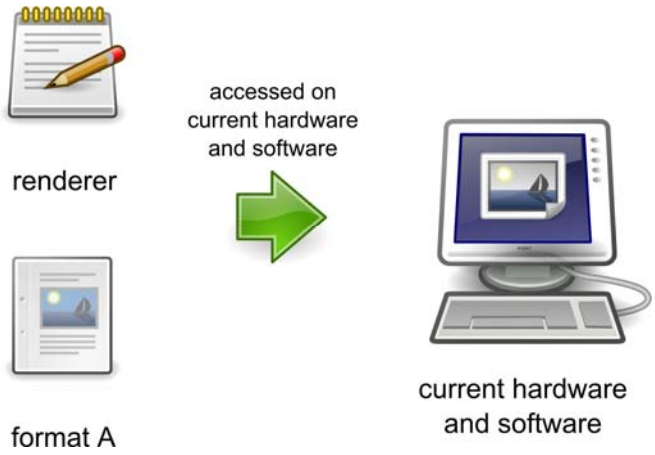


Figure 19: Renderer for Each Format

Pros:

- Potentially most likely way of appropriately accessing each file format.

Cons:

- Potentially requires the most resources.
- If variances in a file format are great enough, that format should be treated as a new format, and so a new renderer would be required.
- In a worst case scenario, each file being preserved would require its own renderer in order to most appropriate conserve appropriate access.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Renderers for Consolidated Formats

In this method, various file formats that share similar characteristics are migrated into a single file format. While there is still a renderer for each format, the total number of renderers is drastically reduced.

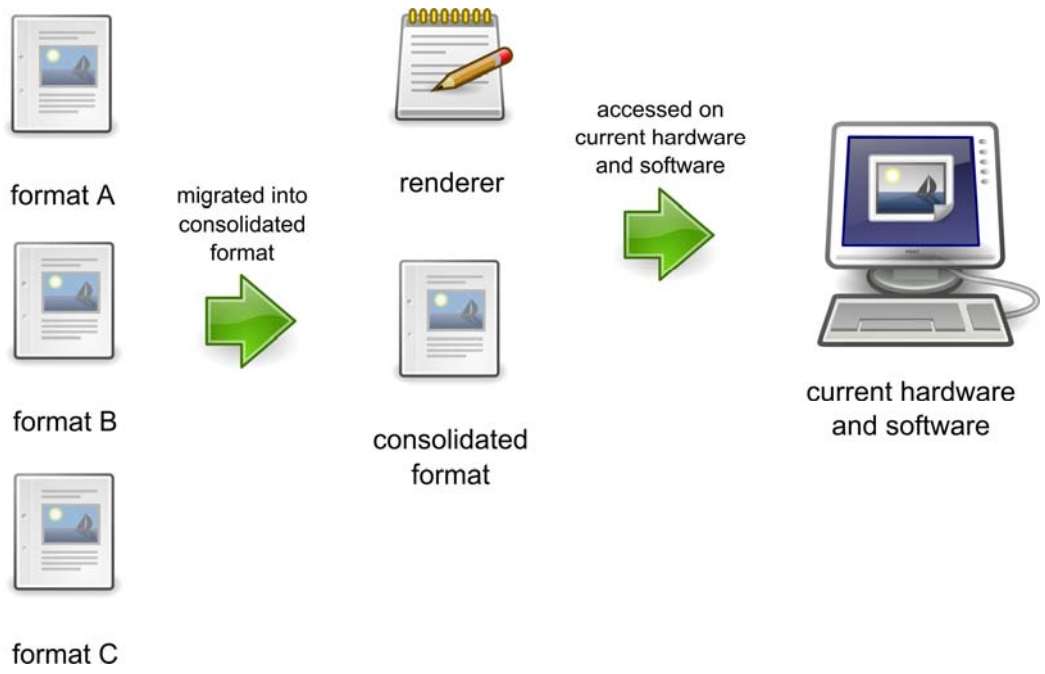


Figure 20: Renderer for Each Consolidated Format

Pros:

- Reduces total number of renderers that an institution is responsible for maintaining.

Cons:

- Requires acceptable level of loss to be articulated for files being migrated to consolidated format.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Generic Renderers

In this method, a renderer is acquired or created that renders the content for the largest possible number of file formats. This could be done in conjunction with consolidated formats, or at the level of individual formats.

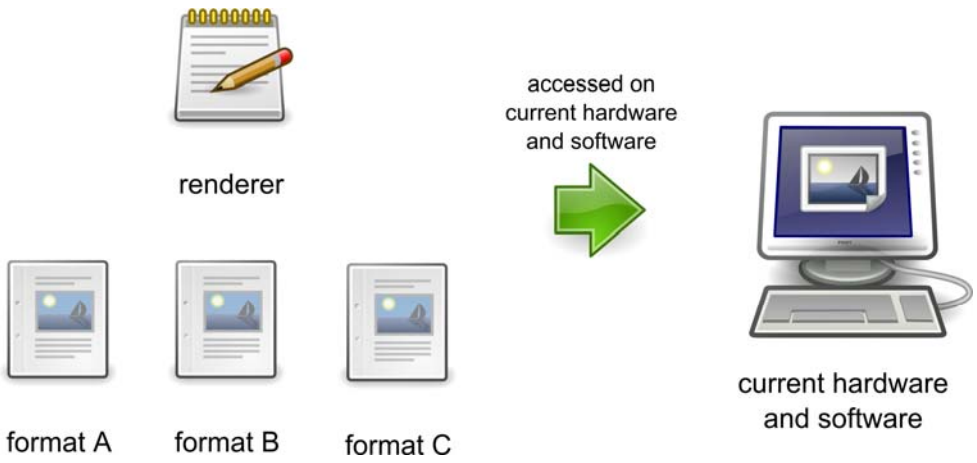


Figure 21: Generic Renderer

Pros:

- reduces number of emulators that an institution is responsible for maintaining.

Cons:

- The more formats a renderer can access, the more complex the renderer, and the more difficult it may be to maintain.
- If access to a single renderer is lost, access to many file format types may also be lost.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

1.5 Technological Museum

What it is:

The institution or individual will collect and maintain the original hardware and software that was used to create or access digital material. This scenario is predicated on understanding and maintaining desired access paths, and their dependencies.

In almost all cases, institutions will be reliant on maintaining hardware for some length of time, even if this is only in the context of providing the current operating platform for other preservation actions. This preservation approach focuses more on the maintenance of hardware over extended periods of time, but will in part be valid even for shorter lengths, such as the 'refresh cycle' for an institution.

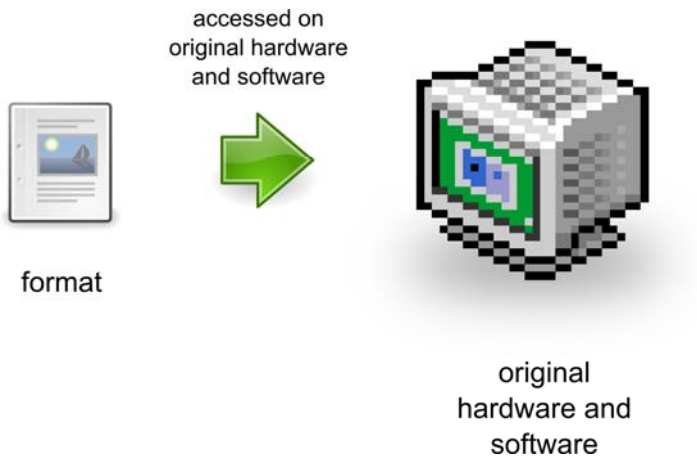


Figure 22: Techology Museum

Pros:

- Using the original environment provides a proven methodology for accessing contemporary digital materials;
- May provide access to physical carriers which are not readable using modern hardware; and
- May provide the only option for reading certain digital materials.
- May be the only viable option for accessing some carriers.

Cons:

- Requires constant preservation planning, and assumes a great deal of understanding of the systems being preserved;
- Assumes a certain level of understanding, knowledge, and documentation of the original access environment, which may be difficult to retain in corporate knowledge;
- May require a great deal of storage real-estate, particularly in the case of older machines;
- Equipment has a life-cycle that can be extended, but which cannot be extended indefinitely. Sooner or later, hardware will fail;
- It may be difficult to implement this solution with any content or preservation management system.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

- Over time, it may become harder to actually do anything meaningful with the digital content, as it will become more difficult for modern hardware to interface with the original hardware;
- Equipment will fail over time (intermittently, and catastrophically), and the older the equipment, the more difficult and cost prohibitive it will be either to find a suitable replacement, or repair;
- The loss of a single dependency may inhibit an entire access path;
- Over time, older equipment may become increasingly hazardous, through the decomposition of chemical components, or from electrical failure, etc.;
- There are innumerable valid variations and permutations for any given access path, which may require even more equipment to be stored and maintained;
- Physical media carriers may degrade at a faster rate than the technology used to access those carriers; and
- Support for hardware will, in some cases, end potentially before the useful life-span of the equipment.

Further reading:

Harvey, R., 2005. *Preserving Digital Materials*. K G Saur, München. pp.127-128.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Technology Museum Access Strategies:

There are a number of ways in which an institution can approach the collection and maintenance of hardware for the purposes of preservation, which are outlined in this section.

Access Paths for Individual Formats

A specific environment is maintained for each file format.

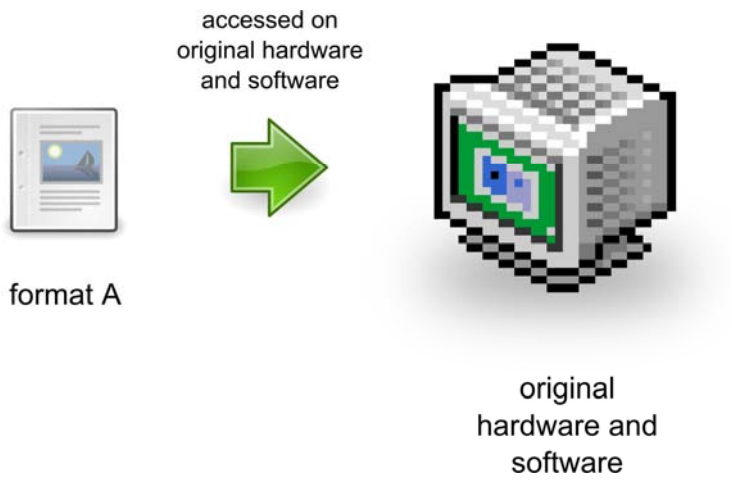


Figure 23: Access Path for Each Format

Pros:

- Probably most reliable way to provide appropriate access to object without change.

Cons:

- Depending on the number of formats for which access must be preserved, can very quickly become unsustainable.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Access Paths for Consolidated File Formats

In this method, various file formats that share similar characteristics are consolidated into a single file format. While there is still an access path for each format, the total number of these is drastically reduced.

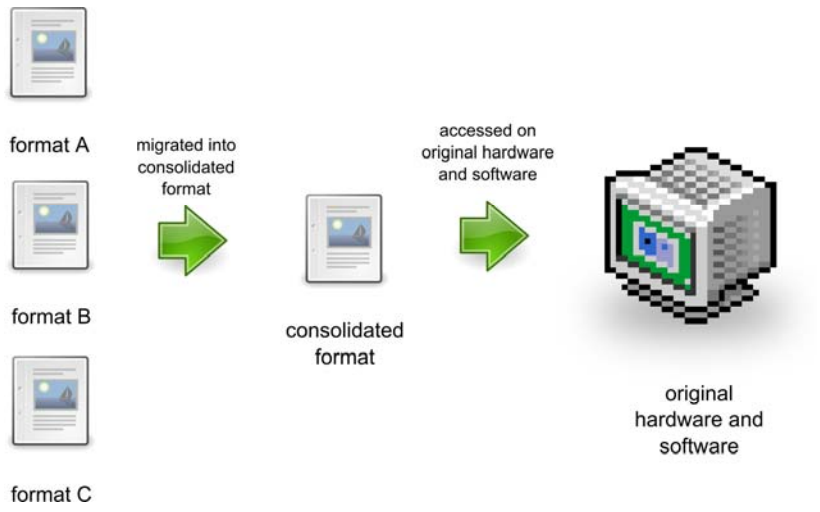


Figure 24: Access Path for Each Consolidated Format

Pros:

- Reduces number of access paths the institution is responsible for maintaining, making this approach overall more practical.

Cons:

- If file formats are being consolidated into a new format, it would be possible to migrate them into formats which would be more easily preserved and accessed on modern computers, thus reducing most of the utility associated with maintaining a technology museum for the sake of access.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Generic Access Paths

In this method, access paths are created and maintained that can access the largest number of file formats possible. This could be done in conjunction with consolidated formats, or at the level of individual formats.

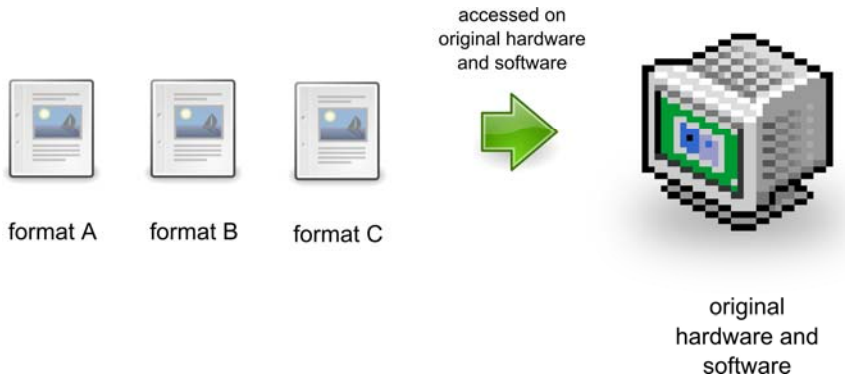


Figure 25: Generic Access Paths

Pros:

- Reduces number of access paths the institution is responsible for maintaining, making this approach overall more practical.

Cons:

- The more file formats a system is capable of accessing, the more dependencies will be inherent in that system. This will make the system more difficult to maintain, and problems more difficult to diagnose; and
- If access to that system should be lost, then access to all the file formats for which that system is accountable is potentially also lost.

Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia

Aknowledgements

Maxine Davis, Andrew Long, Colin Webb.

www.nla.gov.au
Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia