RIOXX: a Modern Metadata Application Profile

Paul Walk
Head of Technology Strategy and Planning, EDINA
p.walk@ed.ac.uk
@paulwalk

EDINA

what is RIOXX?

why build RIOXX?

- new **policies from RCUK and HEFCE** mandate that any journal article funded by research grants be made publicly accessible in a repository

- these policies require that universities make metadata about such papers easily discoverable

- the available metadata formats were inadequate
  - 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

particular concerns

- how to represent the **funder**
- how to represent the **project/grant**
- how to represent **unambiguous licenses**
- how to represent the **persistent identifier** of the item described
- provisions of identifier(s) pointing to related dataset(s)
- how to represent the **rights of use** of the item described

rioxx

the RIOXX metadata profile and guidelines

- an application profile using properties from 4 namespaces:
  - 11 properties from Dublin Core (dc and dcterms)
  - 2 properties from NISO Open Access Metadata and Indicators
  - 8 from a new namespace - ‘rioxxterms’

- constraints imposed through several controlled vocabularies

- **it has one purpose:** to provide a mechanism to help institutional repositories in the UK comply with the RCUK policy on open access.
  - **it is not designed to provide general interoperability!!**

- **Version 2.0 released in January 2015**

RESEARCH COUNCILS UK

UKOLN

EDINA

HIGHER EDUCATION FUNDING COUNCIL hefce FOR ENGLAND

Jisc

very rapid tour of some specific properties

dc:identifier

- identifies the open access item being described by the RIOXX metadata record.
  - regardless of where it is located

- recommended to identify the resource itself, not a ‘splash page’
  - this will not always be possible or desirable

- whatever it identifies, it **MUST** be an **HTTP URI**

- **Example:**

```
<dc:identifier>
 http://oro.open.ac.uk/2/1/LIBARTVICEprints.pdf
</dc:identifier>
```

dcterms:dateAccepted

- this **MUST** be provided
- is more precise than other possible dated events - such as ‘published’

rioxxterms:author & rioxxterms:contributor

- both of these accept an optional ‘ID’ attribute
  - this **MUST** be an HTTP URI
  - use of **ORCID is strongly recommended**

- all authors should be represented as individual *rioxxterms:author* properties

- the ‘first named author’ can be indicated with another optional attribute called, er…, **‘first-named-author’**

- rioxxterms:contributor is for other parties that are not authors but are credited with contributing in some way to the publication

- **Example:**

```
<rioxxterms:author id="http://orcid.org/
0000-0002-1395-3092">
    Lawson, Gerald
</rioxxterms:author>
```

rioxxterms:project

- this expresses funder and project_id in one, slightly more complex, property
- the use of global IDs, e.g. International Standard Name Identifier (ISNI) for funding organisations is recommended
- **Example:**

```
<rioxxterms:project
    funder_name="Engineering and Physical Sciences
Research Council"
    funder_id="http://isni.org/isni/0000000403948681"
>
    EP/K023195/1
</rioxxterms:project>
```

ali:license_ref

- adopted from NISO's *Open Access Metadata and Indicators*
- takes an HTTP URI and a start date
- the URI should identify a license
  - there is a need for a **'white list' of acceptable licenses**
- embargoes can be expressed this way, with a license identified to 'take effect' at some (possibly) future date
- **Example:**

```
<ali:license_ref start_date="2015-02-17">
 http://creativecommons.org/licenses/by/4.0
</ali:license_ref>
```

OpenAIRE Mapping

Crosswalk from RIOXX 2.0 to OpenAIRE 3.0

This crosswalk has been developed by the RIOXX team at EDINA in close collaboration with technical personnel at OpenAIRE. We would like to thank, in particular, Jochen Schirrwagen and Paolo Manghi for their assistance with this.

For guidance on the exact representation of properties within the OpenAIRE 3.0 metadata profile, please consult the documentation OpenAIRE Guidelines: For Literature Repositories

The terms **MUST**, **MUST NOT**, **REQUIRED**, **SHALL**, **SHALL NOT**, **SHOULD**, **SHOULD NOT**, **RECOMMENDED**, **MAY**, and **OPTIONAL** used in the table below should be interpreted as described in RFC 2119.

RIOXX 2.0	OpenAIRE 3.0	Implementation
ali:free_to_read *Zero or one*	**dc:rights** *One or more*	This **MUST** be included, if available
ali:license_ref *One or more*	**dc:rights** *One or more*	This **MUST** be included, if available
dc:coverage *Zero or one*	**dc:coverage** *Zero or more*	This **SHOULD** be included, if available
dc:description *Zero or more*	**dc:description** *zero or more*	This **MUST** be included, if available
dc:format *Zero or more*	**dc:format** *Zero or more*	This **SHOULD** be included, if available If provided in the RIOXX record, values will be in the form of MIME types
dc:identifier *Exactly one*	**dc:identifier** *One or more*	This **MUST** be included This value will be in the form of an HTTP URI

question:

how is RIOXX being developed?

answer: with ruthless pragmatism...

http://images.huffingtonpost.com/2014-05-27-oHOUSEOFCARDSPROMOSfacebook.jpg

principles (with an emphasis on pragmatism)

- **purpose driven**
  - designed to meet a **singe, focussed use-case**
  - solve one problem well, avoid ‘feature creep’

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

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

- **rapid development**
  - (relatively) short iterations

Manifesto for Agile Software Development

We are uncovering better ways of developing
software by doing it and helping others do it.
Through this work we have come to value:

Individuals and interactions over processes and tools
Working software over comprehensive documentation
Customer collaboration over contract negotiation
Responding to change over following a plan

That is, while there is value in the items on
the right, we value the items on the left more.

http://agilemanifesto.org

applying these principles to RIOXX development

- **Individuals and interactions** over processes and tools
  - we concentrated on what worked - & what made sense to the user/sponsor

- **Working software** over comprehensive documentation
  - an application profile is fundamentally a set of documentation!
  - however, RIOXX is *implemented* in software

- **Customer collaboration** over contract negotiation
  - we worked as closely with users as possible, and worked very openly

- **Responding to change** over following a plan
  - iterative - we developed RIOXX in short development cycles punctuated by review

open, community support

- engagement from software suppliers
- community feedback
- good practice starting to be identified and discussed here

work:

```
<base-url>?verb=ListRecords&metadataPrefix=rioxx
```

is actually returning records which the repository managers might not wish it
compatible'. The OAI-PMH standard states that:

> *Records should be included only for items from which the metadata format n*

I have no idea if this is a fault in the ePrints plugin, or if it is more to do with th
wonder if any repository managers - particularly those who have implemente
plugin, there may be emerging good practice in terms of its configuration anc

It may be that we can improve on those compliance rates by simply filtering w

In the spirit of open collaboration, please feel free to leave comments below!

Share this post:

30 Comments RIOXX Website

Recommend Share

Join the discussion…

Will Fyson · 8 days ago
One way of perhaps ensuring EPrints exposes r

Bram Luyten · 12 days ago
In the DSpace patch, items are only vis
https://github.com/atmire/RIOX...
Reply · Share ›

Paul Walk Mod Bram Luyten · 10
Hi Bram,
thanks for the comment and poi
RIOXX record published from a
Edit · Reply · Share ›

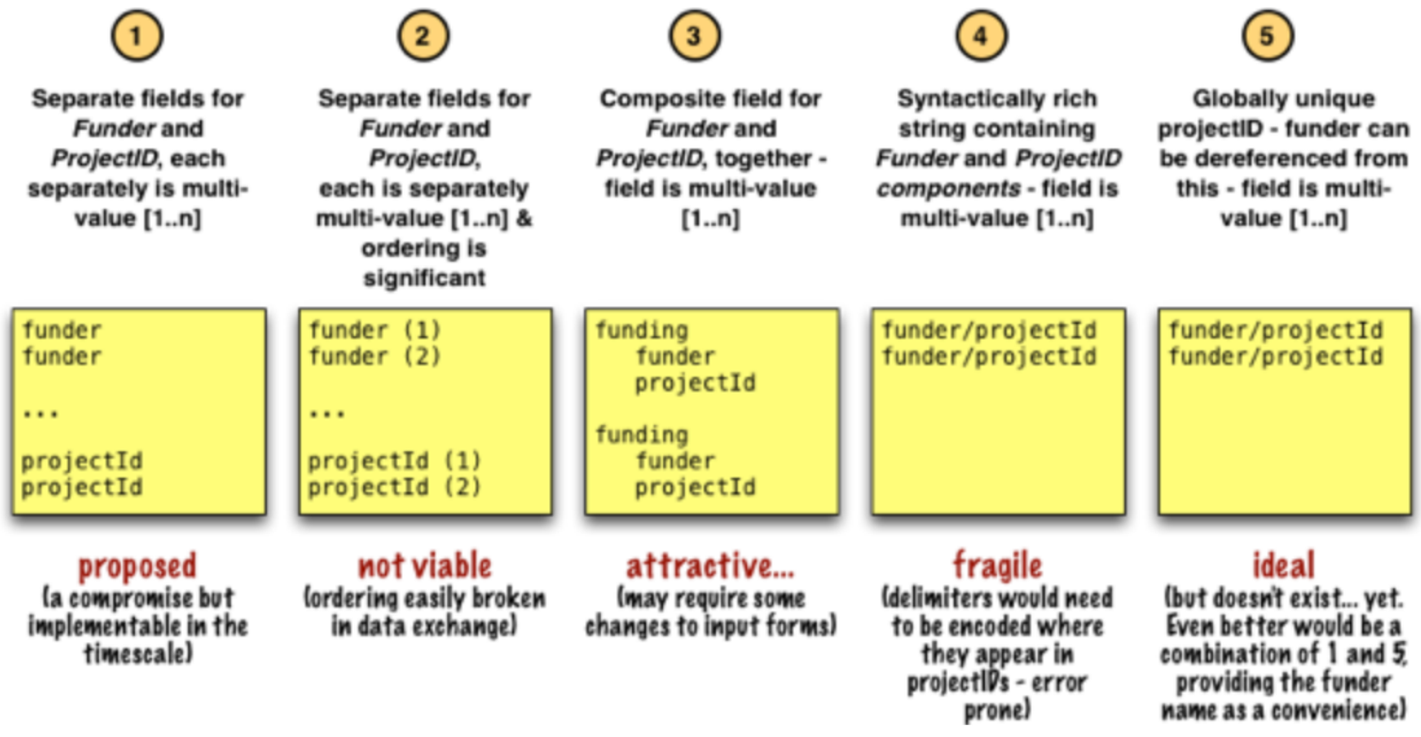
Bram Luyten Paul Walk ·
I don't have an accurate
institutions have taken th
2016 Q1.
Reply · Share ›

working in the open - explaining decisions

The most important requirement for RIOXX is to provide a way to collect basic information about how research outputs (especially papers) have been funded. That is to say, we need to collect two types of information for a given paper:

- the funder
- the project or grant ID under which this research output was funded

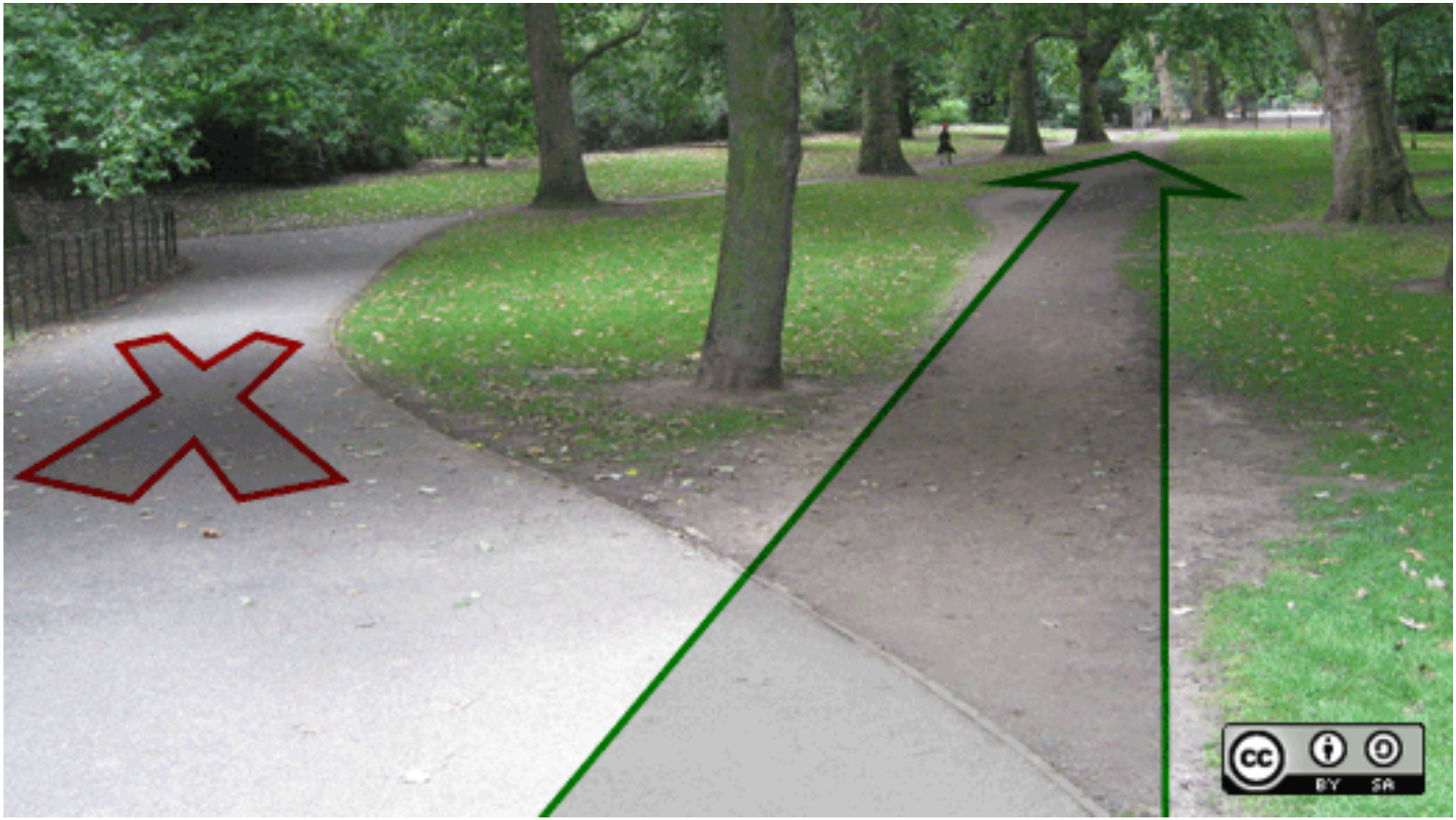
RIOXX has at its heart a simple application profile of Dublin Core, the better to fit naturally with existing technical infrastructure and to be easily implemented. Following from this, it seems that there are the following possibilities for representing funder and projectID in a single record:



Option 1: Separate fields for Funder and ProjectID

This is the ***currently proposed*** approach. It has the virtue of introducing a very low barrier to data collection (the common repository interfaces support this) and provides for the primary use-case which is

‘paving the cowpaths’



www.flickr.com/photos/wetwebwork/2847766967/

'continuous' testing

continuous testing

56 Repositories

Name	OAI-PMH Base URL	Set	Software	Basic RIOXX?	Full RCUK?	Report
Open Repository Preview (TEST)	http://preview.openrepository.com/demo/oai/rioxx		DSpace	**Valid! (100%)**	**Valid! (100%)**	Report
Falmouth University Research Repository (FURR)	http://repository.falmouth.ac.uk/cgi/oai2		EPrints	**Partial (65%)**	**Partial (2%)**	Report
Research SPAce	http://researchspace.bathspa.ac.uk/cgi/oai2	rioxx	EPrints 3.3.10	**Partial (15%)**	**Partial (2%)**	Report
LSBU Research Open	http://researchopen.lsbu.ac.uk/cgi/oai2		EPrints 3	**Valid! (100%)**	**Partial (13%)**	Report
Keele Research Repository	http://eprints.keele.ac.uk/cgi/oai2		EPrints	**Partial (93%)**	**Partial (10%)**	Report
SAS-SPACE	http://sas-space.sas.ac.uk/cgi/oai2		EPrints 3.1.2.1	**Partial (99%)**	**Partial (1%)**	Report
St George's Online Research Archive	http://openaccess.sgul.ac.uk/cgi/oai2		EPrints 3.3.6	**Partial (82%)**	**Partial (1%)**	Report
ESC publication	http://eprints.esc.cam.ac.uk/cgi/oai2		EPrints	**Partial (55%)**	**Partial (1%)**	Report
Institutional Repository	http://irep.ntu.ac.uk/cgi/oai2		EPrints 3	**Valid! (100%)**	**Invalid (0%)**	Report
University of Huddersfield Repository	http://eprints.hud.ac.uk/cgi/oai2		EPrints 3.2.2	**Valid! (100%)**	**Invalid (0%)**	Report
White Rose Research Online	http://eprints.whiterose.ac.uk/cgi/oai2		EPrints 3.3.10	**Valid! (100%)**	**Invalid (0%)**	Report
Bournemouth University Research Online	http://eprints.bournemouth.ac.uk/cgi/oai2		EPrints 3.2.2	**Partial (96%)**	**Invalid (0%)**	Report
LJMU Research Online	http://researchonline.ljmu.ac.uk/cgi/oai2		EPrints 3.3.12	**Partial (96%)**	**Invalid (0%)**	Report
Leeds Beckett University Repository	http://eprints.leedsbeckett.ac.uk/cgi/oai2		EPrints 3.3	**Partial (93%)**	**Invalid (0%)**	Report
Enlighten	http://eprints.gla.ac.uk/cgi/oai2		EPrints 3.1.13	**Partial (91%)**	**Invalid (0%)**	Report
e-Prints Soton	http://eprints.soton.ac.uk/cgi/oai2		EPrints 3.2	**Partial (90%)**	**Invalid (0%)**	Report
Sunderland University Institutional Repository	http://sure.sunderland.ac.uk/cgi/oai2		EPrints 3.2.3	**Partial (90%)**	**Invalid (0%)**	Report

continuous testing - reporting

Record	Validation Reports	Toggle Raw XML
ID: oai:eprints.lse.ac.uk:138 **Datestamp:** 2010-10-01	RIOXX Base RIOXX scheme designed for low-level interoperability **This is not a valid RIOXX record**	Raw XML

Property	Error
dc:identifier	Minimum of 1 value(s) required for dc:identifier - found 0 values
dc:title	Minimum of 1 value(s) required for dc:title - found 0 values
rioxxterms:author	Minimum of 1 value(s) required for rioxxterms:author - found 0 values

```
<?xml version='1.0' encoding='UTF-8'?>
<record>
  <header>
    <identifier>
      oai:eprints.lse.ac.uk:138
    </identifier>
    <datestamp>
      2010-10-01T08:40:52Z
    </datestamp>
    <setSpec>
      736574733D434F4C4C53:4C53455F435F454F
    </setSpec>
    <setSpec>
      736574733D43454E54524553:4C53455F52435
    </setSpec>
    <setSpec>
      736574733D4445505453:4C53452D4543
    </setSpec>
  </header>
  <metadata>
```

analytics

Validation Check Summary Report

This page contains a summary of the validation checks performed against a sample of records harvested from this list of repositories.

This validation is according to the RCUK RIOXX application profile.

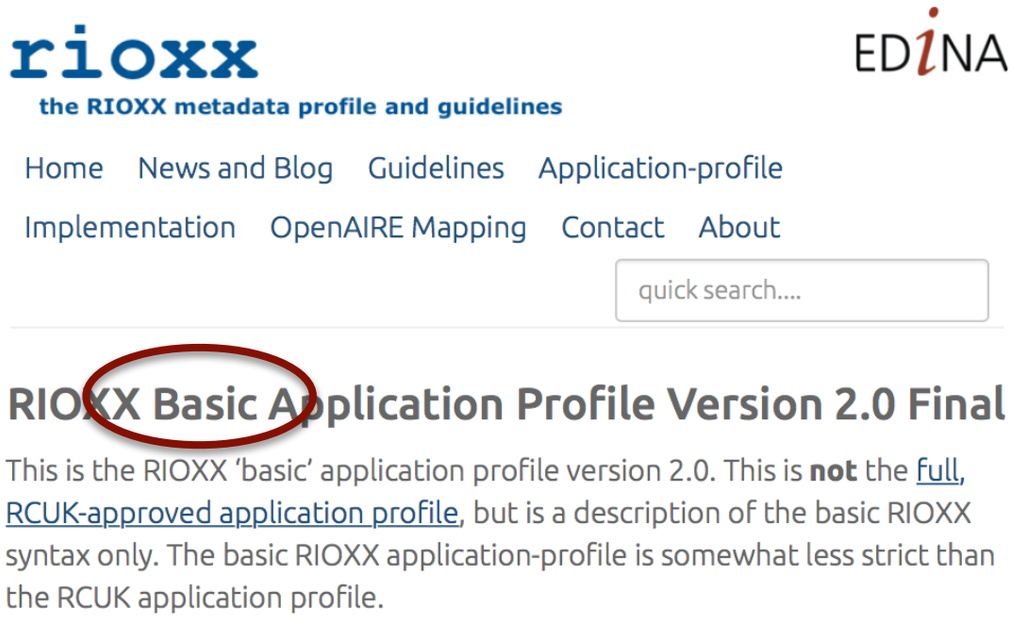
Property	Total Issues	Issues
dcterms:dateAccepted	4979	Type / Count: INVALID_ISO8601_DATE 8; TOO_FEW_VALUES 4971
ali:license_ref	4930	Type / Count: INVALID_ISO8601_DATE 313; TOO_FEW_VALUES 4617
rioxxterms:project	4845	Type / Count: INVALID_HTTP_URI 1; MISSING_VALUE 9; TOO_FEW_VALUES 4835
dc:identifier	1977	Type / Count: TOO_FEW_VALUES 1977

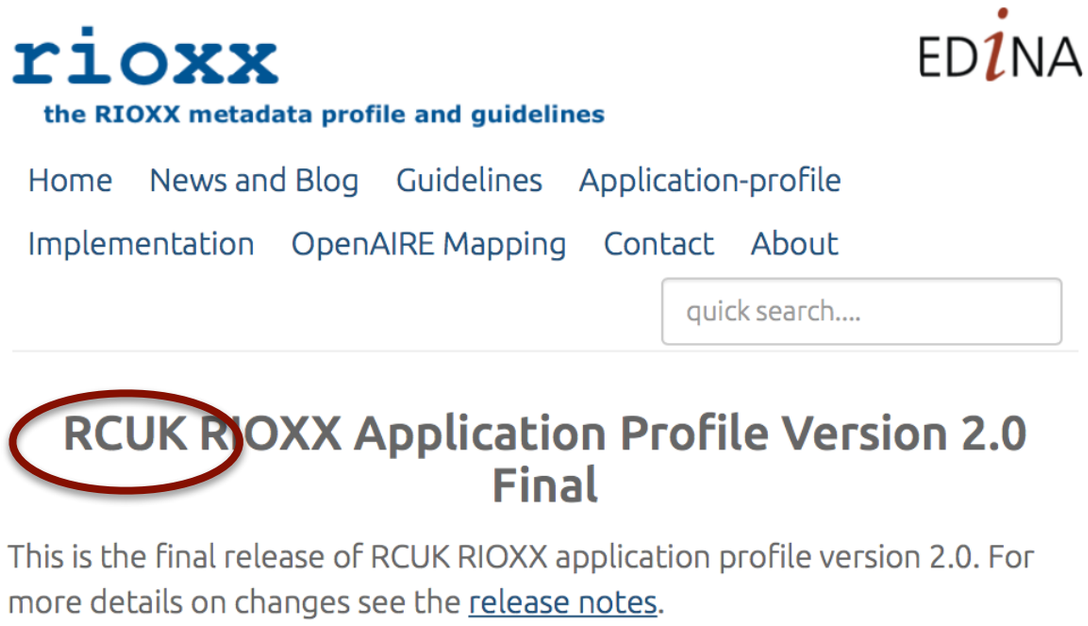
summary

- RIOXX has been created to help universities address open-access reporting requirements from the UK Research & Funding Councils
- it has been developed using agile approaches and techniques borrowed from software-developers
- it has been implemented in 56 known repositories since January 2015
- now also being harvested by CORE

- **adoption of RIOXX is growing steadily :-)**

Future development?





- RIOXX Basic has been used (partially) in two international aggregation initiatives:
  - OneRepo:
    - http://onerepo.net/onerepo-single-page.pdf
  - SHARE
    - https://github.com/CenterForOpenScience/SHARE

thanks for listening!

the RIOXX metadata application profile is maintained & supported by EDINA:

http://www.rioxx.net

Paul Walk
Head of Technology Strategy and Planning, EDINA
p.walk@ed.ac.uk
@paulwalk

EDINA