Metadata Tags for Academic Publications

Contents

Introduction

In order to facilitate electronic searching and cataloguing of publication data — be they citations, abstracts, or the full text — it is very helpful to present the identifying information in a standard format. The identifying information could include, for example, the title of an article, the name of a conference, or the name of an author. Such identifying information can be termed "metadata" — literally meaning 'data that describe the data'.

For example, it would be useful to be able to distinguish searches on the term "water" according to its appearance in the title of an article, the name of a conference, or the name of an author, for example. To do this, the identifying information ("metadata") can be annotated with consistent labels ("tags") *e.g*. `art_title`, `conf_name`, and `auth_name`. However, if the identifying information were haphazardly stored without any annotation, or even using nonstandard annotation (*e.g*. "name-of-article"), that becomes impossible.

As yet there is no *universal* standard.

HTML `meta` tags

The HTML `meta` element

HTML (hypertext mark-up language) is the prevailing standard for formatting text presented on websites. HTML also contains the ability to include metadata — identifying information about the particular web page or content — through the use of the `meta` tag. Metadata contained within the `meta` tag will be discoverable, but not rendered on the screen. The syntax for the HTML `meta` tag is

```
<meta name="author" content="Jane Doe">
```

The `name` and `content` attributes should be set pairwise.

The `meta` tags always go inside the head element.

Previously a `scheme` attribute could be set in the meta tag, such as

```
<meta name="identifier" content="0-2345-6634-6" scheme="ISBN">
```

but this is now *obsolete* in the latest HTML standard (HTML5).

Although the `content` attribute can contain practically any text, to be fully compliant with the HTML5 standard: "The name specified must either be a standard metadata name defined in the HTML5 specification or a registered extension to the predefined set of metadata names".

In principle, the metadata encoded with `meta` would not also be able to be expressed using the `title`, `base`, `link`, `style`, and `script` elements that are already have dedicated components of the `head` element. However, note that the standalone `author` element within the `head` element is designed to specify the author of the web page — not the author of a publication cited on the web page!

The set of defined metadata names in HTML5 is just: application-name, author, description, generator, and keywords. Note that, like the standalone `author` element within the `head` element — the standard `author` metadata name must be "a free-form string giving the name of one of the page's authors." Thus it too is *not* intended to specify the author of a publication cited on the web page!

A registered metadata name is any metadata name registered in the central MetaExtensions registration page.

"Extensions to the predefined set of metadata names may be registered in the WHATWG Wiki MetaExtensions page." Anyone is free to edit the WHATWG Wiki MetaExtensions page at any time to add or amend a metadata name, provided that information consistent with the all of the required definitions (Keyword, Brief description, Specification, Synonyms, and Status) is provided.

In XHTML (extensible HTML) the <`meta`> tag must be properly closed. For example

```
<meta name="author" content="Jane Doe"/>
```

or

```
<meta name="author" content="Jane Doe"></meta>
```

Formally the `meta` tag has no end tag in HTML. (However, it may not cause problems to include one anyway, using the 'minimised' syntax of the first snippet above.)

Names are case-insensitive, and must be compared in an ASCII case-insensitive manner.

Popular standards for HTML `meta` citation tags

There are five commonly used standards for HTML tags relating to academic publications:

1. Highwire Press/Google Scholar `citation_*`
2. Dublin Core `dc.*` (also `dcterms.*`, which is practically equivalent, albeit less common, and semantically preferred)
3. PRISM `prism.*`
4. Eprints `eprints.*`

5. BE Press `bepress_*`

Google Scholar supports all five of these. Mendeley supports all except for BE Press. In both of these cases, providing tags from more than one of the above sets is not a problem — or even recommended. SharePoint 2013 mentions only three of the above namespaces: Highwire Press (citation_*), Eprints (eprints.*), and Dublin Core (DC.*).

`dc:*` and `dcterms:*`, with colons instead of dots, are used in XML (rather than HTML). Such as

```
<dc:creator>Stone J. E.</dc:creator>
```

They may perhaps(?) also be used in a hybrid 'HTML–RDFa' syntax as in

```
<element property="">
<dc:creator content="Stone J. E.">
```

The above list of metadata namespaces is arranged roughly in descending order of popularity. (In 2012 the Dublin Core namespaces were probably most popular, which is unsurprising given that it seems to predate the others and was developed as a public standard, rather than for proprietary use.)

All of these are specifically designed to describe academic publications (especially journal articles), with the exception of Dublin Core, which is a general namespace.

Currently only Highwire Press (`citation_*`) and Dublin Core (`dc.*` and `dcterms.*`) name attributes are listed, and all with the status of "Proposal" (formally should have been "Proposed"). As of November 2018 *none* of the listed extensions to the predefined set of metadata names had the "Ratified" status. Ideally only ratified metadata names would be used. However, both proposed and ratified names are acceptable:

"Conformance checkers [such as HTML validators https://wiki.whatwg.org/wiki/Talk:MetaExtensions#Property_list_revision] may use the information given on the WHATWG Wiki MetaExtensions page to establish if a value is allowed or not: values defined in this specification or marked as "proposed" or "ratified" must be accepted, whereas values marked as "discontinued" or not listed [...] must be reported as invalid. [...]

When an author uses a new metadata name not defined by either this specification or the Wiki page, conformance checkers should offer to add the value to the Wiki, with the details described above, with the "proposed" status."

PRISM 3.0 — version 3.0 of the "Publishing Requirements for Industry Standard Metadata" — includes eight separate metadata specifications:

- PRISM Advertising Metadata 3.0 `prism-ad:*`
- PRISM Basic Metadata 3.0 `prism:*`
- PRISM Dublin Core Metadata 3.0 `dc:*` (with optional `prism:*` attributes)
- PRISM Image Metadata 3.0 `pmi:*`
- PRISM Recipe Metadata 3.0 `prm:*`
- PRISM Usage Rights Metadata 3.0 `pur:*`
- PRISM Crafts Metadata 3.1 `pcm:*`
- PRISM Contract Management Metadata 3.1 `pcmm:*`

Some other namespaces, most notably PSV and PAM, are also associated with PRISM.

HTML `meta` tags for academic publications

Table of `name` attributes

To aid comparison and expedite quick reference, in the following table the `name` attributes are organised by function.

Concept	Highwire Press [1][2][3][4][5][6][7][8]	Dublin Core [1]	PRISM [1][2][3][4]	Eprints [1][2][3]	BE Press [1][2]
	citation_*	**dc.* / dcterms.***	**prism.* and pur.***	**eprints.***	**bepress_***
Names of author(s)	citation_author citation_authors	dcterms.creator / dc.creator		eprints.creators_name	bepress_citation_author
Other author information	citation_author_orcid citation_author_email citation_author_institution citation_dissertation_institution		prism.organization Subsets of dc.creator: ·prism.role ·prism.place ·prism.contactInfo		bepress_citation_author_institution
Title	citation_title	dcterms.title / dc.title dcterms.alternative	prism.alternateTitle prism.subtitle prism.blogTitle	eprints.title	bepress_citation_title
Date(s)	citation_year citation_date citation_online_date citation_publication_date	dcterms.date / dc.date dcterms.created dcterms.dateSubmitted dcterms.dateAccepted dcterms.available dcterms.dateCopyrighted dcterms.issued dcterms.modified dcterms.temporal	prism.creationDate prism.dateReceived prism.copyrightYear prism.coverDate prism.coverDisplayDate prism.publicationDate prism.publicationDisplayDate prism.killDate prism.modificationDate prism.onSaleDate prism.onSaleDay prism.offSaleDate	eprints.datestamp eprints.date eprints.date_type	bepress_citation_date bepress_citation_online_date
Type of work	citation_dissertation_name [PhD, *etc.*]	dcterms.type / dc.type	prism.contentType prism.genre prism.aggregationType	eprints.type	

Concept	Highwire Press [1][2][3][4][5][6][7][8]	Dublin Core [1]	PRISM [1][2][3][4]	Eprints [1][2][3]	BE Press [1][2]
Format & language of work	citation_language	dcterms.medium dcterms.format / dc.format dcterms.extent dcterms.language / dc.language	prism.byteCount prism.wordCount prism.device prism.platform		
Identifier	citation_id citation_doi citation_pmid citation_mjid citation_id_from_sass_path citation_patent_number	dcterms.identifier / dc.identifier	prism.doi	eprints.id_number	bepress_citation_doi
Publisher — name, location	citation_publisher citation_technical_report_institution	dcterms.publisher / dc.publisher	prism.corporateEntity prism.distributor	eprints.publisher	bepress_citation_publisher
SOURCE [journal, conference, book, report series, *etc.*] **— name, identifiers**	citation_journal_title citation_journal_abbrev citation_conference_title citation_inbook_title citation_issn citation_isbn	dcterms.source / dc.source dcterms.isPartOf	prism.seriesTitle prism.publicationName prism.isbn prism.issn prism.eIssn prism.uspsNumber prism.nationalCatalogNumber prism.productCode prism.edition [*e.g.* German/global] prism.versionIdentifier [*e.g.* morning/evening] prism.bookEdition	eprints.publication eprints.issn	bepress_citation_series_title bepress_citation_journal_title bepress_citation_issn
SOURCE [journal, conference, book, report series, *etc.*] **— editor(s), organiser(s), other characteristics**		dcterms.spatial			

Concept	Highwire Press [1][2][3][4][5][6][7][8]	Dublin Core [1]	PRISM [1][2][3][4]	Eprints [1][2][3]	BE Press [1][2]
SOURCE [journal, conference, book, report series, *etc.*] **— part**	citation_technical_report_number citation_volume citation_issue citation_section		prism.seriesNumber prism.volume prism.number prism.issueIdentifier prism.issueName prism.issueTeaser prism.issueType prism.supplementTitle prism.supplementDisplayID prism.section prism.subsection1 prism.subsection2 prism.subsection3 prism.subsection4	eprints.volume eprints.number	bepress_citation_volume bepress_citation_issue
SOURCE [journal, conference, book, report series, *etc.*] **— pages**	citation_firstpage citation_lastpage		prism.startingPage prism.endingPage prism.supplementStartingPage prism.pageCount prism.pageProgressionDirection prism.pageRange prism.samplePageRange	eprints.pagerange	bepress_citation_firstpage bepress_citation_lastpage

Concept	Highwire Press [1][2][3][4][5][6][7][8]	Dublin Core [1]	PRISM [1][2][3][4]	Eprints [1][2][3]	BE Press [1][2]
Subject, Classification, Category, Key words	citation_keywords	dcterms.subject / dc.subject	prism.academicField prism.timePeriod prism.location prism.industry prism.event prism.person prism.object prism.sport prism.profession prism.ticker prism.link prism.keyword		
Abstract		dcterms.abstract dcterms.description / dc.description dcterms.tableOfContents	prism.teaser	eprints.abstract	
Citation		dcterms.bibliographicCitation		eprints.citation	
Related resources — similar, cited, cited by		dcterms.relation/ dc.relation dcterms.hasFormat dcterms.hasPart dcterms.hasVersion dcterms.isFormatOf dcterms.isReferencedBy dcterms.isReplacedBy dcterms.isRequiredBy dcterms.isVersionOf dcterms.references dcterms.replaces dcterms.requires	prism.originPlatform prism.hasAlternative prism.hasCorrection prism.hasTranslation prism.isAlternativeOf prism.isCorrectionOf prism.isTranslationOf		

Concept	Highwire Press [1][2][3][4][5][6][7][8]	Dublin Core [1]	PRISM [1][2][3][4]	Eprints [1][2][3]	BE Press [1][2]
Availability & audience	citation_fulltext_world_readable	dcterms.accessRights dcterms.audience dcterms.educationLevel dcterms.instructionalMethod dcterms.mediator	prism.rating		
Manager(s)/Supervisor(s) **Sponsor(s)** **Contributing editor(s)** **Reviewer(s)**		dcterms.contributor / dc.contributor	Subsets of dc.contributor: ·prism.role ·prism.place ·prism.contactInfo		
Rights/Copyright		dcterms.rights / dc.rights dcterms.rightsHolder dcterms.license	pur.adultContentWarning pur.agreement pur.copyright pur.creditLine pur.embargoDate pur.exclusivityEndDate pur.expirationDate pur.imageSizeRestriction pur.optionEndDate pur.permissions pur.restrictions pur.reuseProhibited pur.rightsAgent pur.rightsOwner		
Reference/Location (*e.g.* URL)	citation_public_url citation_pdf_url citation_fulltext_html_url citation_abstract_html_url citation_abstract_pdf_url		prism.url prism.blogURL	eprints.official_url	bepress_citation_pdf_url bepress_citation_abstract_html_url
Record/metadata		AC.*	prism.complianceProfile		

Concept	Highwire Press [1][2][3][4][5][6][7][8]	Dublin Core [1]	PRISM [1][2][3][4]	Eprints [1][2][3]	BE Press [1][2]
Status				eprints.ispublished	
Miscellaneous	citation_collection_id citation_price citation_patent_country citation_reference	dcterms.accrualMethod dcterms.accrualPeriodicity dcterms.accrualPolicy dcterms.conformsTo dcterms.coverage / dc.coverage dcterms.provenance dcterms.valid	prism.aggregateIssueNumber prism.publishingFrequency prism.channel prism.subchannel1 prism.subchannel2 prism.subchannel3 prism.subchannel4 prism.sellingAgency		bepress_is_article_cover_page

Obviously only some of the above are to be provided for each specific resource.

Suggested `name` attributes to include in a `meta` tag are shaded.

Additional comments on the `citation` namespace

Extension of the `citation` namespace

"Highwire Press, a division of Stanford University, developed its schema for journal articles and GS [Google Scholar] extended the tags to cover additional academic paper types, such as working papers, dissertations, manuscripts, conference papers, books and book chapters." The above table includes all of those extended meta `name` attributes, because they have become *de facto* inclusions in the `citations_*` standard. Having said that, the table may also contain user-defined meta `name` attributes, because there is no authoritative reference as to which were originally defined by Highwire Press, which were added by Google Scholar, and which were adopted by other agencies or individuals.

Schemata for the `citation` namespace

There are apparently no standard URL's in which a schema for the `citation` namespace is defined.

Additional comments on the Dublin Core metadata specification

Author names

It is evident from the official Dublin Core definitions that `creator` is the best match for providing an author's names — *not* `contributor`. Likewise the relevant PRISM standard (a subset of Dublin Core) states: "PRISM recommends that magazine publishers use `dc:contributor` for people who do additional reporting, or individuals who would be called out for special acknowledgments, such as research assistants." Hence the arrangement in the above table.

Dublin Core controlled vocabularies

Valid `dcterms.type` and `dc.type` options

`dcterms.type` is recommended to be set to one of the following:

`Collection`	An aggregation of resources. A collection is described as a group; its parts may also be separately described.
`Dataset`	Data encoded in a defined structure, *e.g.* lists, tables, and databases.
`Event`	A non-persistent, time-based occurrence, *e.g.* an exhibition, webcast, conference, workshop, open day, performance, battle, trial, wedding, tea party, conflagration.
`Image`	A visual representation other than text, *e.g.* images and photographs of physical objects, paintings, prints, drawings, other images and graphics, animations and moving pictures, film, diagrams, maps, musical notation. Note that Image may include both electronic and physical representations.
`InteractiveResource`	A resource requiring interaction from the user to be understood, executed, or experienced, *e.g.* forms on Web pages, applets, multimedia learning objects, chat services, or virtual reality environments.
`MovingImage`	A series of visual representations imparting an impression of motion when shown in succession, *e.g.* animations, movies, television programs, videos, zoetropes, or visual output from a simulation.
`PhysicalObject`	An inanimate, three-dimensional object or substance, *e.g.* a sculpture, fossil, or archæological relic. Note that digital representations of, or surrogates for, these objects should use Image, Text or one of the other types.
`Service`	A system that provides one or more functions, *e.g.* a photocopying service, a banking service, an authentication service, interlibrary loans, a Z39.50 or Web server.

`Software`	A computer program in source or compiled form, *e.g*. a C source file, MS-Windows .exe executable, or Perl script.
`Sound`	A resource primarily intended to be heard, *e.g*. a music playback file format, an audio compact disc, and recorded speech or sounds.
`StillImage`	A static visual representation, *e.g*. paintings, drawings, graphic designs, plans and maps. Recommended best practice is to assign the type Text to images of textual materials.
`Text`	A resource consisting primarily of words for reading, *e.g*. books, letters, dissertations, poems, newspapers, articles, and archives of mailing lists, and including also facsimiles or images of texts.

By default `dc.type` would also be set to one of the above. Note, however, that when used with PRISM a slightly different set of options is suggested.

Valid `*.format`, `*.language` and `*.coverage` options

Metadata in `dc.format` and `dcterms.format` may describe the file format, physical medium, or dimensions (size or duration) of the resource. The Dublin Core specification recommends that file formats be identified using a "controlled vocabulary" such as the list of Internet Media Types [MIME]. Some file formats commonly relevant to academic publications are: `text/plain` (plain text, no formatting *etc*., as in a `*.txt` file), `text/rtf` (RTF, rich text format); `text/html` (HTML); `application/pdf` (PDF); `application/msword` (MS Word); `vnd.oasis.opendocument.text` (OpenDocument text); `application/vnd.ms-powerpoint` (MS Powerpoint); and `application/vnd.apple.keynote` (Apple Keynote). TeX and LaTeX are unregistered media types, but if absolutely necessary they can be denoted with `application/x-tex` and `application/x-latex` (respectively).

Metadata in `dc.language` and `dcterms.language` describes the language(s) of the resource, for which the recommended best practice is to use a "controlled vocabulary" such as RFC 4646. For example, English is "en", and Australian English is "en-AU"; Mongolian written in Cyrillic script as used in Mongolia is represented by "mn-Cyrl-MN"; Serbian written using Latin script as used in Serbia and Montenegro is represented as "sr-Latn-CS".

Similarly, for `dc.coverage` and `dcterms.coverage` the recommended best practice is to use a "controlled vocabulary" such as the Thesaurus of Geographic Names. For example, Munich (Germany) should be listed as München (Deutschland), whereas Beijing (China) should be listed as Beijing (Zhongguo) — *not* 北京 (中国).

Administrative Components

Dublin Core also contains an `ac` namespace to specify so-called administrative metadata designed to assist with interoperability between different systems that have content metadata. As such, several of these names contain information about the metadata used in the respective system: *i.e*. 'meta-metadata'!

Selected comments are included below. For full descriptions see http://biblstandard.dk/ac/ .

Metadata for the entire record

`ac.identifier`	A string or a number, which identifies the metadata record.
`ac.source`	A string or a number, which identifies the recording entity (*e.g*. a library, museum, archive, *etc*.).
`ac.scope`	
`ac.comment`	
`ac.location`	An unambiguous reference to the content metadata within a given context. This element is only used if the content metadata and administrative metadata are not in the same location.
`ac.language`	Language of metadata.

ac.rights	Information about rights held in and over the content metadata.
ac.dateRange	
ac.handling	

Metadata for update and change

ac.activity	This element reflects an action performed on the content metadata. The element functions as a container, which connects an action (of specified type) with further details about that action.

Attributes that refine the activity specification:

ac.action	The action performed on the content metadata by the responsible entity. The actions are taken from a non-exhaustive list including: created, submitted, modified, checked, link-collected, resource-harvested, resource-disappeared, expired, mail-sent and three codes for deleted (delete-error-record, delete-disappearance and delete-out-of-scope).
ac.name	The name of the entity responsible for undertaking a defined action on the content metadata. Examples of Name include a person, an organisation, or a service. Where the person has an affiliation with an organisation, this information may be included. *The name of a person should be provided in reverse order, that is, last name before first name, with a comma separator.*
ac.email	Electronic Mail address for the responsible entity.
ac.contact	Information on how to contact the responsible entity.
ac.date	The date on which the activity took place. This unspecified date must be used in connection with an action, *e.g.* `submitted`.
ac.affiliation	The organization with which the named person was associated when involved with the resource.

Metadata for batch interchange of records

ac.database	
ac.transmitter	
ac.filename	
ac.technicalFormat	
ac.characterSet	
ac.bibliographicFormat	Bibliographic format for data exchange (*e.g.* MARC21, danMARC2, DC)
ac.resultFile	

Dublin Core schemata

It is appropriate to indicate each namespace used for the metadata through inclusion of a dedicated tag in the `head` element.

For instance, for DC

```
<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/">
```

and for DCTERMS

```
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" >
```

For AC there is no clear guide to the appropriate schema, but using the above pattern and inserting a link to a relevant current XML standard yields

```
<link rel="schema.AC"
href="http://biblstandard.dk/ac/schemas/ac_2011-09-01.xsd" />
```

Note that a few references exist setting `href` to a now-defunct URL, which should avoided.

Non-standard extensions in the `DC` namespace

Google Scholar's previous (*circa* 2010–2011) and still current advice is to use extensions to the Dublin Core tags such as `DC.citation.volume`, `DC.citation.issue`, `DC.citation.spage` and `DC.citation.epage`. However, so long as such tags are *not* part of the official specification, they should not be used.

Microsoft SharePoint 2013 looks for several non-standard `attribute` names under the `DC` namespace, such as `DC.citation.volume`, `DC.identifier.issn` and `DC.source.issn`. The duplicate options for the ISSN are just one indication that these are *not* part of the official DUblin Core specification, and so should not be used.

Actual usage of these non-standard `attribute` names is varied.

Additional comments on the PRISM metadata specification

PRISM controlled vocabularies

Vocabularies defined within PRISM

A few of the metadata names defined in the PRISM standard have 'controlled vocabularies'. Below are *a selection* of valid values relevant to academic publications.

`prism.aggregationType`: `book`; `journal`; `magazine`; `manual`; `newsletter`; `newspaper`; `report`; `whitepaper`; and `other` [avoid using].

`prism.contentType`: `article`; `bookChapter`; `introduction`; and `contentBlock` [to be used as 'other', and refined with `prism.genre`].

`prism.genre`: `abstract`; `analysis` [typical of a journal article]; `appendix` [strictly intended for books, but could be used to indicate Supplementary Material or Supporting Information]; `bibliography` [listing for a subject, author, *etc.*]; `chapter`; `correction`; `coverStory`; `essay` [expressing an author's personal point of view]; `feature` [a prominent or special article; may be suitable to indicate keynote or plenary addresses]; `foreword`; `glossary`; `interview`; `legalDocument`; `letters`; `preface`; `qAndA` [historically a common component of conference proceedings]; `references` [list of materials cited]; `reprint`; `response`; `review` [intended for reviews of media or products, but could be adopted for academic reviews too]; and `supplementArticle` [article within a supplement].

`prism.issueType`: `regularIssue`; and `specialIssue`.

`prism.platform`: `email`; `eReader`; `print`; `recordableMedia` [*e.g.* CD or DVD]; `smartPhone`; `tablet`; `web` [viewable with a browser]; and `other` [avoid using]. *None* of these are well-suited to describe the platform intended to read a PDF file: `web` may be the best of these poor options, but others such as `tablet` can additionally be used. This controlled vocabulary also contains an awkward mixture of *hardware* and *software applications* that are not mutually exclusive — *e.g.* email on a smartphone, or web browsing on a tablet.

`prism.presentationType`: `complexBlock` [suitable for some graphical abstracts]; `gallery` [may be suitable for posters comprising graphics with some text]; `infoGraphic` [may be suitable for posters that are "heavily text-oriented"]; `other`; `slideshow`; and `video`. This is intended to describe content that can be contained within an HTML `figure` element.

Sometimes multiple values are relevant to a single `name` attribute, in which case the values should be entered in separate `meta` tags, organised in order of decreasing priority, or "from most inclusive to most specific".

Use of other vocabularies

For some other metadata names in PRISM it is advisable to use a controlled vocabulary, but PRISM neither provides nor references a specific lexicon.

In the case of `prism.academicField`, it would be suitable to adopt a system such as that of the Australian and New Zealand Standard Research Classification (ANZSRC). For example, a publication dealing with "Wastewater Treatment Processes" (code 090409), "Water Treatment Processes" (090410) and "Water Quality Engineering" (090508) could be tagged with each of these three phrases; and/or it could the first two could be grouped as "Chemical Engineering" (0904), and the latter generalised to "Civil Engineering" (0905); and/or all of these can be covered by "Engineering" (09). Adopting the advice alluded to above, best practice would be to include *all* of these phrases in separate tags, organised from the most general to the most specific.

Dublin Core in the PRISM context

PRISM is expected to be used with Dublin Core metadata names in the `DC` namespace. This is especially obvious when observing that PRISM uses `DC.creator` to provide metadata identifying an author, rather than introducing a competing `name` attribute. Thus, PRISM is designed to supplement Dublin Core, not replace it.

As shown in the above table, further information about `dc.creator` or `dc.contributor` can be provided with `prism.role`, `prism.place`, or `prism.contactInfo`. In XML format (*i.e.* Profile 1) the appropriate syntax is:

```
    <dc:creator prism:role="writer" prism:location="England">Jane Doe</dc:creator>
```

There is no evident method for implementation in HTML (or XHTML).

Valid values of the `prism.role` attribute that are potentially relevant to academic articles include: `author`, `commentator`, `correspondant` [could denote a corresponding author, or the author of a letter to the editor], `editor`, `illustrator`, `interviewee`, `interviewer`, `interpreter`, `narrator` [could denote a presenter at a conference], `other`, `photographer`, `researcher`, `researchAssistant`, and `translator`.

It is not clear whether more than one `prism.role`, `prism.place`, or `prism.contactInfo` can be defined for a given entity (e.g. person) comprising a `dc.creator` or `dc.contributor`.

Extension of the PRISM specifications

Users can include their own customised elements alongside the standard PRISM elements: "PRISM is an extensible specification and includes a guide for creating your own namespace."

PRISM Profiles and HTML implementation

PRISM Profiles

Three Profiles are defined in the PRISM metadata specification: **XML-only** (EXtensible Markup Language), **RDF/XML** (Resource Description Framework/XML), and **XMP** (Extensible Metadata Platform). The XML-only profile was defined earliest, and is by far the most frequently used and most comprehensively described.

None of these are provide a clear indication of how they might be embedded in HTML. Strictly speaking, XHTML is a type of XML, and XHTML is often readable by HTML viewers, so use of XHTML may be one approach.

Recommendations for use of PRISM in HTML

Besides the three Profiles for embedding metadata that form a key part of the core PRISM specification, two dependant specifications have been created that each refer to either XHTML or HTML.

PRISM Aggregator Message (PAM) is "an XML tag set that uses PRISM metadata for a very specific purpose". The PRISM Source Vocabulary (PSV) appears to be designed to allow metadata tagging of

multiple content elements *within the body* of a single HTML5 document. Neither of these provides for using the `meta` tag within the `head` element of an HTML document (HTML5 or otherwise).

The PSV Specification explicitly states, "It is the intent that all PSV metadata in the Source be captured within the `<psv:metadata` block using the `<psv:meta` tag and that it not be duplicated or replaced using the HTML5 `<meta` tag. The HTML5 `<meta` tag has therefore not been included in the model for HTML5 `<head` in this PSV Specification." while a related guide adds, "Although there is an optional `<meta` tag in the HTML5 `<head` structure, it is not to be used to store metadata about the article. The recommended PSV HTML5 subset definition for the `<head` only allows structures such as `<link` and `<styles` but does not allow for the encoding of metadata. Metadata is expected to be consolidated in the PSV`<metadata` block."

Nevertheless, the official specification on the nextPub PRISM Source Vocabulary (PSV) Framework also states: "If you need to transform PSV into HTML5 to deliver for browser display, you may wish to transform some of the metadata in the `<psv:metadata` block into `<meta` tags in the HTML5 head. A *PSV to HTML5 Transformation Guide* to document the transformation of PSV XML into HTML5 for delivery to browsers is planned to be added to the PSV Documentation Set in the future."

Compromise implementation in HTML `meta` tags

In the above table all colons (used in the official PRISM standards) were replaced with dots, as apparently applied in practice. The tags would then be used within HTML `meta` elements.

It must be recognised that these practices deviate from the formal PRISM specifications — at least while the foreshadowed "PSV to HTML5 Transformation Guide" remains unpublished.

PRISM schemata

It is appropriate to indicate each namespace used for the metadata through inclusion of a dedicated tag in the `head` element.

For instance, for version 1.2 of PRISM

```
    <link title="PRISM schema" rel="schema.prism"
href="http://prismstandard.org/namespaces/1.2/basic/" />
```

(Notice the XHTML-style syntax.) The `title` attribute is optional.

For version 2.1 of PRISM

```
    <link rel="schema.prism"
href="http://prismstandard.org/namespaces/basic/2.1/" />
    <link rel="schema.pur"
href="http://prismstandard.org/namespaces/prismusagerights/2.1/" />
```

etc.

For version 3.0 of PRISM

```
    <link rel="schema.prism"
href="http://prismstandard.org/namespaces/basic/3.0/" />
    <link rel="schema.pur"
href="http://prismstandard.org/namespaces/pur/3.0/" />
```

etc.

Additional comments on the Eprints metadata specification

There are very few `name` attributes in the `eprints` namespace — apparently only ten. Furthermore, no publicly available official standard could be found.

Additional comments on the BE Press metadata specification

BE Press metadata appears not to be commonly used. Typically the `name` attributes in the `bepress` namespace are practically duplicates of those in the `citation` namespace (prefixed by `bepress_`), except that there are fewer of them.

Rather than expand the contents of the `bepress` namespace, is appears that BE Press is open to adopting labels and/or conventions from other metadata systems including Dublin Core, OpenURL, PubMed.

Recommendations for academic publications

Academic publications should be described with metadata following one of more combinations of the **Dublin Core**, **`citations`**-namespace and **PRISM** conventions.

Dublin Core and PRISM are tightly controlled standards supported by numerous organisations.

Dublin Core is very widely used across diverse applications. Dublin Core and the `citations`-namespace are very commonly used for academic publications; PRISM is also commonly used.

The `citations`-namespace includes numerous `name` attributes specifically relevant to academic publications. PRISM also includes many `name` attributes specifically relevant to publications, albeit more focussed on popular media. Dublin Core also provides several relevant `name` attributes (variously in the `DC`, `DCTERMS` and `AC` namespaces), some of which do not appear in the `citations`-namespace or PRISM.

By far the `DC` namespace is more commonly used than the `DCTERMS` or `AC` namespaces, and — being older — has broader compatibility; therefore, when the same `name` attribute exists in both the `DC` and `DCTERMS` namespaces, it advantageous to use that in the **`DC` namespace**.

Other conventions

Facebook's Open Graph protocol

Common `property` attributes

Various *values* are defined in the Open Graph Protocol created by Facebook developers: `og:title`, `og:url`, `og:site_name`, `og:image`, `og:description`, and `og:type`. These do not meet the requirements for registration as "proposed" or "ratified" *attributes*; rather, they should be used as possible *values* of the enumerated variable `property`. This would follow the general syntax:

```
<meta property="og.*" content="x"/>
```

as in

```
<head prefix="og: http://ogp.me/ns# fb: http://ogp.me/ns/fb# article:
http://ogp.me/ns/article#">
<meta property="fb:app_id" content="302184056577324" />
<meta property="og:url" content="http://www.theage.com.au/story/06546512303540.html" />
<meta property="og:type" content="article" />
<meta property="og:title" content="When Great Minds Don't Think Alike" />
<meta property="og:description" content="How does culture influence thinking?" />
<meta property="og:image" content="http://static.theage.com.au/images/0654646546.jpg" />
```

or

```
<head prefix="og: http://ogp.me/ns# fb: http://ogp.me/ns/fb# book:
http://ogp.me/ns/book#">
<meta property="fb:app_id" content="302184056577324" />
<meta property="og:type" content="book" />
<meta property="og:url" content="http://www.domain.com/pub/Book02103.html" />
<meta property="og:title" content="Sample Book" />
<meta property="og:image" content="http://www.domain.com/pub/images/Book02103.png" />
```

These are sometimes — albeit rarely — used to provide metadata on academic publications.

The complete list of valid object `og:type` values is: `article`, `book`, `books.author`, `books.book`, `books.genre`, `business.business`, `fitness.course`, `game.achievement`, `music.album`, `music.playlist`, `music.radio_station`, `music.song`, `place`, `product`, `product.group`, `product.item`, `profile`, `restaurant.menu`, `restaurant.menu_item`, `restaurant.menu_section`,

`restaurant.restaurant`, `video.episode`, `video.movie`, `video.other`, `video.tv_show`.

Resource-specific `property` attributes

There are further dedicated 'fields' available for specific resource types.

For book types:

`book:author`	An array of the Facebook IDs of the users that authored the book.
`book:isbn`	
`book:release_date`	
`book:tag`	Keywords.

For books.book:

`books:author`	An array of references to the objects representing the authors of the book.
`books:genre`	An array of references to the objects representing the genres of the book.
`books:initial_release_date`	A time representing when the book was initially released.
`books:isbn`	(Required.)
`books:language`	
`books:page_count`	
`books:rating`	The rating of the book.
`books:release_date`	
`books:sample`	A URL of a sample of the book

For article types:

`article:author`	An array of Facebook profile URLs or IDs of the authors for this article.
`article:content_tier`	Specification of whether article is free, locked, or metered.
`article:expiration_time`	
`article:modified_time`	
`article:published_time`	
`article:publisher`	A Facebook page URL or ID of the publishing entity.
`article:section`	The section of your website to which the article belongs, such as 'Lifestyle' or 'Sports'.
`article:tag`	Keywords.

It is evident that these tags are not useful in general for academic publications, given that many authors either won't have a Facebook account, or will have a private Facebook account that they do not wish to refer to in professional circumstances.

Database-specific conventions

Several other databases have their own conventions.

PubMed

For instance, the U.S. **PubMed** website (pubmed.gov) has both abbreviated tags for simple searches and longer tags for more advanced searches — *e.g.*, `[au]` versus `[Author]`, `[Author - First]`, `[Author - Last]`, `[Author - Full]`, `[Author - Identifier]` and

`[Author - Corporate]` when searching author names . However, in their HTML encoding they are much less particular. Thus a seven-page 2018 article by Johnson & Key is tagged with

```
ncbi_uidlist = 30270231
author = Johnson JE , et al.
description = Prog Community Health Partnersh. 2018;12(2):215-221. doi: 10.1353/cpr.2018.0041.
```

Notice that the second author's name is omitted in the `author` tag; the `description` tag is extremely vague, and mixes dates, volumes, issues and page numbers; and the `ncbi_uidlist` tag is quite uninformative outside of PubMed. (Note: NCBI refers to the U.S. National Center for Biotechnology Information, part of the U.S. National Library of Medicine (NLM), which manages PubMed.)

Thomson Reuters

Thomson Reuters, who operate **Web of Science** and **Journal Citation Reports**, provide a specification for queries of their databases using the OpenURL resolver. These include

`rft_id` (`info:doi`, `info:pmid`, `info:ut`), `rft.atitle`, `rft.jtitle`, `rft.btitle`, `rft.issn`, `rft.isbn`, `rft.date` [actually the year], `rft.volume`, `rft.issue`, `rft.spage`, `rft.epage`, `rft.aulast`, `rft.aufirst`, `rft.auinit`, `rft.auinitm`, and `rft.au`.

These are *not* intended to be used in HTML. However, some analogous metadata tags of the form `rft_id`, `rft_issn` *etc*. have been applied on some web pages.