Part 2: Current State of OAR

Interoperability
The COAR Current State of
Open Access Repository Interoperability (2012)

Intended Audience: Institutions, and repository managers, operating at different points in terms of infrastructure, resources, and institutional support

- For institutions new to OA and repositories: Roadmap will provide guidance in getting started – providing ideas on the types of services repositories can offer
  - Which interoperability standards and protocols are necessary to implement in order to provide those services

- For institutions and repository managers already involved in OA and repositories: ideas for additional functionality, services possible to provide
2012 – 2013 Interoperability
Workplan

- State of Open Access Repository Interoperability – October 2012

- Future directions – First Quarter 2013??
Outline of the State of Open Access
Repository Interoperability

- Section I – Introduction and Background Information:
  general overview of the interoperability landscape, identifying main challenges, current state and emerging areas.

- Section 2 – Description of current standards/guidelines/projects
  a) Description – Information about specific interoperability guidelines and standards referenced in the roadmap. Each item summarized in a consistent way.
  b) Applications: concrete examples of services or applications that have been developed using the standards/guidelines/projects.
  c) Additional resources for further readings, particularly focusing on explanatory texts and other useful information.
Areas of Focus for OA Repository Interoperability Initiatives

- Metadata Harvesting
- Managing Compound Objects
- Persistent Identifiers
- Repository Networks
- Usage Statistics
- Author Identification
- Cross-System Content Transfer

OA Repository Interoperability

Confederation of Open Access Repositories

COAR
## Specific Interoperability Initiatives

<table>
<thead>
<tr>
<th>Areas</th>
<th>Associated Initiatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata Harvesting</td>
<td>OAI-PMH</td>
</tr>
<tr>
<td>Repository Networks</td>
<td>DRIVER</td>
</tr>
<tr>
<td></td>
<td>OpenAIRE</td>
</tr>
<tr>
<td></td>
<td>National repository networks</td>
</tr>
<tr>
<td></td>
<td>UK RepositoryNet+</td>
</tr>
<tr>
<td>Usage Statistics</td>
<td>COUNTER</td>
</tr>
<tr>
<td></td>
<td>KE Usage Statistics</td>
</tr>
<tr>
<td></td>
<td>OA-Statistik</td>
</tr>
<tr>
<td></td>
<td>PIRUS/IRUS-UK</td>
</tr>
<tr>
<td></td>
<td>SURE</td>
</tr>
<tr>
<td>Cross-System Content</td>
<td>SWORD</td>
</tr>
<tr>
<td>Transfer</td>
<td>Open Access Repository Junction</td>
</tr>
<tr>
<td></td>
<td>CRIS-OAR Metadata Interoperability</td>
</tr>
</tbody>
</table>
## Specific Interoperability Initiatives

<table>
<thead>
<tr>
<th>Areas</th>
<th>Associated Initiatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author Identification</td>
<td>AuthorClaim, ORCID</td>
</tr>
<tr>
<td>Persistent Identifiers</td>
<td>PersID, Handle System, DataCite</td>
</tr>
<tr>
<td>Managing Compound Objects</td>
<td>OAI-ORE</td>
</tr>
</tbody>
</table>
1. Metadata Harvesting

What is it?

- Repositories expose their metadata to allow other third-party systems to access and “collect” selected metadata records.
- The digital objects remain at the host institution, but metadata from multiple systems is aggregated together in a separate repository.

Associated initiatives:

- Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)
Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)

- The entry point for repository interoperability
- Relies on Dublin Core metadata
- Formed the basis for many later initiatives
- Supported by all major repository platforms including DSpace, ePrints, Fedora, Invenio, ContentDM, and Greenstone
2. Repository Networks

What are they? A commonly-adopted strategy for building national and regional repository networks:

- Use OAI-PMH to harvest metadata records from repositories within the network
- Present contents of these repositories through a single interface
- Requires Dublin Core
2. Repository Networks

Associated initiatives:

- DRIVER
- OpenAIRE
- UK RepositoryNet+
DRIVER

- Initiative to create a pan-European infrastructure for digital repositories
- Guidelines focus on “collection”, metadata, implementation of OAI-PMH, vocabularies
- Mandatory and recommended elements

DRIVER was the first pan-European infrastructure for digital repositories.

Area: Repository Networks
Geographic Focus: Europe
Current status: no longer maintained

Sponsoring Organization: DRIVER consortium

Other Projects in This Area:
- OpenAIRE
OpenAIRE

- Creates an infrastructure for European research outputs
- Designed to provide guidance for local repository managers to comply with the technical requirements of the European Commission OA policies

Towards Repository Interoperability
Berlin 10 Workshop – 6 November 2012
UK RepositoryNet+

1. to provide a socio-technical infrastructure to support deposit, curation & exposure of OA research literature

2. Gap analysis – what’s missing?

3. Provide new repository shared services

http://www.repositorynet.ac.uk/

The RepNet project

Area: Repository Networks
Geographic Focus: United Kingdom, although resulting services may be applied anywhere in the world

Current status: running (until Mar 2013)
Once the original project stage has been completed, RepNet is planned to evolve into a service for the UK repository network

Sponsoring Organization: EDINA (JISC-funded project)

Other Projects in This Area:
- DRIVER
- OA-RJ
- OpenAIRE
- PIRUS2
- SWORD
3. Usage Statistics

What is it?

- Mechanisms for collecting, combining, and compiling usage statistics of various kinds across multiple systems within a network
- Often include statistics from both repositories and open access journal publishers
- Many initiatives use OAI-PMH to harvest data
3. Usage Statistics

Associated Initiatives

- KE Usage Statistics
- OA-Statistik
- PIRUS/IRUS-UK
- SURE
KE Usage Statistics

Guidelines for enabling collection, combination and comparison of usage of OA content in repositories & OA journals

Uses OAI-PMH or SUSHI

Draws on best practices and lessons learned from four national partners – DEFF (Denmark), DFG (Germany), JISC (UK), SURF (Netherlands)

KE Usage Statistics Guidelines


KE Usage Statistics Guidelines describe a metadata format for data usage to be transferred from a repository to a central server.

Area: Usage statistics
Geographic Focus: Global

Current status: currently maintained (formally accomplished)

Sponsoring Organization: Knowledge Exchange Partners (DEFF, DFG, JISC, SURF)

Other Projects in this Area:
- OA-Statistik
- PIRUS2
- SURE

Towards Repository Interoperability
Berlin 10 Workshop – 6 November 2012
4. Cross-System Content Transfer

What is it?

- Initiatives designed to transfer metadata and content from repositories to other types of systems.
- Three initiatives in particular were designed to support researchers – 2 are designed to simplify deposit processes; a third ties into research information systems.

Towards Repository Interoperability
Berlin 10 Workshop – 6 November 2012
4. Cross-System Content Transfer

Associated Initiatives

- SWORD
- Open Access Repository Junction
- CRIS-OAR Metadata Interoperability
SWORD

- Protocol that facilitates the deposit of resources from remote sources into repositories or other similar systems
- Allows researchers to deposit OA materials once and have those materials disseminated into multiple systems

SWORD
swordapp.org

SWORD provides a mechanism for authors to deposit a single article into multiple repositories.

Area: Cross-System Content Transfer
Geographic Focus: Global
Current status: currently maintained
Sponsoring Organization: JISC, UKOLN
Other Projects in This Area:
- OA-RJ
- UK RepositoryNet+

Towards Repository Interoperability
Berlin 10 Workshop – 6 November 2012
Open Access Repository Junction

- Supports automatic deposit of multi-authored, multi-institutional research articles into multiple repositories

OA-RJ aims to simplify the deposit workflow for articles headed to multiple repositories

Area: Cross-System Content Transfer
Geographic Focus: Global
Current status: under development

Sponsoring Organization: EDINA with support from JISC
Related Projects:
- CRIS-OAR Interoperability
- EJ ME
- SWORD
- UK RepositoryNet+

Towards Repository Interoperability
Berlin 10 Workshop – 6 November 2012
CRIS-OAR Metadata Interoperability

- Increase practical interoperability between Current Research Information Systems (CRIS) and Open Access Repositories (OARs)
- Focus was on defining a metadata exchange format for publication information with an associated common vocabulary
- 5 main entities: Publication, Person, Organisation, Event, Project

CRIS-OAR

CRIS-OAR aims to increase the interoperability between CRIS and repositories through metadata exchange

Area: Cross-System Content Transfer
Geographic Focus: Europe
Current status: project has been completed

Sponsoring Organization: Knowledge Exchange

Other Projects in this Area:
- OA-Rj
- SWORD
- UK RepositoryNet+

Towards Repository Interoperability
Berlin 10 Workshop – 6 November 2012
5. Author Identification

What is it?

- Designed to
  - Identify, disambiguate (which John Smith?)
  - Connect authors with their materials in multiple platforms and systems
- A Key component for interoperability
5. Author Identification

Associated Initiatives

- AuthorClaim
- ORCID
AuthorClaim

- Initiative for authors to “claim” their materials
- Authors register and system creates a profile
- The service is based on the REPEC Author Service

AuthorClaim
authorclaim.org

AuthorClaim aims to link scholars with records of the works they have produced.

Area: Author identification
Geographic Focus: Global

Current status: currently maintained; building-up phase

Sponsoring Organization: The development AuthorClaim was funded by an Open Society Institute grant to the ACIS project.

Other Projects in this Area:
- ORCID
ORCID

- Creates a persistent unique identifier for authors to use to associate with their research
- October 2012: authors can now register for their ORCID number

ORCID will create unique identifiers and author records for scholars that will be used to link scholars to their research outputs.

Area: Author identification
Geographic Focus: Global
Current status: in production

Sponsoring Organization: ORCID is an international not-for-profit membership organization for institutions.

Related Projects:
- AuthorClaim
6. Persistent Identifiers

What is it?

- Long-term address for digital objects

Associated Initiatives

- PersID
- Handle System
- DataCite
PersID

- Provides persistent identifiers using Uniform Resource Names (URNs) and National Bibliography Numbers (NBNs)
- Also provides an international infrastructure and framework for applying persistent IDs

Towards Repository Interoperability
Berlin 10 Workshop – 6 November 2012
Handle System

- Resolution services for unique and persistent identifiers of digital objects
- DOI – Digital Object Identifier, uses Handle System
- Dspace repositories implement the Handle System by default

The Handle System
handle.net

Handles are designed to provide unique and persistent identifiers of digital objects
Area: Persistent identification
Geographic Focus: Global
Current status: currently maintained
Sponsoring Organization: Corporation for National Research Initiatives (CNRI).
Other Projects in This Area:
- DataCite
- PersID
DataCite

- Assigns persistent identifiers to published datasets
- Contributes to the establishment of infrastructure(s) that supports simple and effective methods of data citation, discovery, and access.
7. Compound Digital Objects

What is it?

- Ways to move “bundles” of objects from one system to another

Associated Initiatives

- Open Archives Initiative – Object Reuse & Exchange
OAI-ORE

● Defines standards for the description and exchange of web resource aggregations

● Aggregations or compound digital objects – can combine distributed resources with multiple media types – text, images, video, etc.

OAI-ORE

http://www.openarchives.org/ore/

OAI-ORE defines the standards for the description and exchange of aggregations of compound objects.

Area: Managing compound objects
Geographic Focus: Global

Current status: currently maintained;
Current Version: Version 1.0 (2008) is currently in use.
Sponsoring Organization: Open Archive Initiative.

Other Projects in This Area:
- OAI-PMH
- SWORD
8. Linked Data (Emerging)

What is it?

- Connecting datasets with other types of research outputs

Associated Initiatives

- OpenAIREplus – new ways to connect publications and data
- SWORDv2 – dataset transfer among repositories
Part 3: Reflect, Apply, Discuss

eResearch, Open Science and Challenges for Repository Interoperability at Your Institution

Towards Repository Interoperability
Berlin 10 Workshop – 6 November 2012
eResearch Infrastructure at Your Institution

Content | Systems | People

- **Content**: what kinds of materials, resources, documents, digital objects, etc. make up your eResearch infrastructure?

- **Systems**: which systems are involved?
  - What interoperability initiatives are already in place or planned?

- **People**: who are key players -- departments, individuals, system owners?

Towards Repository Interoperability
Berlin 10 Workshop – 6 November 2012
Planning & Prioritization

- Short-term: what are the next steps for your institution?

- Long-term – 3 wishes –
  - If you could sponsor a new interoperability initiative, what would it be?
  - What problem, issue, challenge would you want to solve?
Feedback for us

1. One thing I learned today
2. One thing I liked about today’s workshop
3. One thing I would suggest doing differently in the future
Join COAR

http://www.coar-repositories.org/
Thank you

Eloy Rodrigues
Universidade do Minho
Portugal
eloy@sdum.uminho.pt

Abby Clobridge
Clobridge Consulting
United States of America
aclobridge@clobridgeconsulting.com