Building a Cohesive Repository Infrastructure for the UK

Balviar Notay – Senior Manager for Repository Shared Services
Bringing together key repository services to deliver a connected national infrastructure to support OA
Delivering Open Access Content to Institutions

1. RECEIVE DATA
Customise importers to support supplier format

2. PROCESS DATA
Transform supplier metadata to our format, exposing selected fields including embargo, grant and funders code

3. STORE DATA
Allow transfer of back catalogue when joining

4. SEND DATA
Support SWORD for delivery and email for notification

5. EXPOSE OA DATA
Expose OA Data: GUI and API access to catalogue enables search by author, repository and funder

Author

Publishers

Subject Repository

Institutional Repository

CRIS

Institutional Repository

Subject Repository
Offer to Supplier

Forward your content directly to IRs
Expose your OA content through GUI and API for browsing and downloading
Accept your metadata: metadata only, metadata and OA full-text, or metadata and embargoed full-text
Identify target IRs using affiliation in your metadata
Allow onward tracking of your content
Offer to Consumer

Direct Delivery of OA content
Get OA content from all data suppliers to IR SWORD endpoint
Get all content using one single metadata format
Importer scripts available for Eprints 3.3. & 3.2, DSpace 1.8

Direct Delivery of All content
As above (without DSpace) plus
» Agree to manage embargo periods
» Get OA and embargoed content

http://broker.edina.ac.uk/
Offer to Consumer

Browse
See all OA content from all data suppliers
Web GUI http://broker.edina.ac.uk/
Search by target repository and author

Harvest
Get all OA content from all data suppliers
Use OAI-PMH APIs http://broker.edina.ac.uk/cgi/oai2
Search by target repository, author and funder
Choose from multiple formats

Notification
Receive monthly email alerts of new content from all data suppliers for chosen repositories
Self-register at http://broker.edina.ac.uk/cgi/postcard_registration
Does anyone use the material in your repository?

Institutional Repository Usage Statistics (IRUS) UK

Ross MacIntyre, Mimas, The University of Manchester
June 2014
IRUS-UK: aims & objectives

Collect raw usage data from UK IRs for *all item types* within repositories
  » Downloads not record views
  » Not just articles

Process those raw data into COUNTER-compliant statistics

Return those statistics(+) back to the originating repositories for their own use

Give Jisc (and others) a wider picture of the overall use of UK repositories
  » demonstrate their value and place in the dissemination of scholarly outputs

Offer opportunities for benchmarking/profiling/reporting/

Act as an intermediary between UK repositories and other agencies
  » e.g. global central clearinghouse, national shared services, OpenAIRE
IRUS-UK: gathering data

Considered 2 scenarios for gathering data

» Push: ‘Tracker’ code
  › Whenever a download occurs the repository ‘pings’ the IRUS-UK server with details about the download

» Pull: OAI-PMH harvesting
  › When a download occurs the details of the event are stored on the local repository server
  › IRUS-UK periodically harvests the download data (using the OAI-PMH protocol)

Opted for the Push method – ‘Tracker’

» Just easier - but minimise data pushed
» Patches for Dspace (1.8.x and 3.x) and Plug-in for Eprints (3.3.x)
» Implementation guidelines for Fedora
» Same method used in other contexts too, eg Book downloads from OAPEN
IRUS-UK: Tracker strings ‘pinged’

The OpenURL key/value pairs

» url_ver=Z39.88-2004
» url_tim=2012-07-05T22%3A59%3A59Z
» req_id=urn%3Aip%3A86.15.47.114
» req_dat=Mozilla%2F5.0+(iPhone%3B+U%3B+CPU+iPhone+OS+5_1_1+like+Mac+OS+X%3B+en-us)+AppleWebKit%2F534.46.0+(KHTML%2C+like+Gecko)+CriOS%2F19.0.1084.60+Mobile%2F9B208+Safari%2F7534.48.3
» rft.artnum=oai%3Aeprints.hud.ac.uk%3A8795
» svc_format=application%2Fpdf
» rfr_id=eprints.hud.ac.uk
IRUS-UK: (daily) processing

Step 1: Perl script parses the logs
» Processes entries from recognised IRs
» Sorts and filters entries (incl COUNTER rules)
» Consolidates daily accesses for each item
» Outputs to intermediate file

Step 2: Perl script parses intermediate file
» Looks up each item in the IRUS DB
  › If item is unknown to the system add item with (most) metadata “unknown”
» Updates DB with new statistics (for both ‘known’ & ‘known unknowns’)

Step 3: Obtain “unknown” metadata
» For the ‘known unknowns’ retrieve metadata from Source IR (using OAI GetRecord)
» Updates the metadata to DB

In future will restate data at month-end (including additional filtering)
IRUS-UK: User survey

Used for: Regular reporting to management (86%) & Identifying trends and patterns in usage (73%).

» Eg “Evidence provides case for continued Repository support”

Value: Saves time collecting statistics (65%) & Enables reporting I was previously unable to do (65%)

» Best thing?: “COUNTER compliant reports”; “Fast and easy to use” & “General reporting tool outside of repository platform”

Would you hope to use IRUS-UK for benchmarking: Yes (84%)

» What types of benchmarks?: Previous year (95%) & peer institutions (95%)

Support: Case studies (78%) & Use cases (74%)

“Already a great service, but so much potential with this data”
COnnecting REpositories (CORE)

Petr Knoth
CORE (Connecting REpositories)
Knowledge Media institute
The Open University

@petrknnoth

Open Repositories 2014
Helsinki, Finland
The mission of CORE

Aggregate all open access content distributed across different repository systems worldwide, enrich this content and provide access to it through a set of services ...
The CORE aggregator

Content: 20M+ records, 600+ repositories, 1.8M+ full-texts
Processing pipeline

- Metadata download, extraction and cleaning
- Full-text harvesting
- Text-extraction
- Language detection
- Extraction of citation references from text
- Detection of citation reference targets
- Identification of related content
- Detection of duplicate items
- Parsing of author names
- Indexing
SHERPA Services
RoMEO, JULIET, FACT and OpenDOAR

Bill Hubbard
Centre for Research Communications
University of Nottingham
Outline

- What we do
- Value proposition and benefits we offer
- Example use case
RoMEO
www.sherpa.ac.uk/romeo

Global service of authors’ rights for using repositories, giving details journal by journal.
JULIET
www.sherpa.ac.uk/juliet
Registry of policies on Open Access from research funders worldwide.
FACT

www.sherpa.ac.uk/fact

Advice to UK authors on compliance with funder’s policies in their journal of choice.
OpenDOAR

www.opendoar.org

The world’s authoritative and quality-assured directory of open access repositories.
Value & Benefits
Value & Benefits

- 3 Levels
  - Service
  - Operational
  - Strategic
Service Level
Service Level
Benefits

- Efficiency gains at sector level
- Single points of information
- Quick access to policies and information
- Allows quick comparison of policies and information
- Taking some of the pressure off the OA community
Operational Level
Operational Level

Maintain Central Datasets
Operational Level

Maintain Central Datasets

1. SHERPA RoMEO
   Data on archiving permissions and rights given to authors by journal publishers

2. SHERPA JULIET
   Data on research funder mandates and requirements

3. OpenDOAR
   Data on global institutional repository growth
Some figures

RoMEO
Rights data on over 18,000 journals

JULIET
Mandate and policy data on 133 funders

OpenDOAR
2,600 IR listings
Strategic Level
Strategic Level

Role in developing future OA support services

- SHERPA Services data used to prime system development in response to emerging needs
Use Case
FACT
www.sherpa.ac.uk/fact

Uses data from RoMEO and JULIET
SHERPA Funders’ & Authors’ Compliance Tool (FACT)

01 April 2013; Research Councils UK’s (RCUK) new OA policy came into effect

- The policy required that all peer reviewed research papers, resulting from research funded by the Research Councils, will be published in journals under a Creative Commons Attribution (CC-BY) License or otherwise made available as OA via a repository.
- On the same date, the Wellcome Trust also required that articles with which it is associated – and for which an Article Processing Charge (APC) has been levied – must also be published under a CC-BY license.
SHERPA Funders’ & Authors’ Compliance Tool (FACT)

SHERPA FACT uses data from RoMEO API and JULIET API to:
- Cross-reference the information held in RoMEO and in JULIET on both Funders’ and Publishers’ policies
- Give the author clear information as to whether that journal or publisher offers publication or archiving rights compliant with their funders’ policy
- Note the level of OA fee payable, if any, and if available from the publisher site
- Give guidance to the author as to what action to take next to comply with their funders’ policy, customising guidance according to the stage of the author’s publication – pre-submission, accepted, published, etc
Step 01: Choose a funder and enter a journal name
FACT
www.sherpa.ac.uk/fact

Results: Compliance information and guidance

COMPLIANCE INFORMATION
Information on whether the selected journal's OA policy complies with the selected funder's OA policy

FUNDERS & AUTHORS COMPLIANCE TOOL
Helping you comply with research funders' policies on open access to publications

You can comply with your funder's policy:

✓ You can archive your article compliantly in Europe PMC/Pubmed Central

This journal has a compliant embargo of 6 months for funders that require archiving of articles.

You should self-archive your article within 6 months of publication.

Self-archiving
Please ensure that your article is deposited in Europe PMC/Pubmed Central within 6 months of publication.
To ensure compliance with your funder's policy:
- Confirm that the special arrangements apply before self-archiving.
- If funding agency rules apply, authors may post authors version to their relevant funding body's archive, 6 months after publication.
- Check SHERPA/RoMEO for more details

GUIDANCE FOR AUTHORS
Options on how to comply with funder OA policy

LAST UPDATED
Information on when entry was last updated

DECISION TRAIL
Decision trail detailing how decision was made
Contacts

- juliet@sherpa.ac.uk
- romeo@sherpa.ac.uk
- opendoar@sherpa.ac.uk
- fact@sherpa.ac.uk

@sherpaservices
Standards Implementation
RIOXX Application Profile and Guidelines http://rioxx.net/

- UK focused metadata initiative for research papers.
- Working with UK research funders (RCUK and HEFCE) to align key metadata requirements to support policy compliance
- Initial focus for RIOXX was to apply consistency for project ID (grant number) and funder name.

**Benefit:**
- Facilitates interoperability.
- Will support consistency for key metadata fields to facilitate the tracking of research outputs across scholarly systems.
- Will facilitate reporting from institutions to funders.
- Next version release due end of June 2014
- Software plug-ins for EPrints and Dspace. Repositories to implement these in 2014/2015
Standard Implementation

Vocabularies for Open Access (V4OA) http://v4oa.net

- Mandating the use of particular vocabularies for OA.
  - Embargoes (Derive from Licence)
  - Rights/Licences (LicenceRef)
  - Open access identifier (NISO - Free to read tag)
  - Article Processing Charges (APCs)
  - Versions (NISO)

- V4OA outputs will be incorporated in RIOXX.

- Benefit:
  - Allow improved metadata quality and consistency across key information systems.
  - Allow easy tracking of outputs across scholarly systems
  - Allow institutions and funders to capture and assess the nature and scale of Open Access ‘transactions’ across the scholarly landscape.
Contacts:

RIOXX : Paul Walk @ EDINA
p.walk@ed.ac.uk

V4OA : Sheridan Brown @ Chygrove Ltd
s.brown@keyperspectives.co.uk
Thank you.
Panel Session and Questions

- How do we align infrastructures (what are the pain points, what are the opportunities?)
- International use/re-use of services? How do we facilitate this?
- Further technical integration needed?
- Further automation needed?
- Are there other services that we can plug-into or similar support structures that we should be aware of?
- Gaps?
- Sustainability
- Other questions?