



# The National Archives

[nationalarchives.gov.uk](http://nationalarchives.gov.uk)



The National Archives

## **TNA Digital Preservation Services**

**Adrian Brown**  
The National Archives



- PRONOM services
- Characterisation services



- **PRONOM services**
- Characterisation services



# 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](#)

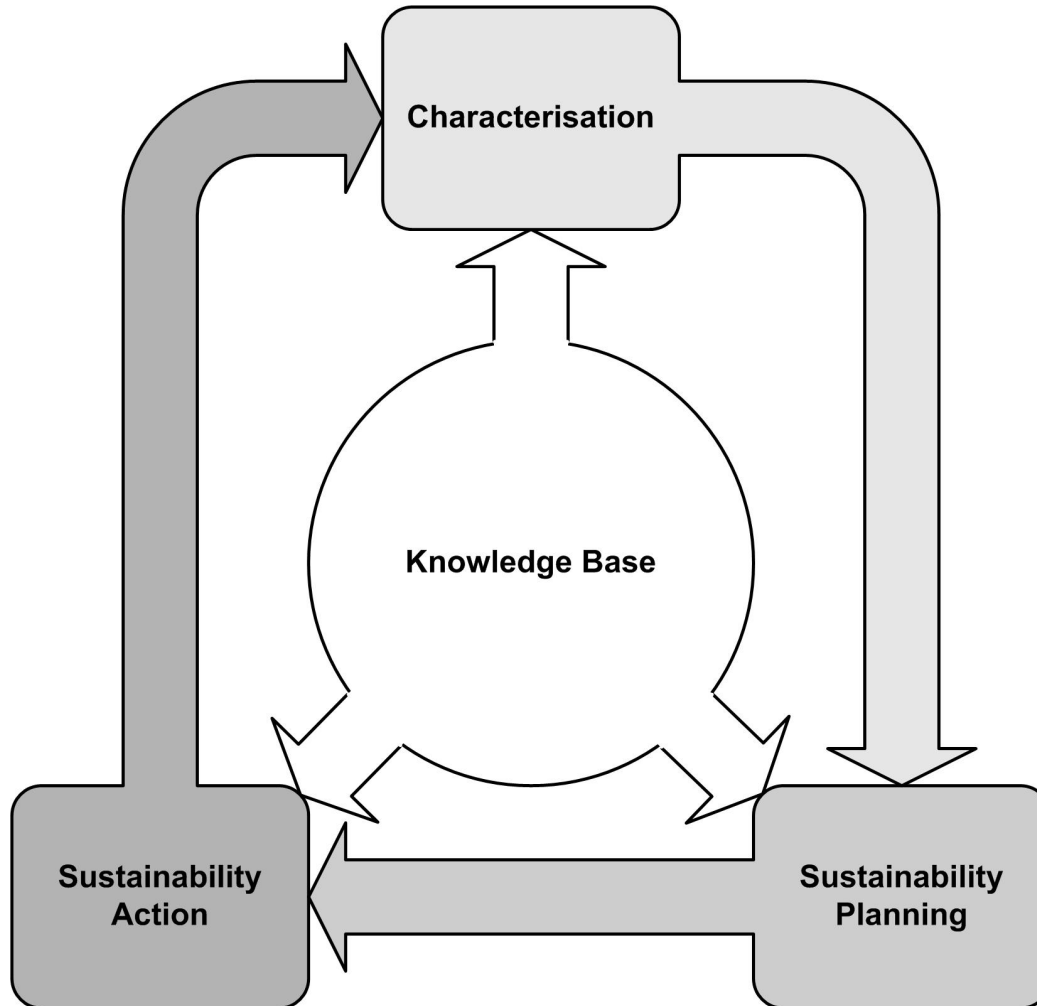
Go to: [Summary](#) | [Documentation](#) > | [Signatures](#) > | [Compression](#) > | [Character encoding](#) > | [Rights](#) > | [Reference files](#) >

## Summary

<b>Name</b>	Portable Document Format
<b>Version</b>	1.6
<b>Other names</b>	PDF (1.6)
<b>Identifiers</b>	PUID: fmt/20 MIME: application/pdf
<b>Family</b>	
<b>Classification</b>	Page Description
<b>Disclosure</b>	Full
<b>Description</b>	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.
<b>Orientation</b>	Binary
<b>Byte order</b>	Big-endian (Motorola)



# Active preservation



# Characterisation

- Identification
  - *DROID*
- Validation
  - *Identification of appropriate validation tools*
- Property extraction
  - *Identification of appropriate extraction tools*





# Preservation planning

- Risk assessment
  - Calculating formats at risk
- Technology watch
  - Monitoring technological change
- Preservation action plans
  - Identifying potential preservation actions (e.g. migration pathways)





# Risk assessment of technologies

- Factors:
  - Adoption
  - Support
  - Disclosure
  - Stability
  - Quality of documentation
  - Intellectual property rights
  - Interoperability
  - Complexity



## Current development

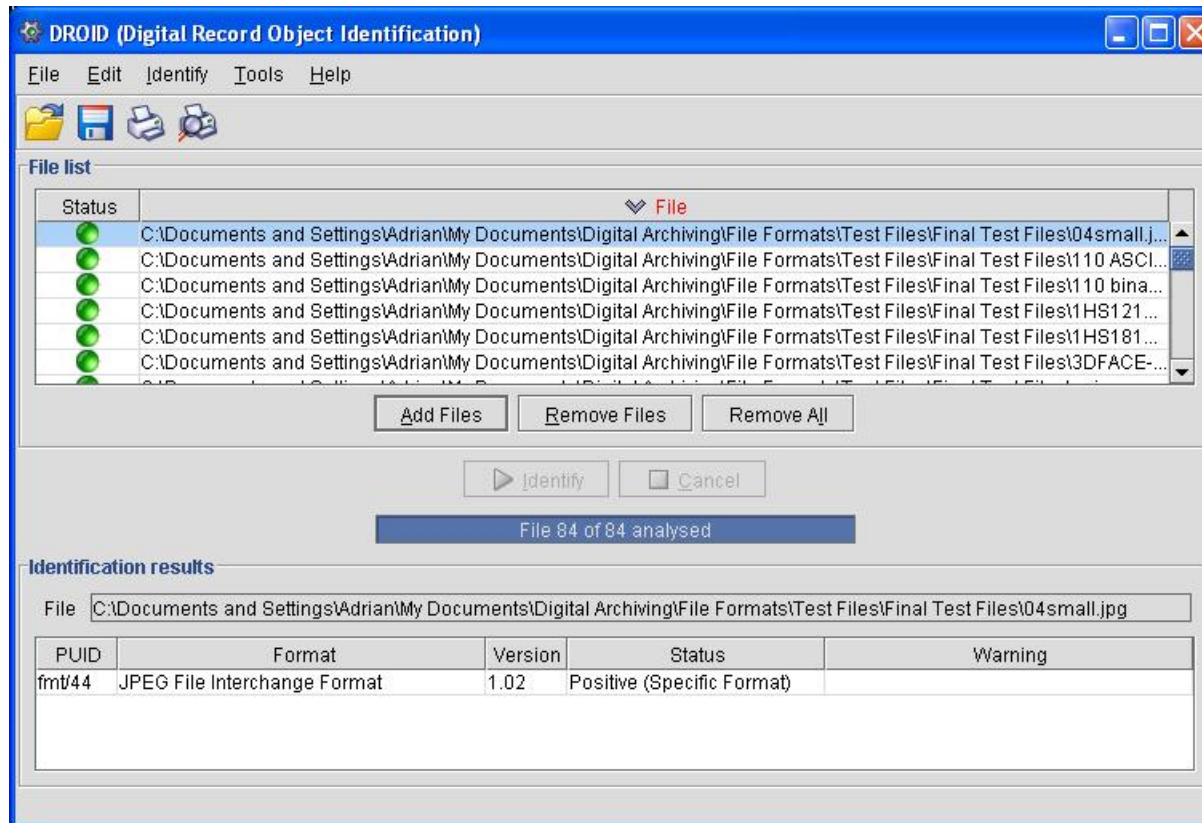
- Planets
  - Characterisation and Preservation Action Registries
  - Development of Java implementation
- Global Digital Format Registry
  - Ongoing discussions to align work



- PRONOM services
- **Characterisation services**



# DROID



<http://droid.sourceforge.net/>



## Collection Profiling

- Format profile
  - *Format distribution by count and volume*
- Age profile
  - *Age distribution by count and volume*
- Volumetric profile
  - *Total volume, smallest, largest and mean file sizes*





# DROID

**Profile**

**Report**

Total readable files: 5163  
Total file size (bytes): 549777970  
Total unreadable folders: 0  
Total unreadable files: 0  
Smallest file's size (bytes): 1  
Largest file's size (bytes): 32034816  
Mean file size (bytes): 106484

**Files by modified date**

Last Modified Year	Number of Files	Total Bytes
2005	58	31062378
2006	133	94554102
2007	4970	424082642
2008	2	78848

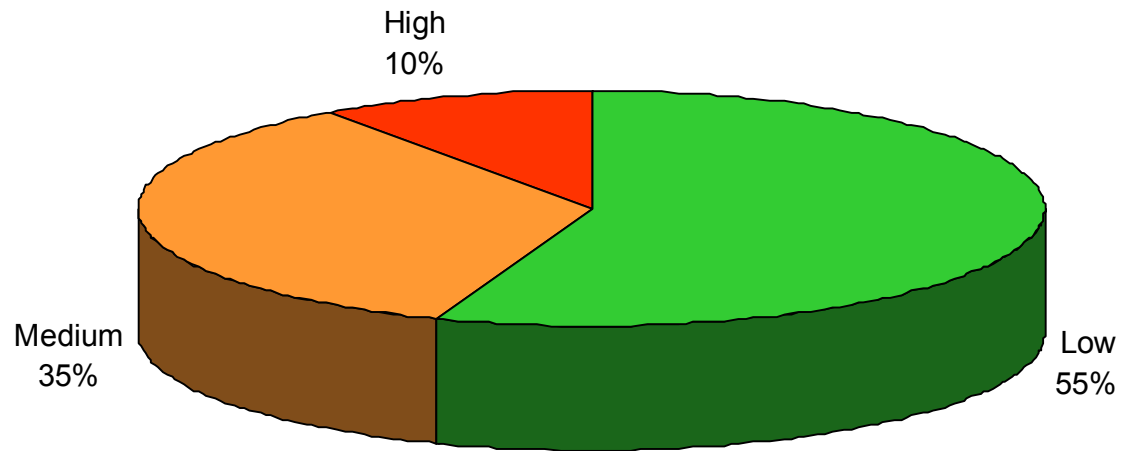
**Files by format**

PUID	MIME	Format	Version	Number of ...	Total Bytes
fnt/59	application...	Binary Inter...	5	2	761856
fnt/99	text/html	Hypertext M...	4.0	8	115046
fnt/136		OpenDocu...	1.0	2	12164
fnt/137		OpenDocu...	1.0	2	13113
fnt/138		OpenDocu...	1.0	2	19444
fnt/139		OpenDocu...	1.0	2	18754
fnt/17	application...	Portable D...	1.3	1	276166
fnt/111		OLE2 Com...		34	6510080
fnt/18	application...	Portable D...	1.4	1	1506613
fnt/11	image/png	Portable N...	1.0	2	151291
fnt/40	application...	Microsoft ...	97-2003	101	104471552

Export to   Stop



### Risk Profile by File Count





## Current development

- Functionality
  - Enhanced reporting
  - GUI improvements
- Performance
  - Scalable to 100 million files
  - Improved error logging
  - Checkpointing and performance throttling





# Preserv Profile

[Home](#) [Browse](#) [Search](#) [Content Search](#) [Register a Repository](#) [Help](#) [Login](#) [Register in ROARMAP](#)

Please sign the [petition](#) in support of the European Commission's proposed [Open Access Self-Archiving Mandate](#)

## 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



The National Archives

---

**Questions?**



**The National Archives**

[nationalarchives.gov.uk](http://nationalarchives.gov.uk)