Rioxx 3: A Modernised Metadata Profile

Paul Walk

Director & Founder, Antleaf

Web: http://www.paulwalk.net
Email: paul@paulwalk.net

www.antleaf.com

What is Rioxx?

- A **metadata application profile** designed to be used in open-access repositories, primarily for exposing metadata to be harvested via the OAI-PMH protocol
  - In OAI-PMH, Rioxx is deployed as a "metadata format"

- A set of documentation, specifying the application profile, with examples of usage given in XML

The history of Rioxx (originally "RIOXX")

- **version 1** released in 2013
  - developed by UKOLN and Chygrove Ltd.
  - sponsored by RCUK, supported by HEFCE & Jisc
  - a metadata application profile for institutional repositories, with some particular features:
  - unambiguous identification of repository full text
  - explicit declaration of licenses
  - expression of APC status
  - use of persistent identifiers for authors (e.g. ORCID) and funders (e.g. DOI or ISNI).

- **version 2.0** (current version) released in 2015

- **version 3** in development

Main Use case in 2013:

Reporting on OA status to UK research funders

Why did we develop a new profile, rather than using an existing one?

- OAI-DC was not rich enough

- OpenAIRE was better but demanded project IDs be encoded in particular syntax not compatible with project IDs from UK Research Councils
  - OpenAIRE syntax
    - info:eu-repo/grantAgreement/Funder/FundingProgram/ProjectID/[Jurisdiction]/[ProjectName]/[ProjectAcronym]
  - RCUK syntax:
    - OpaqueProjectID/version

- There is room for more than one profile. This allows innovation, and mapping between them is possible

Principles of development

- **purpose driven**
  - designed to meet defined use-case(s)
  - solve problems directly & well, avoid 'feature creep'
  - only describe what can be authoritatively described

- **focussed on implementation**
  - has to be relatively easy to implement
  - 'shallow' structure
  - *the simplest thing that can possibly work*

- **treat machines as first-class citizens**
  - use PIDs / URIs where possible
  - use controlled vocabularies

- **open development**
  - public consultation
  - tested openly

radically open development!

The need for a new version

- The original use-case is no longer valid - in fact RCUK no longer exists!

- There has been interest, both in the UK **and internationally**, from service providers who think Rioxx is well-suited for **harvesting and aggregating** metadata about repository resources
  - There have been some suggestions from these service-providers for improvements to Rioxx

- Although the specific RCUK use case is defunct, there is an emergent general use-case that is very similar:

  - exposing metadata about Open Access resources in repositories to provide evidence of compliance with funder mandates

Research into attributes needed for reporting funder compliance (2022)

MORE+BRAINS

Requirement
Acknowledge UKRI funding
Date of deposit
Data Access Statement
ISSN
Article level PID (DOI, URN, Handle)
ORCID ID (all authors)
Is ORCID ID authenticated
Licence (non-proprietary format)
Preservation location (Portico etc)
Self-archiving policy (registered in

Sherpa-Romeo)
Citation data in I4OC (http://opencitations.net/)
PID for funders
PID for research performing orgs
PID for grant
PID for project
Final version or AAM

Rioxx version 3

- Version 3 is significantly different - this will be a <u>major</u> release.

  1. Structure:
    - Introduces a little more structure, so that the main Rioxx record can include references to multiple resources, such as:
      - author's accepted manuscript
      - published version of an article
      - supporting datasets
      - etc...

  2. PIDS:
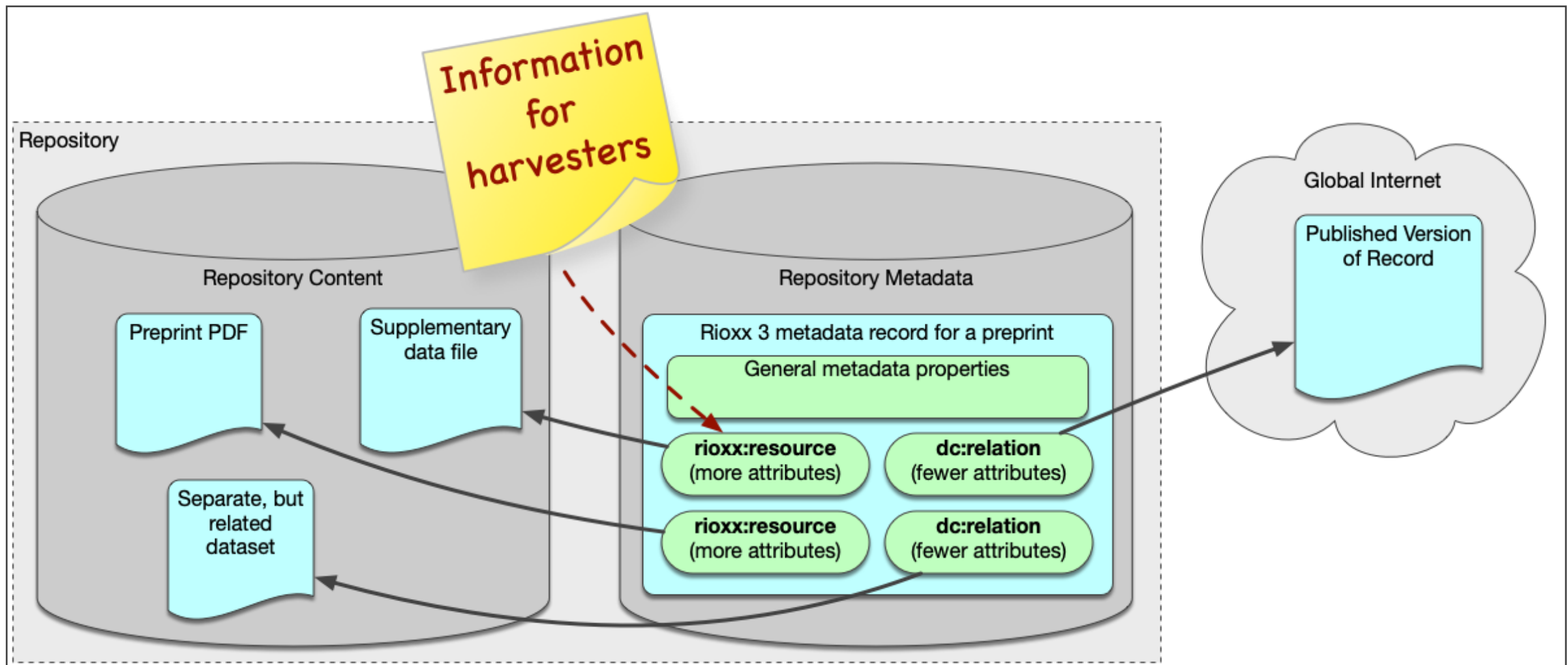   - Makes more use of PIDs (includes these wherever possible)

  3. Events:
    - Captures some life-cycle "event" dates, such as "resource_exposed_date" and "deposit_date" etc.

WORK IN PROGRESS

Structure

Implements a very light entity-relationship model

Structure: rioxterms:resource

- **coar_type** (Controlled vocabulary)

- **coar_version** (Controlled vocabulary)

- **deposit_date** (Date)

- **resource_exposed_date** (Date)

- **cite_as** (Normally a PID)

- **access_rights** (Controlled vocabulary)

- **license_ref** (URL)

- **rel** (Normally "item")

- **format** (Media type / MIME type)

The attributes can be detailed and authoritative because this resource is controlled or hosted by the same party that controls the overall Rioxx record

COAR Controlled vocabularies:

- Resource Types
- Access Rights
- Version Types

https://vocabularies.coar-repositories.org/

"rel" is recommended by signposting.org which aids systems to find resources in other systems by following HTTP link headers

Increased use of PIDs where appropriate

- Examples:

- As URI attribute of dc:publisher

```
<dc:publisher uri="https://isni.org/isni/000000040482455X">
    Public Library of Science
</dc:publisher>
```

- Use of multiple dc:identifier nested in author property

```
<rioxxterms:contributor>
        <dc:contributor>Bhopal, Kalwant</dc:contributor>
        <dc:identifier>https://orcid.org/0000-0003-3017-6595</dc:identifier>
        <dc:identifier>https://isni.org/isni/0000000038079210</dc:identifier>
        <dc:identifier>https://www.wikidata.org/wiki/Q61998297</dc:identifier>
</rioxxterms:contributor>
```

- Grant and funder information

```
<rioxxterms:grant
    funder_name="Wellcome Trust"
    funder_id="https://isni.org/isni/0000000404277672">
    https://doi.org/10.35802/218671
</rioxxterms:grant>
```

- Project identifier

```
<rioxxterms:project>
    https://handle.net/10378.1/1590366
</rioxxterms:project>
```

Capturing significant dates ("events")

- These are important for the "evidence of funder policy compliance" use case.

- *Properties* at the root level:

  - rioxxterms:record_public_release_date
    - when metadata about the resource was first made publicly visible

  - rioxxterms:publication_date
    - publication date of the resource in the form in which it would be cited
    - recommend the ISO 8601 format - e.g. 2011-02-23
    - when publication dates are expressed in terms of seasonal issues, then we use the corresponding month, e.g.
    - "Spring 2020" becomes 2020-04

  - dcterms:date_accepted
    - when the resource was accepted for publication

Capturing significant dates ("events")

- *Attributes* at the level of individual rioxxterms:resource

  - deposit_date
    - when this resource was first deposited, irrespective of embargoes

  - resource_exposed_date
    - when this related resource was made publicly available (which may be following an embargo period)

Current status & immediate schedule

- We will publish a second (hopefully final) release candidate in next 1-2 weeks

- Rioxx v3 will introduce some significant changes

- A (hopefully) final period for public comment

- Intend to publish Version 3.0 release in next 1-2 months

How is Rioxx managed and maintained?

- Rioxx is maintained by UKCORR and Antleaf Ltd.

  - Rioxx website:
    - **http://www.rioxx.net**

  - Rioxx sources (specification, website etc.):
    - **https://github.com/antleaf/rioxx**

- Governance is provided by UKCORR, which has appointed a Rioxx Governing Group