



AONS

AONS II: continuing the trend towards preservation software 'Nirvana' iPRES2007

David Pearson
APSR Project Manager
National Library of Australia
dapearso@nla.gov.au



If only file formats were so durable!



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 (2007-10-03) 03 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 is NLA/APSR building 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; and
- ▶ 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 numbers and diversity of file formats and their obsolescence risks. Hence, the need for a tool like AONS II.

High Level Software Design Principles

AONS II had to conform to several fundamental design principles. We decided it must:

- ▶ Support three different business environments: a national federated infrastructure, enterprise business models, and individual standalone repository sites;
- ▶ Be open source using Java code;
- ▶ Be modular and have a reusable/adaptable design;
- ▶ Be platform independent using a decoupled approach;
- ▶ Be interoperable, using common interfaces, protocols and standards;
- ▶ Provide service interfaces in a Service-Oriented Architecture based on RESTful approach;
- ▶ Provide a core set of functionality, which abstracts repositories and registries functionality away from the core, and would allow new repository and registry adapters to be added without affecting the core; and
- ▶ be demonstrable.

These principles have provided a yardstick and reality check for all development work.

Deployment Modes 2007

In conforming with the principle that the software must work within national federated infrastructure, enterprise business models and individual standalone repository sites, two deployment modes have been implemented:

- ▶ Mode 1: Local/Enterprise deployment (local internal and local networked sites); and
- ▶ Mode 2: Federated/APSR deployment.

AONS II Deployment Mode and Common/ Uncommon Modules

Mode 1: Local/Enterprise Deployment

Mode 2: Federated Deployment

NLA Common Modules Required	<div>Format Service</div> <div>Action Service</div> <div>Obsolescence Service</div>	<div>Format Service</div> <div>Action Service</div> <div>Obsolescence Service</div>
NLA Uncommon Modules	<div>Local Crawl Service</div> <div>Web Interface Service</div>	
APSR Modules		<div>APSR - ORCA</div> <div>APSR - Stats Service Format Summary</div> <div>APSR - Web Interface Service</div>
2007 Out of Scope Modules	<div>Security</div>	<div>Security</div>

David Pearson

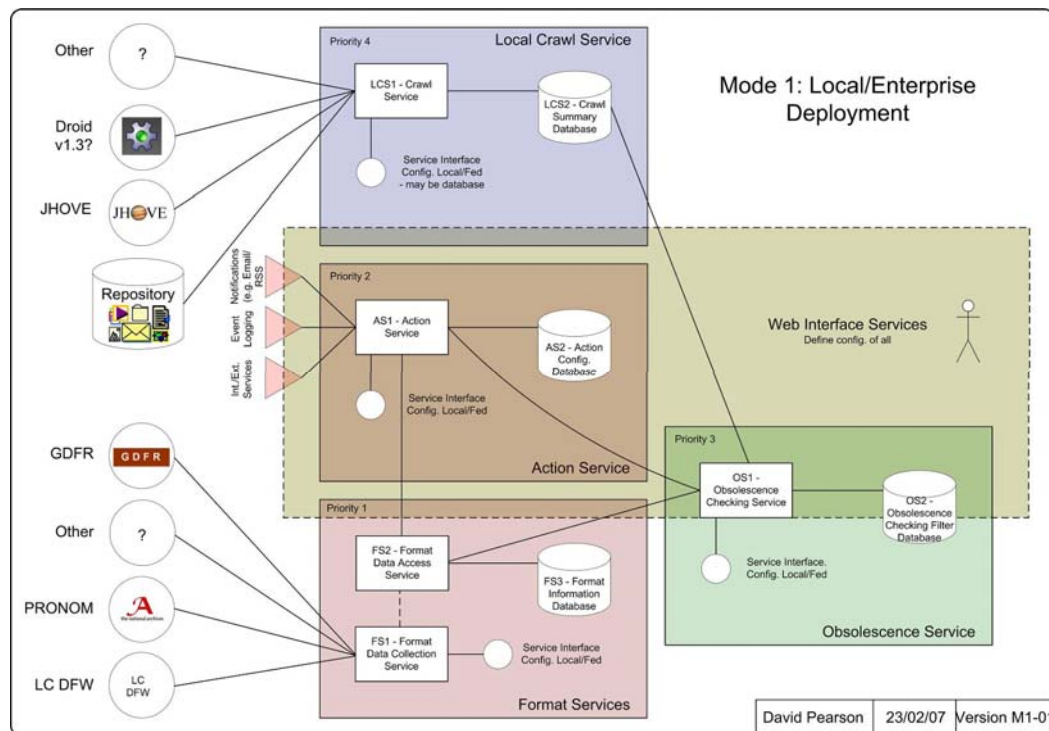
28/02/07

Version 0.1

Mode 1: Local/Enterprise deployment

Mode 1: Local/Enterprise Data deliverable characteristics:

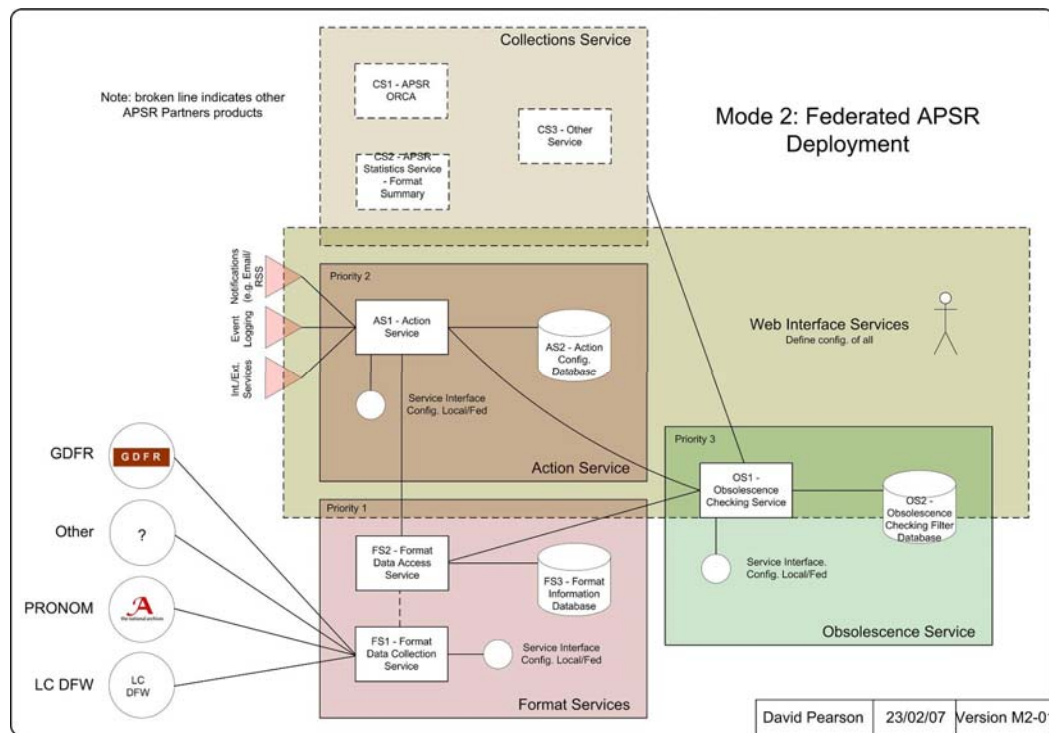
- ▶ No mandatory public access – repository owners can configure this on their own;
- ▶ Access to local and probably unshared data repositories via format summaries produced by a local crawl service;
- ▶ Obsolescence rules will be heavily customised based on client environment;
- ▶ User interface bundled with main components for deployment simplicity;
- ▶ Still utilise access to external format registries, but may augment this information with own data; and
- ▶ Some organisations may wish to integrate the obsolescence tool with their own software.



Mode 2: Federated APSR deployment

Mode 2: Federated (APSR) deliverable characteristics:

- ▶ Public anonymous access for read only operations;
- ▶ Protected access for system owners who can change state of the application;
- ▶ Access to public data repositories via the format summary service and ORCA (Online Research Collections Australia - APSR Collection Registry Product) which will be delivered by APSR;
- ▶ Public viewing of read only obsolescence reports;
- ▶ Public access to static, non-contextual rules for evaluating document obsolescence within public repositories;
- ▶ User interface operating on a separate server to main module deployment;
- ▶ Access to external format registries; and
- ▶ No federated repository crawl due to bandwidth and security factors



AONS II in Action

There are two main schools of thought on where to conduct preservation actions:

- ▶ Preservation on ingest; and
- ▶ Preservation on demand.

Both scenarios aim to address the fact that formats are or will eventually become obsolete. The AONS tool is designed to aid a repository manager to decide when to take action, and can be deployed to respond to both scenarios. To:

- ▶ Check files as they are ingested; or
- ▶ Check files some time after they have been ingested, either on a one-off basis or on a regular monitoring schedule.

This can take place as a part of a workflow or as a stand-alone application.



Recognising File Formats and Building Collection Profiles

AONS II builds a profile of the formats in a repository or a subset such as a collection or even a single file. The profile is constructed from:

- ▶ A compliant XML metadata summary; or
- ▶ A repository crawl using purpose-built AONS adapters designed for a given repository type.

Crawl results are processed using:

- ▶ Metadata from a repository (i.e. DSpace); and or
- ▶ Automated format recognition tools (such as DROID, JHOVE or other plug-in tool).

APSR Collection Services - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

https://pilot.apsr.edu.au/cos/aons/repository/repositorieslist.php

Latest Headlines OurWeb National Library of A...

OurWeb APSR Collection Services

Expand All Menus Collapse All Menus Skip to Menus Skip to Content About APSR Collection Services

APSR AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES

Collection Services

Logout
Change Built-in
Passphrase

Format Obsolescence
Repositories
Registries
Formats Search
Risk Summary

Collections Registry
Search
Browse
Web Services

Logged in as: David Pearson (dpearson) Help

Repositories List

Repository Name	Type	Last Run	Action	Action	Action	Action
ANU DSpace (dev)	ExternalUrl	2007-09-03 12:07	details	update	delete	schedule
ANU DSpace (prod)	ExternalUrl	2007-10-01 12:00	details	update	delete	schedule
UNSW Fedora	ExternalUrl	2007-10-03 13:40	details	update	delete	schedule
USyd eScholarship (prod)	ExternalUrl	2007-09-04 11:00	details	update	delete	schedule
USyd eScholarship (dev)	ExternalUrl	2007-09-04 10:59	details	update	delete	schedule

Register New Repository

* Repository Type:

Create

Copyright © Australian Partnership for Sustainable Repositories (APSR) Partners 2007 | About The Applications | Feedback

Done pilot.apsr.edu.au Disabled

APSR Collection Services - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

https://pilot.apsr.edu.au/cos/aons/registry/registrieslist.php

Latest Headlines OurWeb National Library of A...

OurWeb APSR Collection Services

Expand All Menus Collapse All Menus Skip to Menus Skip to Content About APSR Collection Services

APSR AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES

Collection Services

Logout
Change Built-in
Passphrase

Format Obsolescence
Repositories
Registries
Formats Search
Risk Summary

Collections Registry
Search
Browse
Web Services

Logged in as: David Pearson (dpearson)

Registries List

Registry Name	Status	Last Run	Format Count	Action
Library of Congress	Idle	2007-10-01 02:03	194	details
National Archives (UK)	Idle	2007-09-08 02:04	587	details

Current Registry Utilisation

Library of Congress [9]

National Archives (UK) [17]

Registry Utilisation Over Time

Formats Utilised

Change Date

— National Archives (UK) — Library of Congress

Copyright © Australian Partnership for Sustainable Repositories (APSR) Partners 2007 | [About The Applications](#) | [Feedback](#)

Done pilot.apsr.edu.au Disabled

APSR Collection Services - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

https://pilot.apsr.edu.au/cos/aons/risksummary.php

Latest Headlines OurWeb National Library of A...

OurWeb APSR Collection Services

Expand All Menus Collapse All Menus Skip to Menus Skip to Content About APSR Collection Services

APSR AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES

Collection Services

Logout
Change Built-in
Passphrase

Format Obsolescence
Repositories
Registries
Formats Search
Risk Summary

Collections Registry
Search
Browse
Web Services

Logged in as: David Pearson (dpearson)

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

Done

pilot.apsr.edu.au Disabled

Format Identifiers

AONS II depends on being able to distinguish accurately between different formats, and between different versions of formats, in order to manage and identify relevant risk levels. At times this is problematic.

Based on the repository formats found, AONS II may classify formats as 'identified', and matched with format information held in external registries, or 'unidentified'. As part of this classification process, a repository manager could:


- ▶ Decide to link an unidentified format to an existing AONS internal format;
- ▶ Create a new internal format with links to external format identification;
- ▶ Create a new internal format with no links; or
- ▶ Simply leave the format as unidentified.

AONS II - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost:8080/aons/gui/aons2Workflow.html?_flowExecutionKey=_c2ACDBE6B-2077-BDCC-1BE1-2065C22CA Go

Firefox Help Firefox Support Plug-in FAQ Google AONS II



AONS

Choose Identification Method

Extension: [js]

- ♦ [Search Internal](#): Search and link to an existing internal format
- ♦ [Search External](#): Search for one or more external formats to use as a basis for a new format
- ♦ [Create New](#): Utilise found identification metadata to create a new format

- ♦ [Tasks](#)
- ♦ [Risk Summary](#)
- ♦ [Reports](#)
- ♦ [Registries](#)
- ♦ [Repositories](#)
- ♦ [Tools](#)
- ♦ [Configuration](#)
- ♦ [Notification](#)
- ♦ [Logging](#)
- ♦ [About](#)

AONS 2 Beta Release [20071003], built on 2007/10/03 07:18

© Copyright 2004-2007, <http://www.nla.gov.au>, under the terms of the Apache 2.0 software license.



Done

Adapters

AONS II uses repository/registry adapters which are abstracted from the core software for interfacing to different repository and registry types.

This keeps the core code isolated from the adapters so that the basic business logic does not need to be modified when creating or modifying adapters.

Adapters

Repository Adapters
(completed):

- ▶ DSpace (1.4)
- ▶ Fedora (2.2)
- ▶ NLA Pandora
- ▶ REST/Pull
- ▶ Generic File System

(not completed 2007)

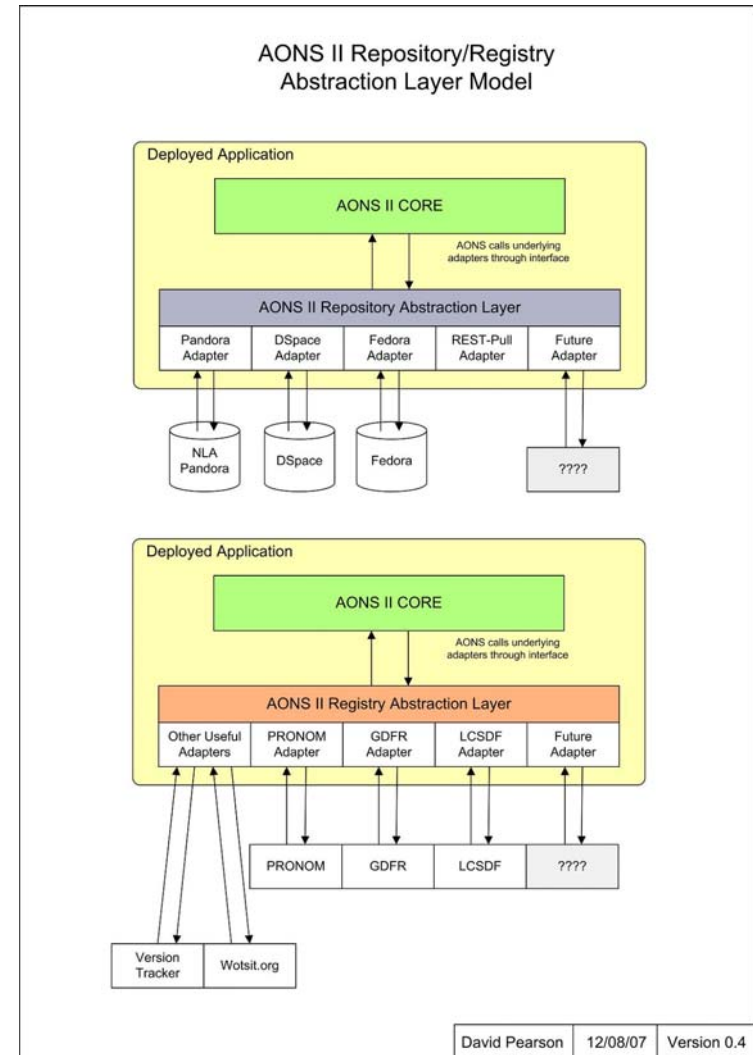
- ▶ EPrints
- ▶ SRB (UniQLD)
- ▶ TRIM

Registry Adapters
(completed):

- ▶ LoC SDF
- ▶ PRONOM

(not completed 2007)

- ▶ GDFR



Notification

The notification part of AONS II is configurable and based on a change in state within the system. For example:

- ▶ end of repository crawl;
- ▶ change in the information about a format in an external registry; or
- ▶ the expiry of a time sensitive trigger, such as format risk re-assessment period ending.

Based on the experiences from AONS I, notification can occur in a number of forms:

- ▶ Email;
- ▶ RSS Feed; and
- ▶ Task boxes via a GUI format summary screen.

AONS II - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost:8080/aons/gui/aons2Workflow.html?_flowExecutionKey=_cD9AEE154-3646-48D4-FAF0-805D075C2f

Firefox Help Firefox Support Plug-in FAQ Google AONS II

AONS

Search

Tasks

Show Completed Tasks

Identify Format Metadata [\(View All\)](#)

Read	Date Created	Status	Cause	Details	Available Actions
*	Outstanding	2007/10/04 09:08	Format identification metadata [extension: [gif]] is not linked to an AONS format	Details	Identify Format Metadata
*	Outstanding	2007/10/04 09:08	Format identification metadata [extension: [js]] is not linked to an AONS format	Details	Identify Format Metadata
*	Outstanding	2007/10/04 09:08	Format identification metadata [extension: [msg]] is not linked to an AONS	Details	Identify Format Metadata

Done

- [Tasks](#)
- [Risk Summary](#)
- [Reports](#)
- [Registries](#)
- [Repositories](#)
- [Tools](#)
- [Configuration](#)
- [Notification](#)
- [Logging](#)
- [About](#)

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). If yes, consider low risk and go to the end of Step 2. 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?
Q2. Is this file format and version referenced in any searched information resources?	Q2. The following hardware and software dependencies have been identified for effective rendering of this format using the original software ... (see Step 1 - Q7) Are these critical dependencies available to you?
Q3. Is there a known support end date for this format version? If yes, how many years to that support end date?	Q3. The following alternative software options have been identified for safe and effective rendering ... (see Step 1 - Q8) How many of these alternative rendering options are available to you?
Q4. How many years since this version was released?	Q4. For the alternative rendering options, the following critical dependencies have been identified ... (see Step 1 - Q9) Are these critical dependencies available to you?
Q5. How many new versions have been released since then?	Q5. Do you have any other alternative means of providing safe and 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? If unknown, state "Unknown".	Q6. Overall, how many access options are effectively available to you (i.e. how many can you make work), including the original rendering software? If none – consider access lost. If one, consider high risk.
Q7. Are there critical hardware and software dependencies for effective use of the original rendering software? What are they? If unknown, state "Unknown".	Q7. Do you have any other information that would exacerbate or mitigate the level of technical obsolescence risk? (i.e. information which might indicate a change in access to this format).
Q8. How many alternative software options for safe and effective rendering can be identified? What are they? If unknown, state "Unknown".	
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".	

AONS II - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost:8080/aons/gui/aons2Workflow.html?_flowExecutionKey=_c968A6150-385C-362B-714C-80ED825EC/ Go

Firefox Help Firefox Support Plug-in FAQ Google AONS II

[\[Help\]](#)

Q2. Is this file format and version referenced in any searched information resources? ☐

[\[Help\]](#)

Q3. Is there a known support end date for this format version? ☐

If yes, how many years to that support end date?

[\[Help\]](#)

Q4. How many years since this version was released? ☐

[\[Help\]](#)

Q5. How many new versions have been released since then? ☐

[\[Help\]](#)

Q6. Is the original primary rendering software for this format version identified?

Done

AONS II - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost:8080/aons/gui/aons2Workflow.html?_flowExecutionKey=_c3A46F959-7696-0140-7261-1D797211F1

Firefox Help Firefox Support Plug-in FAQ Google AONS II

Global Risk Summary as of 2007/10/04 11:08

The global risk assessment covers the following repositories. Other repositories may have been left out if they don't have a repository scan performed.

- ♦ Test File System [view details](#) [view 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

Done

- ♦ [Tasks](#)
- ♦ [Risk Summary](#)
- ♦ [Reports](#)
- ♦ [Registries](#)
- ♦ [Repositories](#)
- ♦ [Tools](#)
- ♦ [Configuration](#)
- ♦ [Notification](#)
- ♦ [Logging](#)
- ♦ [About](#)

Future Work – ‘Nirvana’ ?

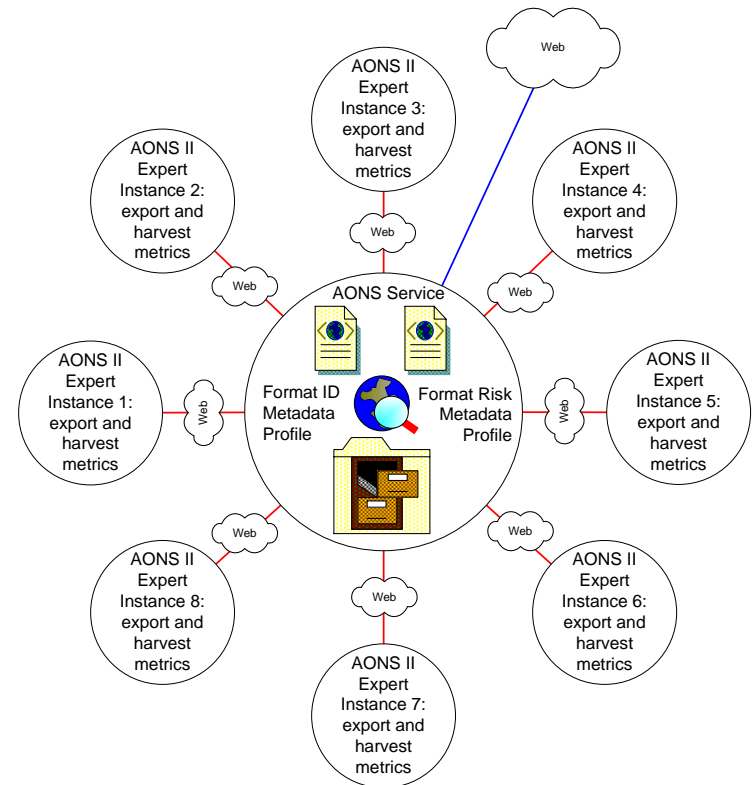
The current work is delivering a Java based, repository/registry agnostic tool which can be deployed in any Service Orientated Architecture. Therefore these foundations could be amenable for future development improvements.

Developing the AONS tool into a central web service based on feedback from multiple AONS expert repositories could:

- ▶ Provide machine- and human-harvestable Format Identification metadata profiles for file formats (new web service); and
- ▶ Provide machine- and human-harvestable Format Risk profiles for file formats (new web service).

If this were to occur, future development work on the AONS software might also include:

- ▶ Development of automated risk workflow; and
- ▶ Development of export functionality to a central web service.



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.

End



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/>