# If God had meant us to understand file formats, He would have told us AONS ago

Matthew Walker
AONS Technical Lead
National Library of Australia
mwalker@nla.gov.au

APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

# AONS headquarters ☺

APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

# AONS II: Automatic Obsolescence Notification System - version 2


Project Manager: David Pearson

Technical Lead: Matthew Walker

Software Developer: David Levy


AONS II – latest beta release (2007-10-23) 23 Oct 2007.
Represents seven months' worth of software development.

# Project Objective (APSR Nov 2006)

*Objective:*

To refine the Automatic Obsolescence Notification System (AONS) developed in an earlier stage of APSR, to a platform-independent downloadable tool that automatically provides information from authoritative international registries to support decisions on preservation action required to retain access to information resources stored in repositories.

# Why AONS II?

Format obsolescence is potentially a major problem for every repository manager. This is particularly true given:

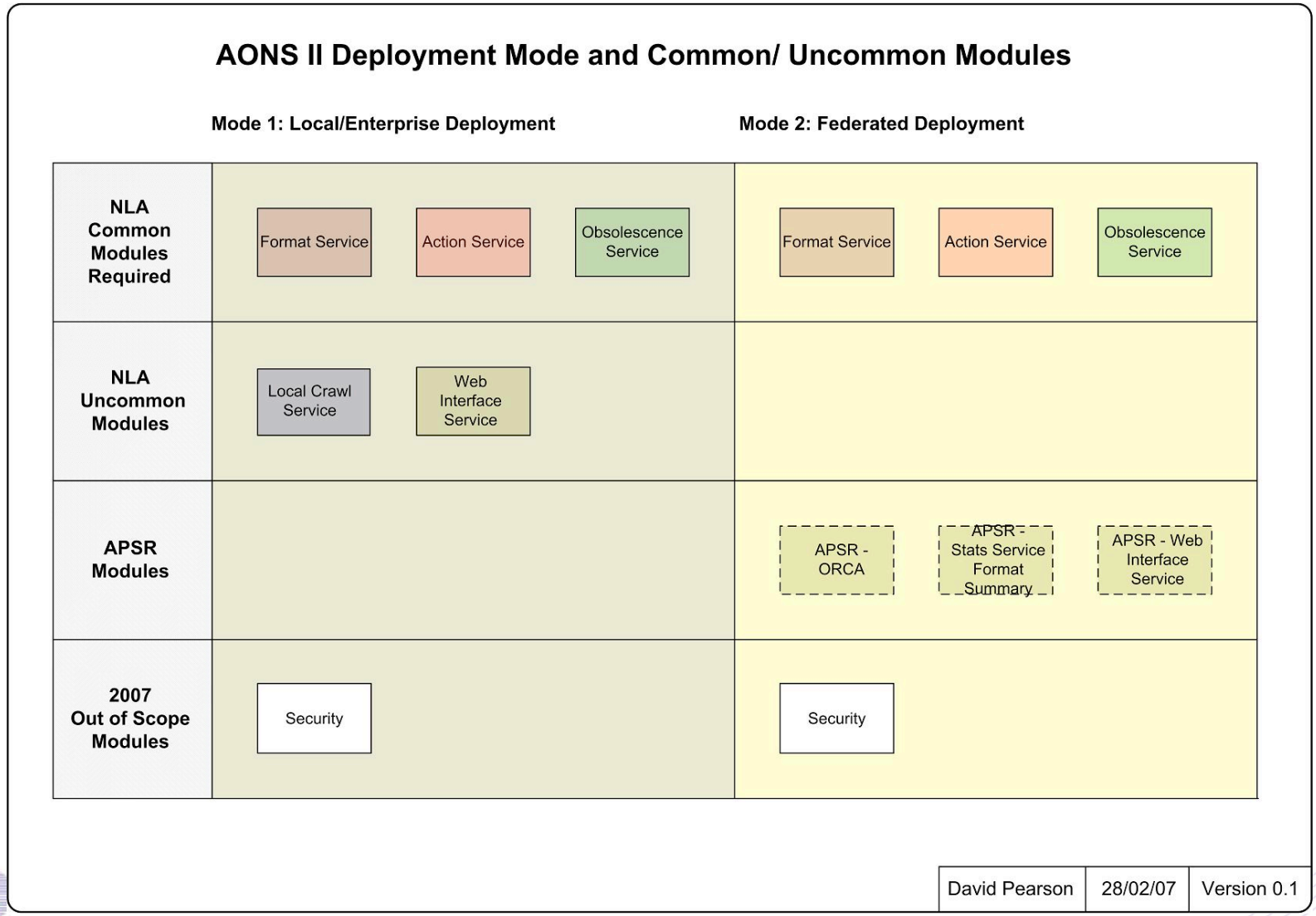
- Ever-increasing volume of digital material
- Plethora of file formats
- Dynamic nature of computing environments
- Rapid and unpredictable drivers that cause formats to become obsolete
- High business value of the specific content of some digital materials or collections can result in policies that mandate that access be maintained to this data for extended periods of time

Repository managers need help to manage the quantity and diversity of file formats and their obsolescence risks.

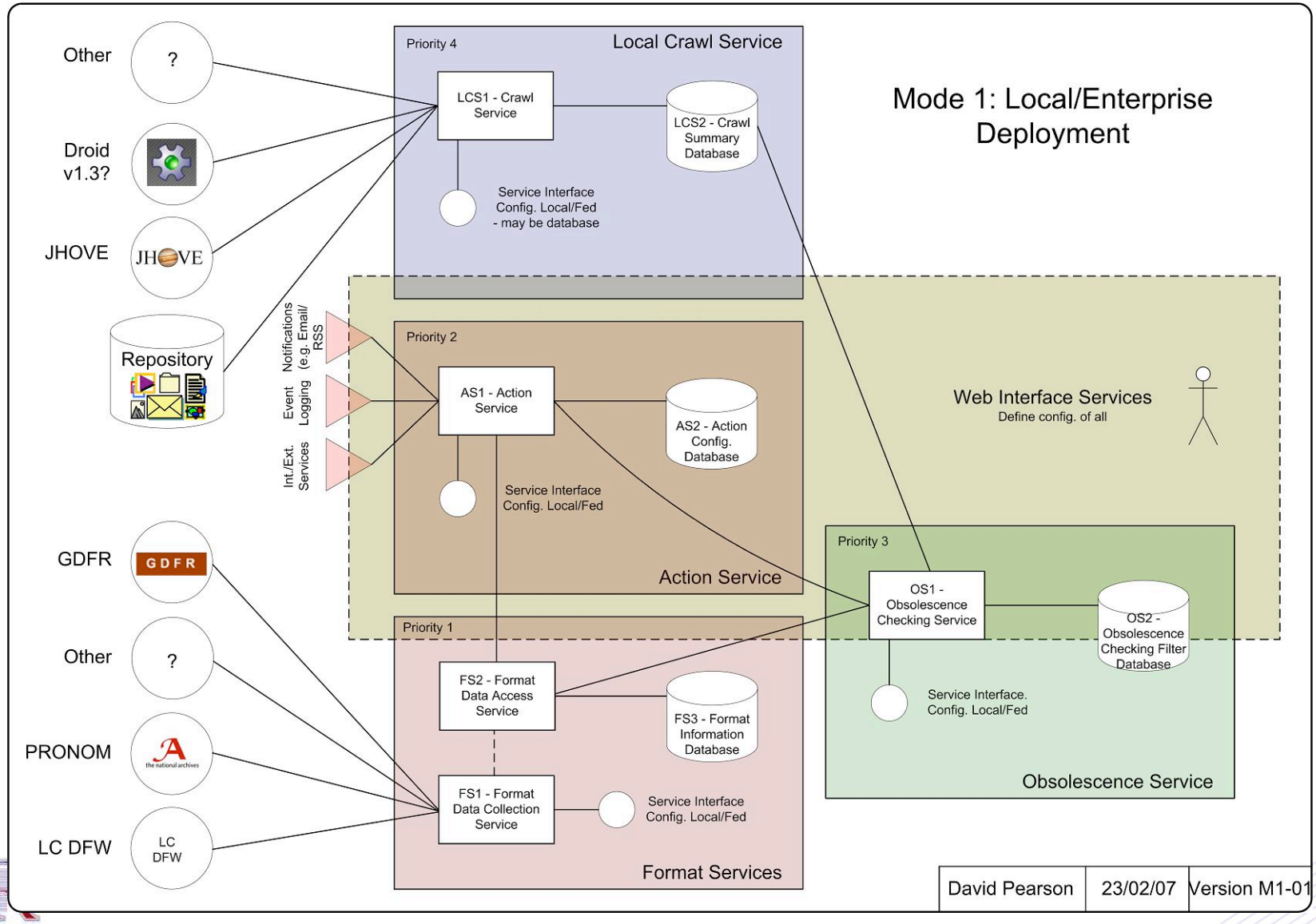
# High-level Software Design Principles

- Support three different business environments
  - a national federated infrastructure
  - enterprise business models
  - individual standalone repository sites
- Open source & use Java code
- Modular & Reusable/Adaptable
- Platform independent
- Interoperable, using common interfaces, protocols and standards
- RESTful Service-Oriented Architecture
- Loosely-coupled configurable design
- Demonstrable

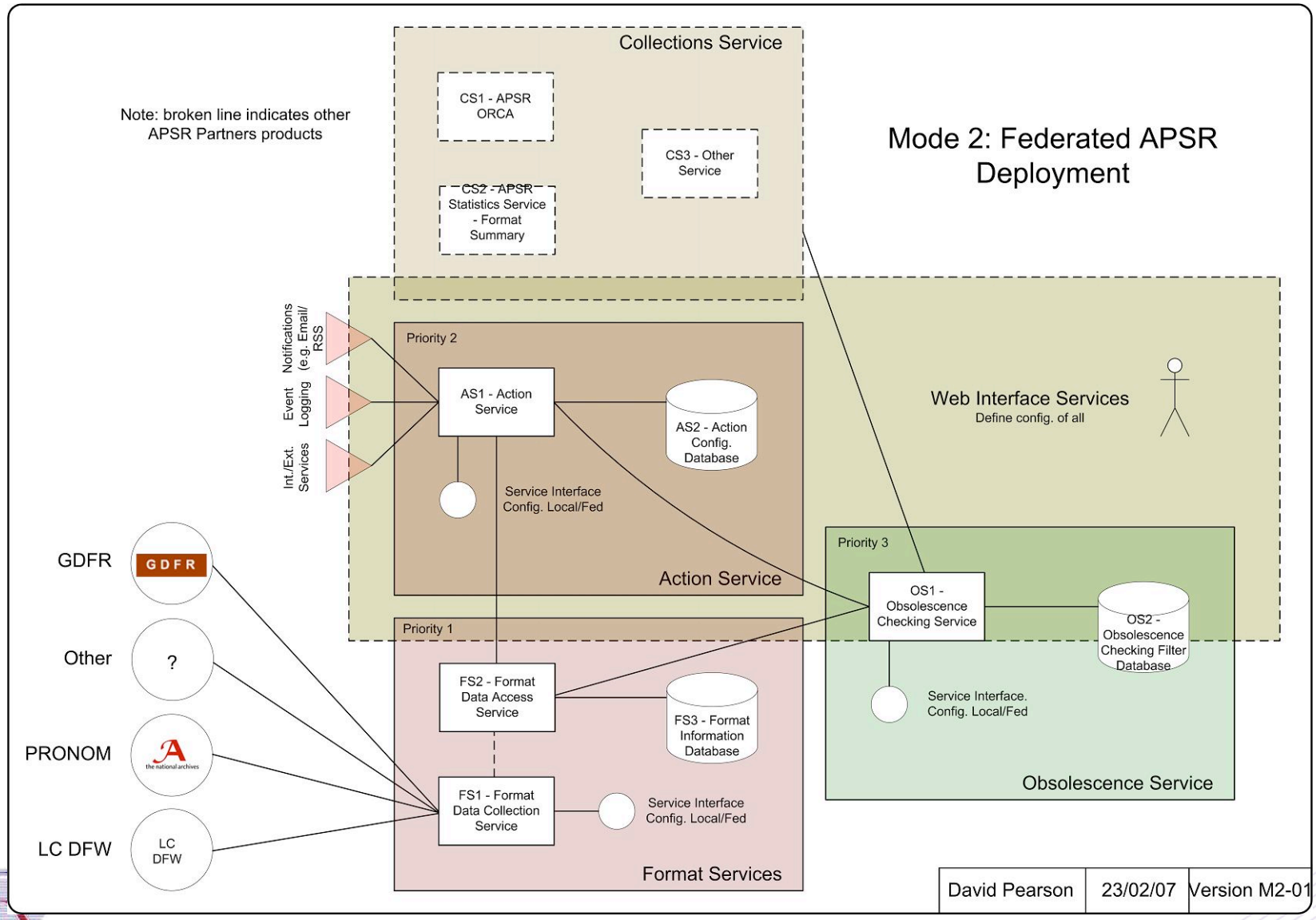
# Deployment Modes



# Mode 1: Local/Enterprise deployment



# Mode 2: Federated APSR deployment



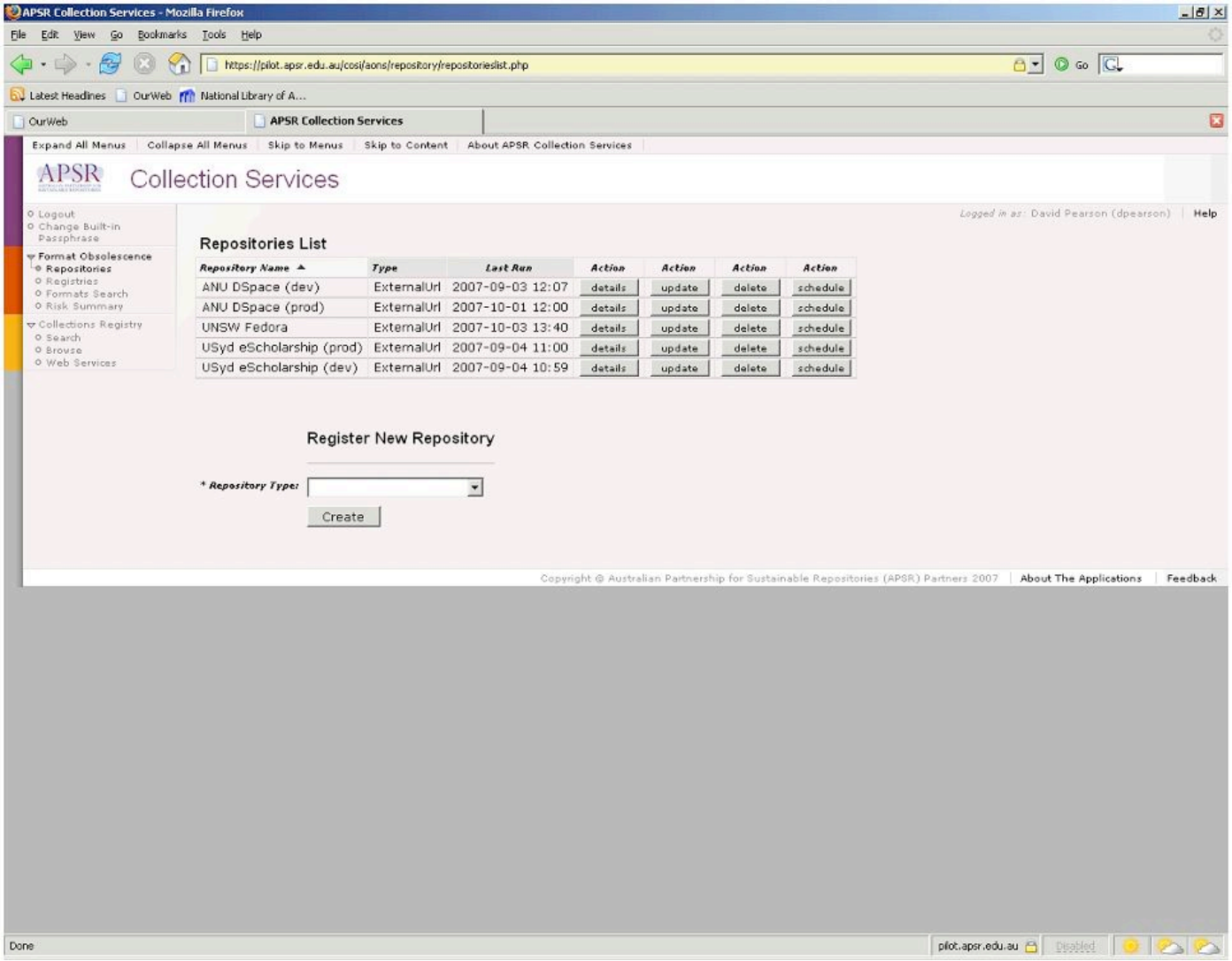
# AONS II in Action

- Check files for format obsolescence
  - On-Ingest
  - On-Demand
    - adhoc or scheduled monitoring

# Recognising File Formats and Building Collection Profiles

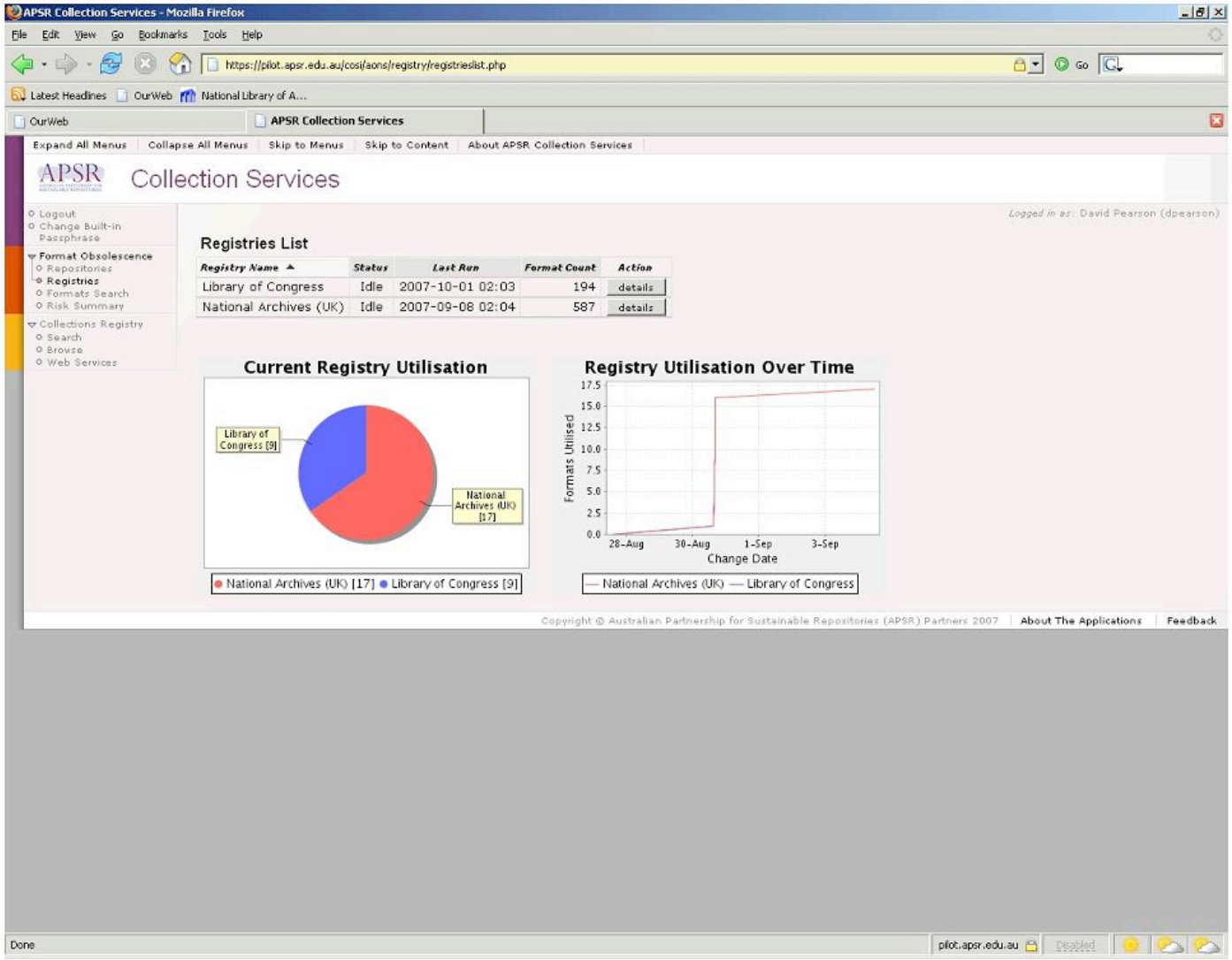
- Collection profile construction
  - Compliant XML metadata summary
  - Repository crawl using adapters purpose-built for a type of repository
- Crawl result processing
  - Repository-supplied metadata (e.g. from DSpace)
  - Automated format recognition tools (such as DROID, JHOVE or other plug-in tool).

# Repository List



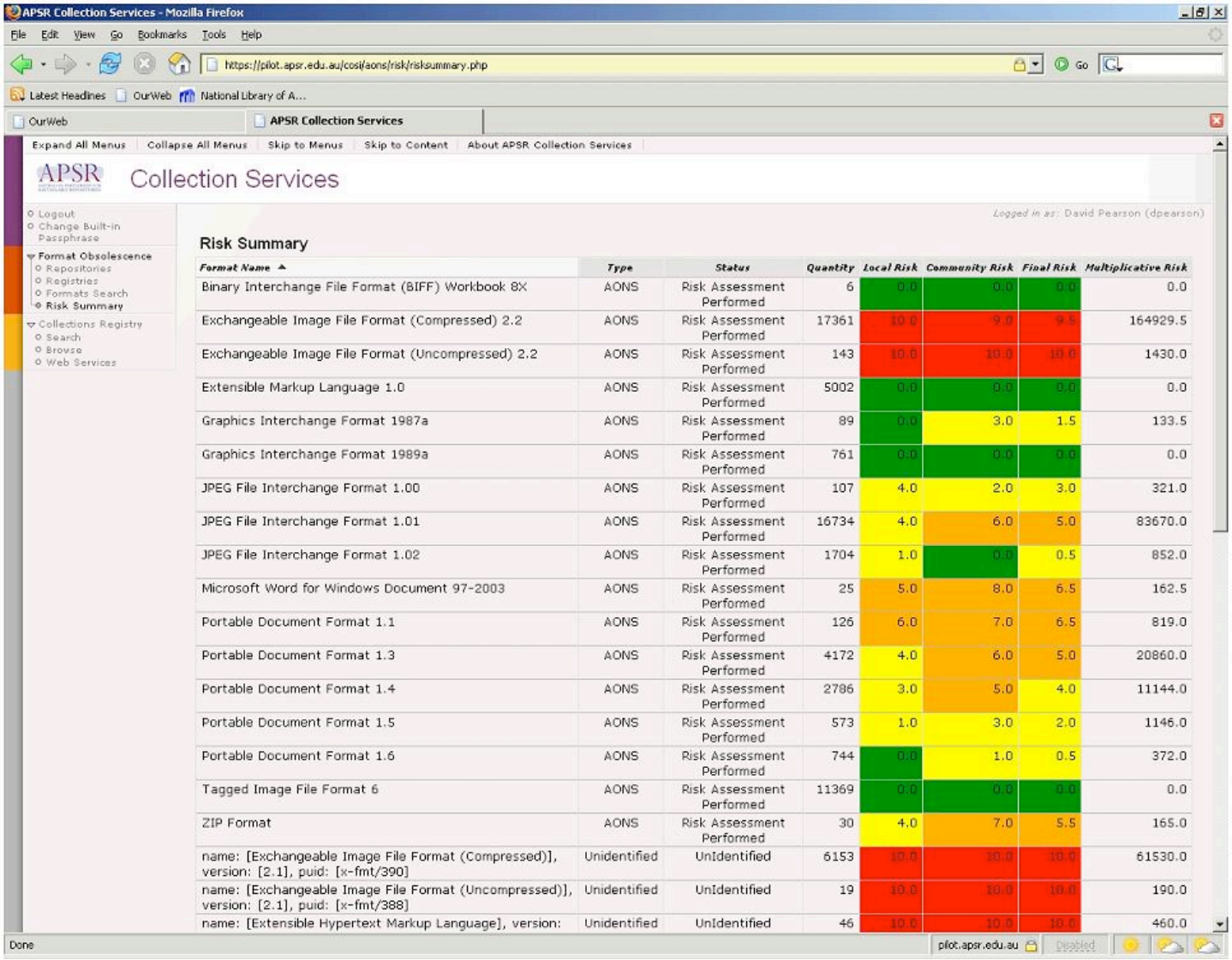

APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

# Registry List




APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

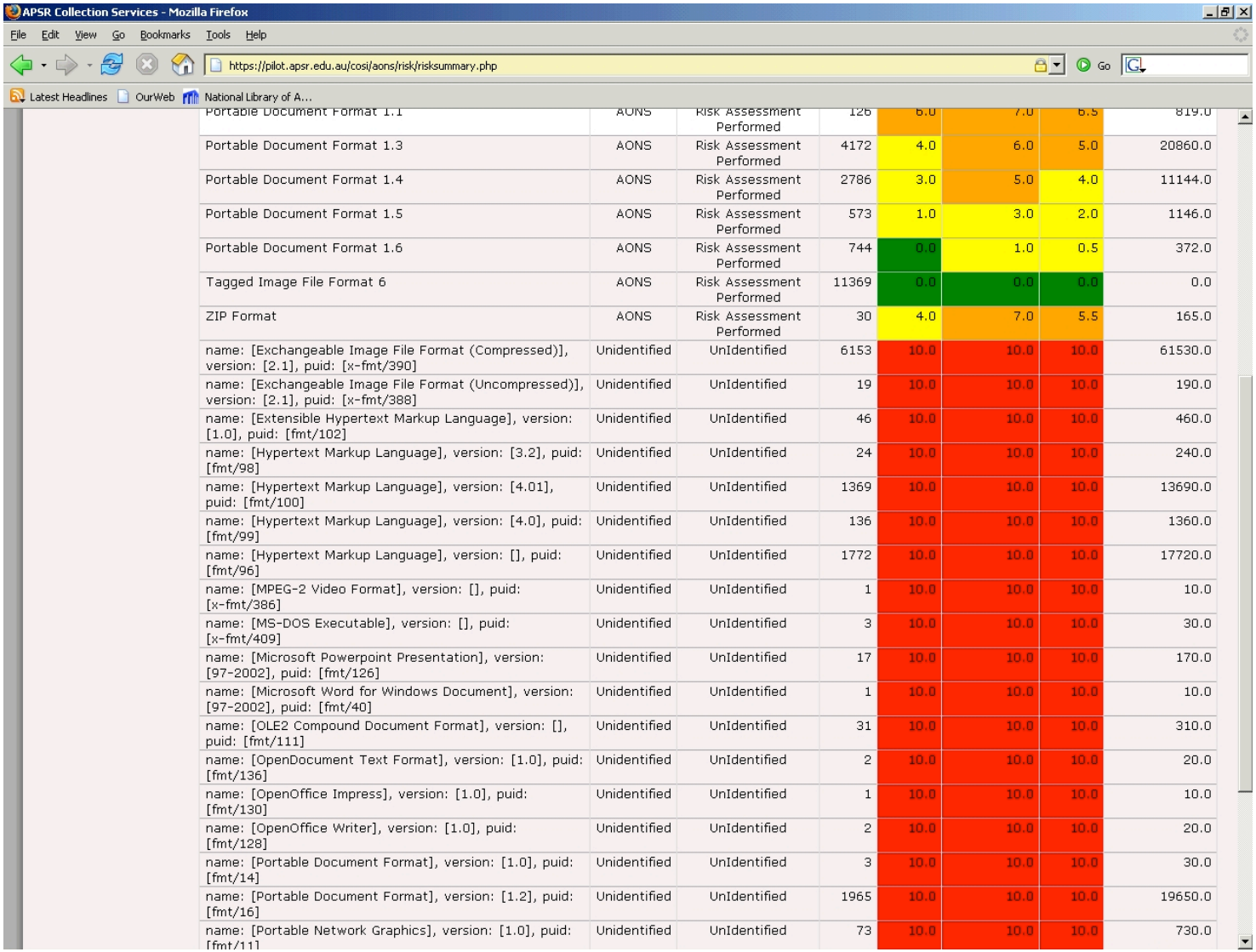
# Risk Summary




| Format Name | Type | Status | Quantity | Local Risk | Community Risk | Final Risk | Multiplicative Risk |
|---|---|---|---|---|---|---|---|
| Binary Interchange File Format (BIFF) Workbook 8X | AONS | Risk Assessment Performed | 6 | 0.0 | 0.0 | 0.0 | 0.0 |
| Exchangeable Image File Format (Compressed) 2.2 | AONS | Risk Assessment Performed | 17361 | 10.0 | 9.0 | 9.5 | 164929.5 |
| Exchangeable Image File Format (Uncompressed) 2.2 | AONS | Risk Assessment Performed | 143 | 10.0 | 10.0 | 10.0 | 1430.0 |
| Extensible Markup Language 1.0 | AONS | Risk Assessment Performed | 5002 | 0.0 | 0.0 | 0.0 | 0.0 |
| Graphics Interchange Format 1987a | AONS | Risk Assessment Performed | 89 | 0.0 | 3.0 | 1.5 | 133.5 |
| Graphics Interchange Format 1989a | AONS | Risk Assessment Performed | 761 | 0.0 | 0.0 | 0.0 | 0.0 |
| JPEG File Interchange Format 1.00 | AONS | Risk Assessment Performed | 107 | 4.0 | 2.0 | 3.0 | 321.0 |
| JPEG File Interchange Format 1.01 | AONS | Risk Assessment Performed | 16734 | 4.0 | 6.0 | 5.0 | 83670.0 |
| JPEG File Interchange Format 1.02 | AONS | Risk Assessment Performed | 1704 | 1.0 | 0.0 | 0.5 | 852.0 |
| Microsoft Word for Windows Document 97-2003 | AONS | Risk Assessment Performed | 25 | 5.0 | 8.0 | 6.5 | 162.5 |
| Portable Document Format 1.1 | AONS | Risk Assessment Performed | 126 | 6.0 | 7.0 | 6.5 | 819.0 |
| Portable Document Format 1.3 | AONS | Risk Assessment Performed | 4172 | 4.0 | 6.0 | 5.0 | 20860.0 |
| Portable Document Format 1.4 | AONS | Risk Assessment Performed | 2786 | 3.0 | 5.0 | 4.0 | 11144.0 |
| Portable Document Format 1.5 | AONS | Risk Assessment Performed | 573 | 1.0 | 3.0 | 2.0 | 1146.0 |
| Portable Document Format 1.6 | AONS | Risk Assessment Performed | 744 | 0.0 | 1.0 | 0.5 | 372.0 |
| Tagged Image File Format 6 | AONS | Risk Assessment Performed | 11369 | 0.0 | 0.0 | 0.0 | 0.0 |
| ZIP Format | AONS | Risk Assessment Performed | 30 | 4.0 | 7.0 | 5.5 | 165.0 |
| name: [Exchangeable Image File Format (Compressed)], version: [2.1], puid: [x-fmt/390] | Unidentified | UnIdentified | 6153 | 10.0 | 10.0 | 10.0 | 61530.0 |
| name: [Exchangeable Image File Format (Uncompressed)], version: [2.1], puid: [x-fmt/388] | Unidentified | UnIdentified | 19 | 10.0 | 10.0 | 10.0 | 190.0 |
| name: [Extensible Hypertext Markup Language], version: | Unidentified | UnIdentified | 46 | 10.0 | 10.0 | 10.0 | 460.0 |

APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

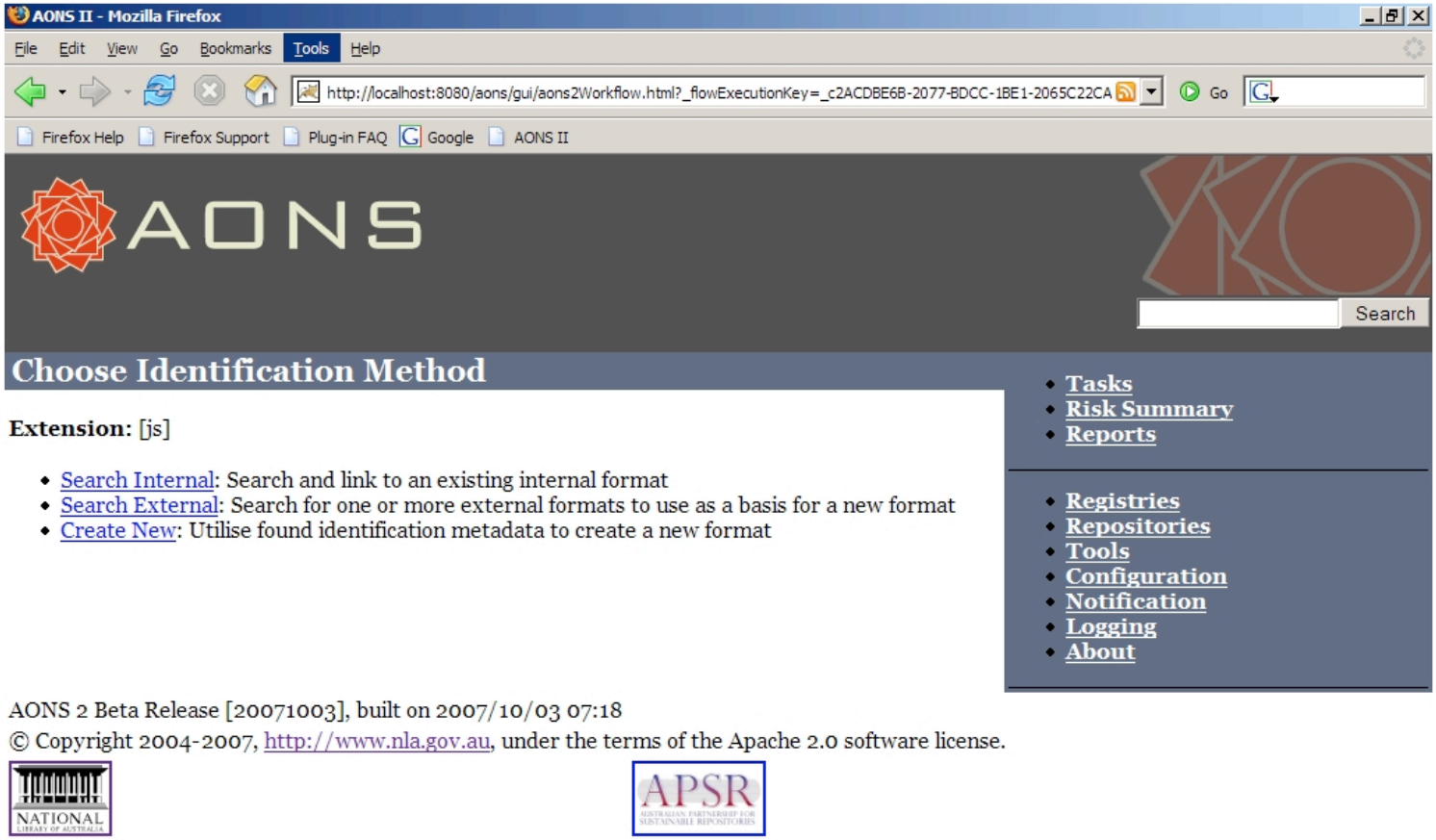
# Format Identification Issues




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Portable Document Format 1.1 | AONS | Risk Assessment Performed | 126 | 6.0 | 7.0 | 6.5 | 819.0 |
| Portable Document Format 1.3 | AONS | Risk Assessment Performed | 4172 | 4.0 | 6.0 | 5.0 | 20860.0 |
| Portable Document Format 1.4 | AONS | Risk Assessment Performed | 2786 | 3.0 | 5.0 | 4.0 | 11144.0 |
| Portable Document Format 1.5 | AONS | Risk Assessment Performed | 573 | 1.0 | 3.0 | 2.0 | 1146.0 |
| Portable Document Format 1.6 | AONS | Risk Assessment Performed | 744 | 0.0 | 1.0 | 0.5 | 372.0 |
| Tagged Image File Format 6 | AONS | Risk Assessment Performed | 11369 | 0.0 | 0.0 | 0.0 | 0.0 |
| ZIP Format | AONS | Risk Assessment Performed | 30 | 4.0 | 7.0 | 5.5 | 165.0 |
| name: [Exchangeable Image File Format (Compressed)], version: [2.1], puid: [x-fmt/390] | Unidentified | UnIdentified | 6153 | 10.0 | 10.0 | 10.0 | 61530.0 |
| name: [Exchangeable Image File Format (Uncompressed)], version: [2.1], puid: [x-fmt/388] | Unidentified | UnIdentified | 19 | 10.0 | 10.0 | 10.0 | 190.0 |
| name: [Extensible Hypertext Markup Language], version: [1.0], puid: [fmt/102] | Unidentified | UnIdentified | 46 | 10.0 | 10.0 | 10.0 | 460.0 |
| name: [Hypertext Markup Language], version: [3.2], puid: [fmt/98] | Unidentified | UnIdentified | 24 | 10.0 | 10.0 | 10.0 | 240.0 |
| name: [Hypertext Markup Language], version: [4.01], puid: [fmt/100] | Unidentified | UnIdentified | 1369 | 10.0 | 10.0 | 10.0 | 13690.0 |
| name: [Hypertext Markup Language], version: [4.0], puid: [fmt/99] | Unidentified | UnIdentified | 136 | 10.0 | 10.0 | 10.0 | 1360.0 |
| name: [Hypertext Markup Language], version: [], puid: [fmt/96] | Unidentified | UnIdentified | 1772 | 10.0 | 10.0 | 10.0 | 17720.0 |
| name: [MPEG-2 Video Format], version: [], puid: [x-fmt/386] | Unidentified | UnIdentified | 1 | 10.0 | 10.0 | 10.0 | 10.0 |
| name: [MS-DOS Executable], version: [], puid: [x-fmt/409] | Unidentified | UnIdentified | 3 | 10.0 | 10.0 | 10.0 | 30.0 |
| name: [Microsoft Powerpoint Presentation], version: [97-2002], puid: [fmt/126] | Unidentified | UnIdentified | 17 | 10.0 | 10.0 | 10.0 | 170.0 |
| name: [Microsoft Word for Windows Document], version: [97-2002], puid: [fmt/40] | Unidentified | UnIdentified | 1 | 10.0 | 10.0 | 10.0 | 10.0 |
| name: [OLE2 Compound Document Format], version: [], puid: [fmt/111] | Unidentified | UnIdentified | 31 | 10.0 | 10.0 | 10.0 | 310.0 |
| name: [OpenDocument Text Format], version: [1.0], puid: [fmt/136] | Unidentified | UnIdentified | 2 | 10.0 | 10.0 | 10.0 | 20.0 |
| name: [OpenOffice Impress], version: [1.0], puid: [fmt/130] | Unidentified | UnIdentified | 1 | 10.0 | 10.0 | 10.0 | 10.0 |
| name: [OpenOffice Writer], version: [1.0], puid: [fmt/128] | Unidentified | UnIdentified | 2 | 10.0 | 10.0 | 10.0 | 20.0 |
| name: [Portable Document Format], version: [1.0], puid: [fmt/14] | Unidentified | UnIdentified | 3 | 10.0 | 10.0 | 10.0 | 30.0 |
| name: [Portable Document Format], version: [1.2], puid: [fmt/16] | Unidentified | UnIdentified | 1965 | 10.0 | 10.0 | 10.0 | 19650.0 |
| name: [Portable Network Graphics], version: [1.0], puid: [fmt/11] | Unidentified | UnIdentified | 73 | 10.0 | 10.0 | 10.0 | 730.0 |

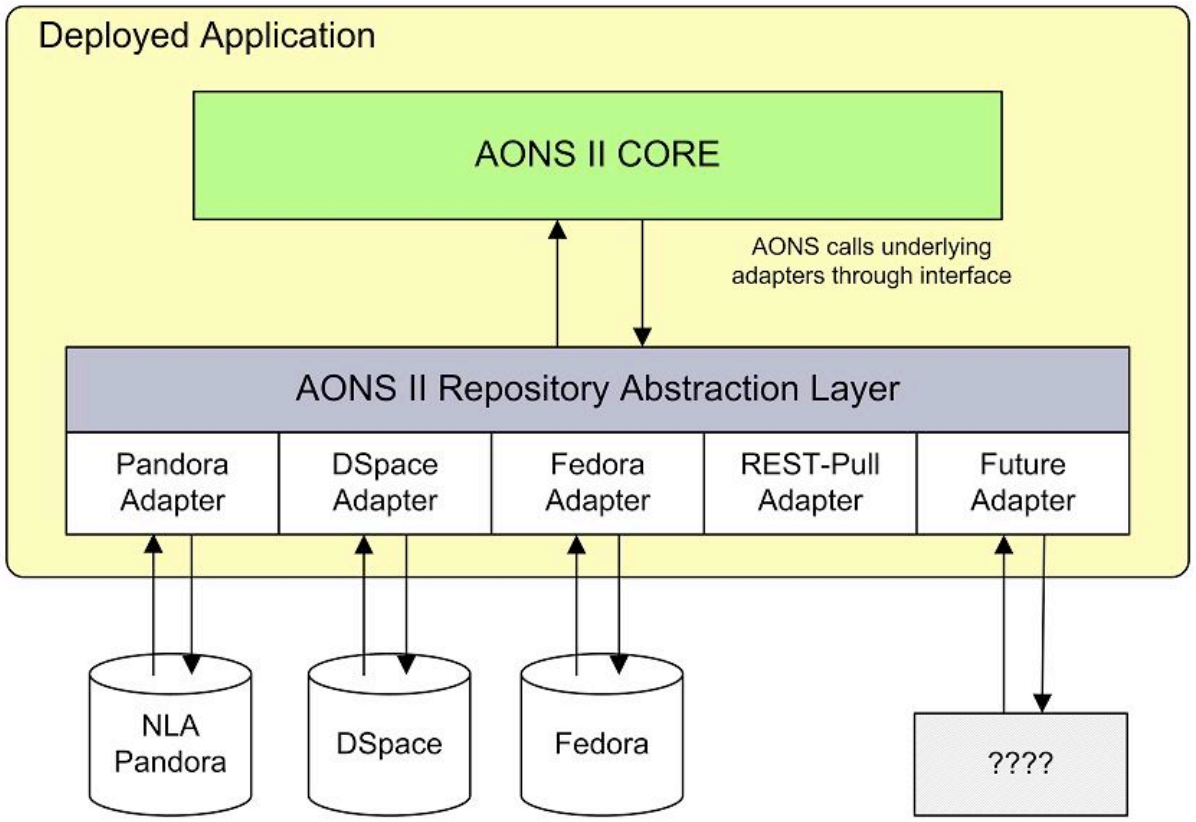
APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

# Format Identification/Linkage



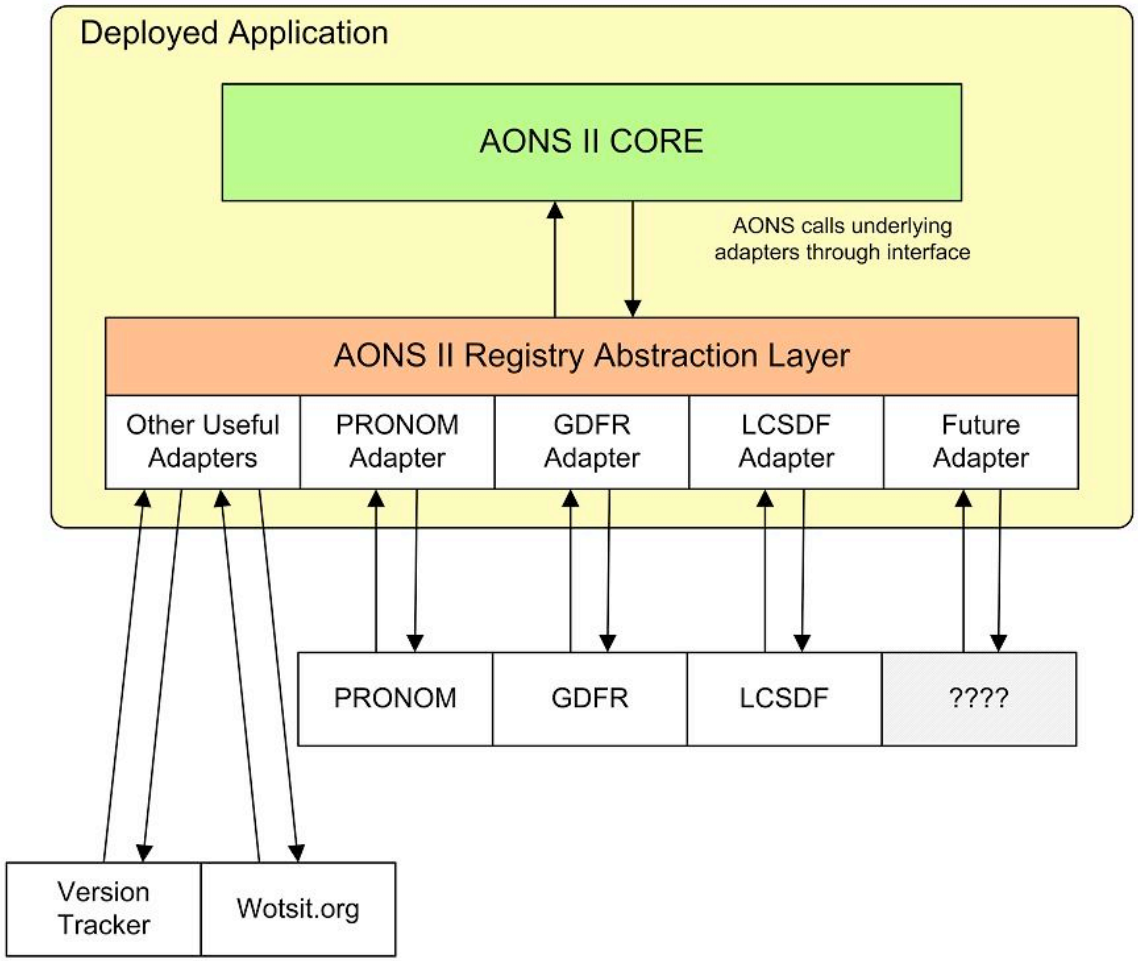
# Repository Adapters

- Completed:
  - DSpace (1.4)
  - Fedora (2.2)
  - NLA Pandora
  - REST/Pull
  - Generic File System
- Not completed 2007
  - EPrints
  - SRB (UniQLD)
  - TRIM

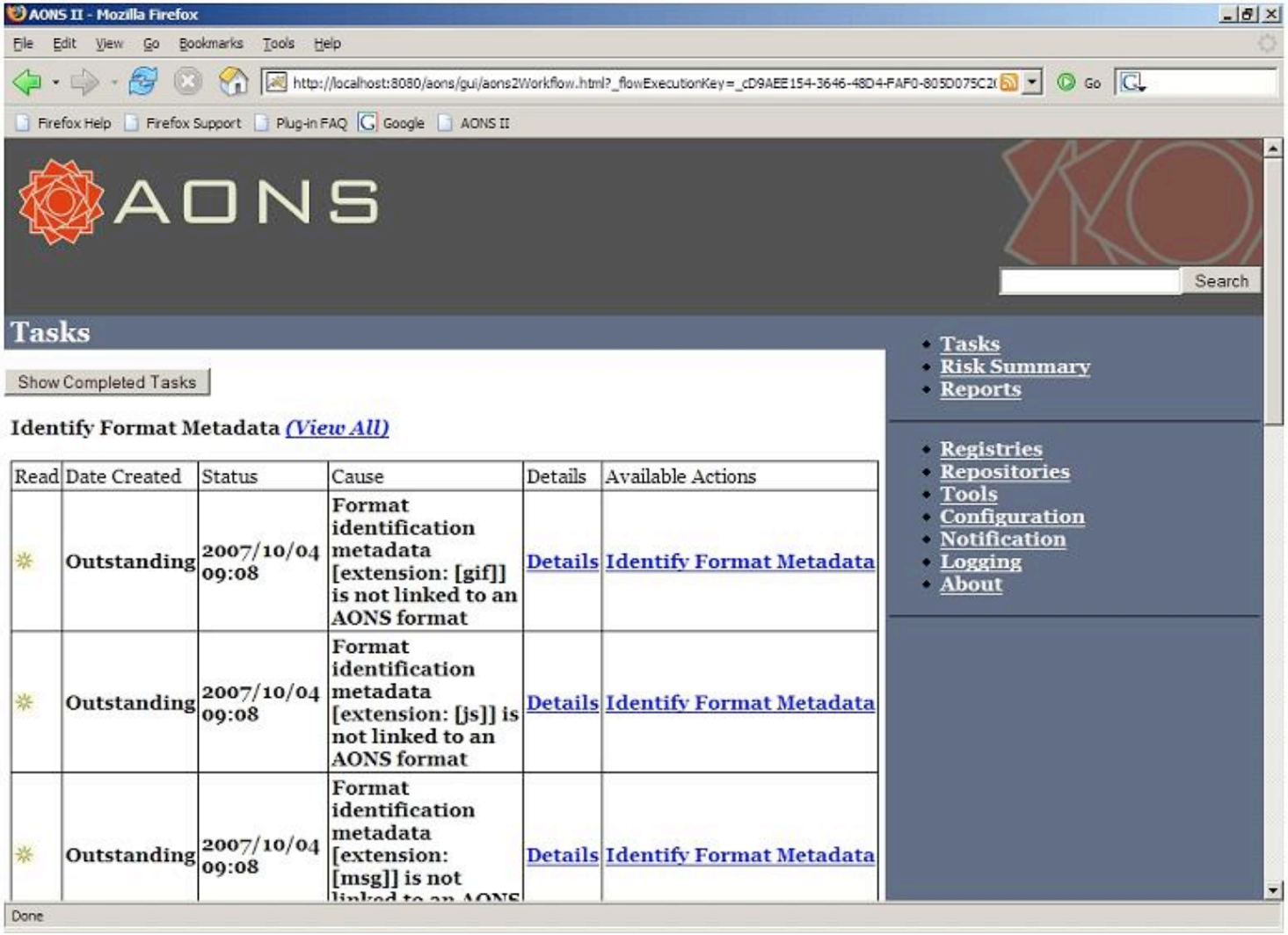


# Registry Adapters

- Completed
  - LoC SDF
  - PRONOM
- Not completed
  - GDFR



# Task List




APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

# RDSS - Risk Questions

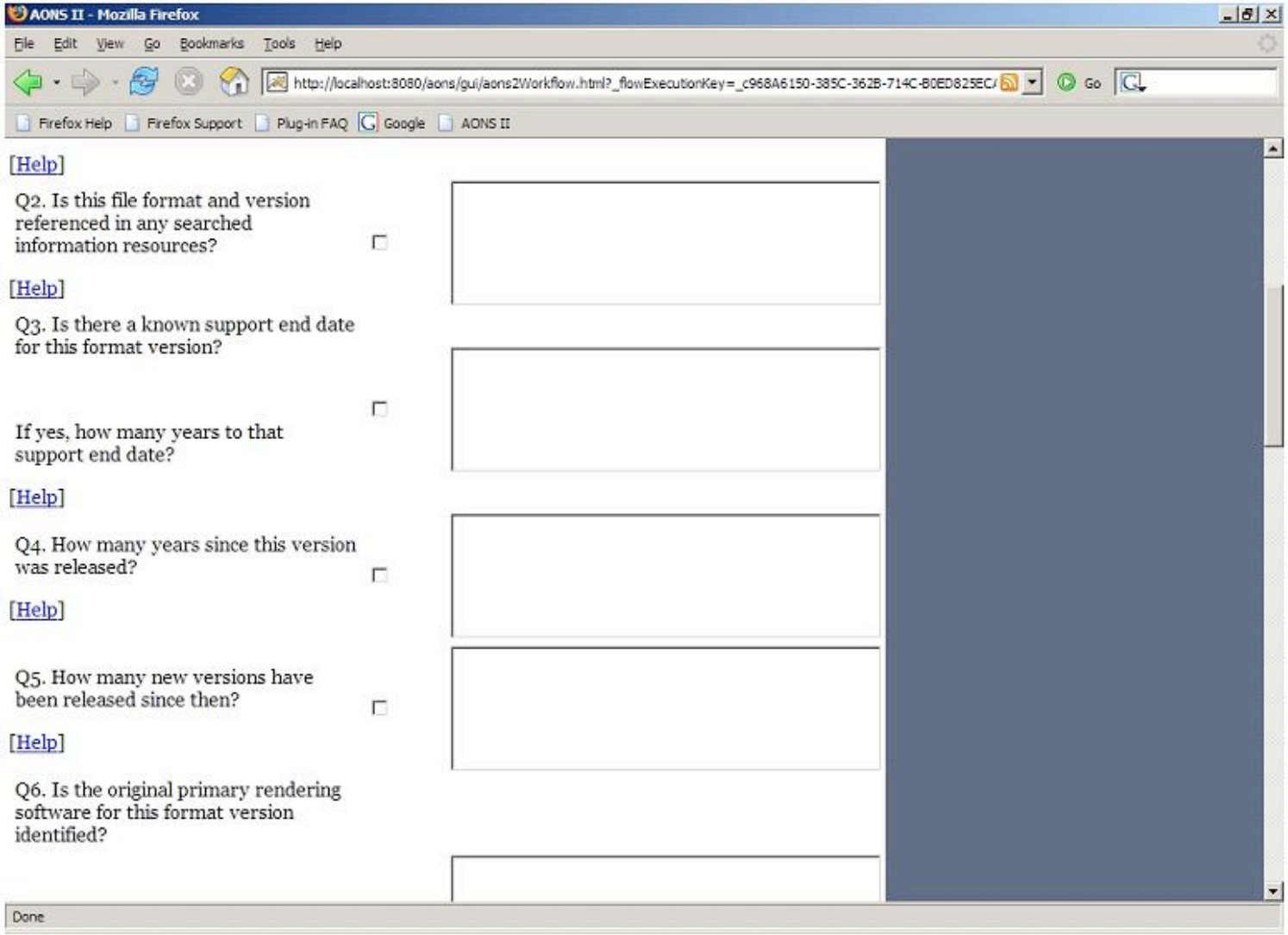
| Step 1 – Risk assessment based on Community Information | Step 2 - Risk assessment based on Collection / Repository Information |
|---|---|
| Community Information Questions | Collection / Repository Information Questions |
| Q1. Is this a base format? A ubiquitous format which is likely to be rendered by most applications (e. g. plain text). | Q1. The original primary rendering software has been identified as ... (see Step 1 - Q6.) |
| | effective access? (i.e. custom designed applications, scripts, emulators). What are they? |
| Q6. Is the original primary rendering software for this format version identified? What is it? | Q6. Overall, how many access options are effectively available to you (i.e. how many can you make work), including the original rendering software? |
| Q9. For each alternative, are there critical hardware and software dependencies for effective use of the alternative rendering software? What are they? If unknown, state "Unknown". | |

Q1. The original primary rendering software has been identified as ... (see Step 1 - Q6.)
Is this primary rendering software available to you?

Q6. Is the original primary rendering software for this format version identified?
What is it?
If unknown, state "Unknown".

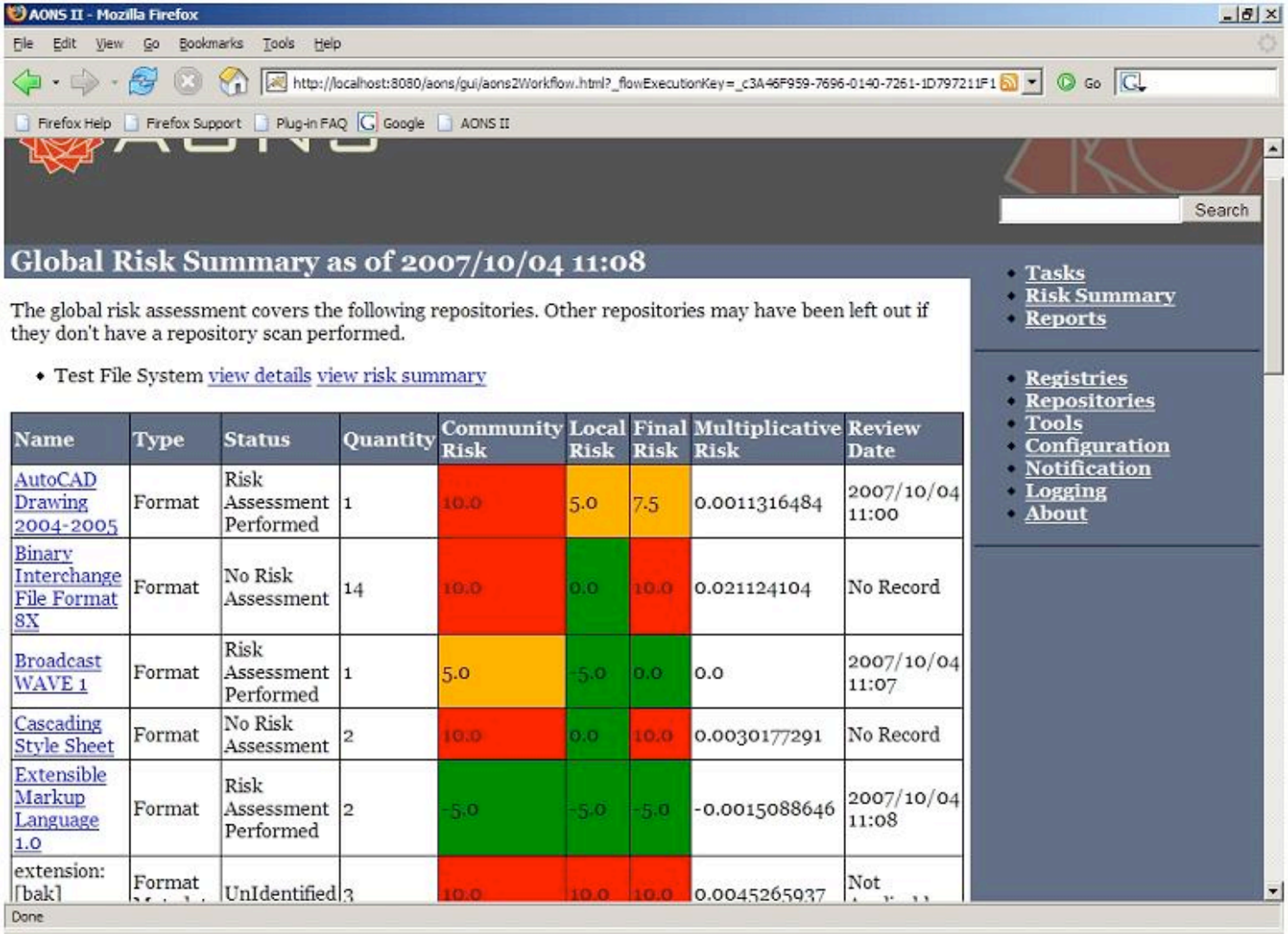
APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

# Format Risk Assessment




APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

# Risk Summary



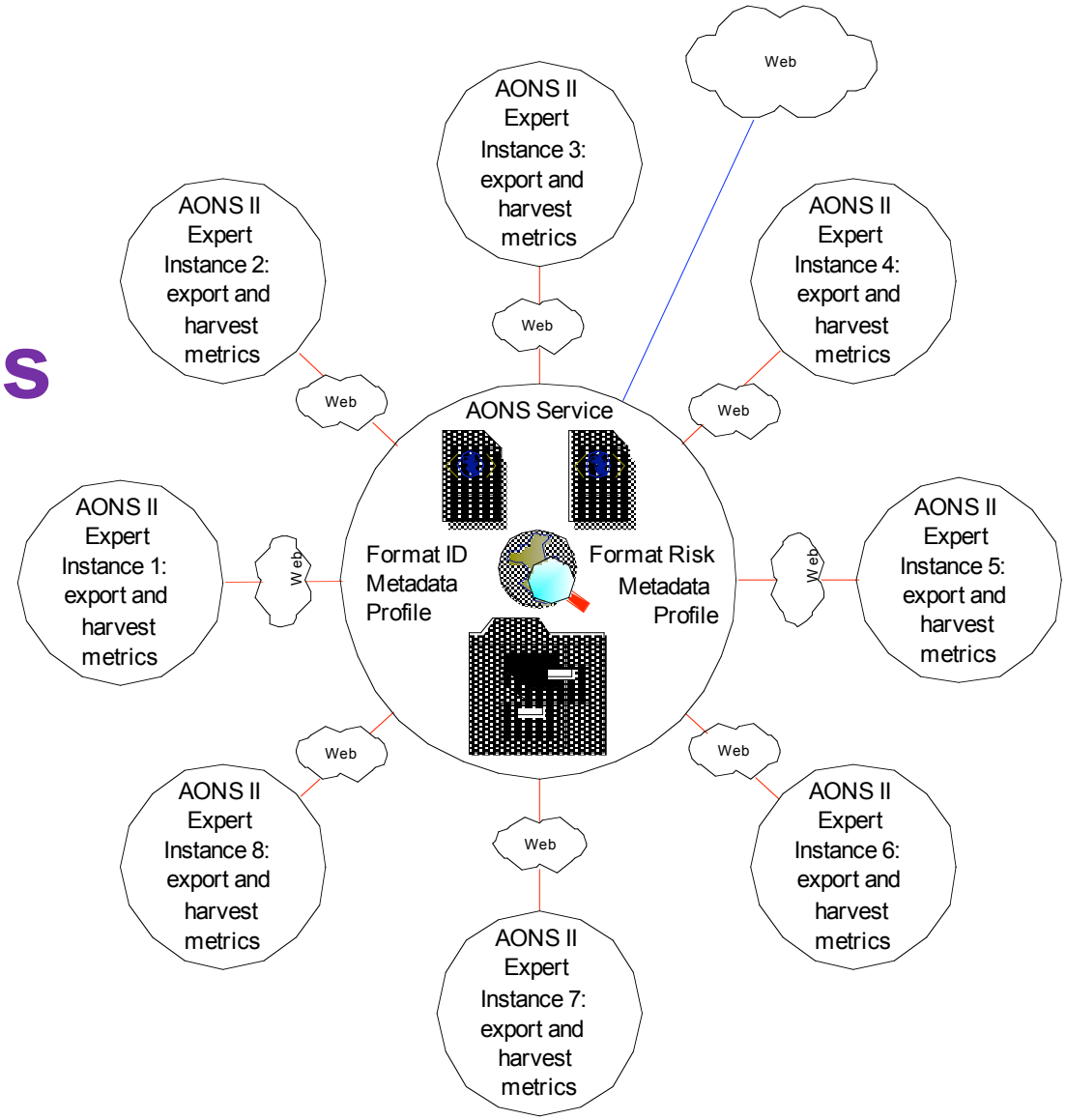

| Name | Type | Status | Quantity | Community Risk | Local Risk | Final Risk | Multiplicative Risk | Review Date |
|---|---|---|---|---|---|---|---|---|
| AutoCAD Drawing 2004-2005 | Format | Risk Assessment Performed | 1 | 10.0 | 5.0 | 7.5 | 0.0011316484 | 2007/10/04 11:00 |
| Binary Interchange File Format 8X | Format | No Risk Assessment | 14 | 10.0 | 0.0 | 10.0 | 0.021124104 | No Record |
| Broadcast WAVE 1 | Format | Risk Assessment Performed | 1 | 5.0 | -5.0 | 0.0 | 0.0 | 2007/10/04 11:07 |
| Cascading Style Sheet | Format | No Risk Assessment | 2 | 10.0 | 0.0 | 10.0 | 0.0030177291 | No Record |
| Extensible Markup Language 1.0 | Format | Risk Assessment Performed | 2 | -5.0 | -5.0 | -5.0 | -0.0015088646 | 2007/10/04 11:08 |
| extension: [bak] | Format | UnIdentified | 3 | 10.0 | 10.0 | 10.0 | 0.0045265937 | Not |

APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

# Demo

- http://localhost:8080/aons

# Possible Futures




AONS II Service (Web)
Providing metrics for:
- Format ID Metadata Profile; and
- Format Risk Metadata Profile.

Information push/pulled from-to other AONS II expert instances. Information available to all.

APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

# AONS Ten Commandments

I: Thou shalt not worship proprietary file formats

II: Thou shalt not discard files of wayward format

III: Remember not to analyse file formats manually on the weekend

IV: Honour “base” file formats

V: Thou shalt not kill digital file contents

VI: Thou shalt not commit to a risky file format

VIII: Thou shalt not falsely assess a file format’s risk

IX: Thou shalt not covet other format rendering infrastructure

X: Thou shalt not covet supported file formats

APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au

# Acknowledgements & Links

- APSR:
  - http://www.apsr.edu.au/
- APSR AONS II Home Page:
  - http://www.apsr.edu.au/aons2/index.htm
- AONS II Wiki:
  - http://www.apsr.edu.au/wiki/index.php/AONS
- AONS II Development Blog:
  - http://aons2dev.blogspot.com
- Download AONS:
  - http://sourceforge.net/projects/aons/

APSR
AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES
www.apsr.edu.au