Infrastructure and Services for Digital Collections:

Automated Obsolescence Notification

Adrian Burton

Leader,
Australian Partnership for Sustainable Repositories (APSR)

DigCCurr 2007
Chapel Hill, April 2007

APSR Partners

The Australian National University

The University of Sydney

The University of Queensland

The National Library of Australia

The Australian Partnership for Advanced Computing

Supported by:

An Innovative Action Plan for the Future

The Systemic Infrastructure Initiative

Obsolescence Notification

Automated Obsolescence Notification System (AONS)

Aim: To make collection managers aware of files in their collections that require preservation action

NATIONAL
LIBRARY OF AUSTRALIA

AONS

AONS

AONS is a loosely-coupled group of services to **identify, notify, and evaluate** the risk of file format **obsolescence**.

It is particularly designed for large repositories that are:

- Long-lived

- Heterogeneous

- Where creators are not responsible for stewardship

- Without options for preservation on ingest

AONS Core Functions

- **Format Information**
- **Repository Information**
- **Notification**
- **Risk Analysis**

AONS Core Functions

Format Info

- GDFR
- the national archives
- LOC

- **Format Information**
- **Repository Information**
- **Notification**
- **Risk Analysis**

AONS Core Functions

Format Information

Repository Information

Notification

Risk Analysis

Repositories

Format Report

Repository

AONS Core Functions

- Format Information
- Repository Information
- Notification
- Risk Analysis

Repositories

Format Report

Repository

Format Info

GDFR

the national archives

LOC
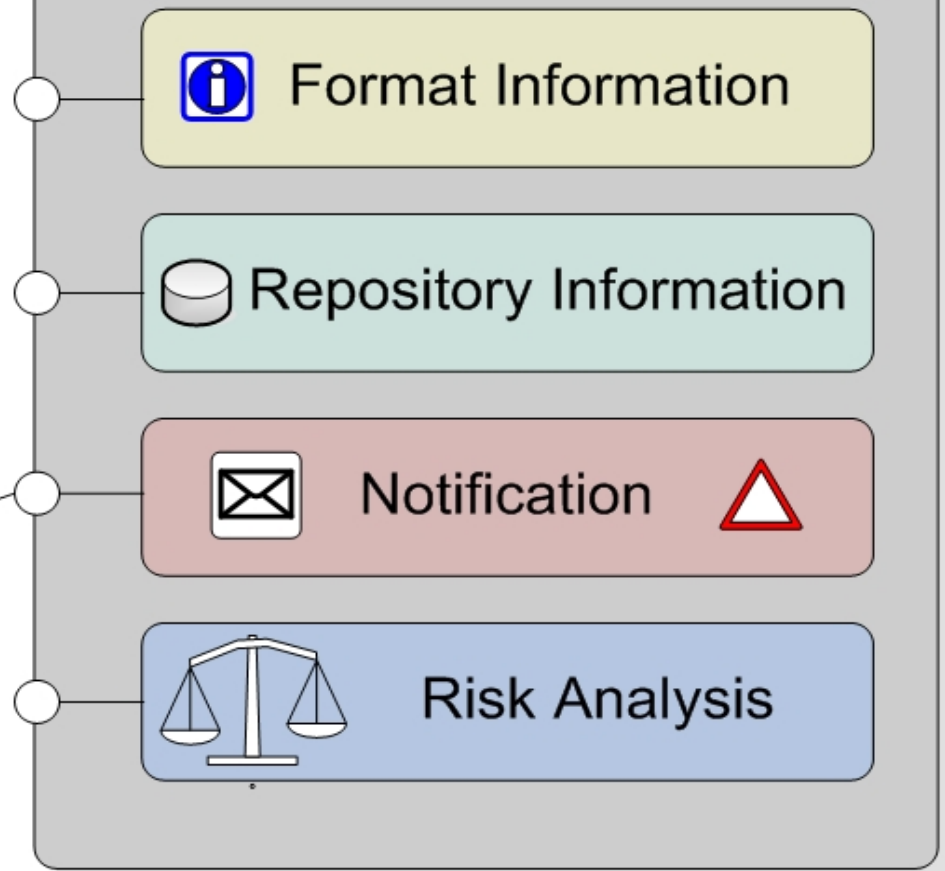
Repositories
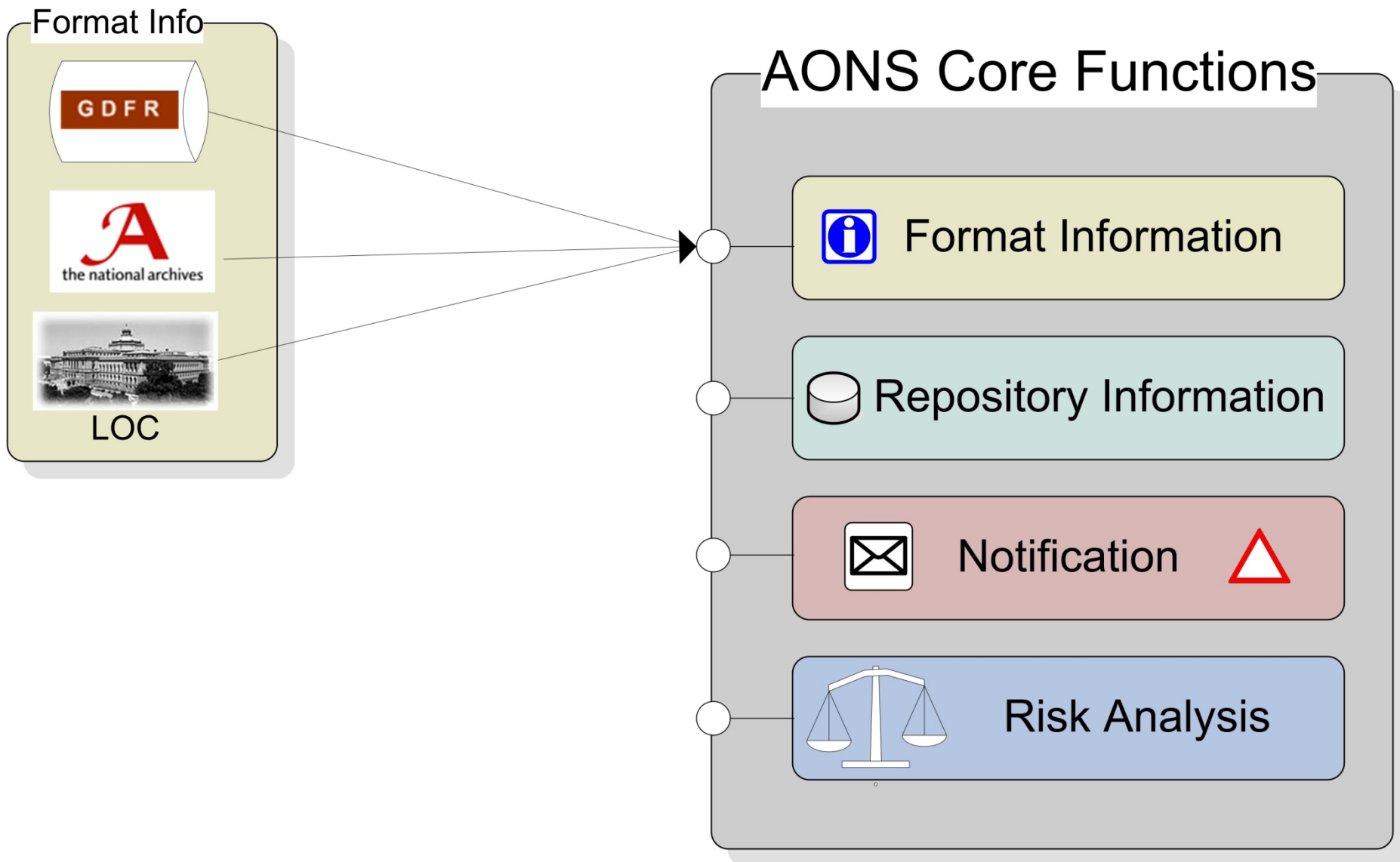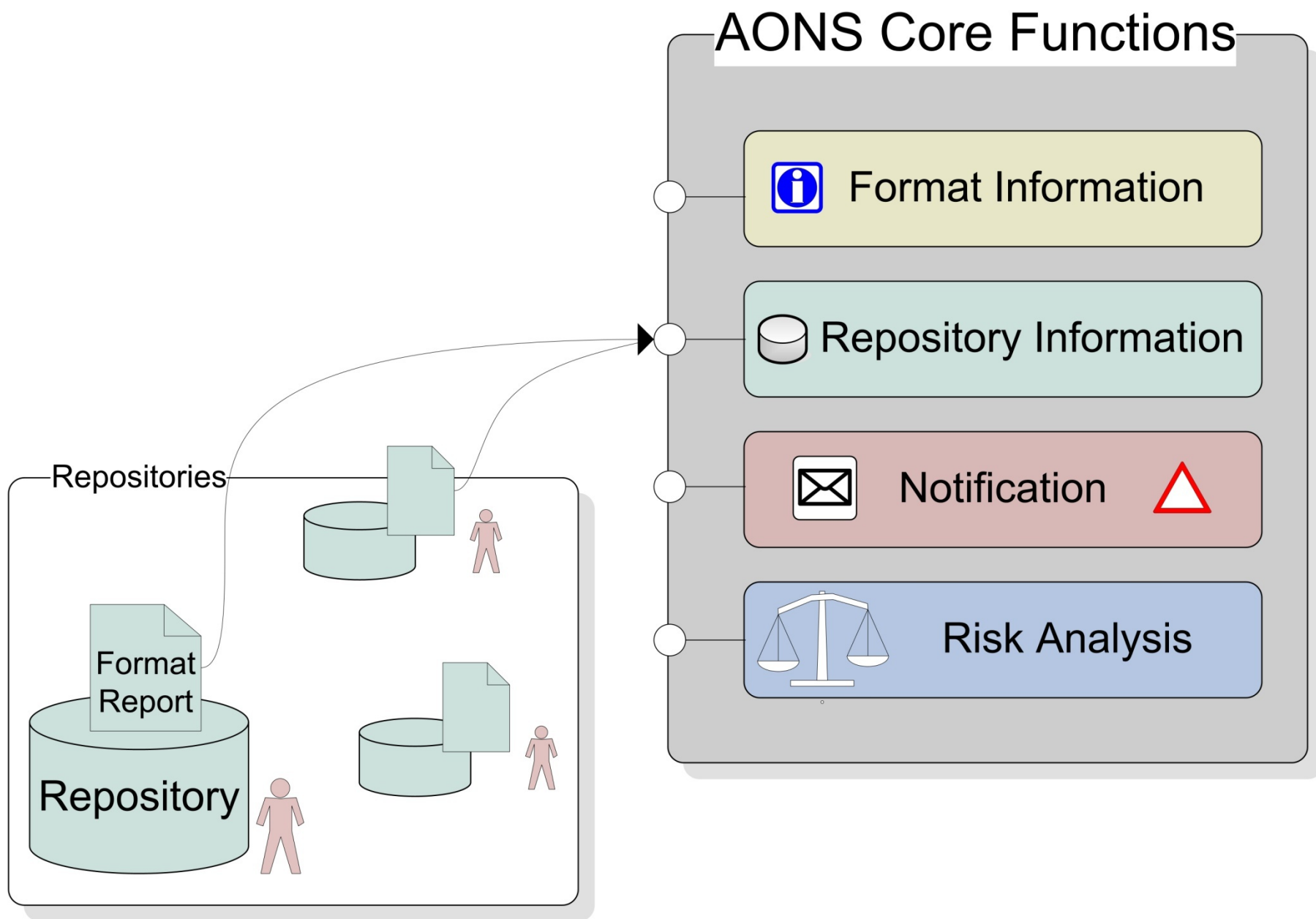
Format
Report

Repository

AONS Core Functions

Format Information

Repository Information

Notification
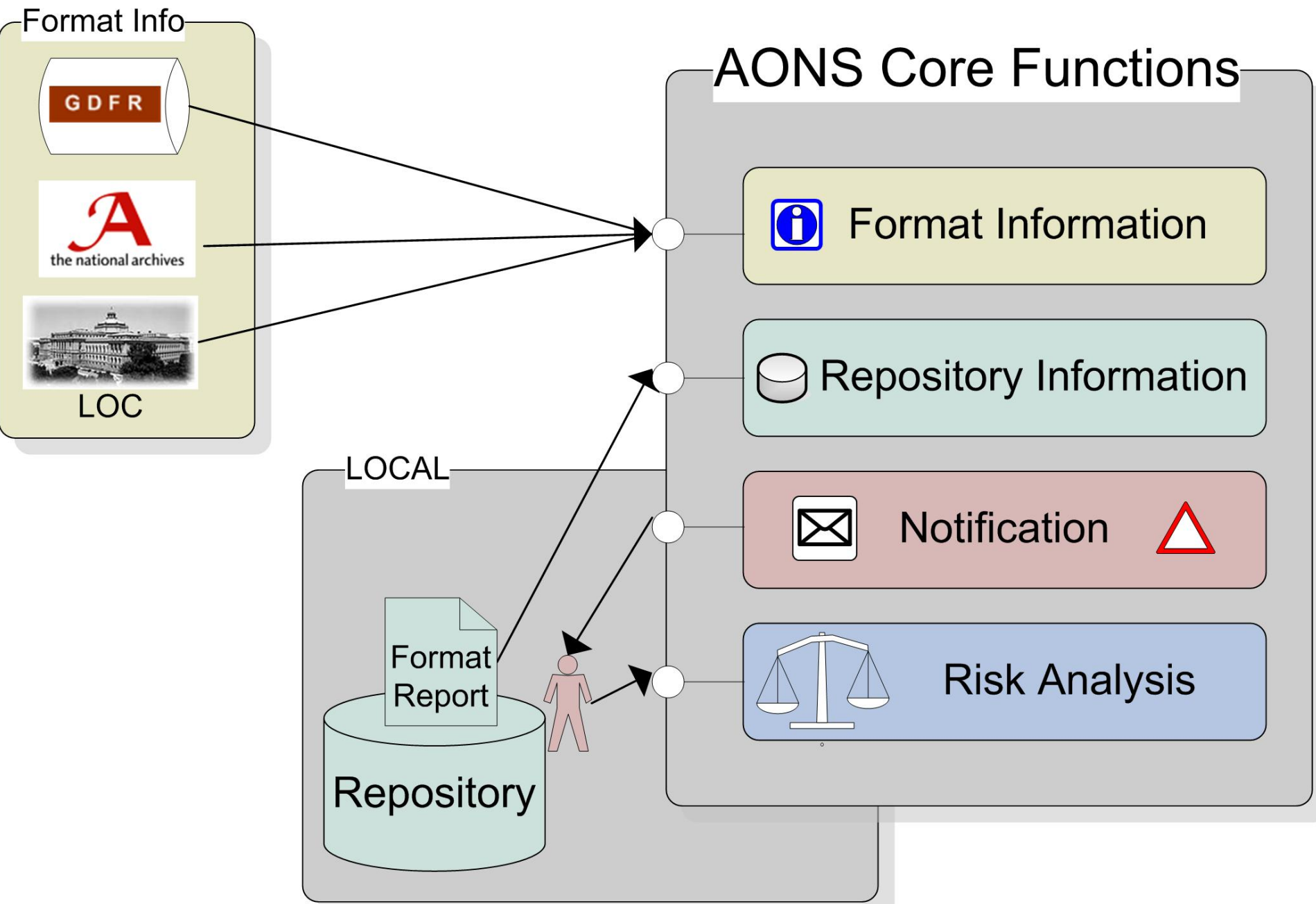
Risk Analysis

Project Background

- 2005: PANIC architecture*

- 2006: AONS 1, proof of concept system

- 2007: AONS 2, enterprise software, risk analysis, and pilot service

*J.Hunter, S.Choudhury, "PANIC – An Integrated Approach to the Preservation of Complex Digital Objects using Semantic Web Services", International Journal on Digital Libraries: Special Issue on Complex Digital Objects, January 2006

Format Information

- PRONOM
- LOC – LCDFW (scraper)
- GDFR (Alpha 1)
- DCC RIF? NGDA?......

AONS Core Functions

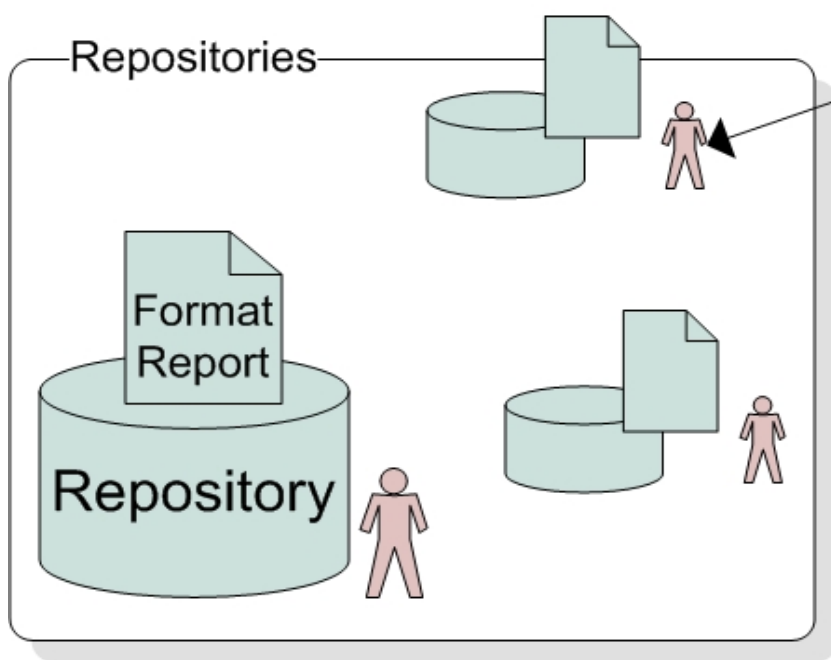- Format Information
- Repository Information
- Notification
- Risk Analysis
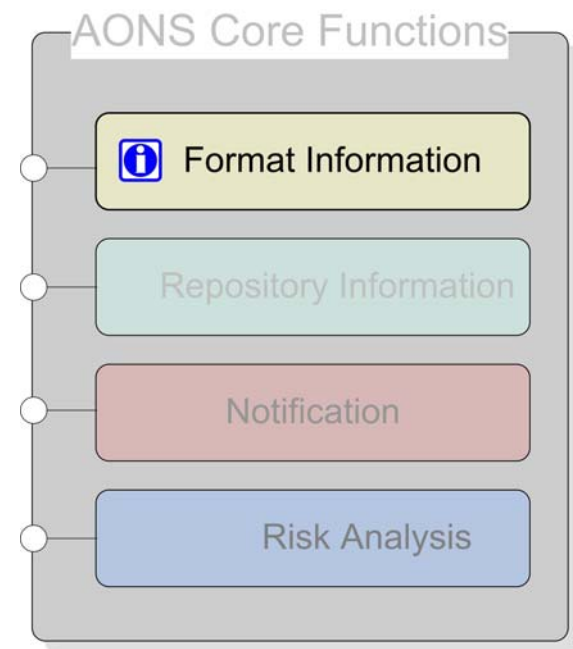
ISSUES

- Standard ways to represent format information
- Unique ID's for formats
- Metric for measuring obsolescence
- Ongoing relationships with information providers

Repository Information

- Managers self-register
- Repositories expose reports
- AONS "gets" the reports

AONS Core Functions

- Format Information
- Repository Information
- Notification
- Risk Analysis
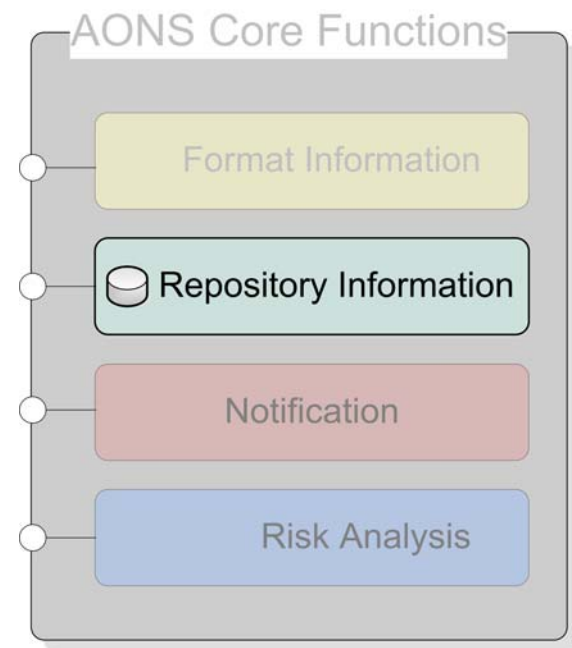
ISSUES

- authorisation/ authentication for registration
- scope of the service
- scaling to meet the needs of (eg) Internet Archive
- third party repository information aggregators

Standard XML Format Report

```xml
<collection>
      <format name="name1" version="v1" puid="fmt1" count="200"/>
      <format name="name2" version="v2" puid="fmt2" count="4000"/>
      <format name="name3" version="v3" puid="fmt3" count="10"/>
      <format name="name4" version="v4" puid="fmt4" count="300"/>
</collection>
```
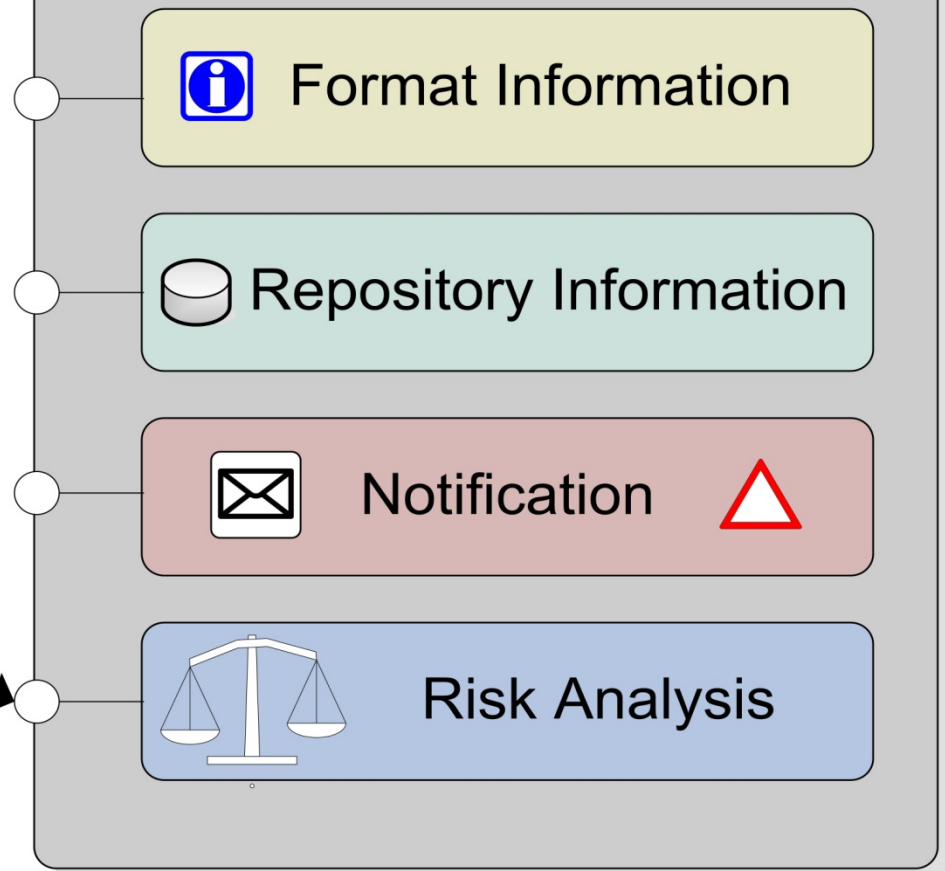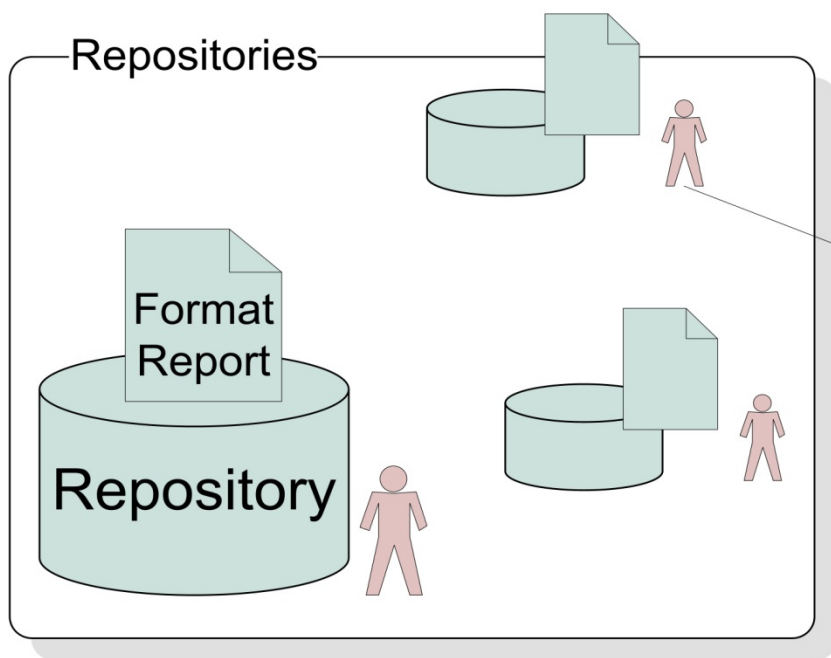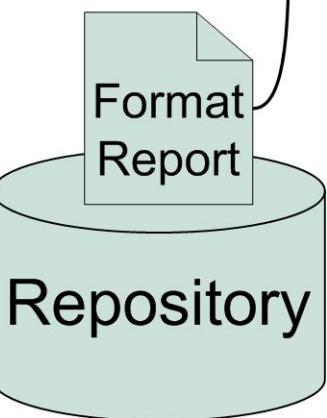
Notification

AONS Core Functions
- Format Information
- Repository Information
- ✉ Notification ⚠
- Risk Analysis

- Email/ RSS....
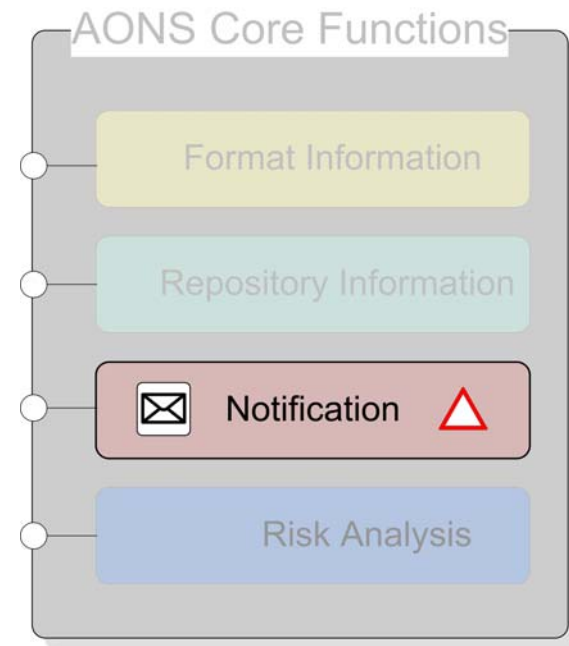
ISSUES

- frequency
- granularity of detail (perhaps just a link)

Risk Analysis

- INFORM/ Deb Woodyard
- General vs individual risk
- Stateful web interface (My AONS)

AONS Core Functions

- Format Information
- Repository Information
- Notification
- Risk Analysis
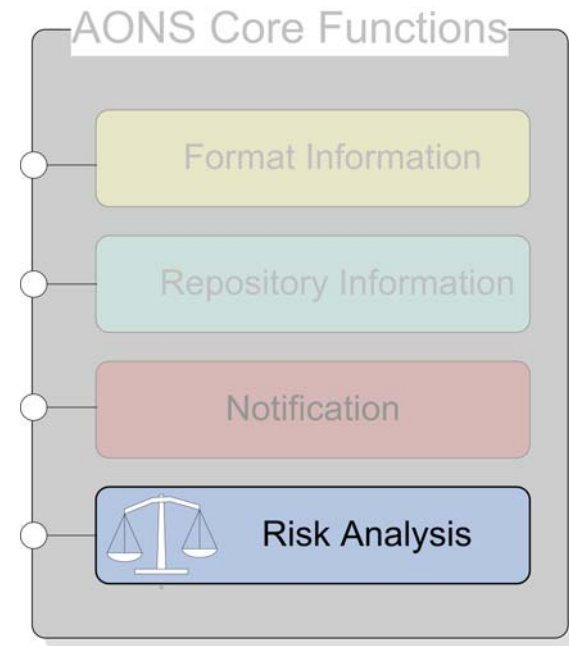
ISSUES

- what digital curators need to know to take action
- testing the metrics
- community risk may be supplied upstream
- other applications of the module

Obsolescence Risk Assessment Flowchart

Score
Add the score to the starting risk of 10

Result

Determine likely creator / vendor support period
- Version support end date known?
  - Y → Years to end date?
    - <= 0 → 0
    - 1-2 → -2
    - 3-5 → -4
    - 5+ → -6
  - N → Version release date known?
    - Y → Years since release date?
      - 7+ → -6
      - 5-7 → -4
      - 3-5 → -2
      - < 2 → 0
    - N ↓

Determine number of new versions of format available
- New version(s) available?
  - Y → Number of later versions
    - 2+ → 0
    - 2 → -2
    - 1 → -4
  - N / Unk → -2

Determine number of view paths available to render format (available support)
- Number of view paths available?
  - 0-1 → 0
  - 2-3 → -2
  - 3-5 → -4
  - 5+ → -6

+10

- 8 - 10 → Very High Risk of Obsolescence
- 6 - 7 → High Risk of Obsolescence
- 3 - 5 → Medium Risk of Obsolescence
- <3 → Low Risk of Obsolescence

AONS

Obsolescence Format Summary

Repository Name: *All*

Repositories contained within:

- RepositoryX
    - RepositoryA
    - RepositoryB
    - RepositoryC
- RepositoryY
- RepositoryZ

Current View of Repository All:

Name	Version	Last Modified	Before Workflow Risk (Community)	Before Workflow Risk (Contextual)	Before Workflow Risk (Net)	After Workflow Risk (Community)	After Workflow Risk (Contextual)	After Workflow Risk (Net)	
AVI	1.1	2007/03/12 13:40	5	10	15	5	10	15	Review
Broadcast Wave	1.1	2007/03/12 13:40	5	10	15	5	10	15	Review
OpenOffice Draw	1.1	2007/03/12 13:40	5	10	15	5	10	15	Review
Windows Media Video	1.1	2007/03/12 13:40	5	10	15	5	10	15	Review
Plain Text file	1.1	2007/03/12 13:40	5	10	15	5	10	15	Review
TIFF		2007/03/12 13:40	5	10	15	5	10	15	Review

AONS Design

- Loosely coupled SOA
- target independent
- Modular
- Robust Java enterprise software

PANIC

http://www.itee.uq.edu.au/~eresearch/projects/panic/index.html
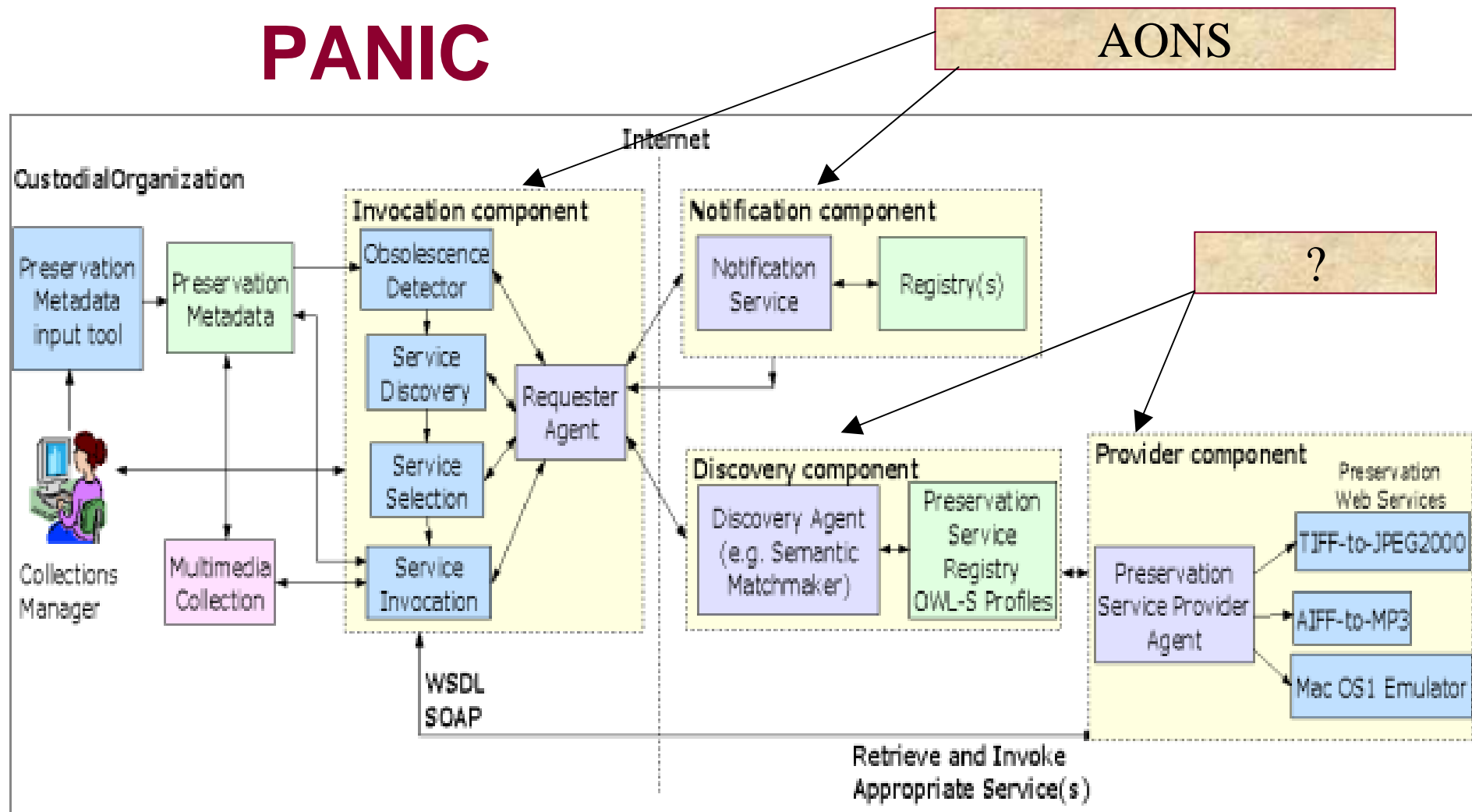
Other Applications

- metadata schema obsolescence

- QA notification

- ….

Data Set Discovery   Service Registry

Dissemination   Curation

Submission

Data Centre   Public Data

Data Store   Reference Set

Repository

| Data Discovery Services | Services Registry | Schema Registry | Identifier Resolution | Etc... |

| Submission Services | Dissemination Services | Curation Services | Etc.... |

| Data Centre | Repository | Reference Collection | Public Data | Etc... |

```
                        ┌──────────┐
                        │   User   │
                        └──────────┘
                       ╱            ╲
              ┌──────────┐      ┌──────────────┐
              │  Portal  │──────│ Application  │
              └──────────┘      └──────────────┘
```

┌─────────────────────────┐ ┌───────────────┐ ┌───────────────┐
│   │ ┊   ┊ ┊   ┊
│   Institutional   │ ┊ Institutional ┊ ┊   ┊
│   Repository,   │ ┊ Repository,   ┊ ┊   ...   ┊
│   Data Centre,   │ ┊ Data Centre, ┊ ┊   ┊
│   Data Grid   │ ┊ Data Grid   ┊ ┊   ┊
│   │ ┊   ┊ ┊   ┊
└─────────────────────────┘ └───────────────┘ └───────────────┘
┌───┐
│   Systems and Networks   │
└───┘

Security (AAA) and Virtual Organisation (VO) Management Framework

User

Portal — Application

Institutional Repository, Data Centre, Data Grid

Institutional Repository, Data Centre, Data Grid

...

Systems and Networks

Info_Env_01.pdf

User

Portal — Application

Security (AAA) and Virtual Organisation (VO) Management Framework

Federated Data Access

Repository Interface Services

Institutional Repository, Data Centre, Data Grid

Institutional Repository, Data Centre, Data Grid

...

Systems and Networks

Info_Env_01.pdf

User

Portal — Application

Security (AAA) and Virtual Organisation (VO) Management Framework

Human & Institut. Resources

Federated Data Access

Repository Interface Services

Institutional Repository, Data Centre, Data Grid

Institutional Repository, Data Centre, Data Grid

· · ·

Systems and Networks

'Meta' Registries & Services

Info_Env_01.pdf

User

Portal — **Application**

A Security (AAA) and Virtual Organisation (VO) Management Framework

A1 AAA Policy-Rights Registry

E Human & Institut. Resources

E1 Education & Training

E2 Expertise & Advice

E3 Standards Activities

E4 ICT Policy Frameworks

...

B Federated Data Access

Discovery Services

B1 Item-level

B2 Collection-level

B3 Workflow Service

B4 Rep. Statistics Service

B5 Services Registry

...

C Repository Interface Services

Repository Services

C1 Submission

C2 Dissemination

C3 Reposition Service

C4 Format & Schema Obsolescence Notification Service

C5 Quality Assurance Service

...

F Public API

| Ingest | Local Stats. | Access | Local Discovery |

Presentation

Local API

Local Management

Storage API

| Data | Metadata |

Institutional Repository, Data Centre, Data Grid

...

Systems and Networks

D 'Meta' Registries & Services

D1 Identifier Service

D2 Schema Registry

D3 Format Registry

D4 Workflow Registry

...

Info_Env_01.pdf