What are some of the things that the ‘Feds’ are doing about Digital ‘Stuff’?

Presentation to
NSW State Library

David Pearson

15 February 2011

NATIONAL LIBRARY OF AUSTRALIA

Collecting framework for Digital Collections

The *National Library Act 1960:*

requires us to maintain and develop "…a comprehensive collection of library material relating to Australia and the Australian people"

The *Collection Development Policy*:

outlines collecting intentions for print and digital resources of all kinds – e.g., pictures, maps, publications …

NATIONAL LIBRARY OF AUSTRALIA

How do we get it?

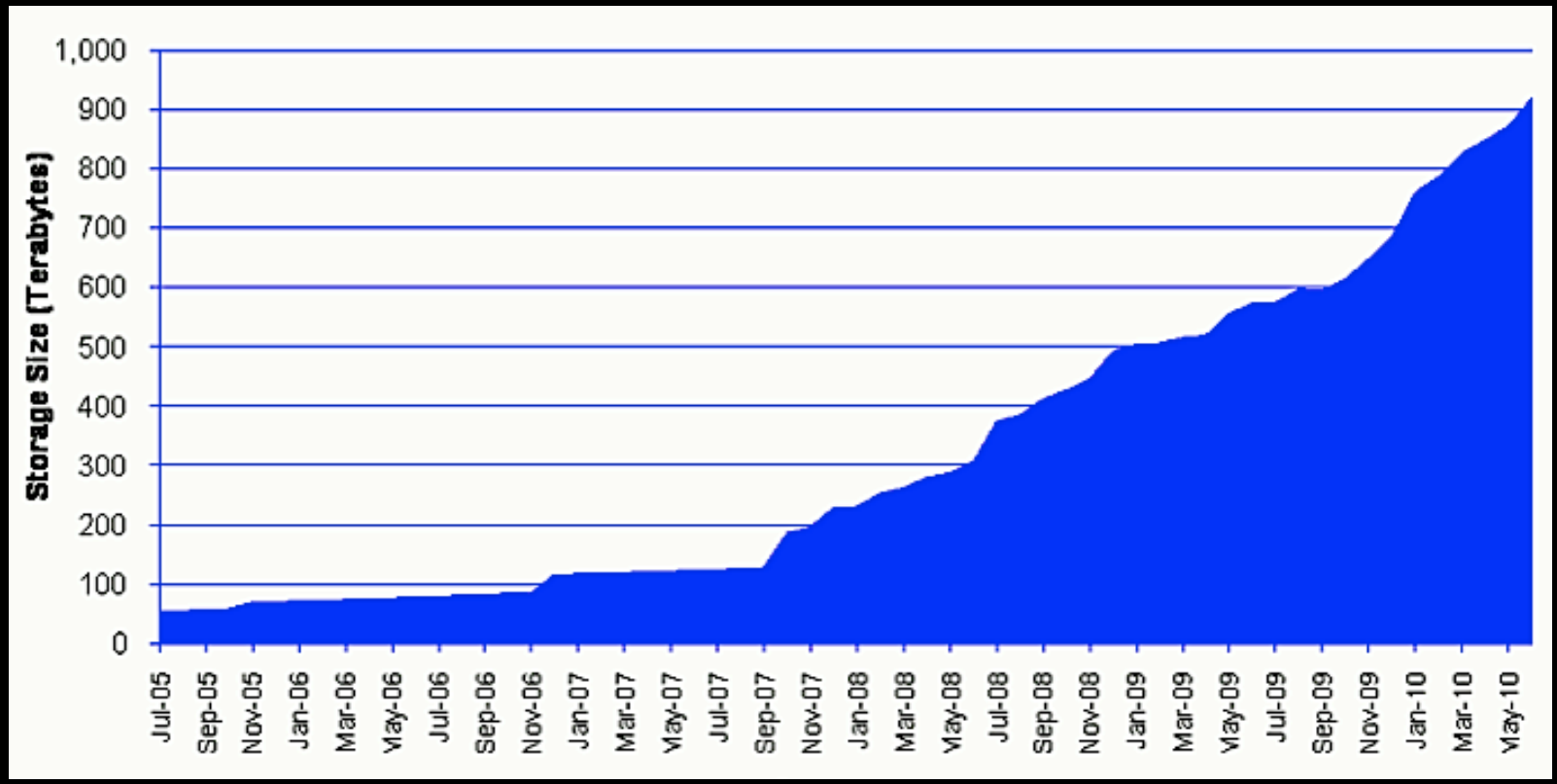
The NLA collects digital ‘stuff’ in many forms and through a number of different ingest mechanisms. Generally this can be categorised as:

- Internally generated;

- External sourced (purchase, donation and future legal deposit).

NATIONAL LIBRARY OF AUSTRALIA

Growth in digital collections stored
July 2005 to June 2010

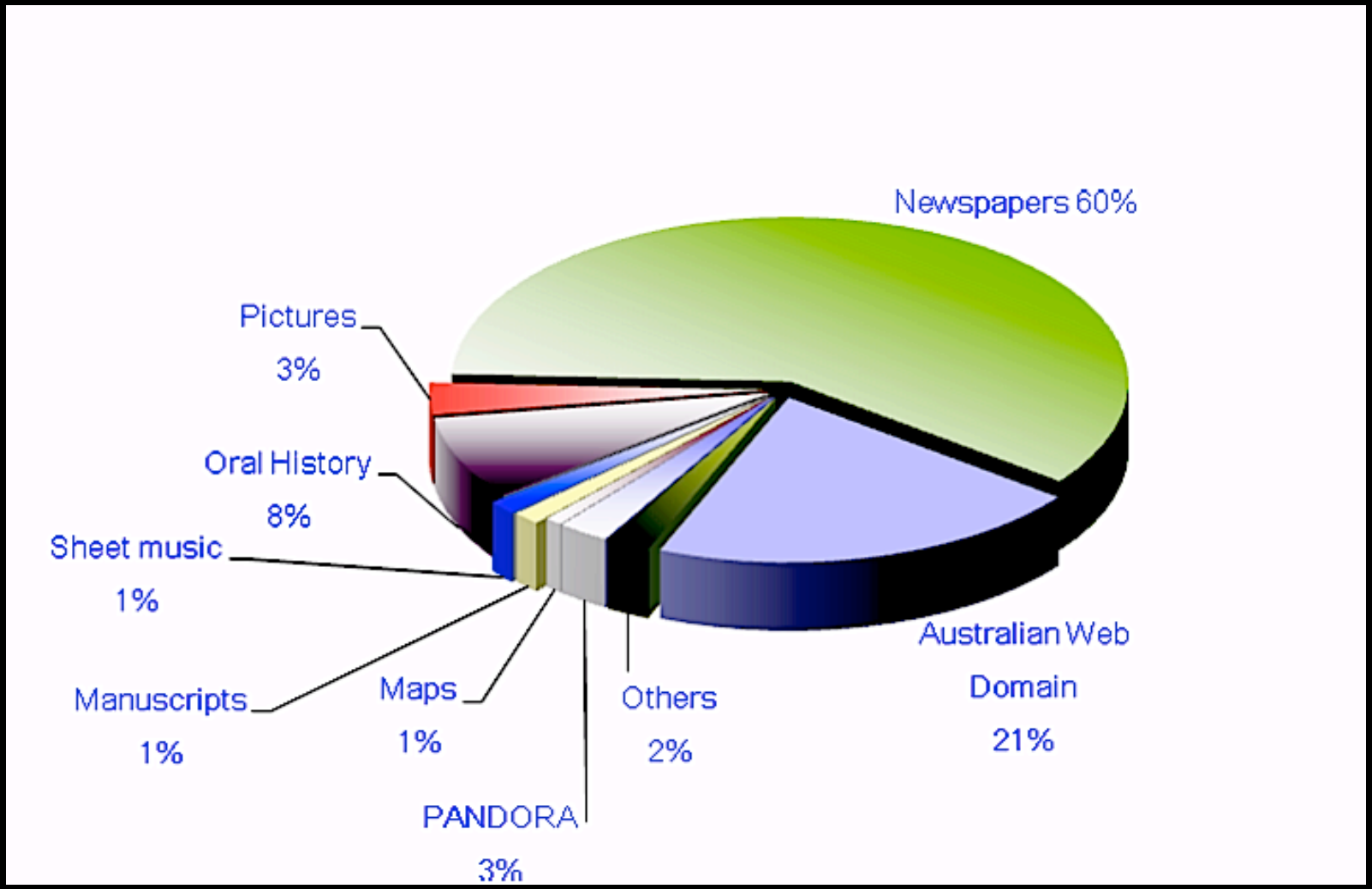


1,000 Terabytes = 1 Petabyte

(= 1,000,000 Gigabytes or 250,000 x 4 GB USB memory sticks)

NATIONAL LIBRARY OF AUSTRALIA

Diversity of digital ‘stuff’ stored by material type - 2009/10



NATIONAL LIBRARY OF AUSTRALIA

Who is responsible for digital ‘stuff’ ?

At the NLA, Digital Preservation is not just the name of a business unit, it is the act of maintaining access to collection materials.

Therefore everyone at the NLA is responsible for digital preservation in some way:

- Collections Owners;
- Digital Preservation; and
- IT.

NATIONAL LIBRARY OF AUSTRALIA

Digital Preservation is about maintaining access to the collection

But this is an ongoing process as ‘digital stuff’ is dependent on technology at all stages:

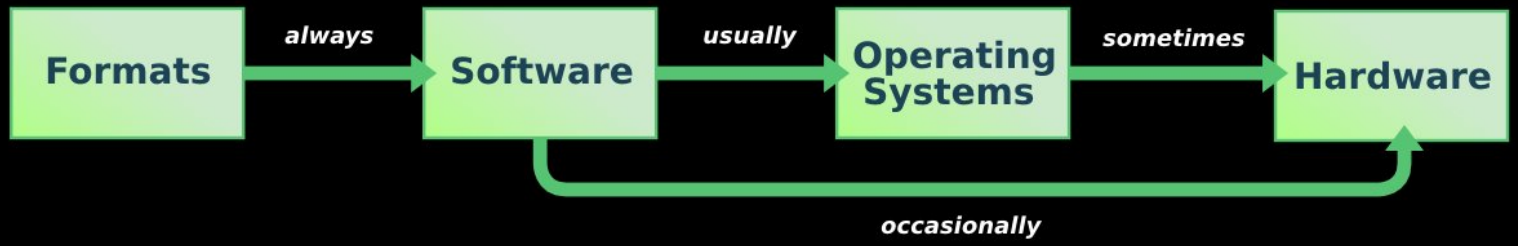
- Creation/capture;
- Storage; and
- Access.

For some of us, the scale of the problem is exacerbated by the scale of our digital ‘stuff’.

NATIONAL LIBRARY OF AUSTRALIA

The general preservation problem

The digital ecosystem is inherently fragile in that our ability to access the dependences for a file change over time (sometimes rapidly, sometimes more slowly).



NATIONAL LIBRARY OF AUSTRALIA

Entropy

- Loss of availability or degradation of the technology over time.
- Loss of skills and understanding of the technology
- This is a constant process

Entropy effects every part of the ecosystem, from the data to the mechanisms to access the data.

NATIONAL LIBRARY OF AUSTRALIA

There are the things that we know.

There are the things we don't know.

Then there are the things we know we don't know.

And things we don't know that we don't know (D. Rumsfeld)

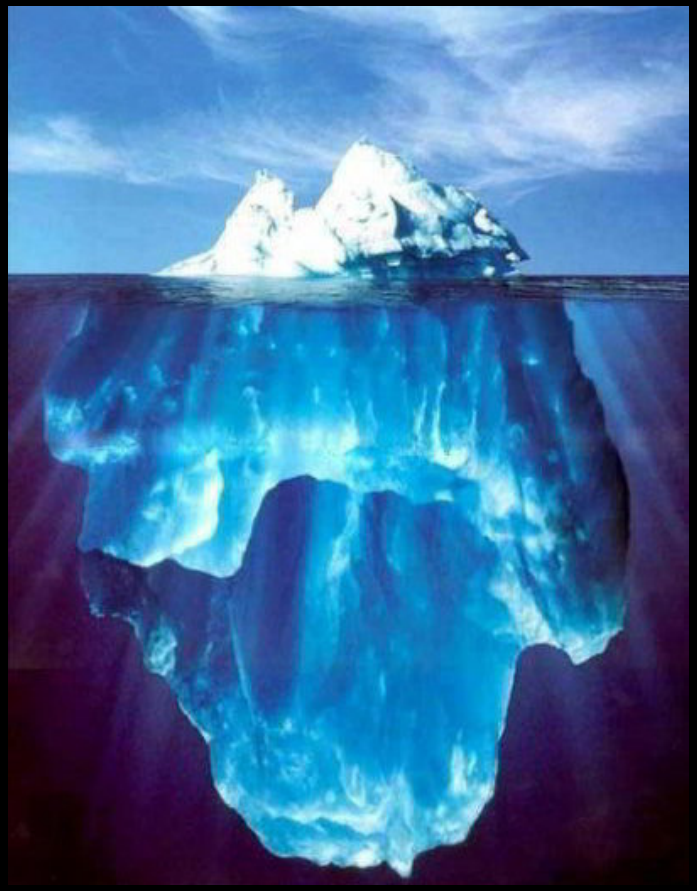


Image from http://media.photobucket.com/image/image%20of%20iceberg%20cross%20section%20water/unenergy/Black%20Beach/iceberg-below-water.jpg

NATIONAL LIBRARY OF AUSTRALIA

Maintaining accessibility: The three headed dog

There are three main facets to our preservation problem:

- Maintaining access to the bit-stream;
- Maintaining access to content;
- Maintaining access to meaning.

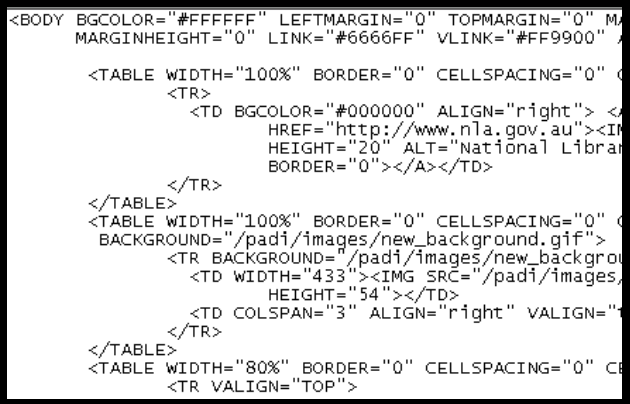
Image from https://supportforums.cisco.com/message/3269168

NATIONAL LIBRARY OF AUSTRALIA

Access to content and meaning

Levels of Representation

Bits



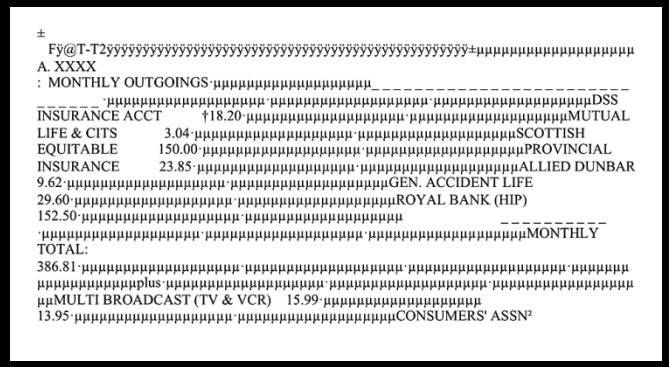
? Meaning

? Meaning

NATIONAL LIBRARY OF AUSTRALIA

Bits

Item from an NLA Manuscripts collection



A. XXXX
: MONTHLY OUTGOINGS

DSS INSURANCE ACCT †18.20
MUTUAL LIFE & CITS 3.04
SCOTTISH EQUITABLE 150.00
PROVINCIAL INSURANCE 23.85
ALLIED DUNBAR 9.62
GEN. ACCIDENT LIFE 29.60
ROYAL BANK (HIP) 152.50

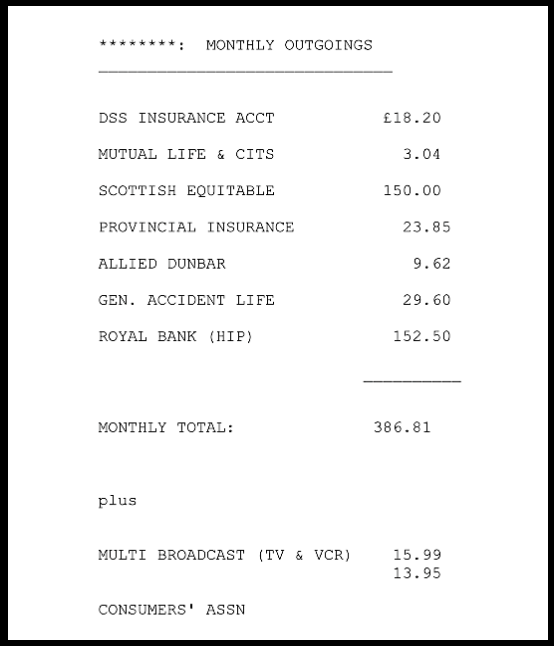
MONTHLY TOTAL:
386.81

plus

MULTI BROADCAST (TV & VCR) 15.99
13.95
CONSUMERS' ASSN

opened with notepad

opened with other software



******** : MONTHLY OUTGOINGS

DSS INSURANCE ACCT	£18.20
MUTUAL LIFE & CITS	3.04
SCOTTISH EQUITABLE	150.00
PROVINCIAL INSURANCE	23.85
ALLIED DUNBAR	9.62
GEN. ACCIDENT LIFE	29.60
ROYAL BANK (HIP)	152.50
MONTHLY TOTAL:	386.81

plus

MULTI BROADCAST (TV & VCR)	15.99
	13.95
CONSUMERS' ASSN	

? Meaning

NATIONAL LIBRARY OF AUSTRALIA

Thus, we have to collect digital ‘stuff’ , we have lots of it, and we need to maintain access to the majority of it for extended periods of time.

Simple, hey?

NLA Image

Google Images

NATIONAL LIBRARY OF AUSTRALIA

The Digital Preservation Business

At the NLA in relation to digital ‘stuff’, we have to:

- Understand what we have;
- Understand what we want to do with it;
- Understand our current level of support and anticipate the impact of change;
- Plan and take appropriate actions on a scale appropriate with the size of the target.

NATIONAL LIBRARY OF AUSTRALIA

Understanding our collections

For all digital objects in the collection we ultimately want to be able to get ‘live’, reliable information from collection metadata about:

- format types and versions,
- file numbers
- file size
- file functions
- business owner
- and other digital object metadata

NATIONAL LIBRARY OF AUSTRALIA

Understanding the preservation intent for all digital ‘stuff’

The NLA needs to be able to group all files in the repository into a preservation profile so that we have an agreed understanding of what we wish to maintain over time. For example:

- What group do these files belong to (could be grouped by genre, format or function etc.)?
- Do we want to preserve them? If so, for how long?
- In what way?

NLA Image

NATIONAL LIBRARY OF AUSTRALIA

Manage the level of support that we wish to maintain

In order to understand the current level of support for formats and more particularly content, the NLA needs to know:

- What software do we have that can access this content?
- Are there any restrictions?
- Based on this how long are we likely to be able to adequately access the file formats .

NATIONAL LIBRARY OF AUSTRALIA

Plan and take action

Thus, we require certain consistent information to be able to preserve a digital object. Based on this information we can then:

- Work out the best preservation action to maintain the preservation intent.
- Plan and test appropriate actions (this could be simple or complex)
- Schedule actions and carry them out
- Repeat as needed

NATIONAL LIBRARY OF AUSTRALIA

We understand what we have to do …

http://www.starstore.com/acatalog/ROTK_Mount_Doom_poster_L.jpg

… doing it is another story

NATIONAL LIBRARY OF AUSTRALIA

Current NLA Strategies

Current digital preservation strategies include:

- making information available about the level of support that the NLA can provide for digital materials;
- helping to process digital content from physical carriers to managed storage for bit level preservation;
- working towards building a holistic understanding of the NLA collection;
- Building knowledgebases about formats, software and dependencies;
- Influencing IT infrastructure:
  - ie. DLIR (Digital Library Infrastructure Redevelopment project)

NATIONAL LIBRARY OF AUSTRALIA

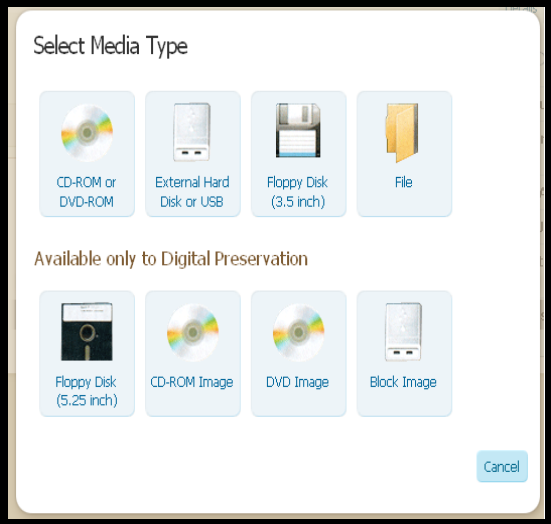
Helping to process digital content from physical carriers to managed storage for bit level preservation – i.e. Prometheus

The NLA built a application called Prometheus to transfer digital content from common media carriers into managed storage system in a systematic manner.

The acquisition of digital materials on carriers is a constantly growing problem. Therefore, the NLA has to address:

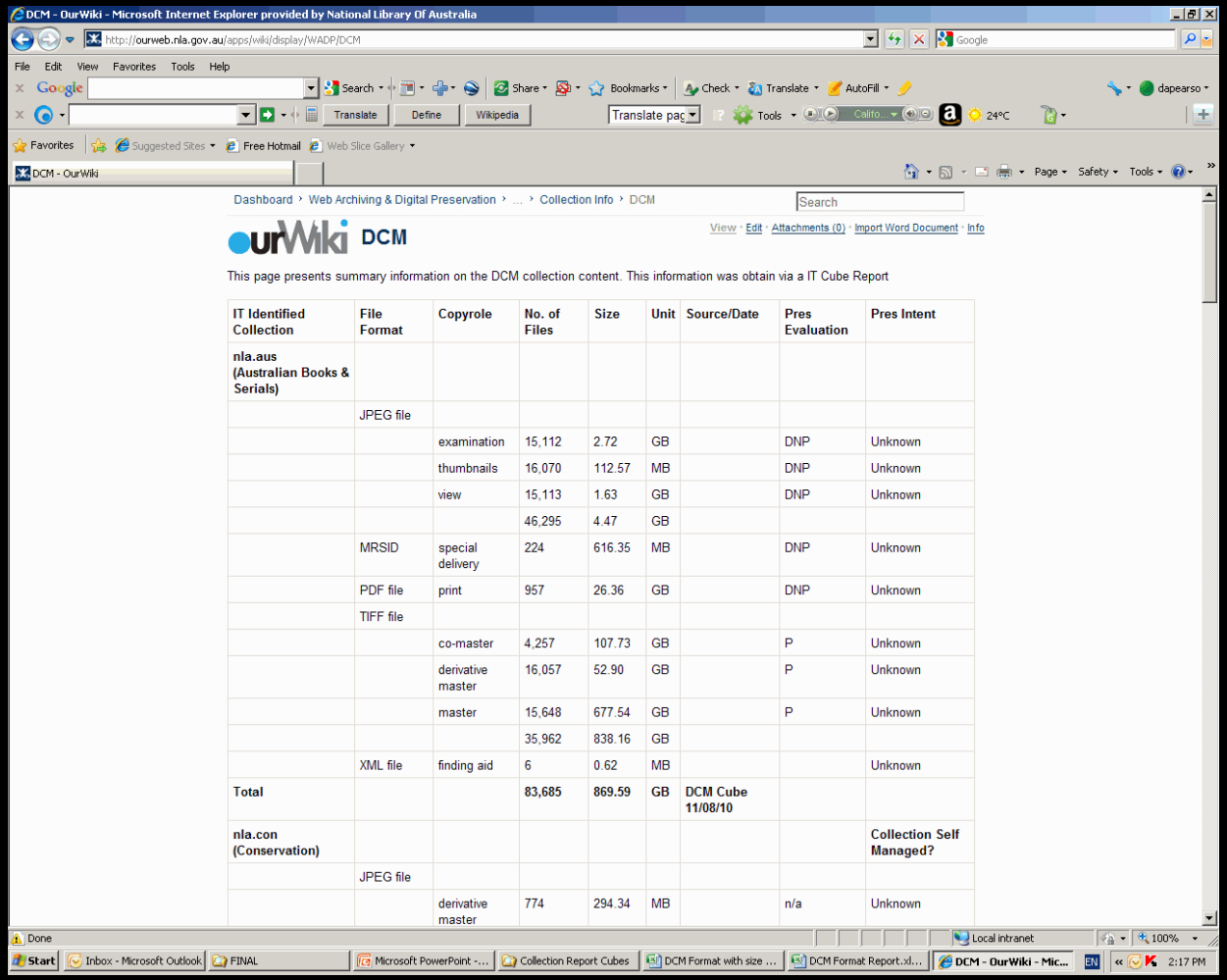
- Processing material that we already have
- Processing new materials as they come in

NATIONAL LIBRARY OF AUSTRALIA



NATIONAL LIBRARY OF AUSTRALIA

Working towards building a holistic understanding of the NLA collections

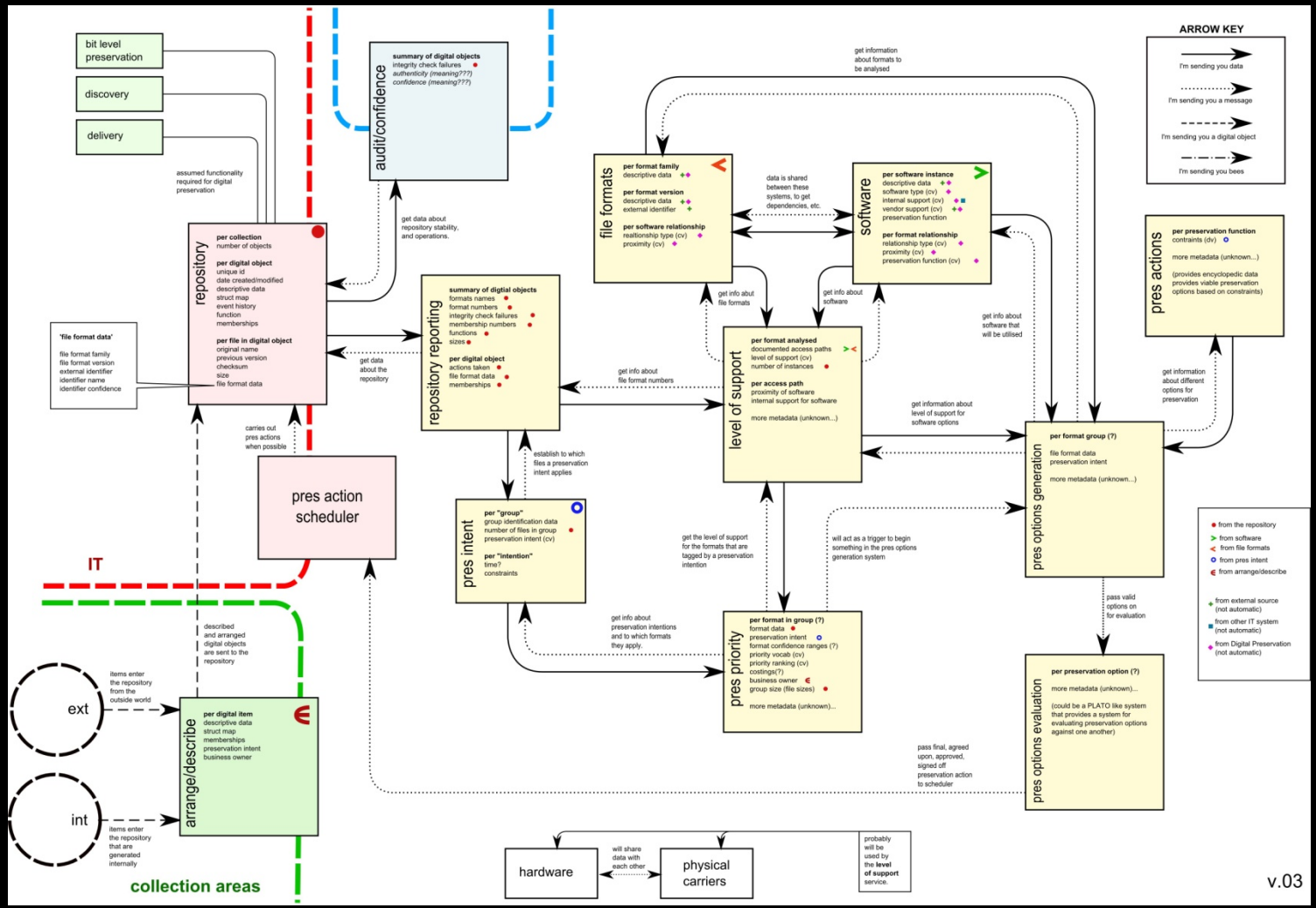


IT Identified Collection	File Format	Copyrole	No. of Files	Size	Unit	Source/Date	Pres Evaluation	Pres Intent
nla.aus (Australian Books & Serials)								
	JPEG file							
		examination	15,112	2.72	GB		DNP	Unknown
		thumbnails	16,070	112.57	MB		DNP	Unknown
		view	15,113	1.63	GB		DNP	Unknown
			46,295	4.47	GB			
	MRSID	special delivery	224	616.35	MB		DNP	Unknown
	PDF file	print	957	26.36	GB		DNP	Unknown
	TIFF file							
		co-master	4,257	107.73	GB		P	Unknown
		derivative master	16,057	52.90	GB		P	Unknown
		master	15,648	677.54	GB		P	Unknown
			35,962	838.16	GB			
	XML file	finding aid	6	0.62	MB			Unknown
Total			**83,685**	**869.59**	**GB**	**DCM Cube 11/08/10**		
nla.con (Conservation)								**Collection Self Managed?**
	JPEG file							
		derivative master	774	294.34	MB		n/a	Unknown

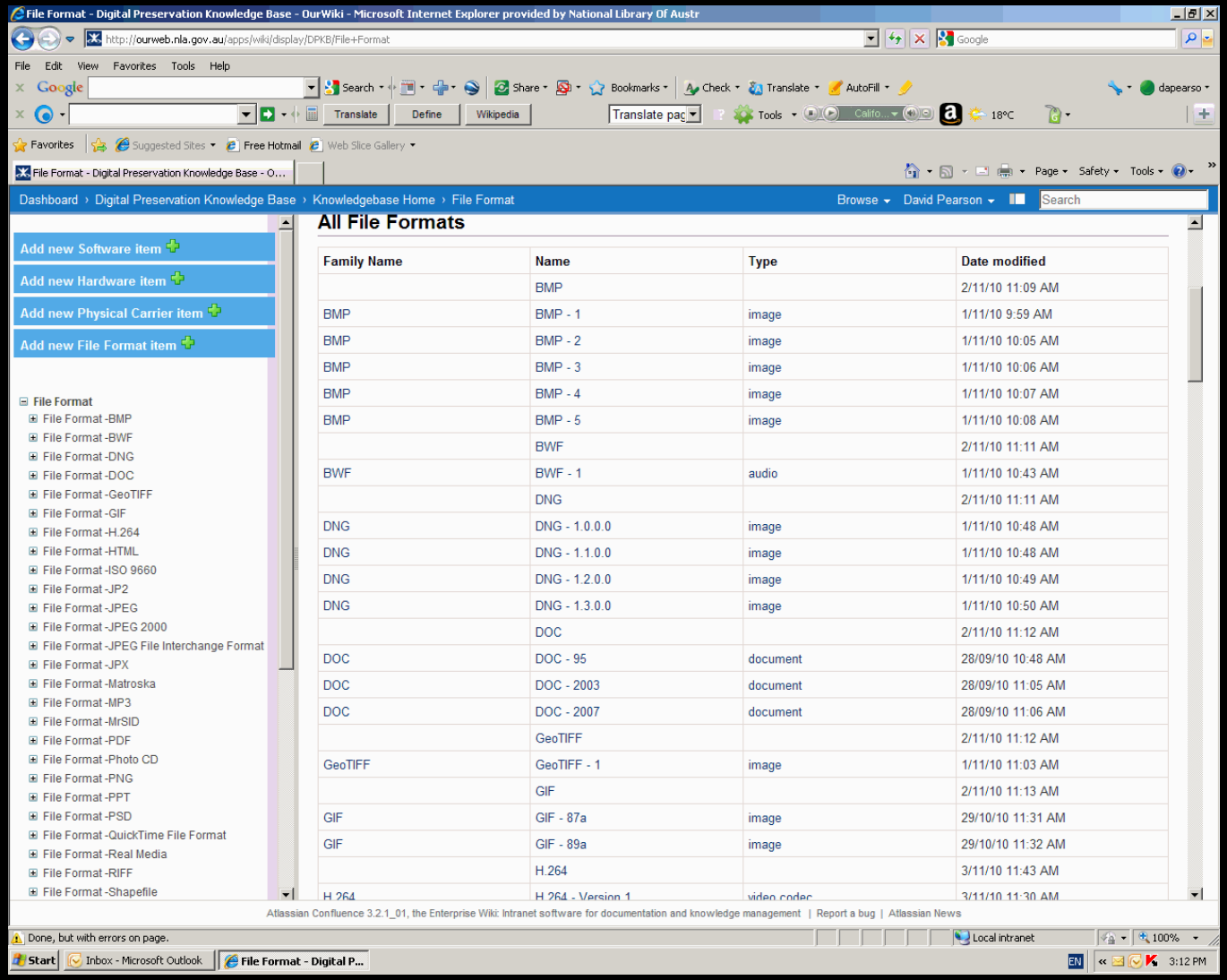
This report for DCM gives us most of this data. But not all NLA systems can provide this data.

NATIONAL LIBRARY OF AUSTRALIA

Build knowledgebases

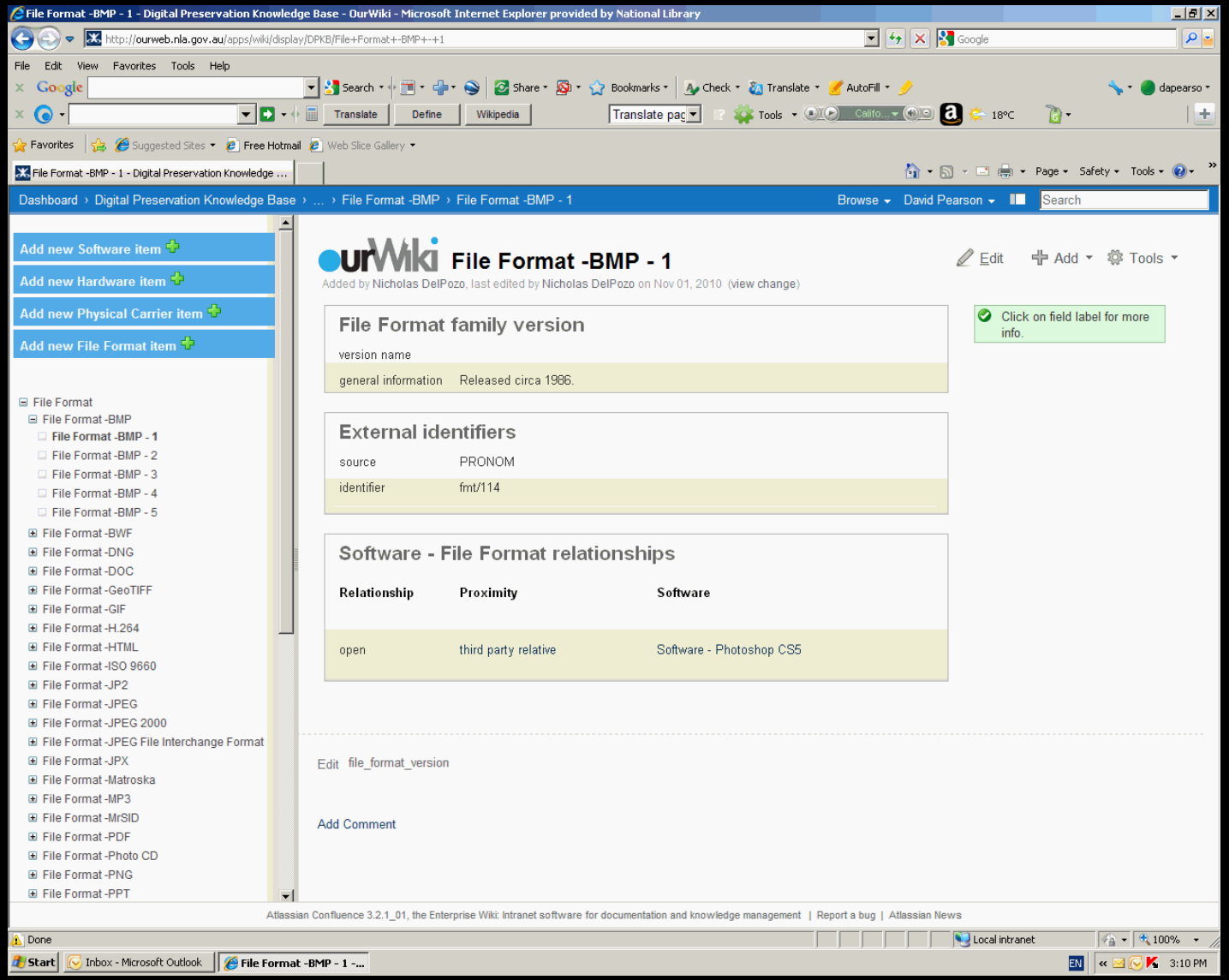


NATIONAL LIBRARY OF AUSTRALIA

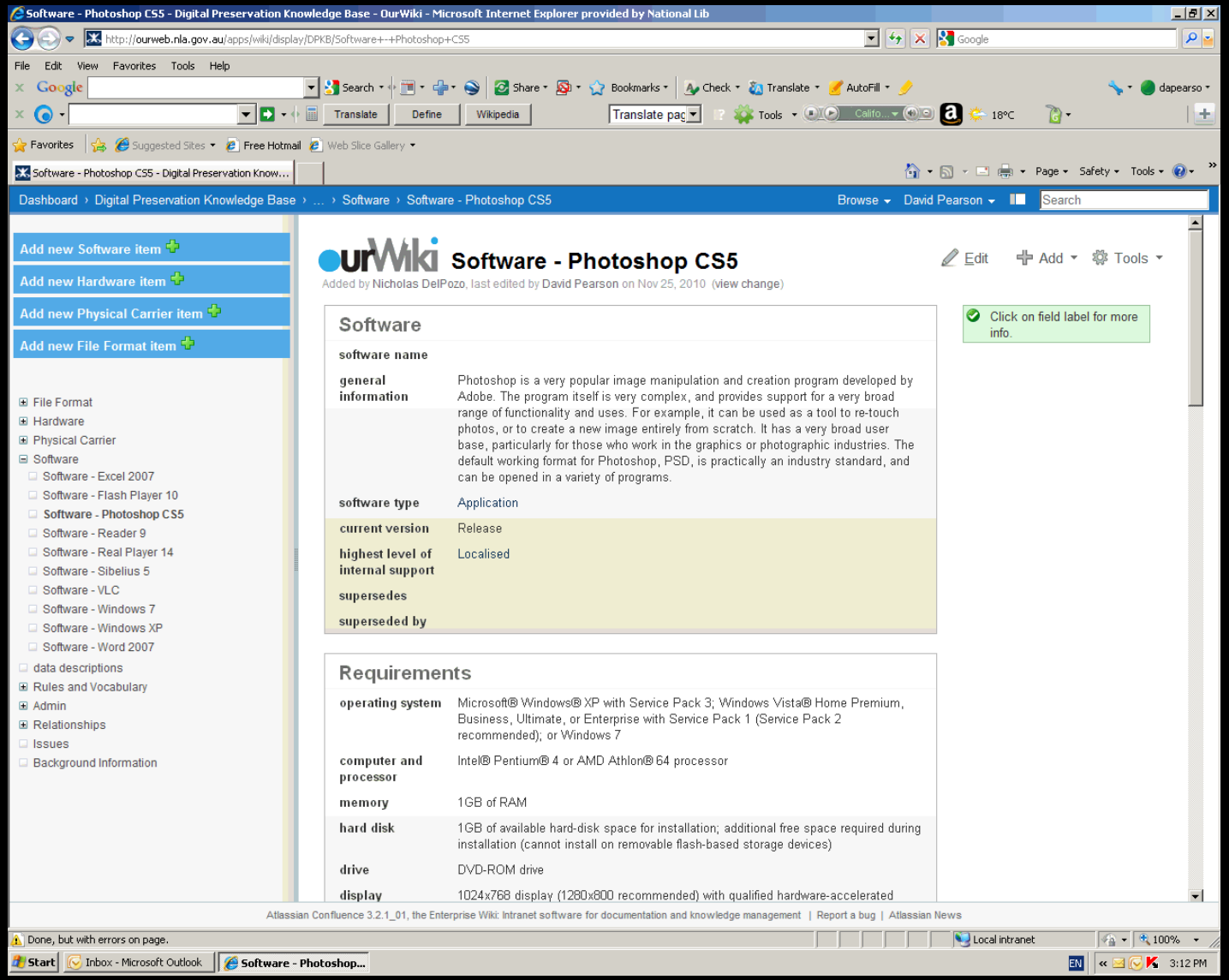


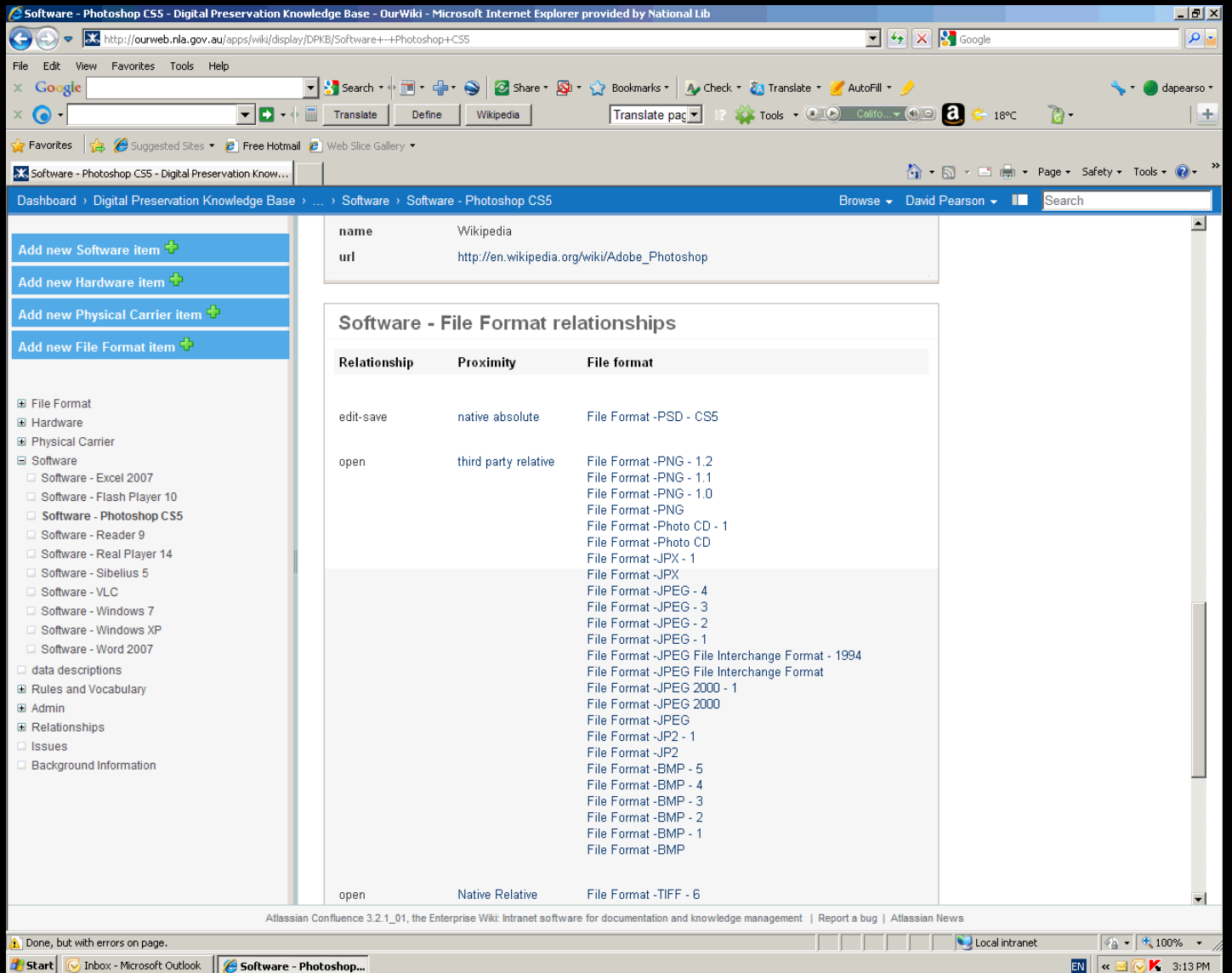
NATIONAL LIBRARY OF AUSTRALIA

NATIONAL LIBRARY OF AUSTRALIA

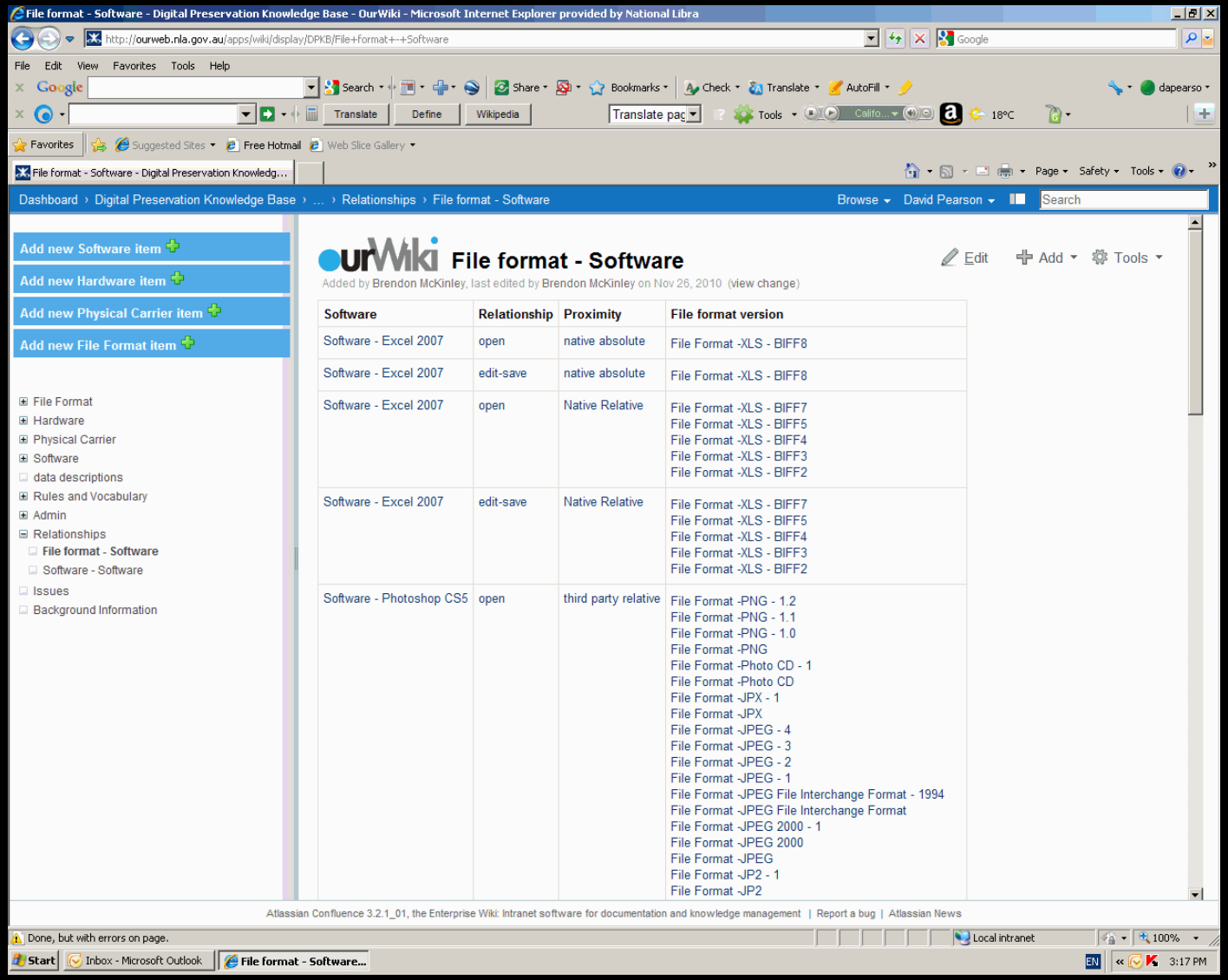


NATIONAL LIBRARY OF AUSTRALIA



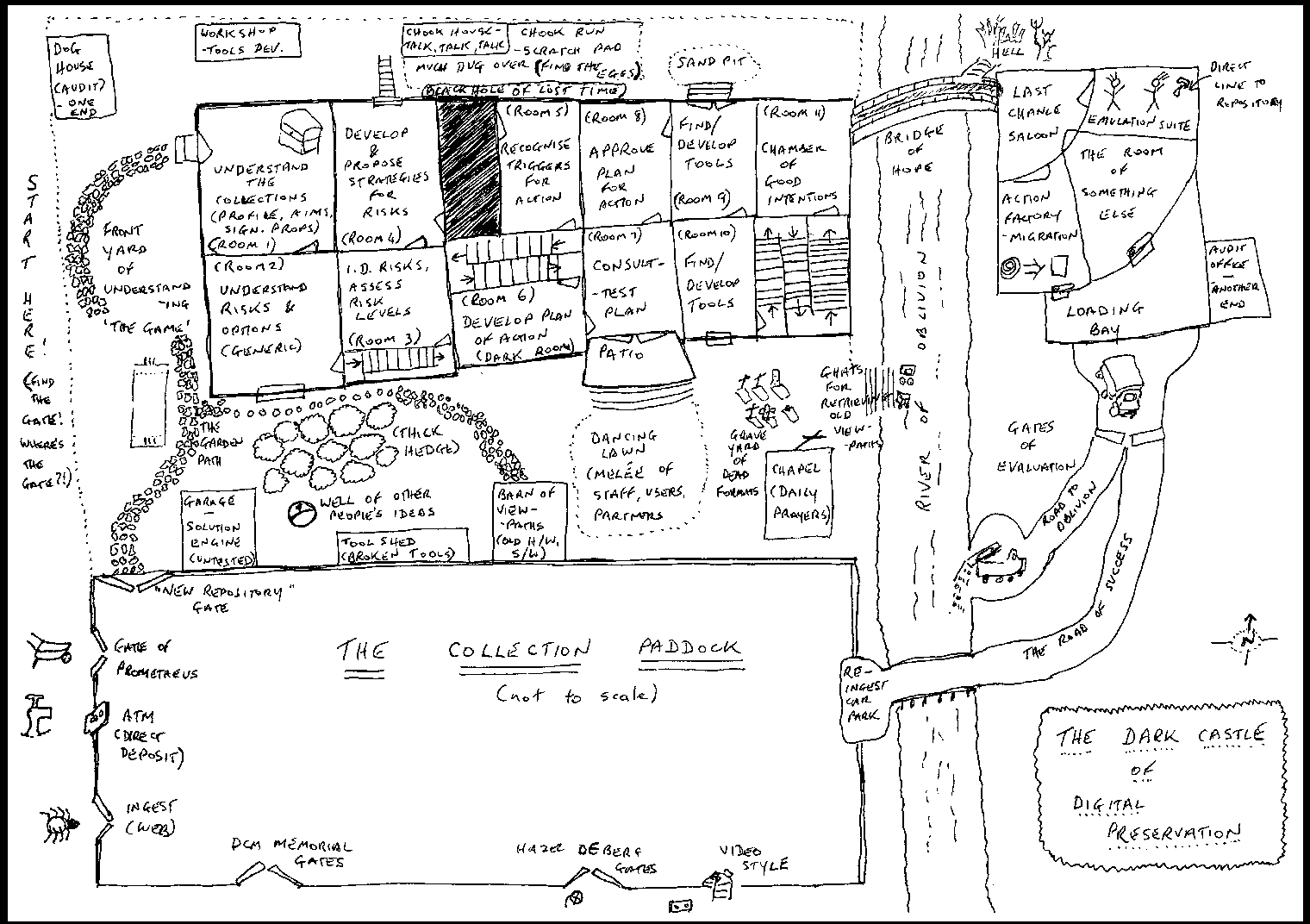


NATIONAL LIBRARY OF AUSTRALIA



NATIONAL LIBRARY OF AUSTRALIA

Influencing IT infrastructure development



Drawing by Colin Webb 2008 NLA

In conclusion:

- The NLA is responsible for a lot of digital ‘stuff’ and we need to actively understand what it is and what we want to do with it in the future;
- If we simply collect and store it, it will become unusable in a relatively short time as technologies change;
- Maintaining the ability to access it requires a lot of good management, planning, & dedicated resources; and
- We have to find and use solutions that can be applied automatically and reliably to billions of digital files.

NATIONAL LIBRARY OF AUSTRALIA

Thank You

Everything, for Everyone Forever

NATIONAL LIBRARY OF AUSTRALIA