DC-SRAP: Metadata Application Profile for Academic Repositories

Dublin Core **S**cholarly **R**esources **A**pplication **P**rofile

Osma Suominen (chair), Jan Ashton, Karen Coyle, Alasdair MacDonald, Freddy Sumba & the DC-SRAP Working Group

Motivation

- To enable the detailed description of scholarly resources, such as **theses**, **doctoral dissertations** and **academic articles** with Dublin Core Metadata Terms
- To base it on existing practice, guidelines and standards (e.g. UKOLN SWAP, British Library ETHOS, MARC21 and Finnish national guidelines)

What is Missing from Dublin Core Terms?

- Author **affiliations**: which institution an author represents in this publication
- Detailed **contributor roles**: editor, illustrator, supervisor, reviewer, opponent...
- Publication **types**: thesis, dissertation, preprint, technical report... (COAR Resource Types vocabulary)
- **Inclusion** in a containing publication: volume, issue, number, page numbers, containing monograph or periodical
- How to express **embargo** dates
- Guidance about **persistent identifiers**: DOI, URN, ISBN, ISSN, ORCID, ISNI...
- **Academic context**: University, faculty, department, project or grant number...
- Linking to **related code** and **data sets**

Use Cases for Metadata about Scholarly Resources

- **Discovery and Findability**: Make the resources searchable and interlinkable using various criteria
- **Aggregation**: Standardized metadata formats and practices make it practical to combine different data sources
- **Harvesting**: Collect metadata about publications using different criteria
- **Statistics**: Calculate publication statistics by type, organization or unit, project...
- **Automation**: Use machine learning and AI techniques to extract standard, structured information from publications

*Elements from outside Dublin Core Terms (e.g. BIBO, FOAF) are **highlighted in blue***
*SRAP-specific additions are **highlighted in red***

Application Profile using DCTAP

shapeID	propertyLabel	propertyID
Scholarly Resource	Type	dct:type
	Creator	dct:creator
	Contributor	dct:contributor
	Language	dct:language
	Publisher	dct:publisher
	Subject	dct:subject
	Title	dct:title
	Format	dct:format
	Date Issued	dct:issued
	Date Modified	dct:modified
	Date Accepted	dct:dateAccepted
	Embargo Date Range	**srap:embargoDateRange**
	Date Retracted	**srap:dateRetracted**
	Volume	**bibo:volume**
	Issue	**bibo:issue**
	Start page	**bibo:pageStart**
	End page	**bibo:pageEnd**
	Is Part Of	dct:isPartOf
	Presented At	**bibo:presentedAt**
	Bibliographic Citation	dct:bibliographicCitation
	Relation	dct:relation
	Abstract	dct:abstract
	Table of Contents	dct:tableOfContents
	Description	dct:description
	Access Rights	dct:accessRights
	License	dct:license
	Rights	dct:rights
	Rights Holder	dct:rightsHolder
	Accessibility statement	**srap:accessibilityStatement**
	Grant number	**srap:grantNumber**
	Identifier	dct:identifier
	ISBN	**bibo:isbn**
	Related code	**srap:relatedCode**
	Related dataset	**srap:relatedDataset**
Periodical	Class	rdf:type
	Title	dct:title
	Publisher	dct:publisher
	ISSN	**bibo:issn**
	E-ISSN	**bibo:eissn**
Book	Class	rdf:type
	Title	dct:title
	Publisher	dct:publisher
	Date	dct:date
	ISBN	**bibo:isbn**
	Contributor	dct:contributor
Person	Class	rdf:type
	Name	**foaf:name**
	Role	**srap:role**
	Affiliation	**schema:affiliation**
	Identifier	dct:identifier
Organization	Class	rdf:type
	Name	**foaf:name**
	Role	**srap:role**
	Identifier	dct:identifier

Property namespaces

dct:	http://purl.org/dc/terms/
bibo:	**http://purl.org/ontology/bibo/**
foaf:	**http://xmlns.com/foaf/0.1/**
schema:	**https://schema.org/**
srap:	**(not defined yet, may go in DCTerms?)**

Creator and Contributor Roles

Recommended practice is to use the Library of Congress MARC Relator codes or other controlled list to express roles.

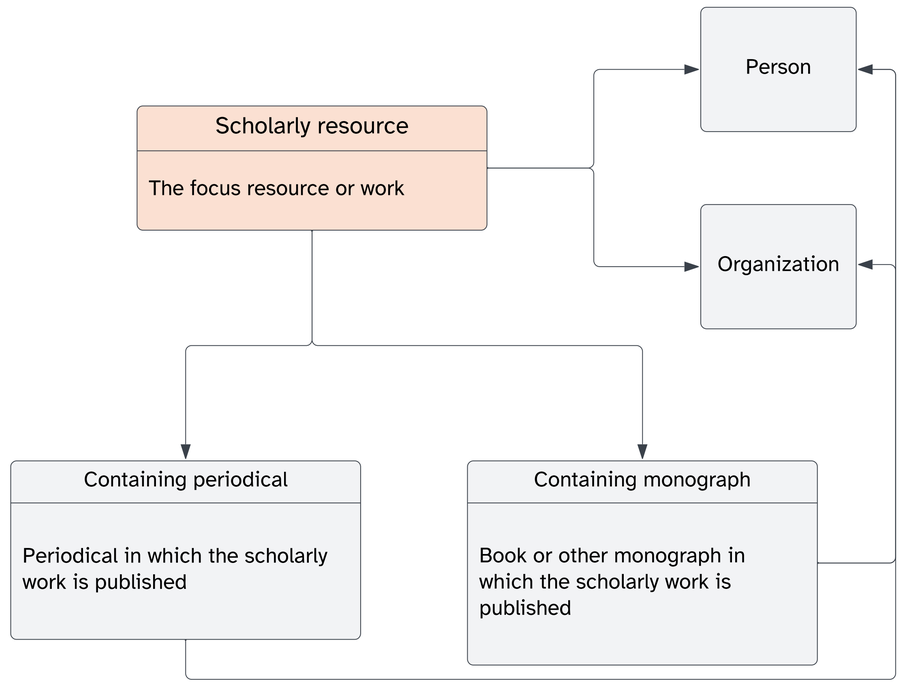
Examples of roles for theses and dissertations

- Degree granting institution
- Degree supervisor
- Praeses
- Respondent
- Thesis advisor

Examples of roles for journal articles

- Editor
- Funder
- Host institution
- Reviser
- Sponsor
- Translator

Data Model Overview



Example 1: Journal Article

a) Rich entity-oriented metadata

```
rdf:type bibo:AcademicArticle ;
dct:title "COVID-19 and multiorgan response" ;
dct:creator "Sevim Zaim", "Jun Heng Chong",
            "Vissagan Sankaranarayanan", "Amer Harky" ;
dct:issued "2020" ;
dct:identifier <https://doi.org/10.1016/j.cpcardiol.2020.100618> ;
bibo:volume "45" ;
bibo:issue "8" ;
bibo:pageStart "100618" .
```

dct:isPartOf

```
rdf:type bibo:Journal ;
dct:title "Current problems in cardiology" ;
bibo:issn "0146-2806" ;
bibo:eissn "1535-6280" .
```

b) Simple, flat metadata

```
rdf:type bibo:AcademicArticle ;
dct:title "COVID-19 and multiorgan response" ;
dc.creator "Sevim Zaim", "Jun Heng Chong",
           "Vissagan Sankaranarayanan", "Amer Harky" ;
dct:.issued "2020" ;
dct:identifier <https://doi.org/10.1016/j.cpcardiol.2020.100618> ;
bibo:volume "45" ;
bibo:issue "8" ;
bibo:pageStart "100618" ;
dct:isPartOf "Current problems in cardiology" .
```

Unable to express in the simple flat format:

- *periodical type, e.g. journal or conference proceedings*
- *periodical identifiers (ISSN, e-ISSN)*
- *inclusion in more than one containing publication*

Example 2: Doctoral Thesis

a) Rich entity-oriented metadata

```
dct:title "Faster than real-time simulation of [...]" ;
dct:creator "Malysheva, J." ;
dct:issued "2021-06-01" ;
bibo:isbn "978-952-335-653-5" ;
dct:publisher "Lappeenranta-Lahti Univ. of Tech." ;
dct:type coar_rt:c_db06 . # doctoral thesis
```

dct:contributor

```
foaf:name "Ellman, A." ;
dct:identifier <https://orcid.org/0000-0002-2158-838X> ;
schema:affiliation "Tampere University" ;
srap:role marcrel:opn . # official opponent
```

dct:contributor

```
foaf:name "Handroos, H." ;
dct:identifier <https://orcid.org/0000-0002-9479-0968> ;
schema:affiliation "Lappeenranta-Lahti Univ. of Tech." ;
srap:role marcrel:dgs .  # degree supervisor
```

b) Simple, flat metadata

```
dct:title "Faster than real-time simulation of [...]" ;
dct:creator "Malysheva, J." ;
dct:issued "2021-06-01" ;
bibo:isbn "978-952-335-653-5" ;
dct:publisher "Lappeenranta-Lahti Univ. of Tech." ;
dct:type coar_rt:c_db06 ; # doctoral thesis
dct:contributor "Ellman, A.", "Handroos, H." .
```

Unable to express in the simple flat format:

- *contributor identifiers (e.g. ORCID)*
- *contributor roles*
- *contributor affiliations*

What about Qualified DC and the dot syntax?

- The flat **Qualified Dublin Core** model (QDC) and the dot syntax used in HTML (*dc.date.issued*, *dc.title.alternative*) have been largely superseded since the 2000s by newer entity-oriented, XML or RDF based styles of DC metadata.
- **QDC is widely used** in repository systems, content management systems etc.
- SRAP WG is looking at possibilities of **expressing SRAP using QDC** to make it possible to use it in legacy systems.

Challenges

- How to extend scholarly metadata elements in a clean and simple way that fits in with RDF practices and existing Dublin Core Terms?
- How to support legacy systems, including repository systems, that only support traditional, flat Dublin Core metadata (Qualified DC, DC-XML, OAI-DC)?
- Avoiding and solving impedance mismatches when reusing existing metadata representations such as BIBO, MARC21, Schema.org and RDA

Current Status of DC-SRAP

- In development since 2021 in the DC-SRAP Working Group.
- The Dublin Core **Usage Board** will consider it later in 2024, including its impact on Dublin Core Terms.
- SRAP will have an indirect impact on the ISO standard **ISO 15836 The Dublin Core metadata element set**

DublinCore

16 40 NATIONAL LIBRARY OF FINLAND

BRITISH LIBRARY

THE UNIVERSITY of EDINBURGH

cedia

Documentation
https://dcmi.github.io/dc-srap/

Development on GitHub
https://github.com/dcmi/dc-srap