# Characterisation in Planets

Adrian Brown

The National Archives, UK

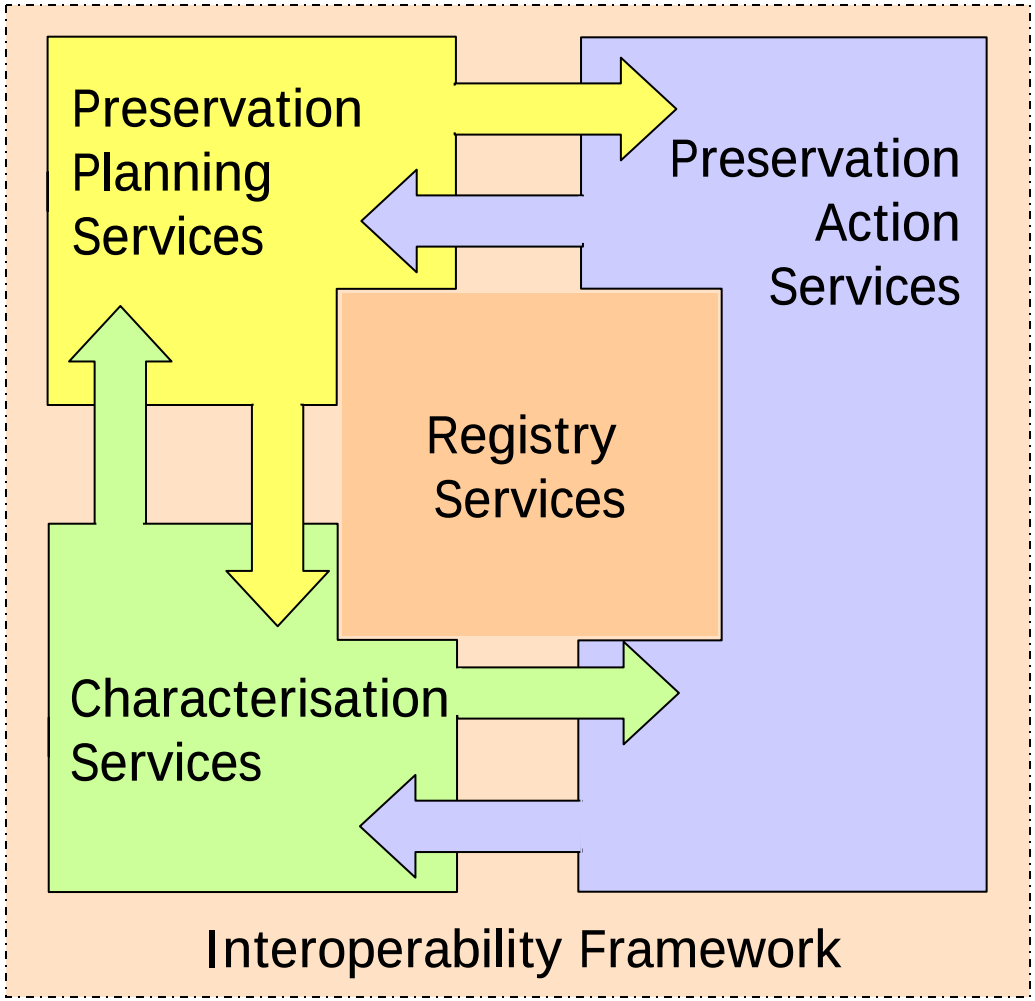
# Planets overview

- 4 year project, started in 2006
- Brings together libraries, archives, research institutions and technology providers
- Developing a scaleable digital preservation infrastructure
- Ensuring wide adoption across user communities and establishing a market place for preservation services and tools

www.planets-project.eu

planets

# Planets architecture



# Characterisation sub-project

- To understand the properties of digital objects which are significant to their long-term accessibility
- To develop methods for expressing and measuring these properties to:
    - inform preservation planning
    - validate the results of preservation actions
    - support the needs of user communities
- To promote the development of object types with ‘preservation-friendly’ characteristics

planets

# Areas of work

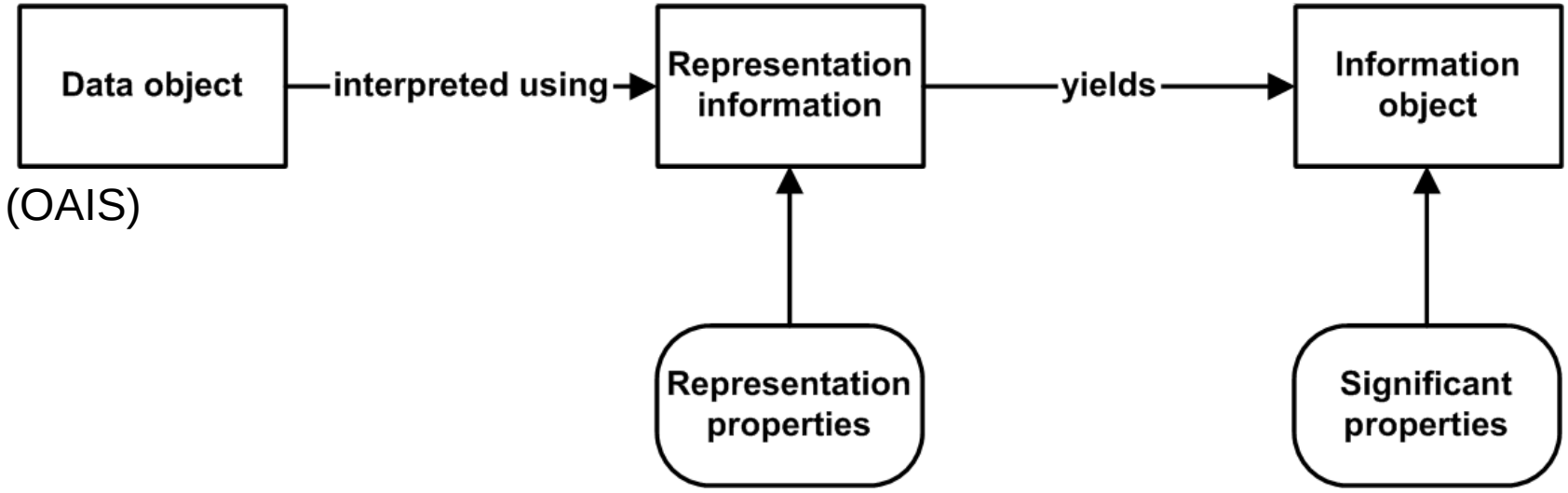
- Characterisation strategies and methodologies
- Characterisation tools
  - Measure properties of objects
- Characterisation registry
  - Describes tools and their capabilities
  - Describes object types and properties

# Characterisation processes

- Identification
  - *The object is in PNG 1.0 format*
- Validation
  - *The object is valid against the PNG 1.0 specification*
- Property extraction
  - *The object uses 24-bit colour and has a resolution of 300 dpi*

# Models


(National Archives of Australia)

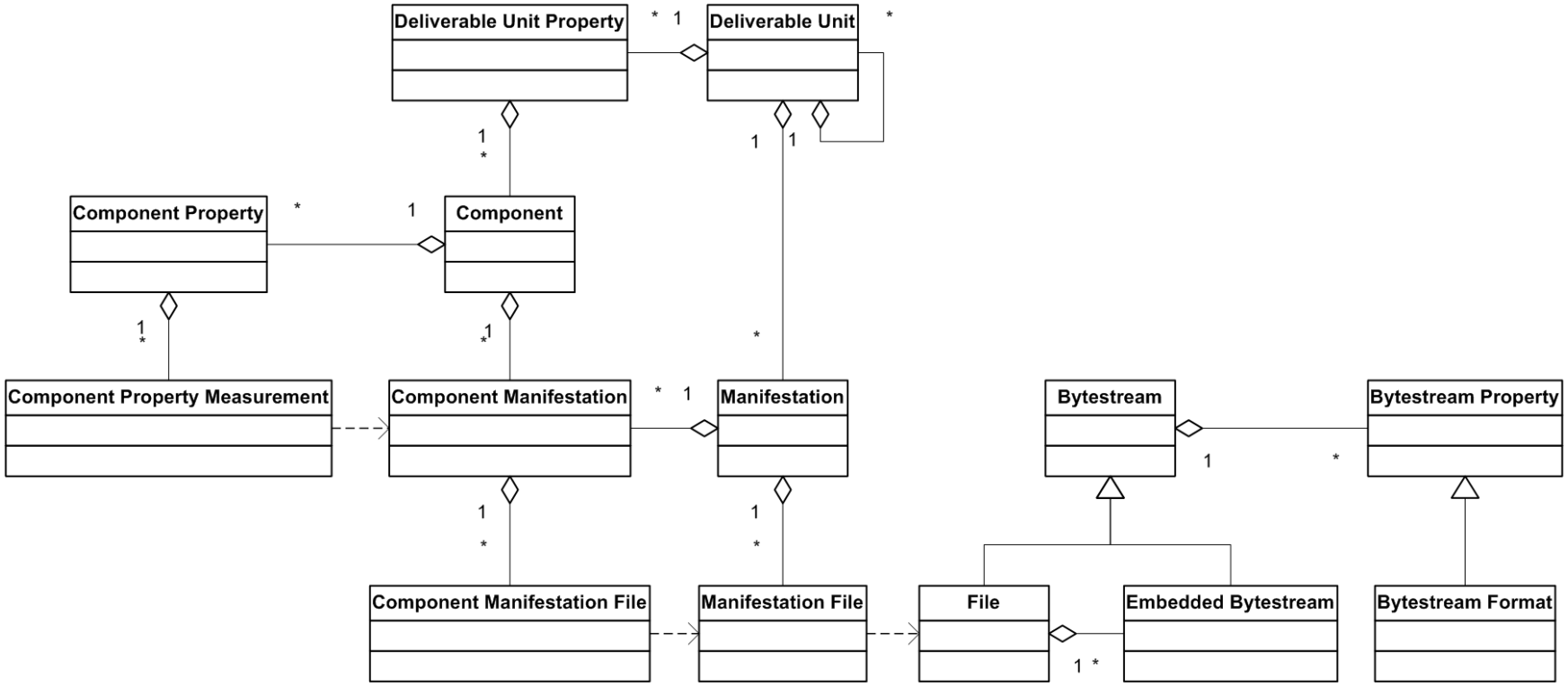



(OAIS)

# Models

- Representation properties:
  - *Describe the data object and its representation network*
- Significant properties:
  - *Describe the information object*

# Planets conceptual model



# Representation properties

- Identified through characterisation
- Described with reference to the characterisation registry
- Use cases:
  - Recursive selection of characterisation tools
  - Populating decision trees for preservation planning using PLATO

# White Paper

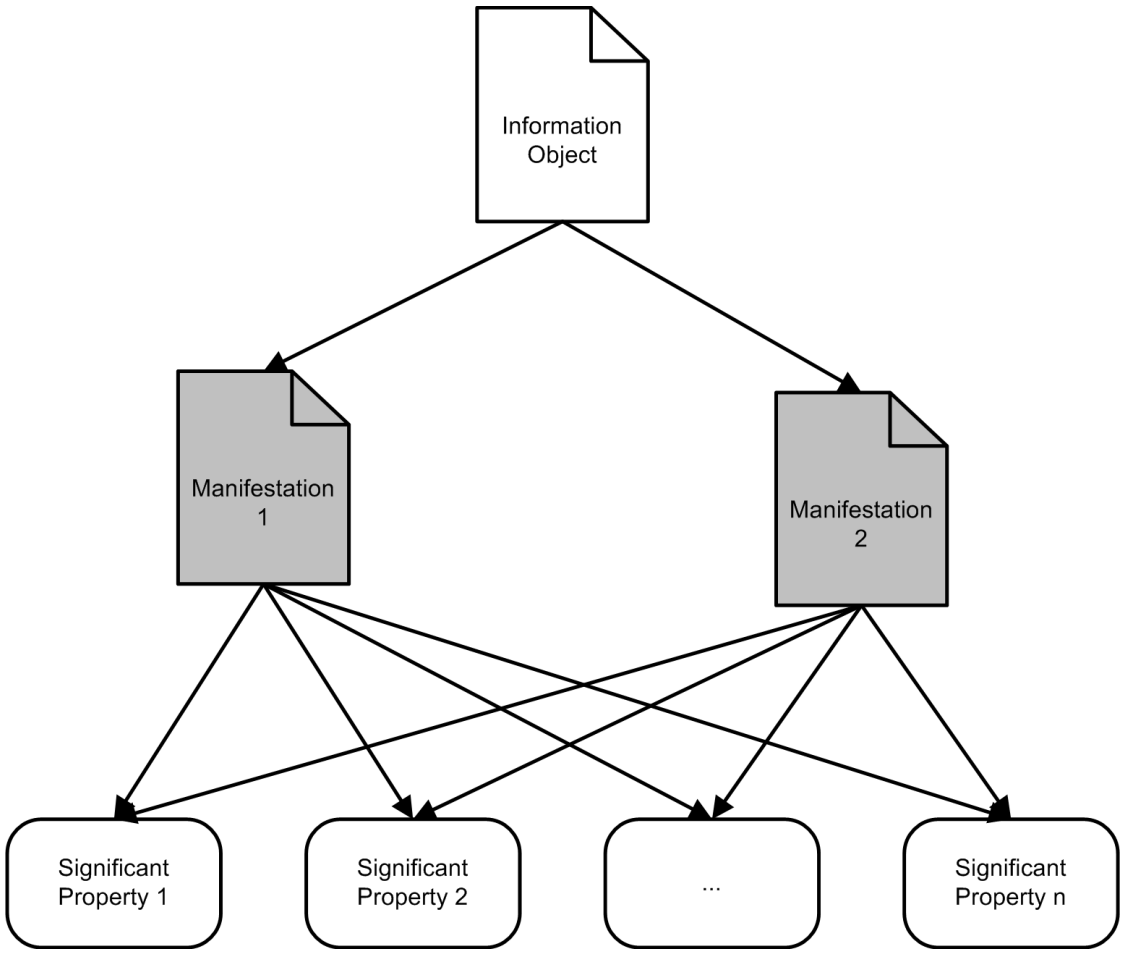
- Discusses concept of Representation Information Registries
- Surveys current state of the art
- Discusses roles of registries in Planets
- Makes recommendations for future research

www.planets-project.eu/docs/reports/Planets_PC3-D7_RepInformationRegistries.pdf

# Significant properties

- Identified through characterisation
- Described with reference to the characterisation registry
- Use cases:
  - Populating decision trees for preservation planning using PLATO
  - Validation of preservation actions

# Validation


Information Object
Manifestation 1
Manifestation 2
Significant Property 1
Significant Property 2
...
Significant Property n

# Categories of significant property

- Content
- Context
- Structure
- Appearance
- Behaviour

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

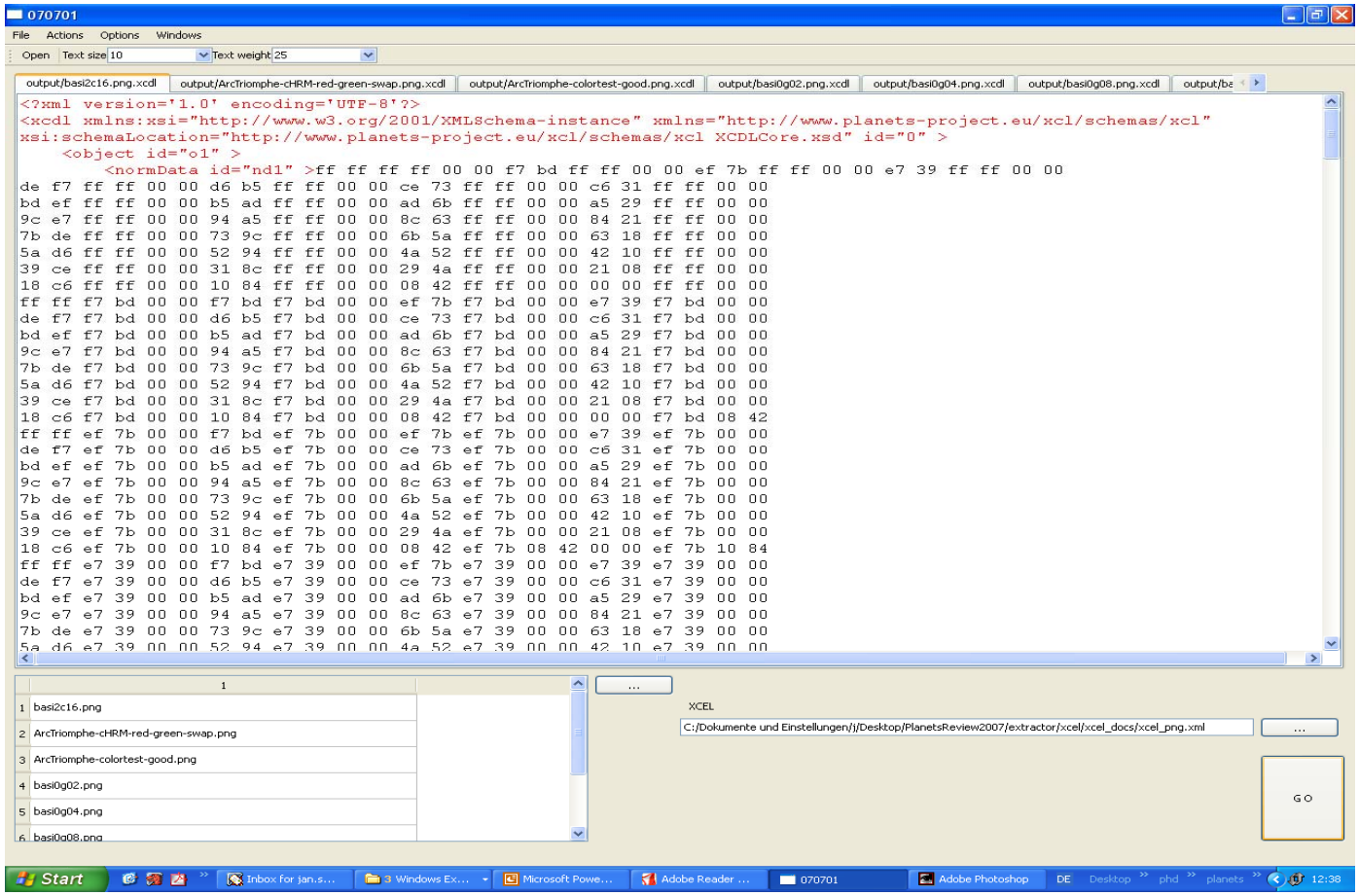
| | |
|---|---|
| **Name** | Portable Document Format |
| **Version** | 1.6 |
| **Other names** | PDF (1.6) |
| **Identifiers** | PUID: fmt/20<br>MIME: application/pdf |
| **Family** | |
| **Classification** | Page Description |
| **Disclosure** | Full |
| **Description** | Portable Document Format is 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) |

www.nationalarchives.gov.uk/pronom

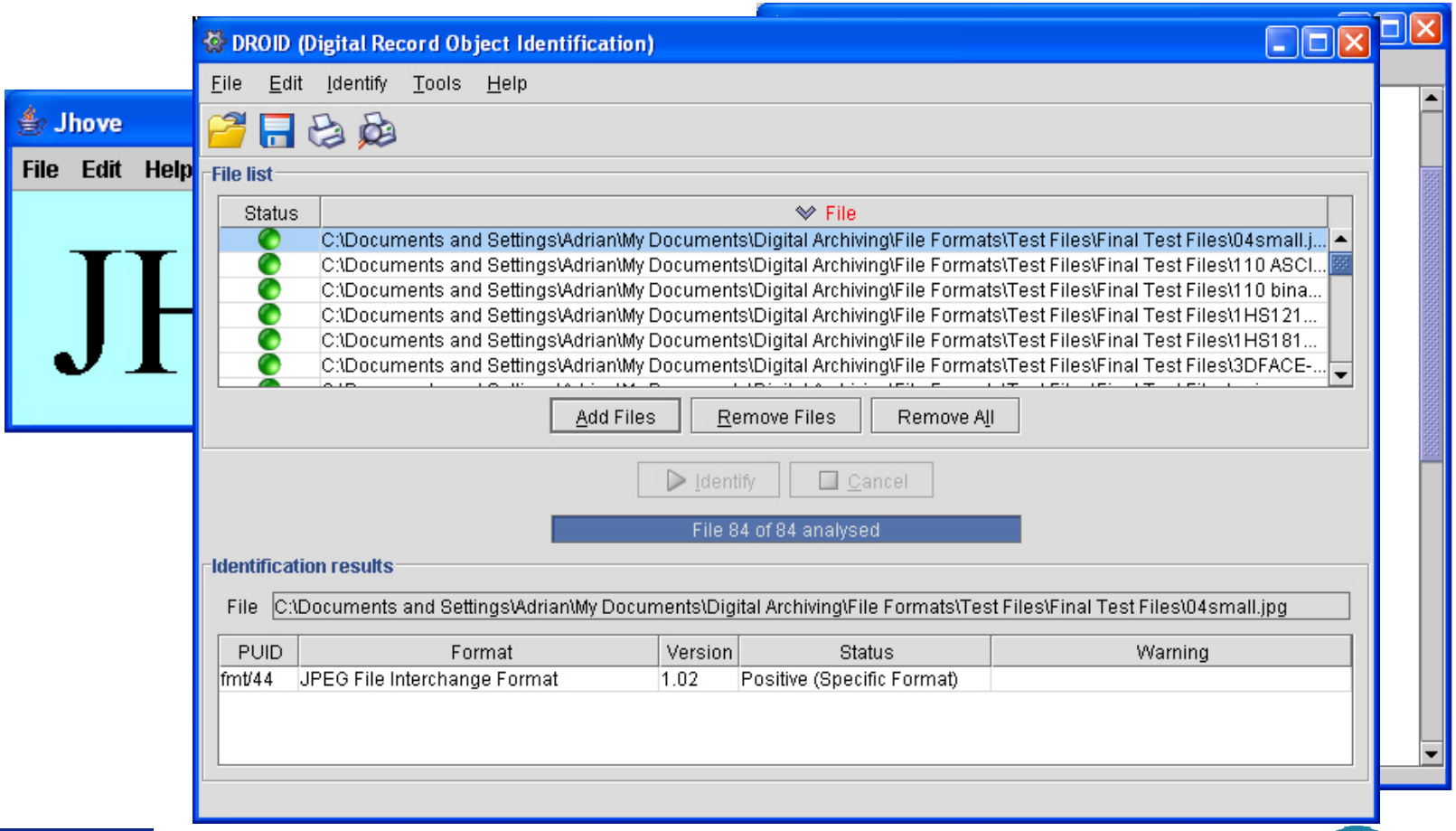
# eXtensible Characterisation Languages

- Description Language (XCDL)
  - Expresses the properties of an object in neutral form
- Extraction Language (XCEL)
  - Expresses how to extract those properties for a given object format
- XCDL extraction tool

# XCDL extractor



# Third-party tools



# Active Preservation Management Console

You are not logged in

Username [ ] login

Password [ ]

| | |
|---|---|
| **Command** | CHR3 |
| **Machine** | potent.tessella.co.uk |
| **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 |

| | |
|---|---|
| **XMLOutputFile** | /home/carj/tmp/F/chr-output.xml |
| **XMLInputFile** | /home/carj/droid-input.xml |

| 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 validation | 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 | | |

# Questions?