Exploiting the value of Dublin Core through pragmatic development
which are you?

• an information scientist/researcher

• an information professional or practitioner

• a software or systems developer
which are you?

• an information scientist/researcher
• an information professional or practitioner
• a software or systems developer
Eric Miller introduced some broad themes in his keynote last year at this conference. I’m going to really narrow the focus...
1. application profiles
2. learning from software development practice
3. working openly
1. application profiles
application profiles

“Application profiles consist of data elements drawn from one or more namespace schemas combined together by implementors and optimised for a particular local application.”

Heery & Patel: Application Profiles: Mixing and Matching Metadata Schemas
range of application profile types

a given standard (e.g. Dublin Core)

more general

application profile which focusses the standard on a particular domain

very specific

application profile which narrows the focus of the standard to a specific problem
• a small 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!!

http://www.rioxx.net
the focus of today’s talk

a given standard (e.g. Dublin Core)

more general

application profile which focuses the standard on a particular domain

very specific

application profile which narrows the focus of the standard to a specific problem
information modelling

software implementation
Dublin Core is infrastructure
most software is not designed as infrastructure
usually, software must evolve, or become extinct....
involve software developers & learn how they work
2. learning from software development
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 this to application-profile development

• **Individuals and interactions** over processes and tools
  • doing what works - and what makes sense to the user ✓

• **Working software** over comprehensive documentation
  • an application profile is fundamentally a set of documentation! ✗

• **Customer collaboration** over contract negotiation
  • working as closely with users as possible ✓

• **Responding to change** over following a plan
  • iterative - short development cycles punctuated by review ✓
transferable Agile techniques

• iterative design and development with users
  • high-bandwidth interaction with users

• short iterations or ‘sprints’
  • documentation can be made this way just as with code

• MVP

• ‘pave the cowpaths’

• continuous testing **during** development (and after!)
  • **testing aids development and understanding**
iterative design and development with users
it’s not a marathon!
Minimum Viable Product (MVP)

• start with the simplest thing that could possibly work

• only revise an application profile in response to real changing requirements

• ‘interoperability’ is not the goal of a focussed application profile
‘paving the cowpaths’
continuous testing

- **extremely** important

- should be mechanistic, or semi-automated, wherever possible
  - so that it actually gets done!

- should deliver immediate and useful feedback

- **not just** the usual XML schema validation - this is often important, but it is not enough
Validation is done in two stages - first a basic check for valid basic RIOXX syntax, and then a strict check for conformance to the full RCUK requirements.

<table>
<thead>
<tr>
<th>Name</th>
<th>OAI-PMH Base URL</th>
<th>Software</th>
<th>Basic RIOXX?</th>
<th>Full RCUK?</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Research Online</td>
<td><a href="http://oro.open.ac.uk/cgi/oai2">http://oro.open.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.1.2.1)</td>
<td>Valid! (100%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>SAS-SPACE</td>
<td><a href="http://sas-space.sas.ac.uk/cgi/oai2">http://sas-space.sas.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.1.2.1)</td>
<td>Valid! (100%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>Enlighten</td>
<td><a href="http://eprints.gla.ac.uk/cgi/oai2">http://eprints.gla.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.1.13)</td>
<td>Partially (90%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>Keele Research Repository</td>
<td><a href="http://eprints.keele.ac.uk/cgi/oai2">http://eprints.keele.ac.uk/cgi/oai2</a></td>
<td>EPrints ()</td>
<td>Partially (90%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>SOAS Research Online</td>
<td><a href="https://eprints.soas.ac.uk/cgi/oai2">https://eprints.soas.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.2.4)</td>
<td>Partially (90%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>Falmouth University Research Repository (FURR)</td>
<td><a href="http://repository.falmouth.ac.uk/cgi/oai2">http://repository.falmouth.ac.uk/cgi/oai2</a></td>
<td>EPrints ()</td>
<td>Partially (40%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>LSTM Online Archive</td>
<td><a href="http://archive.lstmliverpool.ac.uk/cgi/oai2">http://archive.lstmliverpool.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.2.3)</td>
<td>Partially (40%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>Central Archive at the University of Reading</td>
<td><a href="http://centaur.reading.ac.uk/cgi/oai2">http://centaur.reading.ac.uk/cgi/oai2</a></td>
<td>EPrints (3)</td>
<td>Partially (20%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>Insight</td>
<td><a href="http://insight.cumbria.ac.uk/cgi/oai2">http://insight.cumbria.ac.uk/cgi/oai2</a></td>
<td>EPrints (3)</td>
<td>Partially (20%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>RaY - Research at York St John</td>
<td><a href="http://ray.yorksj.ac.uk/cgi/oai2">http://ray.yorksj.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.3.12)</td>
<td>Partially (20%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>St George's Online Research Archive</td>
<td><a href="http://openaccess.sgul.ac.uk/cgi/oai2">http://openaccess.sgul.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.3.6)</td>
<td>Partially (20%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>University of Salford Institutional Repository</td>
<td><a href="http://usir.salford.ac.uk/cgi/oai2">http://usir.salford.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.2.3)</td>
<td>Partially (20%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>WestminsterResearch</td>
<td><a href="http://westminsterresearch.wmin.ac.uk/cgi/oai2">http://westminsterresearch.wmin.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.2.3)</td>
<td>Partially (20%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>Central Lancashire Online Knowledge</td>
<td><a href="http://clok.uclan.ac.uk/cgi/oai2">http://clok.uclan.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.3.7)</td>
<td>Partially (10%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>ROAR at University of East London</td>
<td><a href="http://roar.uel.ac.uk/cgi/oai2">http://roar.uel.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.3.8)</td>
<td>Partially (10%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
<tr>
<td>University of Chichester EPrints Repository</td>
<td><a href="http://eprints.chi.ac.uk/cgi/oai2">http://eprints.chi.ac.uk/cgi/oai2</a></td>
<td>EPrints (3.3.12)</td>
<td>Partially (10%)</td>
<td>Invalid (0%)</td>
<td>Report</td>
</tr>
</tbody>
</table>
continuous testing (RIOXX) - reporting
3. working openly
encouraging (& avoiding barriers to) participation

Public review

**RIOXX 2.0 (beta 1)** is now available for public review. We anticipate making at least one further revision following your feedback - this can be offered either publicly in the comments on this blog post (preferred) or more privately to the rioxx admin email address: admin@rioxx.net (also OK).

We look forward to your comments!

Comments

Comments for this thread are now closed.

45 Comments  RIOXX Website

Valerie McCutcheon  a year ago

Hi Paul,

Just to get it in the official comments remember our tweet conversation:

Identifier - allows direct open of the document, VOR gives link to official version. - Identifier cld be VOR?

You said you had wondered if that could happen and I said I had stuck it on my list to make sure we get a clear explanation - so we know for implementing and for explaining to users.
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:

1. Separate fields for Funder and ProjectID, each separately is multi-value \([1..n]\)
2. Separate fields for Funder and ProjectID, each separately multi-value \([1..n]\) & ordering is significant
3. Composite field for Funder and ProjectID, together - field is multi-value \([1..n]\)
4. Syntactically rich string containing Funder and ProjectID components - field is multi-value \([1..n]\)
5. Globally unique project ID - funder can be dereferenced from this - field is multi-value \([1..n]\)

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
conclusion: we need to work together.

look around you - this is the community to make it happen!
DCMI’s next anniversary (21) is another important birthday in some places…

Saúde!
thanks for listening!
(Obrigado pela atenção)
Credits

• All images are my own except for:

• Sprinters:
  • https://upload.wikimedia.org/wikipedia/commons/b/be/
• Caipirinha:
  • https://upload.wikimedia.org/wikipedia/commons/9/92/
• Pave the cowpaths:
  • www.flickr.com/photos/wetwebwork/2847766967/
• Software developers:
  • https://commons.wikimedia.org/wiki/File:Hackathon_TLV_2013_-_31.jpg
• Dodo:
  • https://commons.wikimedia.org/wiki/File:Dodo_(15574061408).jpg
• Kanban:
  • https://flic.kr/p/gKbsnf