the national archives
www.nationalarchives.gov.uk

the national archives

Active Preservation at The National Archives:

PRONOM and DROID

Adrian Brown

Head of Digital Preservation

The National Archives

the national archives

- **The problem**
- Our solution
- Future priorities

the national archives

Objectives:

- Describe technologies required to access digital objects
- Understand risks arising from these technologies
- Support automated preservation processes
- Support internal and external use
- Be effective and practical

the national archives

- The problem
- **Our solution**
- Future priorities

the national archives

PRONOM services

- Technical Registry content
- Web services interfaces
- Tools
  - e.g. DROID

the national archives

The technical registry
PRONOM

Welcome : About Add an entry
Search Help Information resources

Help : detailed report on file format

Details: File format summary

File format | PRONOM Unique Identifier | Software | Vendor | Lifecycles

Details for: Portable Document Format 1.6

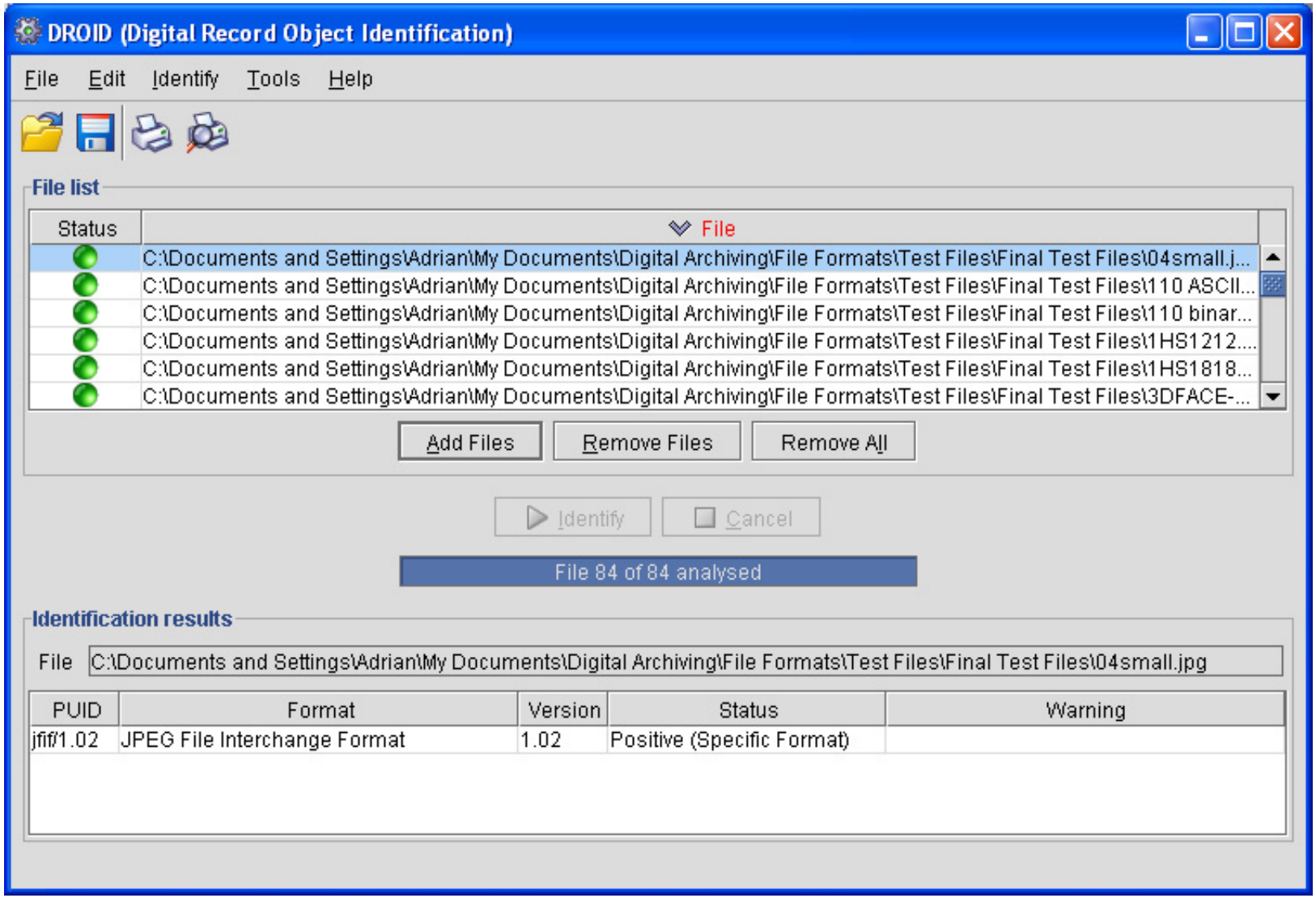
Save as... XML | CSV Print

Go to: Summary | Documentation > | Signatures > | Compression > | Character encoding > | Rights > | Reference files >

Summary

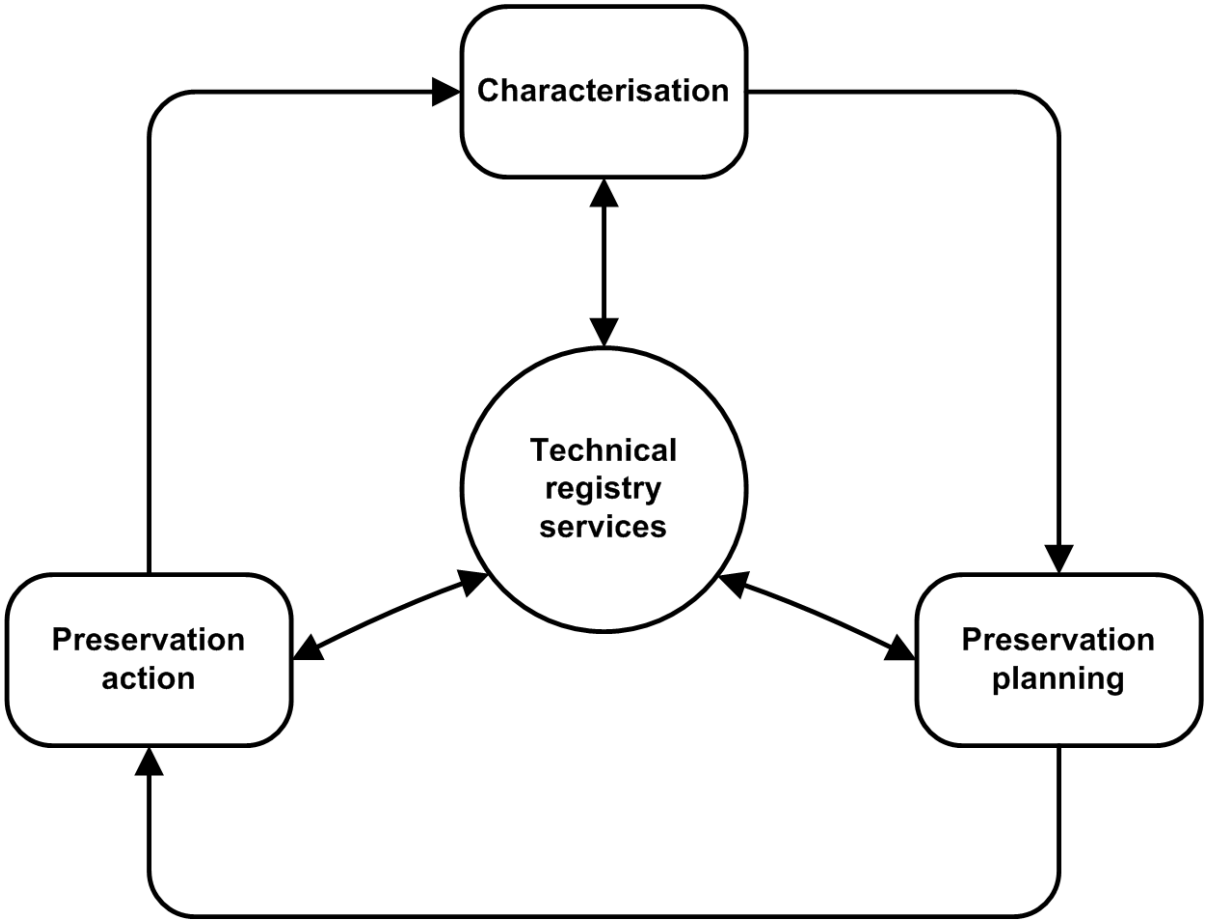
Name	Portable Document Format
Version	1.6
Other names	PDF (1.6)
Identifiers	PUID: fmt/20 MIME: application/pdf
Family	
Classification	Page Description
Disclosure	Full
Description	Portable Document Format is a platform-independent format for representing formatted documents, developed by Adobe Systems Incorporated. It is the native format of Adobe?s Acrobat family of software products, version 1.6 corresponding to the release of Acrobat 7.0. PDF is based on, and shares the same imaging model as, the PostScript page description language. A PDF file comprises a Header section, a Body section containing the objects which make up the document, a Cross Reference Table, and a Trailer section. PDF files can contain a wide variety of content, including text, images, video and audio.
Orientation	Binary
Byte order	Big-endian (Motorola)

the national archives



the national archives

Active preservation



Preserv Profile (ROAR) - Microsoft Internet Explorer provided by National Archives

http://roar.eprints.org/index.php?action=profile&url=http%3A%2F%2Fgtcni.openrepository.com%2Fgtcni%2F

eprints roar

Preserv Profile (ROAR)

Page Tools

Preserv Profile

Summary Preserv Profile

For more information about the Preserv project and repository content profiling see http://preserv.eprints.org/. Pronom's file format identification is currently under development, which may result in some strange format identifications.

The following chart shows the breakdown of all **files** contained in the repository(ies), based on an automated crawler. All files (apart from the abstract 'jump-off' page) were downloaded and then identified using the **Pronom** file format identification tool (Droid). Does not include files that are not hosted on the same server as the OAI interface or files that are over 2 megabytes in size.

You can subscribe to alerts for this profile by going to Celestial's repository page.

OAI Repositories Included (1)

http://gtcni.openrepository.com/gtcni-oai/request

Format	Total Files
Portable Document Format (1.4)	147
Portable Document Format (1.3)	115
Portable Document Format (1.5)	34
Portable Document Format (1.6)	13
Portable Document Format (1.2)	12
Microsoft Word for Windows Document (97-2003)	7
Portable Document Format - Archival (1)	4
OLE2 Compound Document Format	2
[No files found]	1
Binary Interchange File Format (BIFF) Workbook (8)	1

Done Internet 100%

Active Preservation Management Console

You are not logged in

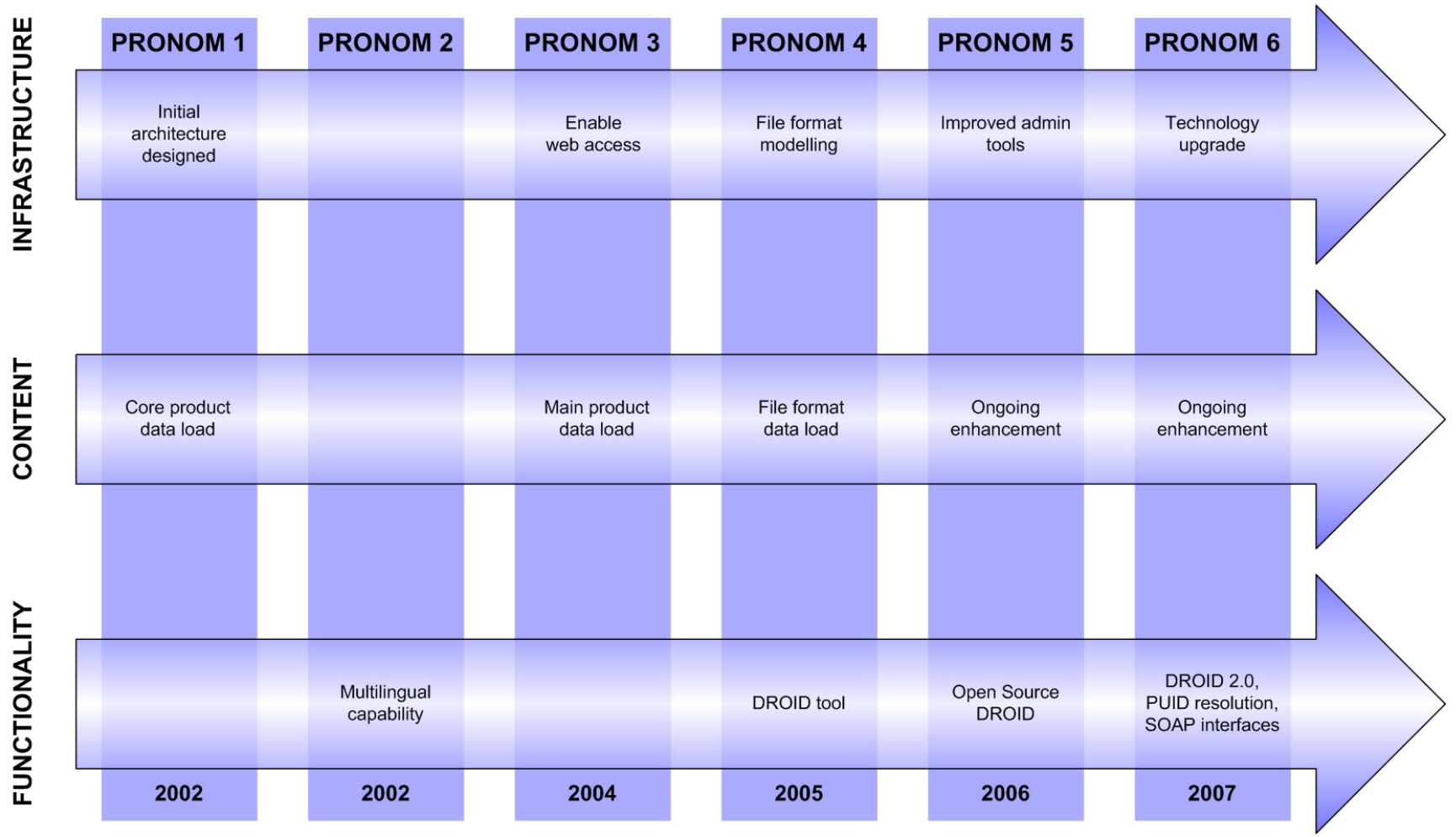
Username [ ] login

Password [ ]

Command	CHR3	**XMLOutputFile**	/home/carj/tmp/F/chr-output.xml
Machine	potent.tessella.co.uk	**XMLInputFile**	/home/carj/droid-input.xml
Date Received	2006-Dec-15 04:50:12		
Working Area	/home/carj/tmp/F/		
Date Completed			
Client Web Service	http://localhost:8080/jobQueueManager/services/JobCompleteService		

Name	Tool Id	Date Started	Date Completed	Priority	% Completed	Status	Action	Message
Characterise File Set	100	2006-Dec-15 04:51:24		Normal Priority	0	Running		Characterise File Set started...
DROID	100	2006-Dec-15 04:51:27	2006-Dec-15 04:51:52	Normal Priority	100	Completed		Tool Completed Normally
Jhove-Jpeg2000	100	2006-Dec-15 04:51:53	2006-Dec-15 04:51:53	Normal Priority	100	Completed		Tool Completed Normally
Jhove-Gif	100	2006-Dec-15 04:51:53	2006-Dec-15 04:51:53	Normal Priority	100	Completed		Tool Completed Normally
Jhove-Html	100	2006-Dec-15 04:51:53	2006-Dec-15 04:51:54	Normal Priority	100	Completed		Tool Completed Normally
Jhove-Tiff	100	2006-Dec-15 04:51:54	2006-Dec-15 04:51:54	Normal Priority	100	Completed		Tool Completed Normally
Jhove-Wave	100	2006-Dec-15 04:51:54	2006-Dec-15 04:51:54	Normal Priority	100	Completed		Tool Completed Normally
XML validatior	100	2006-Dec-15 04:51:54	2006-Dec-15 04:51:55	Normal Priority	100	Completed		Tool Completed Normally
Jhove-Aiff	100	2006-Dec-15 04:51:54	2006-Dec-15 04:51:55	Normal Priority	100	Completed		Tool Completed Normally
Jhove-Pdf	100	2006-Dec-15 04:51:55		Normal Priority	50	Running		ZA544606.800-UK-A.pdf (8/16)
Jhove-Jpeg	100	2006-Dec-15 04:51:57	2006-Dec-15 04:51:57	Normal Priority	100	Completed		

the national archives



PRONOM Development Roadmap

the national archives

Working with others

- Planets
- Global Digital Format Registry
- Digital Curation Centre
- Library of Congress
- Commercial sector (e.g. Microsoft)
- Potential adopters:
  - JHOVE 2
  - Dspace
  - AONS II

the national archives

- The problem
- The solution
- **Future priorities**

the national archives

Future priorities

- More content?
- More tools?
- Enhanced tools?
- New services?
- More integration?

the national archives
www.nationalarchives.gov.uk