
Addition of “metadata” type in relatedInfo

Suggested schema change

Provide the ability for the relatedInfo element to point to richer metadata formats used to describe the collection/party/service/activity.

Problem this suggestion addresses

RIF-CS provides the ability to describe research data at a generic level. As data tools become more prevalent, the need to access more specific metadata becomes more useful. A registry such as RDA provides the ability to quickly locate useful research data. Enhancing this by then providing links to other metadata formats would then provide software tools with the ability to “dig deeper”.

An example user/software workflow might appear as:

1. User searches RDA (or similar registry such as TERN’s Central Portal) for datasets within a spatial area using the web interface or via an API;
2. A result set is returned and the user selects 4 records of interest – three of these records have more specific metadata available (in this example, EML, the Ecological Markup Language)
3. The user’s research tools can ingest EML and use it in synthesis work so the user downloads the three EML files into their tool.

An aggregator may undertake a process such as:

1. Harvest a repository via OAI-PMH for the RIF-CS records
2. For each record that has a link to the EML¹, the individual EML file is then downloaded by the aggregator
3. The EML information can be used to markup the aggregator’s user interface.

Identified by

Duncan Dickinson (QCIF Project Manager – TERN Central Portal, ReDBox & JCU Tropical Data Hub)

RIF-CS schema components affected

¹ Multi-disciplinary repositories will have a mix of discipline-specific metadata formats so not all records will have EML.

RIF-CS Schema Change Suggestion CC-162

This functionality is likely to best sit within the relatedInfo element

A sample entry would appear as:

```
<relatedInfo type="metadata">
  <title>EML</title>
  <identifier=http://data.example.com/meta-eml.xml type="uri"/>
  <notes>This is the EML version of the metadata</notes>
</relatedInfo>
```

Please note that neither the title nor the notes elements contain a specific vocabulary and are not relied on in determining the format or namespace of the identified metadata.

Supplemental option

The system retrieving the RIF-CS record from the registry could then download one or more of the related metadata files for processing. The addition of an optional "schema" property would allow such a system to limit these downloads to only the specific metadata schema required and understood by the software (if it's made available). This would change the XML above to:

```
<relatedInfo type="metadata">
  <title>EML</title>
  <identifier=http://data.example.com/meta-eml.xml type="uri"
  schema="eml://ecoinformatics.org/eml-2.1.0"/>
  <notes>This is the EML version of the metadata</notes>
</relatedInfo>
```

Impact on content providers

As this element is optional, its addition should pose little (if any) change management concerns. The primary change would require the addition of a vocabulary item with the Related Information Type vocabulary.

The supplemental option would require the addition of an optional property in the Identifier element and, again, this is an opt-in.

Pros

RIF-CS Schema Change Suggestion CC-162

This will allow RIF-CS to provide a link to more specific metadata that can then be consumed by software tools either in an automated download fashion or via the user. This will aid in data reuse as the more specific metadata provides a richer context.

The examples given here focus primarily on use within a collection description however, it would be applicable to the other registryObjects:

- Party records could point to EAC-CPF or FOAF versions
- Activity records could point to a DOAP² version
- Service records could point to a SensorML version

Whilst this functionality could be achieved by undertaking an OAI-PMH harvest across the various available metadata formats, this would require increased software development that is easily resolved through the provision of this direct link in the RIF-CS relatedInfo element.

Cons

If a site chooses to provide this relatedInfo, they will need to provide the additional metadata. This is a minor concern as most research data repositories automatically create renditions in various metadata formats. Some small amount of work would be needed to edit the templates and add the new references to other metadata formats. This is no different to the work required in any update to a new schema.

Some repositories may need to provide the functionality to upload the EML for a record, especially if the repository is at an institutional level and the EML is not publicly accessible.

The “schema” property in the Identifier field may be seen as a bit of a kludge but it is an opt-in for ease-of-use at the client end.

Technical options

Option 1: Addition of the “metaData” type to the relatedInfo type vocabulary.

The new relatedInfo type “metaData” will need to be added to the ORCA registry “tbl_terms” database table.

The vocabs.xml file will need to be amended to add the new description types and definitions.

The vocabularies.html will need to be regenerated to reflect the addition of the new description types.

Changes will be required to the Content Providers Guide.

² <https://github.com/edumbill/doap/wiki>

RIF-CS Schema Change Suggestion CC-162

Option 2: Addition of the “metaData” type to the relatedInfo type vocabulary and addition of the “schema” attribute to the relatedInfo element.

The new relatedInfo type “metaData” will need to be added to the ORCA registry “tbl_terms” database table.

The vocabs.xml file will need to be amended to add the new description types and definitions.

The vocabularies.html will need to be regenerated to reflect the addition of the new description types.

Schema change to the identifier element as follows:

```

<xsd:complexType name="identifierType">
  <xsd:simpleContent>
    <xsd:extension base="xsd:string">
      <xsd:attribute name="type" use="required" type="xsd:string">
        <xsd:annotation>
          <xsd:documentation>A value taken from a controlled vocabulary indicating the
            type of identifier.</xsd:documentation>
        </xsd:annotation>
      </xsd:attribute>
      <xsd:attribute name="schema" use="optional" type="xsd:anyURI">
        <xsd:annotation>
          <xsd:documentation>A value describing uri of schema.</xsd:documentation>
        </xsd:annotation>
      </xsd:attribute>
    </xsd:extension>
  </xsd:simpleContent>
</xsd:complexType>

```

Note: if the schema attribute requires a controlled vocabulary then this vocabulary would need to be added to both the tbl_vocab and tbl_terms database tables. The vocabs.xml file will require updating and vocabularies.html will also need to be regenerated.

As the use of both the new vocab and new attribute is optional there are no backward compability issues recognised.