

# Towards a model for a trusted digital repository

*Facilitating access to research information  
and knowledge*



Presented by Ria Groenewald & Ina Smith  
Knowledge, Archives and Records Management Conference

6 May 2008



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA

# Agenda



Part I	Digitization & Preservation	Ria Groenewald <i>ria.groenewald@up.ac.za</i>
Part II	Preservation & Trusted Digital Repositories	Ina Smith <i>ina.smith@up.ac.za</i>



# Part I: Digitization and Preservation

Ria Groenewald



# Alexandria, Egypt







**Demetrius Phalereus (350 - 280 B.C.)**

**The inspirer of the foundation of the Ancient Library**



9/11- 2001



[http://www.cathousechat.com/cathouse\\_chat/WindowsLiveWriter/TwinTowers911.jpg](http://www.cathousechat.com/cathouse_chat/WindowsLiveWriter/TwinTowers911.jpg)

[http://www.alarmingnews.com/archives/Twin-Towers-Reflected\\_1.jpg](http://www.alarmingnews.com/archives/Twin-Towers-Reflected_1.jpg)

[www.arsenalofhypocrisy.com/.../image015.jpg](http://www.arsenalofhypocrisy.com/.../image015.jpg)

## IRAQ 2003



“Rampant looting followed the U.S. occupation of Iraq in 2003. Jeffrey Spurr, a Middle-Eastern librarian at Harvard University, says more than typewriters or desks got lifted. He told us archival material and rare books from the national library and archives was stored in a basement. Then "parties unknown, aware that valuable material were there, stole what they desired, and broke the pipes to flood the rest, covering their tracks completely.”

Pile of documents after looting at a library, Bayt al-Hikma, Iraq. Courtesy [Nabil al-Tikriti](http://www.whyfiles.org/235loot/images/iraqi_library.jpg)  
[http://www.whyfiles.org/235loot/images/iraqi\\_library.jpg](http://www.whyfiles.org/235loot/images/iraqi_library.jpg)



**IRAQ 2003**



Photo: Gleb Garanich/Reuters  
The looted and burned National Library and Archive in Baghdad in April 2003, a week after United States forces seized the capital.



# HHURRICANE KATRINA 2005



NASA Earth Observatory: [http://earthobservatory.nasa.gov/Newsroom/NewImages/images.php3?img\\_id=17017](http://earthobservatory.nasa.gov/Newsroom/NewImages/images.php3?img_id=17017)



<http://www.regent.edu>  
UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA  
[www.regent.edu](http://www.regent.edu)  
Denkeleers • Leading Minds • Dikgopolo Ha Dihaleli



# National Library of Egypt, Cairo







# Future of academic libraries

*No.1 assumption (ACRL, March 2007)*



*There will be an*

- preserving digital archives, and*
- increased emphasis on digitizing collections*
- improving methods of data storage and retrieval*
- *The digitization of unique print collections may emerge as one of the primary missions of academic libraries in the 21st century*
- *Librarians should collaborate with disciplinary colleagues in the **curation** of data as part of the research process*

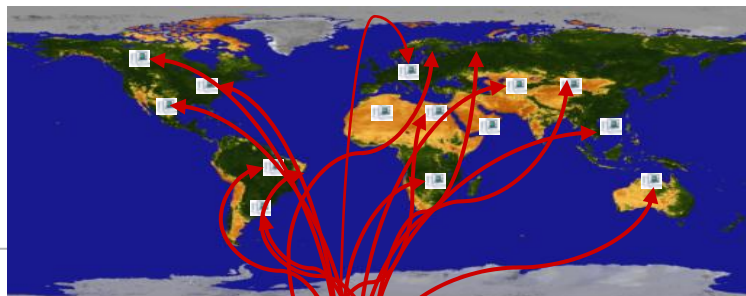
<http://www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2007/april07/tenassumptions.cfm>



# Digital Books workflow



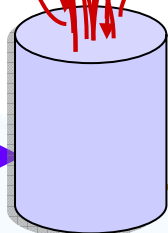
Digital workflow of the Alexandria Library.  
Software for this workflow is available at  
[http://wiki.bibalex.org/DAFWiki/index.php/Main\\_Page](http://wiki.bibalex.org/DAFWiki/index.php/Main_Page)



QA

QA

Metadata Editor



Unique URI created for object

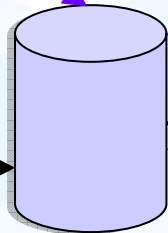
UPSpace | R

QA

QA

Send to submitters via

- email
- External hard drive
- DVD/CD/Flashdrive
- Internal server



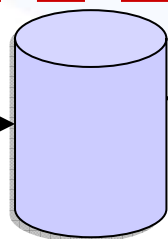
Reviewer

UPSpace | R

QA

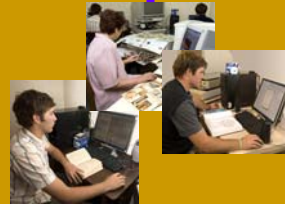
QA

• Scan directly to archival server



Archival server

- Copy from AS
- Quality Control
- Deskew/cleaning/derivating/filter
- Safe webready
- Final QC + Storage





# Standards



- Preservation Metadata Framework Working Group (OCLC, 2003)
- PREMIS (2005)
- OAIS (Open Archival Information System)
- Z39.87 - Standard for Technical Metadata for Digital Still Images(ANSI/NISO)

# Preservation Metadata Framework Working Group (Report 2003)



## Framework for research

- Outline the types of information that should be associated with an archived digital object
- The use of metadata to support the digital preservation process

[http://www.oclc.org/research/projects/pmwg/presmeta\\_wp.pdf](http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf)



# PREMIS Working Group (2005)



## The PREMIS (Preservation Metadata: Implementation Strategies Working Group -

- Develop a data dictionary of core elements for archived objects
- Guide the implementation of element sets in preservation systems
- Suggest best practice for populating the elements

[http://www.oclc.org/research/projects/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/projects/pmwg/pm_framework.pdf)

# OAIS

## (Open Archival Information System)



- The OAIS (Open Archival Information System) reference model was developed under the auspices of NASA's Consultative Committee for Space Data Systems (CCSDS)
- The OAIS reference model is a conceptual framework for a digital archive
- Regarded as the “standard” for digital object repositories



# Z39.87 - Standard for Technical Metadata for Digital Still Images (NISO & AIIM)



Z39.87 is a standard which defines a set of metadata elements for raster digital images

The purpose is to help in the development, exchange and interpretation of digital images

The original DIG35 goals were adapted by the NISO group

# Scanning



- No set resolution can be selected for all projects
- Resolution for a master image range between 300 - 600 dpi
- Colour settings 8-bit greyscale; 24-bit colour
- The most widely adopted format for storing a preservation quality digital master is uncompressed TIFF

# Derivative image

---



- A derivative is a manipulated image derived from the master image, to produce smaller file sizes
- Lossy file formats such as JPEG are used for derivative images
- Resolution ranges between 72 dpi and 150 dpi and up to 800 pixels in width
- ICC (International Colour Consortium) profiles



# Reasons for preservation



- Updated versions of the file format
- Reading device become obsolete
- Updated versions of the software used to create, manage, or access digital content
- Changes in computers
- Movement at vendors level
- Unforeseen errors

# Requirements of data protection



- Visibility/accessibility
- Regular quality control
- Authenticity
- Security
- Performance
- Ease of use
- Interoperability
- Cost of ownership
- Automation

*Web Buyers Guide, 31-03-08*

# Refreshing



- **Refreshing:**  
Copy the same type of digital information from one long-term storage medium to another
- **Modified refreshing:**  
Copy information to another medium of a similar type
- Refreshing is part of a process or program
- Refreshing address issues such as decay and obsolescence



# Migration and Emulation



- **Migration:**  
Move or adapt the objects to another platform
- **Emulating:**  
Environment will be adapted to new platform (the objects themselves will not be tampered with)

# Preserve the usability of a .TIFF file



- A **TIFF viewer**, plus its formal specification and sufficient subsidiary documentation to explain how it work in practice must be preserved
- To run the TIFF viewer - an **operating system** must be preserved
- To run the operating system -
  - the **original hardware** will need to be preserved, or
  - **emulation software** that allows the old hardware to be emulated on new machines needs to be developed



# The technical registry PRONOM

- Welcome
- About
- Add an entry
- Search
- Help
- Information resources

## Welcome to PRONOM

The online registry of technical information. PRONOM is a resource for anyone requiring impartial and definitive information about the file formats, software products and other technical components required to support long-term access to electronic records and other digital objects of cultural, historical or business value.

[Search PRONOM](#)

### Tools and Services

Free PRONOM tools and services to support digital preservation, including DROID, the automatic file format identification tool, together with links to relevant external tools and services. [Tools and Services](#)

### Contribute to PRONOM

Contribute new information to PRONOM via our online [submission form](#). Find out more about [PRONOM's creators](#).

**New to PRONOM?**

- [What is PRONOM?](#)
- [How do I search PRONOM?](#)
- [Who is PRONOM for?](#)
- [How do I find out more?](#)





You are here: Home > Services for professionals > Preservation > PRONOM > Search: PRONOM



The technical registry PRONOM

- Welcome About Add an entry Search Help Information resources

Search : Simple Search

Help : Simple Search

- Simple search File format PRONOM Unique Identifier Software Vendor Lifecycles

1. Search

Enter a simple search string and then click 'search'.

JPEG Search

You searched for: ".JPEG"

Save as... XML | CSV Print  
page 1 2

JPEG File Interchange Format (1.00)

The JPEG File Interchange Format (JFIF) is a file format for storing JPEG-compressed raster images. It was developed by the Independent JPEG Group and C-Cube Microsystems, in the absence of any such format being defined in the JPEG standard, and rapidly became a de facto standard; this is what is commonly referred to as the JPEG file format. A JFIF file comprises a JPEG data stream together with a JFIF marker. It begins with a Start of Image (SOI) marker, immediately followed by a JFIF Application (APP0). This is followed by the JPEG image data, which is terminated by an End of Image (EOI) marker. JFIF supports up to 24-bit colour and uses lossy compression (based on the Discrete Cosine Transform algorithm). Other types of compression are available through JPEG extensions, including progressive image buildup, arithmetic encoding, variable quantization, selective refinement, image tiling, and lossless compression, but these may not be supported by all JFIF readers and writers.

JPEG File Interchange Format (1.01)

The JPEG File Interchange Format (JFIF) is a file format for storing JPEG-compressed raster images. It was developed by the Independent JPEG Group and C-Cube Microsystems, in the absence of any such format being defined in the JPEG standard, and rapidly became a de facto standard; this is what is commonly referred to as the JPEG file format. A JFIF file comprises a JPEG data stream together with a JFIF marker. It begins with a Start of Image (SOI) marker, immediately followed by a JFIF Application (APP0). This is followed by the JPEG image data, which is terminated by an End of Image (EOI) marker. JFIF supports up to 24-bit colour and uses lossy compression (based on the Discrete Cosine Transform algorithm). Other types of compression are available through JPEG extensions, including progressive image buildup, arithmetic encoding, variable quantization, selective refinement, image tiling, and lossless compression, but these may not be supported by all JFIF readers and writers.

JPEG File Interchange Format (1.02)

The JPEG File Interchange Format (JFIF) is a file format for storing JPEG-compressed raster images. It was developed by the Independent JPEG Group and C-Cube Microsystems, in the absence of any such format being defined in the JPEG standard, and rapidly became a de facto standard. A JFIF file comprises a JPEG data stream together with a JFIF marker. It begins with a Start of Image (SOI) marker, immediately followed by a JFIF Application (APP0) marker and one or more optional application extension markers. This is followed by the JPEG image data, which is terminated by an End of Image (EOI) marker. JFIF supports up to 24-bit colour and uses lossy compression (based on the Discrete Cosine Transform algorithm). Other types of compression are available through JPEG extensions, including progressive image buildup, arithmetic encoding, variable quantization, selective refinement, image tiling, and lossless compression, but these may not be supported by all JFIF readers and writers.

<http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=7!>

[PRONOM | Search by format](#)

**Details for: JPX (JPEG 2000 Extended)**

[Save as...](#) [XML](#) | [CSV](#) [Print](#)

Go to: [Summary](#) | [Documentation >](#) | [Signatures >](#) | [Compression >](#) | [Character encoding >](#) | [Rights >](#) | [Reference files >](#) | [Properties >](#)

### Summary

<b>Name</b>	JPX (JPEG 2000 Extended)
<b>Version</b>	
<b>Other names</b>	JPF
<b>Identifiers</b>	PUID: fmt/151
<b>Family</b>	
<b>Classification</b>	
<b>Disclosure</b>	
<b>Description</b>	<p>JPEG 2000 Extended File Format is an optional, lossless, extension of the JP2 format, the main file format for the JPEG 2000 international standard for image coding, created by the Joint Photographic Experts Group in 2000 for the compression of photographic images for storage or transmission. The standard specifies how an image and its metadata is transformed into byte stream data. This format is widely used for storing and transmitting photos and other compressed image data online. Although it is based on JP2 it can support multiple layers, animation and other features, and is a lossless image compression file. Like JP2 the format is made up of a contiguous sequence of boxes, beginning with a signature box and a file type box, which provides version and file type information.</p>
<b>Orientation</b>	Text
<b>Byte order</b>	Big-endian (Motorola)
<b>Related file formats</b>	None.
<b>Technical Environment</b>	

Internet 100%



# Preservation of the format



Digital formats contain texts, databases, still and moving images, audio, graphics, software and web pages. They are fragile and require purposeful production, maintenance and management to be retained

- *Viability* - maintenance of the bitstream
- *Renderability* - viewable by humans and processible by computers
- *Understandability* - interpretable by humans

<http://www.icpsr.umich.edu/dpm/dpm-eng/terminology/preservation.html>



# Part II: Preservation & Trusted Digital Repositories

Ina Smith

# Institutional Repository



“A university-based institutional repository is a **set of services** that a university offers to the members of its community for the **management and dissemination of digital materials** created by the institution and its community members. It is most essentially an **organizational commitment** to the **stewardship** of these digital materials, including **long-term preservation** where appropriate, as well as **organization** and **access** or **distribution**.”

Clifford A. Lynch,

["Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age"](#) ARL, no. 226 (February 2003): 1-7.

# Digitally born & digitized material



## CLINICAL IMAGES

### Awaking a sleeping epidemic

Riaan van Collier, Elna van Raasburg, Claz Schutte, Delese Brink, Gerhard Walthagen, M G Dove

Two patients with African sleeping sickness (SS) presented to the neurology unit, Pretoria Academic Hospital, during 2004 and 2005. SS has shown a recent resurgence, with epidemics in the Sudan, Angola and the Democratic Republic of Congo. The number of infected people in Africa is currently estimated at more than 500 000. According to the World Health Organization (WHO), about 20 *Trypanosoma brucei gambiense* and 30 *T. b. rhodesiense* infections are diagnosed yearly outside endemic areas in Africa. Migration, tourism, peacekeeping and military interventions and the re-emergence of SS epidemics might increase these numbers.<sup>1</sup>

The electroencephalogram (EEG) is often useful in the diagnosis of coma and delirium, but has not been widely used in the diagnosis of SS. The EEG is proposed as a novel way to follow disease progression, treatment response and treatment-induced encephalopathy.

#### Case 1

A 27-year-old man presented with a 4-month history of fatigue, loss of appetite, intermittent severe headaches, excessive daytime sleepiness, loss of concentration and insomnia. He had travelled to Malawi 8 months before admission. His temperature was 38.8°C, he had a palpable hepatomegaly and an unremarkable neurological examination although his cognitive response was slow. Diagnosis of African trypanosomiasis was made on a Giemsa-stained blood smear (Fig. 1).

Shortly after admission the patient had a tonic-clonic seizure, with post-ictal confusion. Treatment with suramin was started and repeat blood smears after 48 hours were clear of trypanosomes. The cerebrospinal fluid (CSF) showed no trypanosomes but a total protein level of 1.2 g/l, glucose 2.1 mmol/l, 4 polymorphs and 82 lymphocytes. WHO-recommended treatment with melarsoprol was started.<sup>1</sup>

Since no trypanosomes were isolated from inoculated mice the diagnosis of West African trypanosomiasis (WAT) was made. Eflornithine was unavailable and treatment with melarsoprol continued. The patient recovered well and returned to the UK.

Five months after discharge he presented to the Hospital for Tropical Diseases in London with fever, sleepiness and an active CSF. Diagnosis of a relapse was made which posed a diagnostic dilemma – recurring *T. b. gambiense*. Treatment with eflornithine was given which cleared his condition.

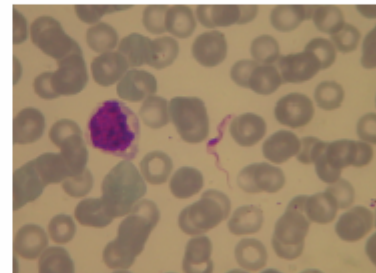


Fig. 1. Giemsa smear (case 1) showing extracellular trypanosomes present in the peripheral blood.

#### Case 2

A 53-year-old man presented with a 2-week history of fever, headache and episodic confusion. He was a farmer from Kariba in Zimbabwe where he had been treated for malaria without any clinical improvement. He gave a history of multiple tsetse fly bites but did not have a chancre. A Giemsa-stained blood smear showed *Trypanosoma* spp. On admission his temperature was 39.4°C, but the general examination was unremarkable. He was very sleepy but easily arousable. The diagnosis of East African trypanosomiasis (EAT) was confirmed by isolating *T. b. rhodesiense* from inoculated mice. Treatment with suramin was started and repeated Giemsa-stained blood smears did not show any trypanosomes. The following day he had a fatal cardiac arrhythmia, probably due to myocarditis.

Our first patient had a series of EEG recordings. These indicated a low-voltage mixed-frequency background with episodic, generalised but frontally dominant irregular delta activity (Fig. 2). Follow-up showed a gradual improvement in the frequency of the background. The second patient had an EEG



©University of Pretoria - Veterinary Science:  
Department Anatomy and Physiology



<https://www.up.ac.za/dspace/>

UPSpace at the University of Pretoria: Home - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.up.ac.za/dspace/>

powered by YAHOO! SEARCH QWeb Search W Wikipedia Spyware Scan 52°F (0) Games

**UPSpace**

UPSpace at the University of Pretoria >

**Welcome to UPSpace, the open access digital research repository of the University of Pretoria!**

**Items archived since January 2006: 3 474**

Inquiries: [upspace@up.ac.za](mailto:upspace@up.ac.za)

**Search**

Enter some text in the box below to search UPSpace.

**Communities in UPSpace**

Choose a community to browse its collections.

[Centre for the Study of AIDS \(CSA\)](#)  
[Economic & Management Sciences](#)  
[Education](#)

**Search UPSpace**

[Advanced Search](#)

[Home](#)

**Browse**

[Communities & Collections](#)  
[Titles](#)  
[Authors](#)  
[By Date](#)

**Sign on to:**

[Receive email updates](#)  
[My UPSpace authorized users](#)  
[Edit Profile](#)

[IRSpace - Search SA research repositories](#)

[Create your own Custom Search Engine](#)

[Gadgets powered by Google](#)

**Open Access Initiatives**

[DOAJ](#)  
[DRIVER](#)  
[eIFL SURFshare](#)  
[Google Book Search](#)  
[Google Scholar](#)  
[Journal Info \(Lund University\)](#)  
[OAISTER](#)  
[Open Archives Institute](#)  
[openDOAR](#)  
[ROAR](#)  
[SHERPA RoMEO](#)  
[Wikipedia](#)

**Become part of the UPSpace community**

Done Internet

100  
1908 - 2008



# Digital Info



- Modern computer technology – barely 50 years old
- Few have seen/used digital objects more than 30 years old
- Lack of experience & consensus on how to proceed with digital preservation processes
- Preserve for 100 years & more – “How old ... / from which era ....” (Jantz & Giarlo 2005)
- Lots of digital info lost already

# Trusted Repository *Defined*



“One whose mission is to provide **reliable, long-term access** to managed digital resources to its designated community, now and in the future.”

(RLG-OCLC Report 2002)

# Attributes of a Trusted Repository



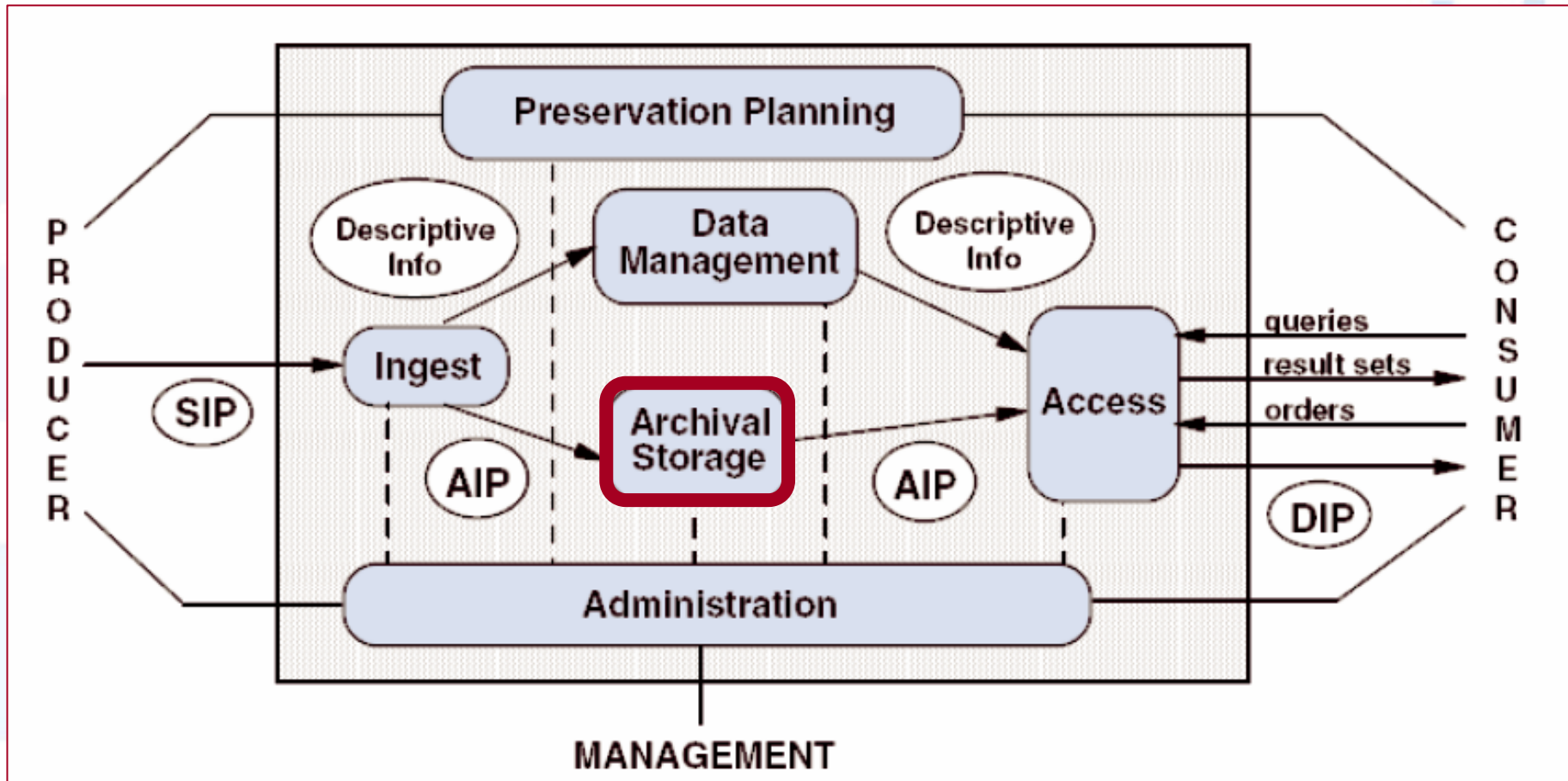
- Compliance with the *Reference Model for an Open Archival Information System (OAIS)*
- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technological & procedural suitability
- System security
- Procedural accountability

Source: Trusted Digital Repositories: Attributes and Responsibilities

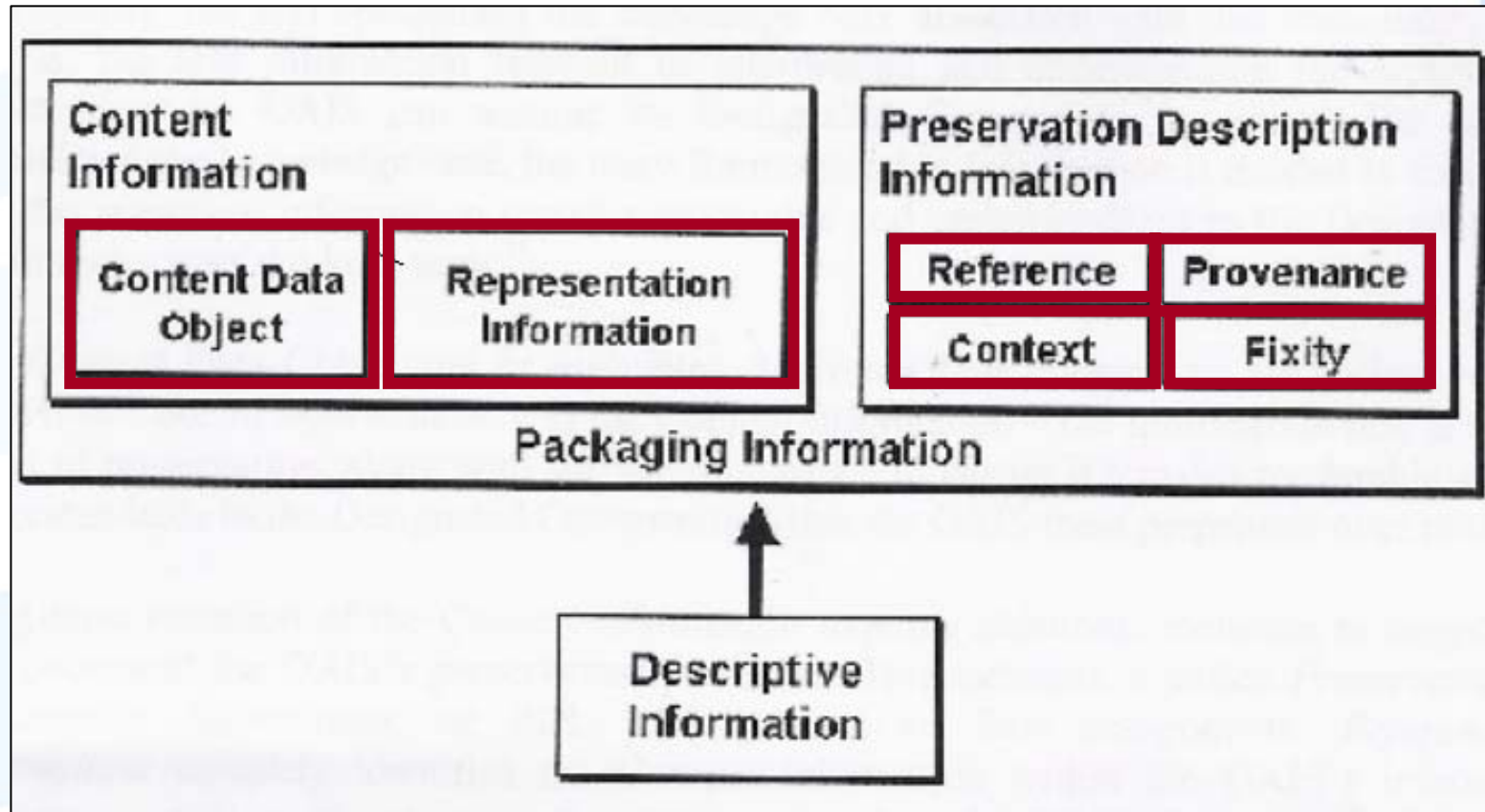
An RLG-OCLC Report <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>



# OAIS Functional Model – Archival Storage



# Archival Information Package *(Digital item submitted)*





# Technologies for enabling trust

# Digital Repository Software



- Proquest Digital Commons (*proprietary*)
- DSpace (*open source*)
- ContentDM (*proprietary*)
- Fedora (*open source*)
- E-Prints (*open source*)
- Greenstone (*open source*)



# DSpace Commitment to Preservation



- 2 levels of preservation: Bit & Functional
- Three levels of preservation for a given file format:
  - **Supported:** The format will be fully supported and preserved using either format migration or emulation techniques.
  - **Known:** The format can be recognised by DSpace, but full support cannot be guaranteed.
  - **Unsupported:** The format cannot be recognised by DSpace; these will be listed as "application/octet-stream", aka Unknown.
- Bit-level preservation will be done so that digital archaeologists of the future will have the raw material to work with if the material proves to be worth that effort.

- Groups
- Items
- Dublin Core Registry
- Bitstream Format Registry
- Workflow
- Authorization
- Edit News
- Supervisors
- Statistics
- Help
- Log Out

### Bitstream Format Registry

**Extensions** are comma-separated lists of file extensions. Do not include the dot.

When you add a bitstream format, it is not immediately visible. After you've finished editing the format metadata. Be sure to update the format. When you've updated the format, the format should appear in the submission UI list of formats.

[Help...](#)

E.g. Adobe PDF, XML, Text, HTML, MSWord - **Known**

ID	MIME Type	Name	Location	Support Level	Internal?	Extensions	Actions
1	application/octet-stream	Unknown	Unknown data	Unknown	<input type="checkbox"/>		Update
2	text/plain	License	Item-specific license agreement	Known	<input checked="" type="checkbox"/>		Update Delete...
3	application/pdf	Adobe PDF	Adobe Portable Document Format	Known	<input type="checkbox"/>	pdf	Update Delete...
4	text/xml	XML	Extensible Markup Language	Known	<input type="checkbox"/>	xml	Update Delete...
5	text/plain	Text	Plain Text	Known	<input type="checkbox"/>	txt asc	Update Delete...
6	text/html	HTML	Hypertext Markup Language	Known	<input type="checkbox"/>	htm html	Update Delete...
7	application/msword	Microsoft Word	Microsoft Word	Known	<input type="checkbox"/>	doc	Update Delete...

# DSPACE



UPSpace at the University of Pretoria: Item 22632273 - Microsoft Internet Explorer

Address: <http://www.up.ac.za/dspace/handle/22632273>

**Metadata**

**Title:** Elephant caudal vertebrae 18, left lateral view

**Author/s:** Seegers, Christine D.

**LC Subjects:** African elephant; veterinary anatomy; *Loxodonta africana*

**Keywords:** Elephant; C018, Left lateral view; Vertebrae, Animal; Spine, Animal

**Issue Date:** 2008-07-22

**Creation Date:** 2008-07-22

**Abstract:** Black/white wash painting technique. Original canvas size: (w)24.5 x (h)22.0 cm. Original scanned size in pixels: 6312 x 7872 pixels (800 dpi). Final size in pixels: 550 x 436 (150 dpi). Estimate download time: 37 sec @ 28.8 kbps.

**URI:** <http://hdl.handle.net/22632273>

**Rights:** ©University of Pretoria - Faculty of Veterinary Science, Dept of Anatomy and Physiology (original), ©University of Pretoria, Dept of Library Services (digital). Provided for educational purposes only. It may not be downloaded, reproduced, or distributed in any format without written permission of the original copyright holder. Any attempt to circumvent the access controls placed on this file is a violation of copyright laws and is subject to criminal prosecution. Please contact the collection administrator for copyright issues.

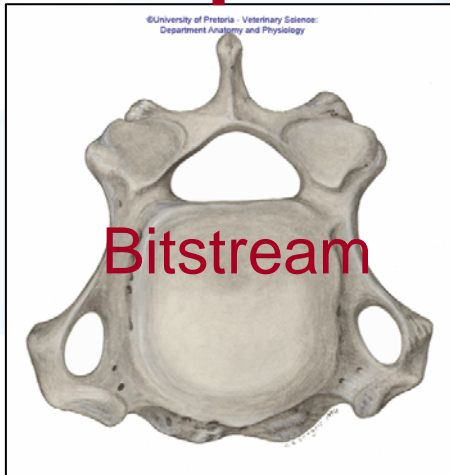
**Type:** Image

**Language:** English

**Appears in Collections:** The Elephant

**Files in This Item:**

File	Description	Size	Format
oil_040b.pdf		97kb	Adobe PDF



UPSpace at the University of Pretoria: Edit item policies - Microsoft Internet Explorer

Address: <http://www.up.ac.za/dspace/dspace-admin/authorize>

**Relationships stored between components in a bundle (METS Metadata Standard)**

Bitstream 6783 (license.txt)

ID	Action	EPerson	Group	
118949	READ	...	Anonymous	Edit Delete

Policies for Bundle 6001

ID	Action	EPerson	Group	
118852	READ	...	Anonymous	Edit Delete

Bitstream 6782 (oil\_040b.pdf)

ID	Action	EPerson	Group	
118851	READ	...	Anonymous	Edit Delete

All items in UPSpace are [copyright](#) protected, with all rights reserved. Inquiries: [upspace@up.ac.za](mailto:upspace@up.ac.za) - [Feedback](#)

# File formats



- Recognise the preservation risks of file formats
- Store content in **open format** on IR – pdf + additional
- Specify restricted range of deposit formats
- Investigate use of XML to describe data and metadata
- For verification purposes original copy should be available on Archival Server (TIFF)
- Plan for migrating rare and obsolete file formats
- Maintain file format information



# UPSpace Policy for file formats



- **Everything** put in UPSpace will be **retrievable**
- As many **files formats** as possible will be **recognised**
- As many known **file formats** as possible will be **supported** through UPSpace
- **Formats and techniques** will be continuously **monitored** to ensure needs can be accommodated as they arise
- The **size of a bitstream** allowed for submission is currently **unlimited**, but this will be revised over time
- The same **file** can be submitted in **more than one format**, of which one must be **pdf** (does not apply to media files)

# Metadata

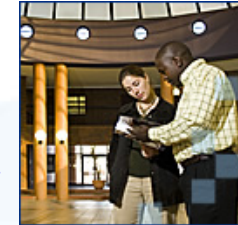


- Data about data
- Qualified Dublin Core Metadata Schema
- DSpace supports the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) v2.0 as a data provider
- Enhance descriptive metadata
- Capture technical & administrative metadata, preservation metadata

“Preservation metadata is the information necessary to maintain the viability, renderability, and understandability of digital resources over the long-term.”

Source: Feasibility and Requirements Study on Preservation of E-Prints/ Hamish et al.

# Preservation Metadata



identifier	uri	<a href="http://hdl.handle.net/2263/4726">http://hdl.handle.net/2263/4726</a>
description	abstract	In this paper, an attempt will be made to examine the concept ataraxia as it appears in the works of Pyrrho of Elis, Sextus Empiricus and
description	provenance	Scanned in 24-bit descreened colour 100% on DigiBookRGB10000 scanner at 400 dpi.
description	provenance	Submitted by Ina Smith (ina@ais.up.ac.za) on 2008-03-13T11:59:46Z No. of bitstreams: 1
description	provenance	Approved for entry into archive by Julene Vermeulen(julene.vermeulen@up.ac.za) on 2008-03-14T06:03:00Z (GMT) No. of bitstreams: 1
description	provenance	Made available in DSpace on 2008-03-14T09:00:22Z (GMT). No. of bitstreams: 1 PHv1_Wilkinson-009.pdf: 145337 bytes, checksum:
format	extent	145337 bytes
format	mimetype	application/pdf

# Checksums in DSpace



## Error detection techniques

File	Size	File Format	Checksum
<a href="#">artikel.pdf</a>	157,611 bytes	Adobe PDF ( <a href="#">known</a> )	3eb8f0dcb136f073944f8e9106529a02 (MD5)

Comparing the checksum displayed above with a checksum worked out on your local computer. They should be exactly the same.

DSpace generates an MD5 checksum for every file it stores; we use this checksum internally to **verify the integrity of files over time** (a file's checksum shouldn't change). You can use this checksum to be sure what we've received is indeed the file you've uploaded.



# Storage Management



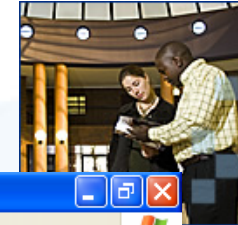
- Storage hardware is a key component of a repository
- SAN (Storage Area Network) vs NAS (Network Attached Storage)
  - Increased scalability: up to 16 million devices can be added
  - All other participants on SAN can connect and see each other
  - High-speed throughput: carry traffic between devices at 2 Gb/s
  - Independent of other network operations – functions separate from any LAN

# Persistent Identifiers



- Web references are untrustworthy; telephone numbers, IP addresses, Social Security numbers share properties of PID's – more trustworthy
- **Persistent Identifiers:** globally unique name assigned to a digital object that can be used in perpetuity, to refer to and to retrieve the digital object
- CNRI Handle System

# Persistent Identifiers




UPSpace at the University of Pretoria: Item 2263/5103 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Print Mail Stop

Address <https://www.up.ac.za/dspace/handle/2263/5103> Go Links Search Encarta

VMN powered by YAHOO! SEARCH Web Search Wikipedia Spyware Scan N/A 61°F (0) Games

 University of Pretoria

**UPSpace**

Search UPSpace  Go **Advanced Search**

Home

**Browse**

- Communities & Collections
- Titles
- Authors
- By Date

**Sign on to:**

- Receive email updates
- My UPSpace authorized users
- Edit Profile

[UPSpace at the University of Pretoria >](#)  
[Veterinary Science >](#)  
[Companion Animal Clinical Studies >](#)  
[Research Articles \(Companion Animal Clinical Studies\) >](#)

**Please use this identifier to cite or link to this item: <http://hdl.handle.net/2263/5103>**

**Title:** Eustachian tube diverticulum chondroids and neck abscessation in a case of *Streptococcus equi* subsp. *equi*

**Inquiries:** [ann.carstens@up.ac.za](mailto:ann.carstens@up.ac.za)

**Author/s:** Furniss, C.  
Carstens, A.  
Cilliers, I.

**LC Subjects:** *Streptococcus equi*  
Radiology  
Horses -- Surgery

**Keywords:** Carrier  
Chondroids  
Equine  
*Streptococcus equi* subsp. *equi*

**Issue Date:** Sep-2007

**Publisher:** South African Veterinary Association

**Citation:** Furniss, C, Carstens, A & Cilliers, I 2007, 'Eustachian tube diverticulum chondroids and neck abscessation in a case of *Streptococcus equi* subsp. *equi*', *Journal of the South African Veterinary Association*

100 1908 - 2008 Done Internet

# Digital Signatures




- **Digital signatures** added to full text
- Compute a digital signature for digital masters & store signature in technical metadata of object
- Compute signature for complete item and store externally to repository

**Document Status**

This document has special status or special features.

**Document Status**

 Certification VALID.

This document was Certified with the Digital Signature of:

Name: Broadgun Server Paid Sign  
Organization: Broadgun  
Issued by: Broadgun Server Paid Sign

The Digital Signature is valid and the document has not been subsequently tampered with.

Click 'Signature Properties' to view more information about the signature and its validity. Reliance upon this Certified document requires your acceptance of the terms described when you click 'Legal Notice'.

To access signature information later, open the Signatures Tab on the left, select the Certifying Signature and choose 'Properties' from the Options menu.

Do not show this dialog next time this document is opened.

Document Status icons are always located at the bottom-left corner of the document window. Click a Document Status icon to view this dialog again.



# Tools



## DRAMBORA <http://www.repositoryaudit.eu> Digital Repository Audit Method Based on Risk Assessment Toolkit

The screenshot shows the DRAMBORA website in a Microsoft Internet Explorer browser window. The address bar displays <http://www.repositoryaudit.eu/about/>. The website header features the title "DRAMBORA interactive" and the subtitle "Digital Repository Audit Method Based on Risk Assessment". Logos for the Digital Curation Centre (DCC) and Digital Preservation Europe (DPE) are visible in the top right.

On the left side, there is a registration form titled "Register a New Repository" with fields for "Username:" and "Password:", a "Remember:" checkbox, and a "Log In" button. Below the form are links for "Forgot Password?", "Home", "Latest News", "Get Expert Help", "Download Offline Version", and "Submit Feedback". A navigation menu includes "About", "Objectives", "Benefits", "The DRAMBORA Team", "Dissemination", and "DRAMBORA Users".

The main content area is titled "DRAMBORA: About" and includes a sub-heading "About DRAMBORA". It features an image of a hand holding a pen over a document with a checkmark and the word "YES". The text describes the toolkit as developed jointly by the Digital Curation Centre (DCC) and Digital Preservation Europe (DPE). It states that DRAMBORA represents the main intellectual outcome of a period of pilot repository audits undertaken by the DCC throughout 2006 and 2007. It presents a methodology for self-assessment, encouraging organisations to establish a comprehensive self-awareness of their objectives, activities and assets before identifying, assessing and managing the risks implicit within their organisation.

Below the main text, a paragraph explains that within DRAMBORA, digital curation is characterised as a risk-management activity; the job of digital curator is to rationalise the uncertainties and threats that inhibit efforts to maintain digital object authenticity and understandability, transforming them into manageable risks. Six stages are implicit within the process. Initial stages require auditors to develop an organisational profile, describing and documenting the repository's mandate, objectives, activities and assets. Latterly, risks are derived from each of these, and assessed in terms of their likelihood and potential impact. Finally, auditors are encouraged to conceive of appropriate risk management responses to the identified risk. The process enables effective resource allocation, enabling repository administrators to identify and categorise the areas where shortcomings are most evident or have the greatest potential for disruption. The process itself is an iterative one and therefore subsequent recursions will evaluate the effectiveness of prior risk management implementations.

At the bottom of the main content area, it lists "DRAMBORA Consortium Sponsors:".

On the right side, there is a "News" section with two entries: "Improving DRAMBORA: Your Help Required" dated 10th Apr 2008, and "DRAMBORA Interactive Released!" dated 9th Apr 2008. Below the news is a "Download" section with the text "Please register to download the offline copy of the DRAMBORA Toolkit". There is also a "Comment" section with the text "Submit feedback via our feedback form or email us." and a "Get Help" section with the text "Get help in the DRAMBORA evaluation process by volunteering to host a DRAMBORA facilitated assessment."

# Institutional Repository Workshop

A to Z of digital preservation within an Institutional Repository  
Business Plans, Policies, Digitization, Metadata, Implementation,  
Marketing & Buy-in and many more ...

**1 – 3 October 2008**

University of Pretoria

[www.library.up.ac.za/irtoolbox/workshop.htm](http://www.library.up.ac.za/irtoolbox/workshop.htm)

OR

**E-mail:**

[ria.groenewald@up.ac.za](mailto:ria.groenewald@up.ac.za)

[ina.smith@up.ac.za](mailto:ina.smith@up.ac.za)

# Join our IRSpace CoP!

---



E-mail us:

[ina.smith@up.ac.za](mailto:ina.smith@up.ac.za)

**Will your work withstand the test of  
times to come?**



**Questions?**



# Sources

1. Caplan, P., and Guenther, R. (Summer 2005). Practical Preservation: The PREMIS Experience. Library Trends, Vol. 54, No.1
2. Data Migration (Wikipedia). [http://en.wikipedia.org/wiki/Data\\_migration](http://en.wikipedia.org/wiki/Data_migration). Retrieved May 2008
3. DIG35. <http://xml.coverpages.org/dig35.html>
4. Digital Preservation Research. <http://www.kb.nl/hrd/dd/dd-en.html>
5. Dioscuri, KB Nationaal Argief. <http://dioscuri.sourceforge.net/> Retrived May 2008
6. Draft Charter on the Preservation of the Digital Heritage, Item 8.6 of the provisional agenda. 19 August 2003. General Conference 32<sup>nd</sup> session, Paris
7. Emulation [http://www.kb.nl/hrd/dd/dd\\_projecten/projecten\\_emulatie-en.html](http://www.kb.nl/hrd/dd/dd_projecten/projecten_emulatie-en.html)
8. Metadata Framework Working Group. [http://www.oclc.org/research/projects/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/projects/pmwg/pm_framework.pdf). Retrieved April 2008
9. Migration [http://www.kb.nl/hrd/dd/dd\\_projecten/projecten\\_migratie-en.html](http://www.kb.nl/hrd/dd/dd_projecten/projecten_migratie-en.html)
10. Mullins, J.L., Allen F.R., and Hufford, J.R. (April 2007). Top ten assumptions for the future of academic libraries and librarians: A report from the ACRL research committee. C&RL News, Vol. 68, No. 4
11. NDHA - Business Requirements Specification, July 2005 [www.natlib.govt.nz/ndha](http://www.natlib.govt.nz/ndha)
12. NISO. [http://www.niso.org/kfile\\_download?pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90\\_d5\\_ymGsj\\_IKVa86hjP37r\\_hM9t9qad1BrrORLqssvegis%3D](http://www.niso.org/kfile_download?pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hM9t9qad1BrrORLqssvegis%3D). Retrieved May 2008
13. OAIS Reference Model for an Open Archival Information System <http://public.ccsds.org/publications/archive/650x0b1.pdf>
14. PLANETS <http://www.planets-project.eu/>
15. PREMIS Working Group. [http://www.oclc.org/research/projects/pmwg/presmeta\\_wp.pdf](http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf). Retrieved April 2008
16. PRONOM <http://www.nationalarchives.gov.uk/pronom/> Retrieved May 2008
17. Rog, J., van Wijk, C. (2008) Evaluating File Formats for Long-term Preservation, National Library of the Netherlands; The Hague, The Netherlands. [http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/publicaties/KB\\_file\\_format\\_evaluation\\_method\\_27022008.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/KB_file_format_evaluation_method_27022008.pdf)
18. Weinberger, E., Clayton, R., Anderson, R. A security policy for a digital repository. Cambridge University. <http://www.cus.cam.ac.uk/~ew206/security-policy.html>
19. Wikipedia on Z39.87. <http://en.wikipedia.org/wiki/Z39.87>
20. Wikipedia on NISO. [http://en.wikipedia.org/wiki/National\\_Information\\_Standards\\_Organization](http://en.wikipedia.org/wiki/National_Information_Standards_Organization)
21. <http://www.nyu.edu/its/humanities/ninchguide/VI/> Retrieved March 2007