RIOXX

UK Metadata Guidelines for Open Access Repositories

January 2015

Version 3.0

Summary

- The RIOXX Application Profile and Guidelines have been developed in conjunction with RCUK and HEFCE. The development of RIOXX has been funded by Jisc and supported by EDINA.

- The development of RIOXX was based on a central use case: the ability to track research outputs across systems. More specifically, RCUK requires the means to monitor compliance with its open access policies. Certain metadata elements within RIOXX will be useful for compliance with HEFCE's open access policy and REF requirements.

- Both RCUK and HEFCE endorse and encourage compliance with RIOXX by UK higher education and research institutions

- RIOXX was developed with the OpenAIRE Guidelines and EThOS in mind, the objective being to minimise the deviation from these existing approaches as far as reasonably possible.

- There is a particular focus on two new metadata elements: project ID (or grant/award number) and funder name. This information is not routinely exposed in institutional repositories at present. Collecting and exposing this information is a key requirement.

- The key outputs from NISO's recently launched *Access License and Indicators* Recommended Practice have been incorporated into the RIOXX Application Profile[1], namely the free_to_read and license_reference metadata elements.

- As well as introducing new metadata elements, it is hoped that the introduction of RIOXX will help normalise peoples' interpretation of common metadata elements. Analysis of RepUK, one of the UK's metadata aggregations, indicated that consistency is lacking with respect to how different institutions interpret metadata standards.

- The sponsors of RIOXX encourage UK higher education and research institutions to begin the process of adopting RIOXX as soon as possible. In any event RIOXX compliance should be demonstrated by April 2016.

- To help with the compliance process, a plugin for EPrints repositories (versions 3.3.x) and an add-on for DSpace repositories (versions 3, 4 and 5) will be freely available. The RIOXX EPrints plugin is currently being trialed by around ten early adopters; the DSpace plugin will be developed by @Mire and is scheduled to be available around the end of March 2015. CRIS users should address their RIOXX-related requirements to their software vendors.

---

[1] http://www.niso.org/apps/group_public/download.php/14226/rp-22-2015_ALI.pdf

- Technical support for the EPrints plugin is available from the following organisations:

  - EPrints Services, Southampton University: for its hosted customers
  - ULCC, London: for its hosted customers
  - Digital Repository Solutions (Peter West): for self-hosted EPrints users

- The authors of the Application Profile and Guidelines are Paul Walk (EDINA) and Sheridan Brown (Key Perspectives/Chygrove). They can be contacted directly via the RIOXX website.

These Guidelines complement the RIOXX Application Profile. Whereas the Application Profile is not designed to be updated frequently, the Guidelines may be updated more frequently in response to user feedback and issues of interpretation that may by raised. In that sense, these Guidelines are dynamic in nature. The RIOXX development team has tried to make the Application Profile as clear and unambiguous as possible but if you think there are points that merit further clarification please do contact the RIOXX development team. All feedback will be considered in partnership with the prime sponsors, Jisc, RCUK and HEFCE, and the Guidelines will be augmented as necessary.

1. Introduction

The successful development of open access repositories in very many of the UK's higher education and research institutions is testament to the efforts of repository managers and their information management colleagues. The result of these efforts is a growing body of research information that can be freely discovered and re-used by people around the world. The foundations of the UK's repository infrastructure are firmly established but there remain opportunities for the community to build and improve services that provide additional value to a variety of stakeholders in the research communication chain.

Those involved in the collection and management of information understand the central role played by metadata in the success of their institutional information management systems. Accurate, rich, high quality metadata enhances not only discoverability, re-usability and interoperability but also the extent to which different stakeholders can use outputs for different purposes.

Analysis of the UK's aggregations of metadata collected from open access repositories indicate that, at present, there are inconsistencies in the ways in which metadata is managed. For example, a recent snapshot of the content of UK open access repositories in the tertiary sector clearly showed a significant disparity between the number of full-text pdfs indicated by metadata and the actual number of actionable pdfs in those repositories. The development of a national Application Profile like RIOXX for open access repositories aims to reduce ambiguity regarding the implementation of metadata standards and improve the overall quality and consistency of metadata.

The key impetus for the development of these national guidelines is the government-driven need for Research Councils to be able to identify the research outputs from projects they have funded. At present there is no straightforward or systematic way for these funders to identify when relevant articles appear in open access repositories. The introduction of two new core metadata fields is designed to address this particular problem, namely a field describing a project's identity – such as a grant number - and a field describing the identity of the funder. This information is not routinely collected in open access repositories at present.

As well as address RCUK's need to be able to monitor the extent to which institutions are adhering to the terms of its open access policy, the RIOXX Application Profile has been adapted to accommodate some of HEFCE's metadata requirements with respect to their recently-announced open access policy and the requirements for the post-2014 REF.

Scope of RIOXX

It is important to note that the RIOXX Application Profile and Guidelines are specifically designed to encompass *publications*. These are specifically identified in the rioxxterms:type metadata element: the list of types includes books, book

chapters, edited books, conference outputs, journal articles and reviews, manuals and guides, monographs, policy briefing reports, technical reports, technical standards, theses, consultancy reports, working papers and other publications not specifically listed. Metadata standards for other types of digital object are being addressed by other initiatives and organisations including CASRAI. In terms of geographical scope, because RIOXX exists primarily to support UK funders the focus of RIOXX is the UK. That said, many of the 21 metadata elements have much wider application.

Timelines and helping with compliance

All institutions that receive funding from RCUK or HEFCE are expected to have implemented processes to capture at least the mandatory metadata described in RIOXX by April 2016. It is recommended that institutions aim to implement RIOXX well before that date. To help with the implementation process Jisc is funding the development of applications for two of the most widely used open source repository platforms in the UK, EPrints and DSpace. A RIOXX plugin for EPrints (versions 3.3 and higher) has already been created and is in the process of being tested by up to ten early adopters. A similar application for DSpace (versions 3, 4 and 5) will be available in the spring.

In addition, Jisc is funding technical support for users of the EPrints plugin for twelve months. Support for installation and technical issues to do with operating the plugin will be provided by EPrints Services for the repositories they host, ULCC for their hosted repositories and Digital Repository Services (Peter West) for non-hosted repositories. Jisc-sponsored workshops are planned which will provide further technical information on the plugins and support.

Access to repositories by robots

In order to aid the development of infrastructure to support open access, repositories are encouraged to provide the same level of access to metadata and full text content to external automated systems (variously known as "robots", "crawlers" or "harvesters") as they provide to human users. In addition, repositories should allow robots to harvest the entire metadata and full text content of the repository in a reasonable timeframe. It is recommended that repositories use the Robots Exclusion Protocol (via a "robots.txt" file) to make clear the level of access offered to external systems.

2. Detailed description of the metadata elements

The UK-specific RIOXX Application Profile has been developed with reference to the Driver and OpenAIRE Guidelines (which are related to the OpenAIRE project[2]) and UKETD_DC, the metadata core set recommended by the British Library's Electronic Theses Online Service (EThOS)[3]. RIOXX also draws on the wealth of information provided by the Dublin Core Metadata Initiative[4] largely because, in the quest for accurate, appropriate and consistent use of metadata, it is important that the RIOXX Application Profile is rooted in standards that have been developed over many years and which have been widely adopted around the world.

Whether you are creating metadata through a manual process or setting up the automatic conversion of existing records to new ones, these Guidelines exists to help with the organisation and management of those metadata. Care should be taken to attribute the information you collect to the most appropriate metadata element. There may be occasions where the choice of element is not clear-cut so you will need to make a judgment. The key is to make these judgments on a consistent basis for your repository.

Please note that for the moment these Guidelines are designed primarily with publications in mind. The comments below often refer to a "resource" which for now should be taken to mean "publications". This semantic constraint may be amended in future versions of the Guidelines as other types of research outputs are considered for inclusion.

ali:free_to_read
Zero or one instance

Use of this element is: **Optional**

This element is defined in the <u>NISO Open Access Metadata and Indicators</u>. This element does not take a value - the semantics of *ali:free_to_read* are conveyed by its presence or absence. This element may be modified by two optional attributes
. start_date
. end_date

Each of these attributes, if present, takes a date value which **MUST** be encoded using ISO 8601 (post–2004 versions) which follows the following format: YYYY-MM-DD. Examples:
<ali:free_to_read start_date="2013-03-28" end_date="2014-04-30">
<ali:free_to_read start_date="2013-03-28">
<ali:free_to_read>

---

[2] http://www.openaire.eu
[3] http://ethos.bl.uk/Home.do
[4] http://dublincore.org/

The absence of a *start_date* attribute implies that the meaning conveyed by the ali:free_to_read element is current and immediate, unless an *end_date* attribute, which is a date in the subjective past, is present in the element.

Note that NISO have yet to specify a namespace for this element - RIOXX will adopt the NISO recommendation when it is made

Comments
On the face of it the free_to_read tag should provide a mechanism for quickly establishing whether an item is "open access". It would be prudent to be cautious since in fact this element simply describes whether a work is accessible to read online without charge or authentication; no information is offered or implied about other forms of access. It does not imply that a resource is free to download. Since HEFCE's open access policy states that, to be eligible for the REF, deposited material should be free to read *and download* the free_to_read metadata element of itself will not be sufficient to demonstrate policy compliance. Reference to the publisher's license is a more certain means to establish policy compliance.

The period of time which a resource is "free to read" may be limited; for example a publisher may allow free access for certain period of time for promotional purposes. This element is potentially useful in terms of offering a shortcut to identify resources that may, at least, be read online, but for the time being funders are likely to rely more on other metadata fields which is why the free_to_read metadata element is optional.

The NISO Working Group recommends that the free_to_read metadata is included in existing metadata distribution channels and with the content itself where appropriate. It is for the creator, publisher or other primary rights holder to declare whether a resource is free to read or not and they should provide that information. Where such a declaration is not available then clearly this metadata element should not be used.

ali:license_ref
One or more instance

Use of this element is: **Mandatory**

This is defined in the <u>NISO Open Access Metadata and Indicators</u>. This element **MUST** take an HTTP URI for its value. This HTTP URI **MUST** point to a resource that expresses the license terms specifying how **the resource** may be used.

This element **MUST** include the attribute:
. start_date
This attribute takes a date value that **MUST** be encoded using ISO 8601 (post–2004 versions) that follows the following format: YYYY-MM-DD.

This attribute is used to indicate the date upon which this license takes effect. Multiple *ali:license_ref* elements may be included. Where several such elements

are included, the one with the *start_date* attribute indicating the most recent date takes precedence.
Example:
<ali:license_ref start_date="2015-02-17">http://creativecommons.org/licenses/by/4.0</ali:license_ref>

This approach allows the expression of 'embargoes', where a particular license takes effect at a date in the subjective future.

In the absence of any other license, the copyright holder reserves all rights automatically. As a convenience, RIOXX provides two URLs, which may be used to explicitly convey this state:
• http://www.rioxx.net/licenses/all-rights-reserved
• http://www.rioxx.net/licenses/under-embargo-all-rights-reserved

Comments
As with the free_to_read element, the license_reference element (normally shortened to license_ref) has the potential to be challenging to implement at least to begin with since it depends on the primary rights holder – often a publisher since RIOXX is concerned primarily with publications – providing a suitable reference to a URI that carries the license terms specifying how a work may be used. It may be some time before all publishers become familiar with NISO's Recommended Practice and amend their workflows to routinely supply this metadata. In the meantime, where information is not available, to comply with the mandatory nature of this element you should use one of the two default URL's noted above. For resources published by the "Gold" route using RCUK funding specifically allocated for this purpose, there should be no problem providing the metadata since RCUK's policy specifies a Creative Commons license.

Some examples of publishers' license_reference URIs, taken from the Recommended Practice, are listed below.

Publisher	Possible existing URIs that might be used
American Chemical Society	http://pubs.acs.org/userimages/ContentEditor/1218220609981/authorchoice_form.pdf
University of California Press	http://ucpressjournals.com/assets/ucp_sample_auth_agr.pdf
BioMed Central (CC BY v 2.0)	http://creativecommons.org/licenses/by/2.0/
Rockefeller University Press	http://www.rupress.org/site/subscriptions/terms.xhtml and http://creativecommons.org/licenses/by-nc-sa/3.0/
British Institute of Radiology	http://www.bjrpublications.org/page/copyright

There has been some concern in the community about how best to address the issue of metadata for embargoes. There has been a focus on three particular aspects: embargo start date (often the publication date), embargo duration and embargo end date. Knowing two of these three pieces of information allows the computation of the third, but there is ample scope for ambiguity and error, particularly in relation to the embargo start date or publication date which are often, reportedly, not always clearly communicated by the rights holder. The use of license_ref offers a means by which to communicate unequivocally the start date of an open license and therefore the end of an embargo period.

The example given in the NISO Recommend Practice, shown below, indicates that the open license, in this case a Creative Commons license, begins on 2015-02-03, signifying the end of a twelve month embargo period.

```
<license_ref start_date="2014-02-
03">http://www.psychoceramics.org/license_v1.html</license_ref>

<license_ref start_date="2015-02-
03">http://creativecommons.org/licenses/by/3.0/</license_ref>
```

The free_to_read and license_ref metadata elements are the only ones in RIOXX taken from the NISO Recommended Practice. That document has a section entitled "Recommended Mechanisms for Distributing Metadata". That section bears repetition below because it makes clear that NISO expects publishers,

aggregators and other content providers to disseminate these metadata elements.

> "To ensure the widest dissemination of metadata, publishers, aggregators, and other content providers are encouraged to include the free_to_read and license_reference elements in all of their standard metadata sets. Wherever possible, creation and population of these elements should become part of standard editorial/production workflows. The metadata should be made an integral part of the feeds to CrossRef and other DOI Registration Agencies, included alongside (or within) article/chapter content on hosting websites, and delivered in content feeds to third parties.
>
> The metadata should be embedded in the content itself along with other metadata; for example, in HTML META tags and in PDF files where bibliographic and other metadata are being included.
>
> It may also be worthwhile for content providers to consider including the metadata elements within other alerting channels, such as e-ToCs and RSS subscription feeds as well as information provided directly to abstracting and indexing services. Whatever channel is used, wider distribution of this (and other) article, chapter, or book metadata is likely to be helpful in driving discovery and usage for the materials concerned." [5]

There may be resources created by members of your institution that will not be published such as conference contributions and working papers. In these instances it will be necessary to establish with the creator which license they propose to reference. This may be a Creative Commons license or your institution may already have or need to create a license that can be referenced appropriately.

dc:coverage
Zero or more instances

Use of this element is: **Optional**
The Openaire Guidelines recommend the inclusion of this element.

Comments
This refers to the scope or extent of the content of the resource. It may include jurisdictional, temporal or spatial information. It is **recommended** that, where possible, a recognised globally unique identifier is used, such as the Thesaurus of Geographic Names, but free text may be used. For example, the place of publication may be recorded.

---

[5] http://www.niso.org/apps/group_public/download.php/14226/rp-22-2015_ALI.pdf, see p.9

dc:description
Zero or more instances

Use of this element is: **Recommended**

This field may be indexed and its contents presented to people conducting searches. The goal is to describe the content of ***the resource*** using free text. It is **RECOMMENDED** that an English language abstract be used where available. HTML or other markup tags **SHOULD NOT** be included in this field.

dc:format
Zero or more instances

Use of this element is: **Recommended**

This refers to the format of ***the resource***. The MIME type of the object pointed to by this RIOXX record's *dc:identifier* element **MUST** be entered here. Note that this element should not be confused with *rioxxterms:type*

Comments
If more than one category is needed to describe a single resource, use separate instances of the dc:format element.

dc:identifier
Exactly one instance

Use of this element is: **Mandatory**

This field **MUST** contain an HTTP URI that is a persistent identifier for ***the resource***. The purpose of this field is, through direct identification of ***the resource***, to allow access to it, therefore it is **RECOMMENDED** that this identifier should point to the *actual resource being described by the RIOXX record* (typically a file in MS Word or PDF format), rather than to an intermediary resource such as a repository web page. Note that RIOXX does not require any require particular file format to be used for ***the resource***.

To describe another version of ***the resource*** with a different identifier, a completely separate RIOXX record should be created.

Comments
It is important to note that one RIOXX record may only describe one resource (or publication). It is recognised that this may cause issues since typically one repository record may have several resources or publications associated with it. It would be prudent for repository administrators to bear in mind this limitation.

dc:language
One or more instance

Use of this element is: **Mandatory**

This refers to the primary language in which the content of ***the resource*** is presented. The element **MAY** be repeated if ***the resource*** contains multiple languages. Values used for this element **MUST** conform to ISO 639–3. This offers two and three letter tags e.g. "en" or "eng" for English and "en-GB" for English used in the UK.

dc:publisher
Zero or more instances

Use of this element is: **Recommended**

This element contains the name of the entity, typically a 'publisher', responsible for making the ***version of record*** of ***the resource*** available. This could be a person, organisation or service.

Where available, the name of the publisher entered here **SHOULD** be from a controlled list.

Comments
Repository administrators in the UK often use the SHERPA RoMEO database that can perform the role of a controlled list for publisher names (as well as journal titles and ISSNs).

dc:relation
Zero or more

Use of this element is: **Optional**

The format of this element **MUST** be an HTTP URI that points to a related resource, e.g. a research data set that underpins ***the resource***. An exception to this is the DOI identifying the related 'version of record' - this **MUST** be recorded in the *rioxxterms:version_of_record* element.

Each related resource **MUST** appear as a separate instance of the field.

dc:source
Zero or one instance

Use of this element is: **Mandatory where applicable**

The source label describes a resource from which ***the resource*** is derived (in

whole or in part). It is **RECOMMENDED** that the source is referenced using a unique identifier from a recognised system e.g. the unique 8-digit International Standard Serial Numbers (ISSN) assigned to *electronic* periodicals, or the 13 digit International Standard Book Number (ISBN13) assigned to books. In the latter case, the ISBN13 for the electronic version of the book ***SHOULD*** be used if available.

Use of this element is applicable where ***the resource*** is to be published as part of a larger resource. Examples might include a journal article, a conference paper or a chapter of a book, but not a complete book for example.

Comments
It is important to reiterate that the ISSN or ISBN (or other recognised unique identifier) should be that issued for the ***electronic*** version of the source. SHERPA/RoMEO offers a controlled list of ISSNs.

dc:subject
Zero or more instances

Use of this element is: **Recommended**
The Openaire Guidelines recommend the inclusion of this element.

Comments
Normally keywords, phrases or classification codes are used to describe the topic of ***the resource***. If using free text, avoid using general keywords. The *recommendation* is to use a formal classification scheme or controlled vocabulary e.g. Library of Congress Classification Headings or Medical Subject Headings (MeSH).

When including terms from multiple vocabularies, use separate element iterations. If multiple vocabulary terms or keywords are used, either separate terms with semi-colons or use separate iterations of the Subject element.

dc:title
Exactly one instance

Use of this element is: **Mandatory**

This refers to the title, and any sub-titles, of ***the resource***. The title should be represented using the original spelling and wording. The **RECOMMENDED** format for expressing subtitles is:

Title:Subtitle

Note that where ***the resource*** is a chapter in a book, the chapter title **MUST** be entered here, with the ISBN13 of the book being recorded in the *dc:source* element.

dcterms:dateAccepted
Exactly one instance

Use of this element is: **Mandatory**

The date on which *the resource* was accepted for publication. The date **MUST** be encoded using ISO 8601 (post–2004 versions) using the following format: YYYY-MM-DD.

Comments
The HEFCE open access policy states "to be eligible for submission to the post-2014 REF, authors' final peer-reviewed manuscripts must have been deposited in an institutional or subject repository on acceptance for publication"[6]. It is necessary, therefore, to have a means of unequivocally recording the date of acceptance. Publishers normally provide this information.

It will also be necessary to record the date on which the resource described in HEFCE's policy was deposited into the repository. That date may be captured by your repository software automatically but at present there is no metadata element in RIOXX designed for that particular purpose.

rioxxterms:apc
Zero or one instance

Use of this element is: **Optional**

This element expresses whether or not *the resource* has an associated 'article processing charge'. The value of this element **MUST** be one of the following:
. Paid
. Partially waived
. Fully waived
. Not charged
. Not required
. Unknown

Comments
It is acknowledged that the subject of APCs is complex and the administration and recording systems vary across different institutions. The Jisc APC pilot project and the follow-on Jisc Monitor project is currently investigating how best to capture APC data on a wide scale. In recognition of the fluidity of this particular field, this element is optional.

---

[6] http://www.hefce.ac.uk/whatwedo/rsrch/rinfrastruct/oa/policy/

rioxxterms:author
One or more instances

Use of this element is: **Mandatory**

The author of ***the resource*** may be a person, organisation or service, but is most commonly a person. This element **SHOULD** take an optional attribute called *id*, which **MUST** contain an HTTP URI that uniquely identifies the author. Where there is more than one author, a separate *rioxxterms:author* element **MUST** be used for each. As many authors may be entered as required. The ideal use of this element is to include both an HTTP URI in the *id* attribute, and a text string in the body of the element, thus:

<rioxxterms:author id="http-uri-for-this-author-entity">
    name-of-this-author-entity
</rioxxterms:author>

Where the author is a person, the **RECOMMENDED** format is to add text in the form Last Name, First Name(s), and to include an ORCID ID, if known, in its HTTP URI form, e.g.

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

Where the author is an organisation, the **RECOMMENDED** format is to add the official name of the organisation, and to include an ISNI ID, if known, in its HTTP URI form, e.g.

<rioxxterms:author id="http://isni.org/isni/0000000419367988">
    University of Edinburgh
</rioxxterms:author>

Where the *rioxxterms:author* element appears multiple times for one record, it **CAN** be assumed that the order is significant, in that the first element describes the 'first named author' of ***the resource***. In order to make this more explicit, an extra attribute, *first-named-author*, **SHOULD** be used to indicate which of the *rioxxterms:author* elements describes the first named author of ***the resource***, thus:

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

rioxxterms:contributor
Zero or more instances

Use of this element is: **Optional**

This field is designed to describe an entity – for example the name of a person, organisation or service – responsible for making contributions to the content of *the resource*. As many rioxxterms:contributor elements may be entered as required. This element **SHOULD** take an optional attribute called *id*, which **MUST** contain an HTTP URI that uniquely identifies the contributor. The ideal use of this element is to include both an HTTP URI in the *id* attribute, and a text string in the body of the element, thus:

```
<rioxxterms:contributor id="http-uri-for-this-contributor-entity">
   name-of-this-contributor-entity
</rioxxterms:contributor>
```

Where the contributor is a person, the **RECOMMENDED** format is to add text in the form Last Name, First Name(s), and to include an ORCID ID, if known, in its HTTP URI form, e.g.

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

Where the contributor is an organisation, the **RECOMMENDED** format is to add the official name of the organisation, and to include an ISNI ID, if known, in its HTTP URI form, e.g.

```
<rioxxterms:contributor id="http://isni.org/isni/0000000419367988">
   University of Edinburgh
</rioxxterms:contributor>
```

rioxxterms:project
One or more instances

Use of this element is: **Mandatory**

This is designed to collect the project ID(s), issued by the funder(s), that relate to *the resource*, together with the name and/or global identifier for the funder(s).

The element **MUST** contain one project ID, an alphanumeric identifier provided by the funder in its original format. In cases where *the resource* has been funded internally, an appropriate internal code might be used.

The element takes two attributes, *funder_name* and *funder_id*. One or both of *funder_name* and *funder_id* **MUST** be supplied.

funder_name
The canonical name of the entity responsible for funding *the resource* **SHOULD**
be recorded here as text.

funder_id
A globally unique identifier for the funder of *the resource* **SHOULD** be recorded
here. An HTTP URI **MUST** be used for this. It is **RECOMMENDED** that one of the
following identifier schemes is used:
. An ISNI ID
. A DOI (in its HTTP URI form) made available by FundRef
.
Example
```
<rioxxterms:project
   funder_name="Engineering and Physical Sciences Research Council"
   funder_id="http://isni.org/isni/0000000403948681"
>
   EP/K023195/1
</rioxxterms:project>
or
<rioxxterms:project
   funder_name="Engineering and Physical Sciences Research Council"
   funder_id="http://dx.doi.org/10.13039/501100000266"
>
   EP/K023195/1
</rioxxterms:project>
```

Where *the resource* has been funded by more than one funder a separate
rioxxterms:project element **MUST** be added for each. Similarly, where several
project IDs provided by the same funder have been attached to *the resource*, a
separate *rioxxterms:project* element **MUST** be added for each.

This means that it is quite normal for a given *funder_name*, *funder_id* or *project_id*
to appear in multiple instances of the *rioxxterms:project* element in a single
RIOXX metadata record.

Comments
Project IDs are perhaps not typically collected in repositories; they may exist on
systems operated primarily by a university's research office. This element is
particularly important since it will enable funders to identify which grants are
associated with particular publications.

rioxxterms:publication_date
Zero or one instance

Use of this element is: **Optional**

This element takes the publication date of ***the resource*** in the form in which it would be cited. This element is not used in a RIOXX context but allows for a RIOXX record to become a reasonable bibliographic record for ***the resource***. This is a free-text field.

Examples:
<rioxxterms:publication_date>
   2011-02-23
</rioxxterms:publication_date>

<rioxxterms:publication_date>
   Spring, 2010
</rioxxterms:publication_date>

As RIOXX is primarily concerned with such issues as compliance with funders' mandates and licensing of open access publications, the critical dates for the assertion of compliance are those held in the *start_date* attributes of the *ali:license_ref* elements.

Comments
There is continuing uncertainty in the community about how to capture accurate publication date details since it is not always straightforward to find this information. This element provides the space to capture publication dates (which may then be used in combination with other data to produce relevant embargo information). The start date of the first license referenced via the license_ref element should equate to the publication date.

rioxxterms:type
One or more instances

Use of this element is: **Mandatory**

Type refers to the 'type' - the nature or genre of the content of ***the resource***. Take care not to confuse this with *dc:format*.

Values recorded here **MUST** be from the following controlled list of types:
. Book
. Book chapter
. Book edited
. Conference Paper/Proceeding/Abstract
. Journal Article/Review
. Manual/Guide
. Monograph

- Policy briefing report
- Technical Report
- Technical Standard
- Thesis
- Consultancy Report
- Working paper
- Other

rioxxterms:version
Exactly one instance

Use of this element is: **Mandatory**

This element indicates which 'version' of **the resource** is being described. The value of this element **MUST** be one of the following:
- AO
- SMUR
- AM
- P
- VoR
- CVoR
- EVoR
- NA

These terms are adopted from the <u>Journal Article Versions (JAV):</u> <u>Recommendations of the NISO/ALPSP JAV Technical Working Group</u> and have the following meanings:

- AO = Author's Original
- SMUR = Submitted Manuscript Under Review
- AM = Accepted Manuscript
- P = Proof
- VoR = Version of Record
- CVoR = Corrected Version of Record
- EVoR = Enhanced Version of Record
- NA = Not Applicable (or Unknown)

rioxxterms:version_of_record
Zero or one instance

Use of this element is: **Recommended**

This field **MUST** contain an HTTP URI that is a persistent identifier for the published version of **the resource**. If a DOI has been issued by the publisher then this *MUST* be used. Such a DOI **MUST** be represented in its HTTP form, for example:

```
<rioxxterms:version_of_record>
   http://dx.doi.org/10.1006/jmbi.1995.0238
</rioxxterms:version_of_record>
```

3. Helping you adopt these guidelines

The sponsors of these guidelines are committed to helping you adopt them. You are likely to be already collecting most of the mandatory metadata but you may need to think about the two additional fields (ProjectID and Funder Name) and where to source the information. If you do not already have this information your institution's Research Office may be able to supply this it. The RIOXX project team is working to agree access for the community to a new directory of unique funder names.

The RIOXX project is working with EPrints and DSpace developers to develop the applications necessary to facilitate the efficient capture of the required metadata. The goal is to make compliance with these metadata Guidelines as simple as possible.

The EPrints RIOXX plugin

The RIOXX plugin gives your EPrints repository the capability to expose its publication data using the RIOXX Application Profile. It works by adding the twenty-one metadata fields required by RIOXX to an EPrints repository. These fields allow you to capture the necessary data about each publication in order to create a RIOXX representation of a publication. The plugin will check the data entered in these fields against the Application Profile and provide warnings of any problems. You won't need to enter data twice: the plugin checks your existing data and derives a suitable value for RIOXX. Where your repository does not currently have a field for metadata required by RIOXX, such as dcterms:dateAccepted, the plugin creates a new dedicated field for capturing the date of acceptance and automatically inserts this field into the workflow when the plugin is installed.

Although the plugin can be installed from the EPrints Bazaar, we recommend you contact the organisation contracted by Jisc to provide technical support since some minor metadata mapping corrections may be required when the plugin is first installed. The relevant organisations are EPrints Services, ULCC and Digital Repository Services (for self-hosted repositories).

4 Frequently Asked Questions

The project team will collate questions, comments and other feedback and respond via the RIOXX blog; commonly recurring themes may be incorporated in these formal Guidelines periodically.

Why do we need RIOXX

The RIOXX Application Profile is designed to mitigate the detrimental effects of divergent interpretation of the standards that exist in the open access repository space by advocating a common approach. Adopting a common approach through the use of commonly-used guidelines has the potential, therefore, to reduce ambiguity, boost the extent to which metadata can be harvested efficiently, enhance the accuracy and value of services built on metadata harvesting and aggregation processes and improve confidence in the veracity of reports based on metadata.

Do I have to adopt the RIOXX Application Profile?

The development of RIOXX Application Profile was instigated and is being strongly supported by RCUK and more recently by HEFCE. The benefits for many stakeholders – researchers and other information consumers, funders and institutions – are attractive. Beyond the task of demonstrating compliance with funders' open access policies, better information discovery, higher quality statistical reporting, higher quality aggregations and the possibility of building new services will all flow from a consistent approach to collecting and exposing metadata in the UK's open access repositories. Working together, the UK's information management community can continue to promote the importance and usefulness of their open access repositories both within and beyond their own institutions. The Application Profile simply provides the means to help the research information management community pull in the same direction for the common good.

Are these guidelines supported by the community?

The RIOXX application profile and guidelines have been developed in consultation with interested parties in the community including the cooperation of UKCoRR. Jisc's V4OA project tested many of the key metadata elements with industry representative bodies; drafts of the RIOXX Application Profile have been made available for public comment. Discussions have taken place between people responsible for looking after other similar initiatives such as OpenAIRE.

Is RIOXX set in stone?

The final version of the Application Profile was published on 23rd January 2015 but suggestions to improve these Guidelines are always welcome. Feedback on any aspect of the Application Profile may be sent to the RIOXX development team. We would be very interested in hearing about your experience of implementing RIOXX. Given the dynamic nature of the sector and the initiative to develop vocabularies and associated metadata elements for open access, these Guidelines will in any case evolve over the course of 2015. There will be further opportunities for people to contribute to the ongoing development of the Guidelines. The basic elements will not change but where there is a need for perhaps greater clarity with respect to interpretation or additional examples, such issues will be addressed through updates to these Guidelines.