

RIOXX

RIOXX is a metadata application profile developed by EDINA and Chygrove Ltd (with support from Jisc) to assist UK higher education institutions to comply with the Research Councils UK Policy on Open Access. RIOXX is designed to aid the supply of precise metadata describing those scholarly publications which are generated by UK Research Councils funding. The RIOXX metadata application profile, version 2.0, was released on 22 January 2015.

This article outlines the application profile and its purpose, and offers a little more detail on some particular metadata properties. It goes on to describe some of the principles behind its development, together with some of the challenges encountered and the solutions (and compromises) made to address these. A particular innovation has been the adoption of some of the practices associated with modern software development, notably the continuous testing regime. The article concludes with a brief account of progress to date.

The requirement for RIOXX

The Research Councils UK (RCUK) Policy on Open Access¹ requires that institutions in receipt of UK Research Council funding deposit, into a repository, all research papers which acknowledge that funding. Most institutions use their own repository for this purpose. In some cases a paper is embargoed by the publisher – that is to say, it may not be made freely available for a period of time after publication. In any case, RCUK requires that the institution be able to demonstrate that the policy requirements have been adhered to. Effectively, this means that the institution should report the deposit of a paper, and its associated metadata, in a standard way, linking the paper to the funding which supported its production.

The UK is fortunate in that the majority of its higher education institutions have their own repository. These repository systems have, for some time, been used to manage research papers and associated metadata. Recognizing this, RCUK has worked with Jisc and the RIOXX team to develop the RIOXX metadata profile so that institutions might adopt a standard way of describing their open access (OA) papers. The supported approach is for repositories to be enhanced (normally via the use of a plug-in or similar extension mechanism) to allow metadata records to be exported in the RIOXX format.

Standards for structuring and exchanging metadata can be large and complex, especially where the goal is to be comprehensive, to cover a range of use cases and to promote general interoperability. The concept of the metadata application profile (a defined collection of metadata 'properties' and constraints) is used to cater for more closely defined, well-understood use cases.

A handful of existing application profiles are in fairly common use, mainly based on the Dublin Core Terms². One such is the OpenAIRE³ metadata profile, which is used to describe OA publications funded by the European Commission. Initially, it was hoped that the OpenAIRE profile could be used to satisfy the RCUK requirement. However, in addition to the conventional metadata which repositories already manage, RCUK requires information about the source of funding for the paper, and the licence under which it is made available. While OpenAIRE does address these concerns, it does so in a way which is incompatible in the context of the UK Research Councils, due to its strict requirements for the formatting of the strings used to identify projects.



PAUL WALK

Head of Technology
Strategy & Planning,
EDINA
University of
Edinburgh

'Standards for structuring and exchanging metadata can be large and complex'

161 For this reason, RIOXX was conceived as a metadata application profile which, while reusing existing application profiles where possible, would nonetheless need to include some new 'properties'.

Examining RIOXX 2.0

RIOXX 2.0 was released in January 2015. It is a simple metadata application profile of 21 properties including 11 properties from Dublin Core, two properties from the NISO Open Access Metadata and Indicators⁴, with the further eight properties newly defined especially for RIOXX. The complete profile is formally defined and documented on a page on the website⁵, alongside XSD schemas to support validation of metadata, and a guidelines document to aid implementation.

General characteristics

RIOXX was developed in accordance with some design principles which were decided upon in advance, and has the general characteristics listed below.

Simplicity

RIOXX has, intentionally, a 'flat' structure, with no 'nesting' of properties. This aspect ensures that RIOXX is easily documented and understood, and has allowed RIOXX to be implemented rapidly. While RIOXX does specify some properties which are optional, every specified mandatory property is essential to meeting the RCUK requirement. The design approach was to define the 'minimum viable product' which could address the RCUK use case, and to develop just this.

'RIOXX is easily documented and understood'

Factual metadata

The OA domain is, unfortunately, beset with imprecise terms. Indeed, even the phrase 'open access' is, itself, not subject to a precise definition – at least, not one which is universally recognized. Furthermore, some metadata properties, used to describe research papers, are similarly imprecise. For example, the concept of 'publication date' is problematic: often, the precise publication date is not known and it is not always clear, in the case of a research paper, to what the publication date refers. The treatment of embargoes, set by publishers to restrict access to so-called green OA papers for a period of time, is also an issue. One particular challenge for RIOXX was to handle the precise reporting of these embargoes.

'the phrase "open access" is, itself, not subject to a precise definition'

RIOXX has a core design principle to avoid or at least reduce the use of imprecise terms and, instead, constrain its properties to contain factual metadata values so that people and systems consuming RIOXX metadata records are able to make decisions based on information which is *reliable*.

'RIOXX has a core design principle to avoid or at least reduce the use of imprecise terms'

Machine-actionable metadata

Metadata which is *factual* lends itself to being used systematically. Factual metadata properties may be thought of as *assertions*, creating the possibility for decisions to be made based on them. Another important factor in ensuring that metadata can be used systematically is the use of precise formats for properties such as dates and identifiers. In the case of identifiers, the RIOXX specification requires the use of HTTP URIs, and these are used in several places in the RIOXX metadata profile. This is a very important aspect: while RIOXX does not mandate the use of particular identifier schemes (although it does recommend the use of some), this insistence on the use of HTTP URIs means that the identifier *system*, as well as the identifier itself, can be inferred from the URI. Commonly used examples would include ORCIDs, identifiable as such by the prefix

'http://orcid.org', DOIs, with the prefix 'http://dx.doi.org' and ISNIs, prefixed with 'http://isni.org'. This feature allows third-party software to take a RIOXX metadata record and to link it automatically and unambiguously, through the use of HTTP URI identifiers, to other data sources. This is what is meant by 'machine-actionable' metadata – metadata which can be automatically processed by software.

“machine-actionable” metadata – metadata which can be automatically processed by software’

Specific properties

The development of RIOXX has been largely predicated on the development of some particular metadata properties which did not already exist, and the reuse of others which were not already present in an existing profile. Some of these properties are described here. It may be useful, while reading this section which describes RIOXX in more detail, to have open the web page at <http://www.riox.net/profiles/v2-0-final/>.

A property to describe licences and embargo periods: `ali:license_ref`

RCUK requires that institutions provide an explicit licence statement for OA papers. The issue of how to describe embargo periods was much discussed. NISO provides the *license_ref* property, which expresses a licence by simply presenting the URL which locates the definitive documentation of that licence, together with an optional *start-date* – the date from which that licence takes effect. It is permitted to have several instances of this property in one RIOXX record (although not with the same *start-date* attribute). This offers a good, machine-readable and actionable solution to the problem of expressing a licence, and so RIOXX has adopted this as one of its mandatory properties.

A property to describe funding: `rioxterms:project`

In order to meet the RCUK requirements, institutions need to link their OA papers to the funders and funding streams that support their production. In practical terms, this means including an identifier for the funder together with the project ID. Fortunately, RIOXX is only concerned with those papers originating from projects which have received funding from one of the seven UK Research Councils. This reduces the burden considerably as, for the common case, one of seven possible identifiers will be applicable (although multiple funders may fund a given project, and not *all* of those will necessarily be a UK Research Council). Furthermore, all UK Research Council-funded projects are given a unique identifier. Institutions are able to use this project ID, allocated to a project by the funder, to identify the funding stream used to fund that project and, consequently, the project's outputs – including its research papers.

A new property, *rioxterms:project*, was defined. This property is slightly more involved than the others in the RIOXX profile, although it is still quite simple. Because the use of HTTP URIs to identify funders is not yet established practice, the ID element is not mandated. It should be recognized, however, that use of identifiers for funders is likely to become established practice, and so implementers of RIOXX are encouraged to adopt this.

‘use of identifiers for funders is likely to become established practice’

Properties to describe dates

Compared to other bibliographic metadata schemas, RIOXX is unusual in that it deprecates the importance of the publication date of the paper in favour of the date that the paper was accepted for publication. A property called *rioxterms:publication_date* is included in RIOXX, since the bibliographic record seems incomplete without it. However, this property is not important to the primary use case for RIOXX and is therefore optional.

Of more importance to RIOXX is the *dcterms:dateAccepted* date property, which is intended to capture the exact date on which the paper was accepted for publication. This date is significant in RCUK's Policy on

‘the exact date on which the paper was accepted for publication ... is significant in RCUK's Policy on OA as the policy focuses on accepted manuscripts’

163 Open Access as the policy focuses on accepted manuscripts, rather than published articles. This property, therefore, is mandatory.

Properties to describe people and organizations

RIOXX makes use of HTTP URIs as identifiers where possible, in order to promote the creation of machine-actionable metadata. This is certainly true in those properties used to describe or identify people and organizations. In the RIOXX profile these are the *rioxterms:author* and *rioxterms:contributor* properties, which allow an extra attribute to carry such an identifier.

The ORCID system, which provides a globally unique and persistent identifier for researchers, is growing in popularity. For this reason, RIOXX recommends the use of ORCID IDs to identify authors and contributors where they are individuals.

'RIOXX recommends the use of ORCID IDs'

Properties to identify the paper

It is important to recognize that a RIOXX record does not describe the published paper – it describes the locally held copy of the accepted manuscript. The primary identifier in the RIOXX profile is held in the *dc:identifier* property. This property must contain an HTTP URI which is the URL to the version of the paper held, typically, in the institution's repository. Ideally, this URL should locate the actual paper itself, usually in PDF form, but often will point to an intermediary web page from which the paper can be downloaded.

'a RIOXX record ... describes the locally held copy of the accepted manuscript'

The other property in RIOXX which can be used to identify the paper is called *rioxterms:version_of_record* and is used for the identifier of the published version of the paper. This property is optional, since it will normally be populated later, after the RIOXX record has been created. If used, this property will most often contain a DOI in its HTTP URI form.

Properties with vocabularies

Three of the RIOXX properties, *rioxterms:apc*, *rioxterms:type* and *rioxterms:version*, have associated with them defined vocabularies. Essentially, these properties require a value from a short, controlled list of possible values. For example, the *rioxterms:version* property may contain values from a list including codes such as 'AO' (meaning 'Author's Original') or 'CVoR' (meaning 'Corrected Version of Record'). This approach is designed to improve consistency in the resulting metadata, supporting its use in automated processes.

Other properties

The RIOXX profile includes ten other properties, many of which, such as '*dc:title*', are already commonly used in bibliographic metadata. In several cases, RIOXX introduces further constraints on their use in order to support the provision of better quality metadata.

Open and 'ruthlessly pragmatic' development

EDINA and Chygrove Ltd worked very closely with RCUK to develop an initial specification for RIOXX, which was then refined through a process of iterative development. Some principles were established such as (radically) open development, a focus on the requirement with a determination to avoid complexity where possible, and an emphasis on supporting immediate implementation. This approach to development was influenced by the principles behind agile⁶ software development.

Previous application profiles have been developed openly, in the sense that anyone could participate – typically, by joining a mailing list. However, RIOXX took the more radical approach of developing 'in the open' on the website's blog⁷ explaining, for example, the

rationale behind decisions taken about which properties to include in the profile and how they should be structured. As a result, the development of RIOXX has been informed by an unusually wide range of people from across the repository community, as can be seen in the public comments on the website.

At various stages in the development of RIOXX the development team resisted suggestions to expand RIOXX to address other requirements. The team adopted an approach of 'ruthless pragmatism', keeping only what was strictly necessary for RIOXX to meet the RCUK requirements. The single exception to this is, arguably, the inclusion of the *rioxxterms:publication_date* property, which is defined to allow a complete RIOXX record to function as a simple bibliographic record.

From the outset, RIOXX was developed with implementation in mind. Because RIOXX is designed to meet a constrained and well-defined use case, the systems (and even, to an extent, the people) likely to be involved in implementation were predictable. Representative software developers and systems administrators were engaged throughout the development process, and provided significant input. The intention was to develop a metadata profile that could be easily and rapidly implemented. Recognizing that development might benefit from an iterative approach, a system of 'continuous testing' – another concept borrowed from agile software development – was developed to support this. The results⁸ of testing all known implementations of RIOXX are published openly on the RIOXX website, with detailed reports showing precisely the extent to which individual records comply with the RIOXX profile. As more test results are accumulated, patterns of implementation are emerging, highlighting issues with particular properties or constraints. This will lead to clarification of the metadata profile's documentation or, if necessary, further revision of the profile itself.

'the development of RIOXX has been informed by an unusually wide range of people'

'The team adopted an approach of "ruthless pragmatism", keeping only what was strictly necessary'

Progress

At the time of writing, RIOXX is known to have been implemented in 52 institutional repositories, according to the OpenDOAR⁹ registry. Nearly all of these repositories are based on the ePrints¹⁰ software and have implemented RIOXX by installing a plug-in. Similar support for the DSpace¹¹ repository software is in development, and is expected to be released soon. When this happens, the number of repositories supporting RIOXX is likely to increase rapidly. In any case, this is a very encouraging rate of adoption over a period of less than 18 months.

'RIOXX is known to have been implemented in 52 institutional repositories'

As repositories supporting RIOXX have come on stream, the RIOXX team has regularly harvested sample records from them for testing. The results of testing indicate that repositories are, increasingly, able to meet the basic metadata quality requirements demanded by RIOXX. However, repositories are generally not yet meeting the full RCUK requirements. Some reasons for this are offered below, under 'Challenges'.

In addition to the rapid adoption in institutional repositories, elements of RIOXX have been implemented in *consuming* systems such as The One Repo¹², and SHARE¹³, and the Open University's COncecting REpositories (CORE) aggregator¹⁴, which is part-funded by Jisc to harvest RIOXX records from institutional repositories. Furthermore, some of the design thinking and development methodology behind RIOXX has been picked up by others. For example, the approach to documenting the application profile has received very positive reviews, and the source-code for the software used to test and validate RIOXX records was requested by and shared with the development team behind the CORE Dashboard¹⁵.

Challenges

The issue of how OA papers should be licensed continues to be problematic. Institutional repositories are expected to be able to reconcile a number of requirements on

165 licensing – from the funders, from publishers and, in some cases, even from the institution itself. RIOXX demands that an HTTP URI, which unambiguously identifies the licence that the institution has declared to apply to the locally held paper, be included in the *ali:license_ref* property. RCUK does not mandate a particular licence, requiring only that the manuscript is ‘made available without restriction on non-commercial reuse’. The use of some form of the *Creative Commons CC-BY* licence is recommended and, in such cases, the RIOXX requirement for an HTTP URI is easily accommodated. However, a common case where this is not possible is when the paper is under embargo. In this case, the paper is not licensed at all by the institution for reuse until such time as the embargo period is ended. Although ‘All rights reserved’ is the default status for creative works, RIOXX includes a simple web page stating this in order to provide an HTTP URI which might be used to make this status explicit. In this way, institutions can use two instances of the *ali:license_ref* property in the same record, indicating both an embargo period and a declaration of the licence for use of the paper at a future date, after the embargo period.

‘The issue of how OA papers should be licensed continues to be problematic’

While there is now a clear mechanism for dealing with licensing, it must be said that there is still some way to go before clear licensing of OA papers becomes the norm.

‘there is still some way to go before clear licensing of OA papers becomes the norm’

Another challenge revealed in the course of testing the implementation of RIOXX concerns the ePrints software. It has become apparent that, by default, an ePrints repository system will offer *all* of its metadata records when an external system makes a request to it (via OAI-PMH) for RIOXX-formatted records. Consequently, a normal ‘harvest’ of RIOXX records from an ePrints system will retrieve records which were never meant to be offered as RIOXX records. This results in an apparent but misleading indication that the level of compliance with the strict RCUK requirements is poor, explained by the fact that many records being tested should not actually have been provided by the ePrints system. The ePrints supplier has been informed of this and is investigating. The DSpace plug-in supplier has asserted that this will not be a problem with DSpace, so when that plug-in becomes available and adopted, a better level of compliance is expected.

Why RIOXX matters

There are a number of reasons why RIOXX might be considered to matter. Most importantly, it supports institutional repositories in the UK in helping their institutions to comply with the RCUK Policy on Open Access. However, in addition to this, it has innovated around the process of developing metadata profiles. The introduction of radically open development coupled with continuous testing shows how the community can play a very active part in the development of solutions of this kind. The rate of adoption of RIOXX indicates that this is a successful development approach.

‘the community can play a very active part in the development of solutions of this kind’

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the ‘Abbreviations and Acronyms’ link at the top of the page it directs you to: <http://www.uksg.org/publications#aa>

Acknowledgments

The RIOXX team would like to acknowledge:

- RCUK, and particularly Ben Ryan, for their excellent sponsorship and support for the development of RIOXX
- EDINA (and the University of Edinburgh) for continuing to host the RIOXX website and technical documentation
- Jisc, for funding the initial development of RIOXX
- the participants in the HHuLOA project for their valuable analysis and input
- the enthusiastic and knowledgeable community of developers and repository managers for their active engagement with the development of RIOXX.

Competing Interests

The author has declared no competing interests.

References

- 1 RCUK Policy on Open Access:
<http://www.rcuk.ac.uk/research/openaccess/policy/> (accessed 25 April 2016).
- 2 Dublin Core Terms:
<http://dublincore.org/documents/dcmi-terms/> (accessed 28 April 2016).
- 3 OpenAIRE Guidelines for Literature Repositories:
<https://guidelines.openaire.eu/en/latest/literature/index.html> (accessed 28 April 2016).
- 4 NISO Open Access Metadata and Indicators:
<http://www.niso.org/workrooms/ali/> (accessed on 28 April 2016).
- 5 RIOXX 2.0:
<http://www.riox.net/profiles/v2-0-final/> (accessed on 28 April 2016).
- 6 Manifesto for Agile Software Development:
<http://agilemanifesto.org> (accessed 1 May 2016).
- 7 RIOXX website blog:
<http://www.riox.net/post/> (accessed 28 April 2016).
- 8 RIOXX test results:
<http://www.riox.net/implementation/> (accessed 1 May 2016).
- 9 OpenDOAR:
<http://www.opendoar.org> (accessed 1 May 2016).
- 10 ePrints:
<http://www.eprints.org/uk/index.php/eprints-software/> (accessed 2 May 2016).
- 11 DSpace:
<http://www.dspace.org> (accessed 1 May 2016).
- 12 The One Repo:
<http://onerepo.net> (accessed 2 May 2016).
- 13 SHARE:
<http://www.share-research.org> (accessed 2 May 2016).
- 14 CORE:
<https://core.ac.uk> (accessed 2 May 2016).
- 15 CORE Dashboard:
<https://core.ac.uk/intro/dashboard> (accessed 1 May 2016).

Article copyright: © 2016 Paul Walk. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



Paul Walk

Head of Technology Strategy & Planning, EDINA
University of Edinburgh, UK
E-mail: p.walk@ed.ac.uk | Personal Web: <http://www.paulwalk.net>

ORCID iD: <http://orcid.org/0000-0003-1541-5631>

To cite this article:

Walk, P, RIOXX, *Insights*, 2016, 29(2), 160–166; DOI: <http://dx.doi.org/10.1629/uksg.303>

Published by UKSG in association with Ubiquity Press on 05 July 2016