☞SWAD-Europe Thesaurus Activity

Deliverable 8.5

☞SWAD-Europe

☞W3C

☞

RDF Encoding of Classification Schemes

An example encoding of the PACS scheme, with some recommendations for classification schemes in general

N.B. This document refers to the SKOS Core Vocabulary, which has been modified since the publication of this document series. Please see ☞http://www.w3.org/2004/02 /skos/core/guide/ for the latest version of the ☞SKOS Core Guide.

Abstract:

This report describes an RDF encoding of the Physics and Astronomy Classification Scheme (PACS). How to correctly capture the semantics of the PACS scheme, and of a PACS resource classification, is discussed. Some recommendations for the RDF encoding of classification schemes in general are offered.

Authors:
☞Alistair J. Miles, CCLRC

Status of this document

This section describes the status of this document at the time of its publication. This is a draft document and may be updated, replaced, or obsoleted by other documents at any time. The latest status of this document series is maintained at the W3C.

This document is a public DRAFT for discussion. This document and the SKOS-Core schema are an output of the research work of the ☞Semantic Web Advanced Development for Europe Project, which is associated with the ☞W3C Semantic Web Activity. This document is made available by W3C for discussion only. Publication of this document by W3C does not imply endorsement by W3C, including the Team and Membership.

Comments on this document are welcome and should be sent to the authors or to the ☞public-esw-thes@w3.org list. An archive of this list is available at ☞http://lists.w3.org /Archives/Public/public-esw-thes/.

Contents

1. Introduction   [☞back to contents]

This document describes an RDF encoding of the Physics and Astronomy Classifiation Scheme (PACS) [☞PACS], as an example of RDF encoding of classification schemes.

There is a significant degree of structural variation between the many classification schemes that are in the public domain. This document does not attempt to make any sort of exhaustive survey of classification scheme types, nor does it attempt to provide specific schema classes and properties to support the multitude of variations that exist. It does, however, illustrate the use of a core schema [☞SKOS SCHEMA], which is suitable as a basis for an RDF encoding of PACS, and which may be extended in specific instances wherever additional features are required.

2. The Fundamental Features of PACS   [☞back to contents]

2.1. Structure -

PACS consists of a set of classification codes, each corresponding to a single term. These terms are also known as 'classification headings'. A PACS code/term pair is hereafter referred

to as a 'PACS value'.

Extract from PACS

```
02.        Mathematical methods in physics

02.10.-v   Logic, set theory, and algebra

02.10.Ab   Logic and set theory

02.10.De   Algebraic structures and number theory

02.10.Hh   Rings and algebras

02.10.Kn   Knot theory

02.10.Ox   Combinatorics; graph theory

02.10.Ud   Linear algebra
```

The PACS values are organised hierarchically, into a tree with maximum depth of 4. Each classification code serves both as a unique local identifier for a PACS value, and as an indicator of the hierarchical location of that value.

For example, a top level value has a classification code conforming to the following regular expression ...

Regexp for code of top level value

```
[0-9]0\.
```

... i.e. '00.' '10.' 20.' are examples of codes for top level values.

A second level value has a classification code conforming to the following regular expression ...

Regexp for code of second level value

```
[0-9][1-9]\.
```

... i.e. '01.' '23.' '67.' are examples of codes for second level values. Furthermore, '01.' is a child of '00.', '23.' is a child of '20.', '67.' is a child of '60.' etc.

A third level value has a classification code conforming to the following regular expression ...

Regexp for code of third level value

```
[0-9][1-9]\.[0-9][0-9]\.[+\-][a-z]
```

... i.e. '68.15.+e' '71.15.-m' '82.30.-b' are examples of codes for third level values. Furthermore, '68.15.+e' is a child of '68.', '71.15.-m' is a child of '71.', '82.30.-b' is a child of '82.' etc.

A fourth level value has a classification code conforming to the following regular expression ...

Regexp for code of fourth level value

```
[0-9][1-9]\.[0-9][0-9]\.[A-Z][a-z]
```

... i.e. '82.35.Gh' '98.20.Jp' '45.20.Dd' are examples of codes for fourth level values. Furthermore, '82.35.Gh' is a child of '82.35.-x', '98.20.Jp' is a child of '98.20.-d', '45.20.Dd' is a child of '45.20.-d' etc.

2.2. Semantics -

In the overwhelming majority, the PACS values are obviously concepts in the domain of physics and astronomy, as in the above extract and additional examples below:

PACS values that are domain concepts

```
03.        Quantum mechanics, field theories, and special relativity

03.30.+p   Special relativity
```

```
03.50.-z   Classical field theories

03.50.De   Classical electromagnetism, Maxwell equations

03.50.Kk   Other special classical field theories
```

However, a small number of PACS values are obviously **types of resources**, and **not** domain concepts. Examples of these are below:

PACS values that are resource types

```
01.30.-y   Physics literature and publications

01.30.Bb   Publications of lectures (advanced institutes, summer schools,
           etc.)

01.30.Cc   Conference proceedings

01.30.Ee   Monographs and collections

01.30.Kj   Handbooks, dictionaries, tables, and data compilations

01.30.Mm   Textbooks for graduates and researchers

01.30.Pp   Textbooks for undergraduates

01.30.Rr   Surveys and tutorial papers; resource letters

01.30.Tt   Bibliographies

01.30.Vv   Book reviews

01.30.Xx   Publications in electronic media
```

Whether a PACS value is a domain concept or a resource type is not indicated by its classification code.

3. RDF Encoding of PACS [☞back to contents]

The SKOS-Core schema [☞SKOS SCHEMA] was used here as the basis for an RDF encoding of PACS. The SKOS-Core Guide [☞SKOS GUIDE] describes this schema in more detail. The SKOS-Core Migration Guidelines document [☞SKOS MIGRATE] describes the process of generating RDF encodings for thesauri and designing schema extensions.

An extract from PACS in RDF is below:

Extract from PACS in RDF

```
<rdf:RDF
    xmlns:pacs="http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs/schema-ext#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:skos="http://www.w3.org/2004/02/skos/core#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xml:base="http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs/" >

  <skos:ConceptScheme rdf:about="http://www.w3.org/2001/sw/Europe/reports/thes/ns/p
    <dc:title>Physics and Astronomy Classification Scheme</dc:title>
    <dc:creator>American Institute of Physics</dc:creator>
    <rdfs:seeAlso rdf:resource="http://www.aip.org/pacs/"/>
    <rdfs:seeAlso rdf:resource="http://publish.aps.org/PACS/"/>
  </skos:ConceptScheme>

  <pacs:Value rdf:about="90.">
    <skos:prefLabel>GEOPHYSICS, ASTRONOMY, AND ASTROPHYSICS</skos:prefLabel>
    <pacs:code>90.</pacs:code>
    <skos:inScheme rdf:resource="http://www.w3.org/2001/sw/Europe/reports/thes/ns/p
    <skos:narrower rdf:resource="93."/>
    <skos:narrower rdf:resource="92."/>
    <skos:narrower rdf:resource="91."/>
    <skos:narrower rdf:resource="97."/>
    <skos:narrower rdf:resource="98."/>
```

```
      <skos:narrower rdf:resource="95."/>
      <skos:narrower rdf:resource="96."/>
      <skos:narrower rdf:resource="94."/>
      <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#TopConcept"/>
  </pacs:Value>

  <pacs:Value rdf:about="91.">
    <skos:prefLabel>Solid Earth physics</skos:prefLabel>
    <pacs:code>91.</pacs:code>
    <skos:inScheme rdf:resource="http://www.w3.org/2001/sw/Europe/reports/thes/ns/p
    <skos:narrower rdf:resource="91.30.-f"/>
    <skos:narrower rdf:resource="91.45.-c"/>
    <skos:narrower rdf:resource="91.50.-r"/>
    <skos:narrower rdf:resource="91.10.-v"/>
    <skos:narrower rdf:resource="91.60.-x"/>
    <skos:narrower rdf:resource="91.70.-c"/>
    <skos:narrower rdf:resource="91.90.+p"/>
    <skos:narrower rdf:resource="91.35.-x"/>
    <skos:narrower rdf:resource="91.65.-n"/>
    <skos:narrower rdf:resource="91.25.-r"/>
    <skos:narrower rdf:resource="91.40.-k"/>
    <skos:broader rdf:resource="90."/>
  </pacs:Value>

  <pacs:Value rdf:about="91.10.-v">
    <skos:prefLabel>Geodesy and gravity</skos:prefLabel>
    <pacs:code>91.10.-v</pacs:code>
    <skos:inScheme rdf:resource="http://www.w3.org/2001/sw/Europe/reports/thes/ns/p
    <skos:narrower rdf:resource="91.10.Pp"/>
    <skos:narrower rdf:resource="91.10.Vr"/>
    <skos:narrower rdf:resource="91.10.Tq"/>
    <skos:narrower rdf:resource="91.10.By"/>
    <skos:narrower rdf:resource="91.10.Sp"/>
    <skos:narrower rdf:resource="91.10.Rn"/>
    <skos:narrower rdf:resource="91.10.Qm"/>
    <skos:narrower rdf:resource="91.10.Da"/>
    <skos:narrower rdf:resource="91.10.Fc"/>
    <skos:narrower rdf:resource="91.10.Nj"/>
    <skos:narrower rdf:resource="91.10.Ws"/>
    <skos:narrower rdf:resource="91.10.Lh"/>
    <skos:narrower rdf:resource="91.10.Kg"/>
    <skos:narrower rdf:resource="91.10.Jf"/>
    <skos:broader rdf:resource="91."/>
  </pacs:Value>

  <pacs:Value rdf:about="91.10.Pp">
    <skos:prefLabel>Gravimetric measurements and instruments</skos:prefLabel>
    <pacs:code>91.10.Pp</pacs:code>
    <skos:inScheme rdf:resource="http://www.w3.org/2001/sw/Europe/reports/thes/ns/p
    <skos:broader rdf:resource="91.10.-v"/>
  </pacs:Value>

</rdf:RDF>
```

The full RDF encoding of PACS is linked from the reference [☞PACS RDF]. The source code of the Java program used to parse the PACS text format and generate the RDF encoding is linked from the reference [☞PACS JAVA].

The choice to re-use the main features of the SKOS-Core schema, and not design an entirely PACS-specific schema, was made because PACS is essentially a hierarchy of concepts, each with a single preferred label, and thus fits into the SKOS model. Re-using schema features wherever possible is highly desirable, as it promotes interoperability and sharing of data.

A small schema extension was used to capture some PACS specific features:

PACS schema extension (RDF/N3)

```
@prefix pacs: <http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs/schema-ext#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .

pacs:Value
```

```
    a    rdfs:Class;
    rdfs:subClassOf    skos:Concept.

pacs:code
    a    rdf:Property;
    rdfs:label    'classification code';
    rdfs:domain    pacs:Value;
    rdfs:subPropertyOf    skos:externalID.

pacs:classification
    a    rdf:Property;
    rdfs:label    'PACS classification';
    rdfs:domain    rdf:Resource;
    rdfs:range    pacs:Value.
```

The property `pacs:code` was created to capture the semantics of the PACS classification code, which acts both as a concept's scheme-local identifier and a hierarchy position indicator (see above). The `pacs:classification` property is explained in the next section. The `pacs:Value` class was created to support the use of these two PACS-specific properties.

N.B. To this author's knowledge no definitive URIs for PACS itself or any of the PACS values have been published by the American Institute of Physics, the authority responsible for PACS. In the RDF encoding of PACS published with this report, URIs under the SWAD-E namespace were used to refer to PACS and the PACS values - *these should not be considered as the definitive URIs for these resources*.

4. RDF Encoding of Resource Classifications [☞back to contents]

Most PACS values are concepts from the domain of Physics and Astronomy, and the classification of a resource (such as a book or web page) by that value means that the intended concept is the **subject** of the resource.

The `dc:subject` property carries appropriate semantics to capture this type of resource classification. However, for these values, a PACS classification implies not just that the indicated concept is the subject of a resource, but is in fact the **primary subject** of that resource. To capture these additional semantics, it would be suitable to define a property that extends the `dc:subject` property, for example `ex:primarySubject`. A property carrying these semantics is unfortunately not currently present in any standard vocabulary.

Example of use of a proposed property ex:primarySubject to capture typical resource classification

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:es="http://www.example.org/classification#" >

  <rdf:Description rdf:about="http://www.example.org/aWebPage.html">
    <ex:primarySubject rdf:resource="http://www.w3.org/2001/sw/Europe/reports/thes,
  </rdf:Description>

</rdf:RDF>
```

A small number of PACS values are not domain concepts, but are in fact **resource types** (see above). For this small minority of PACS values, a classification of a resource means that the PACS value indicates the **type of the resource** (and **not** the subject of the resource).

For these PACS values, the `rdf:type` property carries the appropriate semantics to capture a resource classification.

Example of use of rdf:type to capture exceptional resource classifications

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" >

  <rdf:Description rdf:about="http://www.example.org/aBookReview.html">
    <rdf:type rdf:resource="http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs/(
  </rdf:Description>
```

```
</rdf:RDF>
```

However, it is not possible to determine without human interpretation which of the PACS values are used as in the first case, and which as in the second. Therefore, as part of a first automated step to generating an RDF encoding of **existing** resource classifications, it is recommended that a PACS-specific property (e.g. `pacs:classification`) is used to capture the classification of resources. Disambiguation of the exact meaning of each classification may be performed manually at a later date, if desired.

<div align="center">Example of use of pacs:classification to capture ambiguous resource classifications</div>

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:pacs="http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs/schema-ext#"

  <rdf:Description rdf:about="http://www.example.org/aResource">
    <pacs:classification rdf:resource="http://www.w3.org/2001/sw/Europe/reports/the
  </rdf:Description>

</rdf:RDF>
```

5. Discussion: Classification Schemes and Thesauri [☞back to contents]

Essentially and for the most part classification schemes consist of sets of 'concepts' or 'subjects', against which resources (e.g. books) may be grouped. In this sense, there is fundamentally no difference between a classification scheme and a thesaurus which has been designed for subject-based indexing, except that a classification scheme does not usually include any alternative labelling of concepts.

There is, however, a fundamental difference between the application of these two types of scheme for subject-based organisation of resources: when using a thesaurus, a resource may usually be indexed against one or more concepts, but when using a classification scheme, a resource may only be classified (indexed) against a single concept.

This restriction of one classification (subject) per resource is rooted in the necessity to create some sort of meaningful spatial ordering of physical resources (i.e. books on shelves). However, the use of metadata in an electronic enviroment makes creating a virtual organisation of resources extremely simple. I would argue that this restriction on the use of classification schemes within a semantic web environment is not necessary or useful.

What would be useful, however, would be to be able to state which concept (whether from a classification scheme or a thesaurus) is the **primary subject** of a resource such as a book or web page. Other concepts may then be applied as **secondary subjects** of a resource.

This proposal is similar to the approach taken by `foaf:topic` / `foaf:primaryTopic` property pair [☞FOAF].

<div align="center">Example property carrying extended semantics</div>

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ex: <http://www.example.org/classification#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

ex:primarySubject
    a  rdf:Property;
    rdfs:label  'primary subject';
    rdfs:subPropertyOf  dc:subject.
```

However, as the case of PACS has demonstrated, the meaning of a resource classification is not always consistent. Therefore, caution must be excercised in the choice of which RDF property(ies) to use to describe resource classifications, and the choice needs to be evaluated on a case by case basis.

N.B. It will almost always be incorrect to assume that a 'classification scheme' consists of

a set of 'classes' arranged in a 'class hierarchy'. In the experience of this author, most 'classification schemes' are best described as consisting of a set of concepts, arranged into a semantically ambiguous hierarchy, and which are used to categorise a set of resources according to their primary subject.

Because this is exactly the type of structure that the SKOS-Core schema is suited to describing, it is likely that the SKOS-Core schema, or a simple extension of it, will be suitable and sufficient to support the RDF description of the majority of classification schemes.

However, this will not always be the case. Therefore, before generating an RDF encoding of a classification scheme, it is strongly recommended to carefully evaluate what the scheme's values represent, and what is intended by a hierarchical relationship between values and a resource classification.

References

[PACS]
Physics and Astronomy Classification Scheme. American Institute of Physics.
☞http://www.aip.org/pacs/
[SKOS SCHEMA]
SKOS-Core 1.0 RDF Schema. Miles, A.J., Rogers, R., Beckett, D. SWAD-Europe Thesaurus Activity.
☞http://www.w3.org/2004/02/skos/core.rdf
[SKOS MIGRATE]
SKOS-Core 1.0 Guidelines for Migration. Miles, A.J., Rogers, R., Beckett, D. SWAD-Europe Thesaurus Activity.
☞http://www.w3.org/2001/sw/Europe/reports/thes/1.0/migrate/
[SKOS GUIDE]
SKOS-Core 1.0 Guide. Miles, A.J., Rogers, R., Beckett, D. SWAD-Europe Thesaurus Activity.
☞http://www.w3.org/2001/sw/Europe/reports/thes/1.0/guide/
[FOAF]
The Friend of a Friend Project (FOAF).
☞http://www.foaf-project.org/
    See also ☞SKOS (Simple Knowledge Organisation System).

Associated Files

[PACS RDF]
☞http://www.w3.org/2001/sw/Europe/reports/thes/8.5/pacs.rdf.xml
[PACS JAVA]
☞http://www.w3.org/2001/sw/Europe/reports/thes/8.5/Parser.java

Appendix: PACS Schema Extension

```
@prefix pacs: <http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs/schema-ext#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .

pacs:Value
    a     rdfs:Class;
    rdfs:subClassOf    skos:Concept.
```

```
pacs:code
    a     rdf:Property;
    rdfs:label     'classification code';
    rdfs:domain     pacs:Value;
    rdfs:subPropertyOf     skos:externalID.

pacs:classification
    a     rdf:Property;
    rdfs:label     'PACS classification';
    rdfs:domain     rdf:Resource;
    rdfs:range     pacs:Value.
```

```
pacs:code
    a     rdf:Property;
    rdfs:label     'classification code';
    rdfs:domain     pacs:Value;
    rdfs:subPropertyOf     skos:externalID.

pacs:classification
    a     rdf:Property;
    rdfs:label     'PACS classification';
```