PIDs in Dublin Core

PIDapalooza 2019, Dublin

Paul Walk

Founder and Director, Antleaf
Director, Dublin Core Metadata Initiative (DCMI)

Web: http://www.paulwalk.net
Email: paul@paulwalk.net

www.antleaf.com

www.dublincore.org

innovation in metadata design, implementation & best practice

Join DCMI!

Dublin Core Metadata Initiative

The Dublin Core Metadata Initiative supports innovation in metadata design and best practices. DCMI is supported by its members and is a project of ASIS&T.

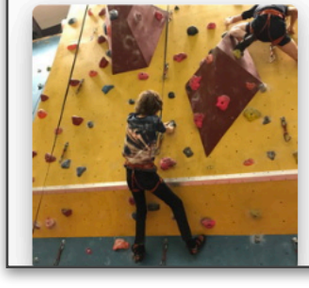
Stewardship

For more than twenty years, the DCMI community has developed and curated Dublin Core Specifications. More recently, DCMI has become recognised as a trusted *steward* of metadata vocabularies, concept schemes and other metadata artefacts, and has taken responsibility for other community-created specifications. DCMI remains committed to this important work, and is actively developing more efficient and sustainable approaches to the stewardship of these standards, through the work of the DCMI Usage Board.

Community

DCMI is defined by its community which is responsible for the innovative developments and evolving good practices which DCMI shares with the world. Much of DCMI's work is organised in working and interest groups. DCMI's community is and has always been international, with active participants from around the world. The primary community event is the DCMI Annual Conference. DCMI also organises regular webinars (open to DCMI members), given by members of the community wishing to share their expertise with like-minded peers. Finally, DCMI collaborates with a number of other organisations.

Learning



DCMI supports teachers and learners of modern metadata technologies and practices. An updated Metadata Basics page highlights current trends in descriptive metadata in the style of

Development

DCMI has a long history of fostering and supporting technical development and innovation through the activities of its community, often in partnership with other

News

DCMI to Maintain the Bibliographic Ontology (BIBO)

The Dublin Core Metadata Initiative is pleased to announce that it has accepted responsibility for maintaining the Bibliographic Ontology (BIBO). BIBO is a long-established ontology, is very stable and is widely used in bibliographic linked-data. We are also very pleased to announce that Bruce D'Arcus has joined the DCMI Usage Board, and will act as the point of contact for issues relating to BIBO. Bruce had this to say about the decision:

read more...

This morning I will talk about....

1. The problem
2. The process to find a solution
3. The recommended solution

1. The problem

How to represent PIDs in XML-based metadata using Dublin Core properties?

2. The investigation

We asked the community for their ***user stories***

As an **<ACTOR>**,

I would like to **<DO SOMETHING>**,

in order to achieve **<SOME BENEFIT>**

crowd-sourcing **requirements**,
rather than ~~opinions~~

Filters | is:issue is:open | Labels | Milestones | New issue

16 Open | 0 Closed | Author | Labels | Projects | Milestones | Assignee | Sort

- **Alignment with Draft recommendation: Supporting ORCID in repository systems** — 1
  #16 opened on 31 Oct 2018 by bram-atmire
- **PIDs in machine-actionable DMPs** Consumer Supplier user-story — 3
  #15 opened on 6 Sep 2018 by TomMiksa
- **Person identifiers in respositories exported in metadata to third parties** Consumer Supplier candidate solution question or issue user-story — 3
  #14 opened on 4 Sep 2018 by MonicaDukeJisc
- **An updated DC standard would allow vendors and customers easier pathways to 'properly' share data** Supplier user-story — 1
  #13 opened on 3 Sep 2018 by jesusbagpuss
- **Ensuring Dublin Core metadata can be expressed as Linked Data** Consumer Supplier user-story
  #12 opened on 3 Sep 2018 by paulwalk
- **Digitisation of historical thesis collections** Supplier user-story
  #11 opened on 3 Sep 2018 by AlasdairMacDonald
- **New additions to the institutional repository** Supplier user-story
  #10 opened on 3 Sep 2018 by AlasdairMacDonald
- **Catalogue user** Consumer user-story
  #9 opened on 3 Sep 2018 by AlasdairMacDonald
- **Human- and machine- readable data** Supplier candidate solution question or issue user-story — 3
  #8 opened on 30 Aug 2018 by jesusbagpuss
- **Standards for implementation** Supplier user-story
  #7 opened on 30 Aug 2018 by jesusbagpuss

http://bit.ly/pids-in-dc

Synthesis of these Requirements

1. **Must allow the expression of a PID together with a string literal**
   - It may be useful to express both an identifier and a string - e.g. both the PID and name of an author - using the same property.

2. **PIDs must be easy to parse from the XML**
   - The PID must be expressed either in the content or in an attribute of an XML element.

3. **No restrictions on which PIDs can be used**
   - However, there is a strong trend towards using HTTP URIs and some significant advantages in doing so.

4. **The provider of the PID must be identifiable**
   - Not only must PIDs be identifiable as such, but the provider of the PID must also be identifiable.

5. **The provenance and authenticity of PIDs should be indicated**
   - Some application profiles require and provide this kind of information.

The DCMI position

1. **Supporting semantic interoperability**
   - has always promoted interoperability as a core concern
   - semantic interoperability through RDF as a common data model
   - metadata expressed in XML should be translatable into RDF as necessary

2. **Using HTTP URIs**

3. **Being namespace agnostic (/elements/1.1/ or /terms/).**

4. **Simplicity**
   - The recommended approach should be as simple as possible

5. **Conventional**
   - The recommended approach should be generally applicable, rather than be tied to specific domains. This gives the approach a better chance of being widely adopted as a convention, which will increase the rate of implementation.

It becomes apparent that the problem is better expressed as:

How to express PIDs together with associated string-literals?

3. The recommended solution

Candidate solutions

- 2 candidate solutions identified
- these were discussed at a special session at the 2018 DCMI annual conference in Porto

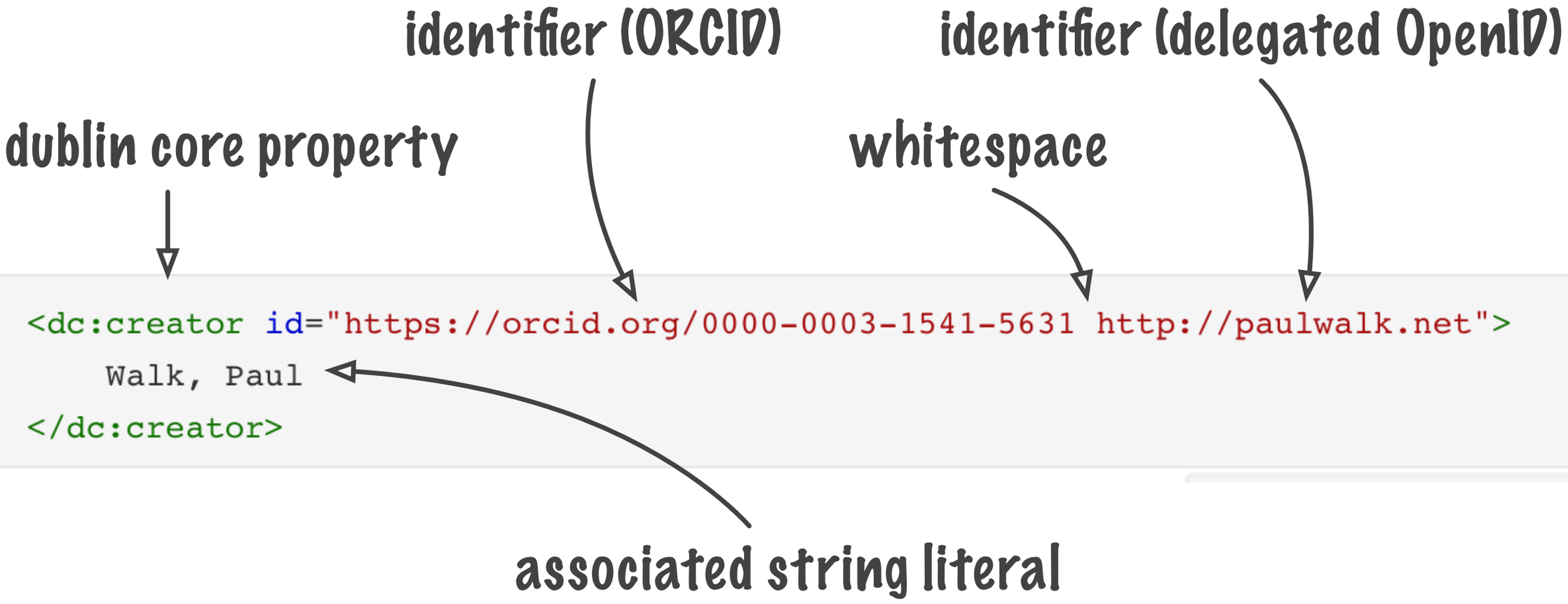
1. To use an attribute (`id=`) to hold a PID for an XML element with a literal value:

```
<dc:creator id="https://orcid.org/0000-0003-1541-5631">Walk, Paul</dc:creator>
```

2. To use one or more DC Identifier elements in a nested description of the entity in question:

```
<dc:creator>
    <dc:identifier>https://orcid.org/0000-0003-1541-5631</dc:identifier>
    <dc:identifier>http://paulwalk.net</dc:identifier>
    <foaf:name>Walk, Paul</foaf:name>
</dc:creator>
```

Recommended solution



Next steps

- Draft specification detailing how the proposed construct can be expressed with JSON, or with JSON-LD or RDF/XML for publication as Linked Data.
- The draft specification will be posted for public comment and, ideally, publication as a DCMI Recommendation.
- Some work needed to examine the requirement for provenance tracking and 'authentication' of PIDs

Thank you!

Paul Walk

Founder and Director, Antleaf
Director, Dublin Core Metadata Initiative (DCMI)

Web: http://www.paulwalk.net
Email: paul@paulwalk.net

www.antleaf.com

www.dublincore.org