Transforming the Integrative Levels Classification to SKOS: representation of numbers, dates and people via parallel facets

Claudio Gnoli, Douglas Tudhope
Science and Technology Library, University of Pavia, Italy
Hypermedia Research Group, University of South Wales

claudio.gnoli@unipv.it
douglas.tudhope@southwales.ac.uk

NKOS Consolidated Workshop, 22 Sept 2021

Integrative Levels Classification (ILC)

- A general KOS
- Developed since 2004 by an international team
- Phenomenon-based
  (lists “birds” not “ornithology”)
- Freely faceted
  (any class may be combined with any other)
- Arranged systematically by notation
  (starts with *A* “reality” not *aardvarks*)

Notation

- Controls mechanical sorting in classification schemes:

aardvarks

…

...

zebras

A ...

F “zebras”

Q “aardvarks”

Z ...

ILC notation

- In ILC, some classes are expressed by reusing notation for other classes (parallel facets)

- Example:

numbers → dates → persons

ILC notation for **numbers**

an “quantities”
anb “negative quantities”
ank “thousandths”
ann “units”
*ann**p*** “one; 1”
*ann**q*** “two; 2”
ano “tens”
*ano**q*** “two tens”
*ano**qp*** “twenty-one; 21”

ILC notation for **numbers**

*anbn**g*** “minus seven; -7”
*anbn**j*** “minus four; -4”

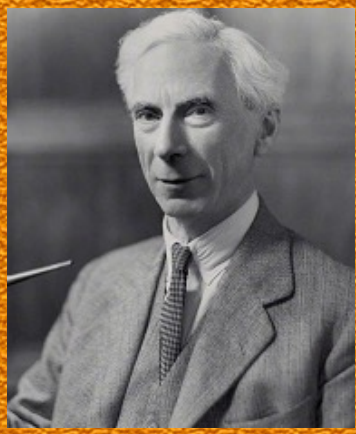
*ann**s*** “plus four; 4”
*ann**v*** “plus seven; 7”

NB: ***-n*** “negative zero” is also needed !

ILC notation for **dates**

rab “historical periods”
*rab**e*** “9999 to 9000 BCE”
*rab**j*** “4999 to 4000 BCE”
rab***n*** “999 to 1 BCE; 1st millennium BCE”
rab***o*** “1 to 999 CE; 1st millennium CE”
*rab**q*** “2000 to 2999 CE; 2nd millennium”
*rab**qo*** “2000 to 2099 CE; 21st century”
*rab**qoqp*** “2021 CE”
*rab**qoqp**jw* “22 September 2021”

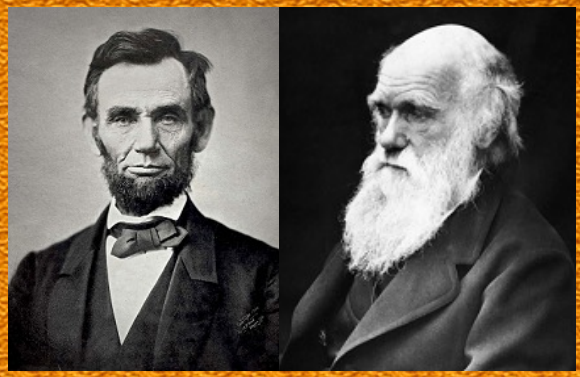
How to sort persons?...



- By occupation (Ranganathan's solution)
  but some have several occupations,
  e.g. mathematician and philosopher…
- By birth place
  but sometimes this is occasional
- By birth date
  possibly the most neutral
- Facets for occupation, place etc. may be added

ILC notation for **persons**

*px91**p*** “persons born in 1st millennium CE”
*px91**pw*** “persons born in 19th century”
*px91**pwo*** “persons born from 1800 to 1809”
*px91**pwox*** "persons born in 1809"
*px91**pwox**pt* "born on 19/1/1809; EA Poe"
*px91**pwox**qme* “born on 12/2/09 3AM; Darwin”
*px91**pwox**qmh* “born on 12/2/09 6AM; Lincoln”



A parallel facet in SKOS

<http://www.iskoi.org/ilc/2/class/**p91**>
<http://www.w3.org/2000/01/rdf-schema#**domain**>
<http://www.iskoi.org/ilc/2/class/**p**> .
<http://www.iskoi.org/ilc/2/class/**p91**>
<http://www.w3.org/2000/01/rdf-schema#**range**>
<http://www.iskoi.org/ilc/2/class/**rab**> .

Conclusions on notation

- Notation can control sorting (and retrieval) by highly expressive devices
- Faceted classifications may have complex structures,
  e.g. categories, parallel facets
- Expressing these in SKOS requires special solutions, e.g. RDF extensions

Complex classification schemes are complex

Complex classification schemes have not been straightforward to represent in SKOS.

Early papers on the development of the SKOS standard show examples from taxonomies, *simple* classifications and thesauri. However, the features of complex classification schemes go beyond the functionality of the SKOS standard.

Our JDoc 2021 paper (see References) discusses work by Baker, Isaac, Miles, Panzer, Prasad, Summers, Zeng and others on the challenges for SKOS expression of complex classifications, such as Colon, DDC, LCSH.

Challenges for SKOS include the representation of facets, spans of classes, coordinated subject headings or synthetic compound descriptors, concepts considered non-assignable for indexing, expressing ordering principles.

MADS/RDF (mapped to SKOS) was developed by the Library of Congress to address coordination, eg in LCSH – see later.

Main elements of ILC in SKOS conversion

- Classes – *skos:Concept*, *skos:ConceptScheme*
- Identifying top level classes - *skos:hasTopConcept*
- Class hierarchy - *skos:broader* (and *skos:narrower*)
- Facet indicators – *rdf:Property* and sub-properties
  with the top level fundamental categories *(0-9)*
  as sub-properties of *skos:related*
- Special facets - *rdf:Property* and sub-properties

- *Concept attributes:*
- Notation – *skos:notation* (ILC notation sorting opaque to SKOS but available to ILC aware applications)
- Main caption – *skos:prefLabel*
- Synonyms – *skos:altLabel*
- Description – *skos:note*

Conversion of ILC (ed. 2) to SKOS

ILC data was provided as a CSV (Comma Separated Values) file, exported from the editorial MYSQL database, with rows for ILC entities and columns for the different elements.

Building on experience in converting UK Heritage Thesauri, CSV data was converted to SKOS RDF using the open source STELETO transformation tool, not previously applied to complex classification schemes.

Due to the complexity of a faceted classification system such as ILC, bespoke rules were added to the transformation process. Crucially, the hierarchical structure of the classification was inferred from the letter sequence of the notational codes. Other CSV fields provided synonyms and labels (sometimes requiring the combination of fields.

Data cleaning was one of the first requirements of the conversion process. After parsing the data into separate fields each field value required normalization to ensure consistency of whitespace and case; other more specialized processing could also be undertaken.

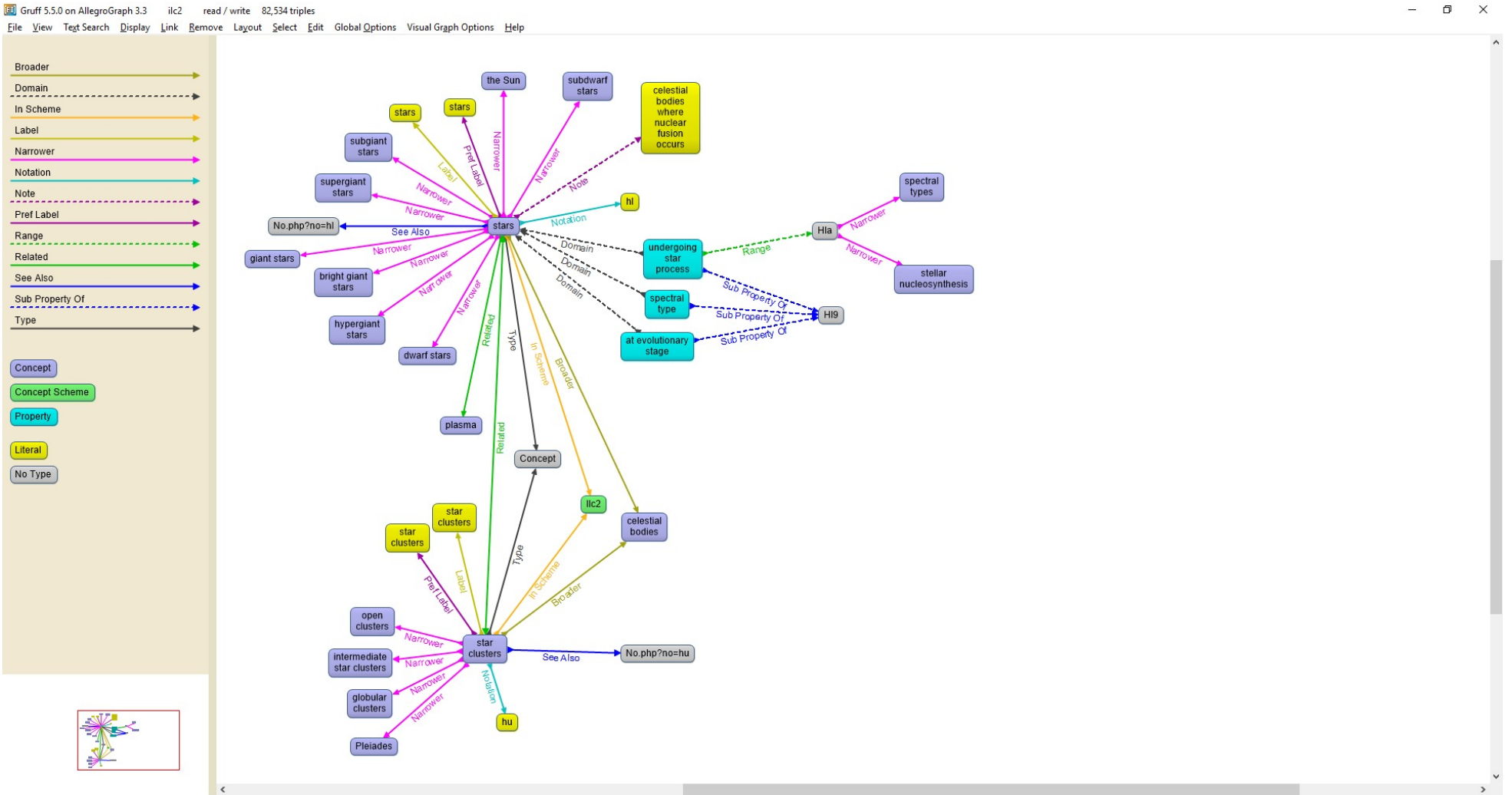
Mappings to external DDC concepts were included where the DDC class number mappings were present in the CSV file.

Extracts of SKOS conversion for ILC class Stars

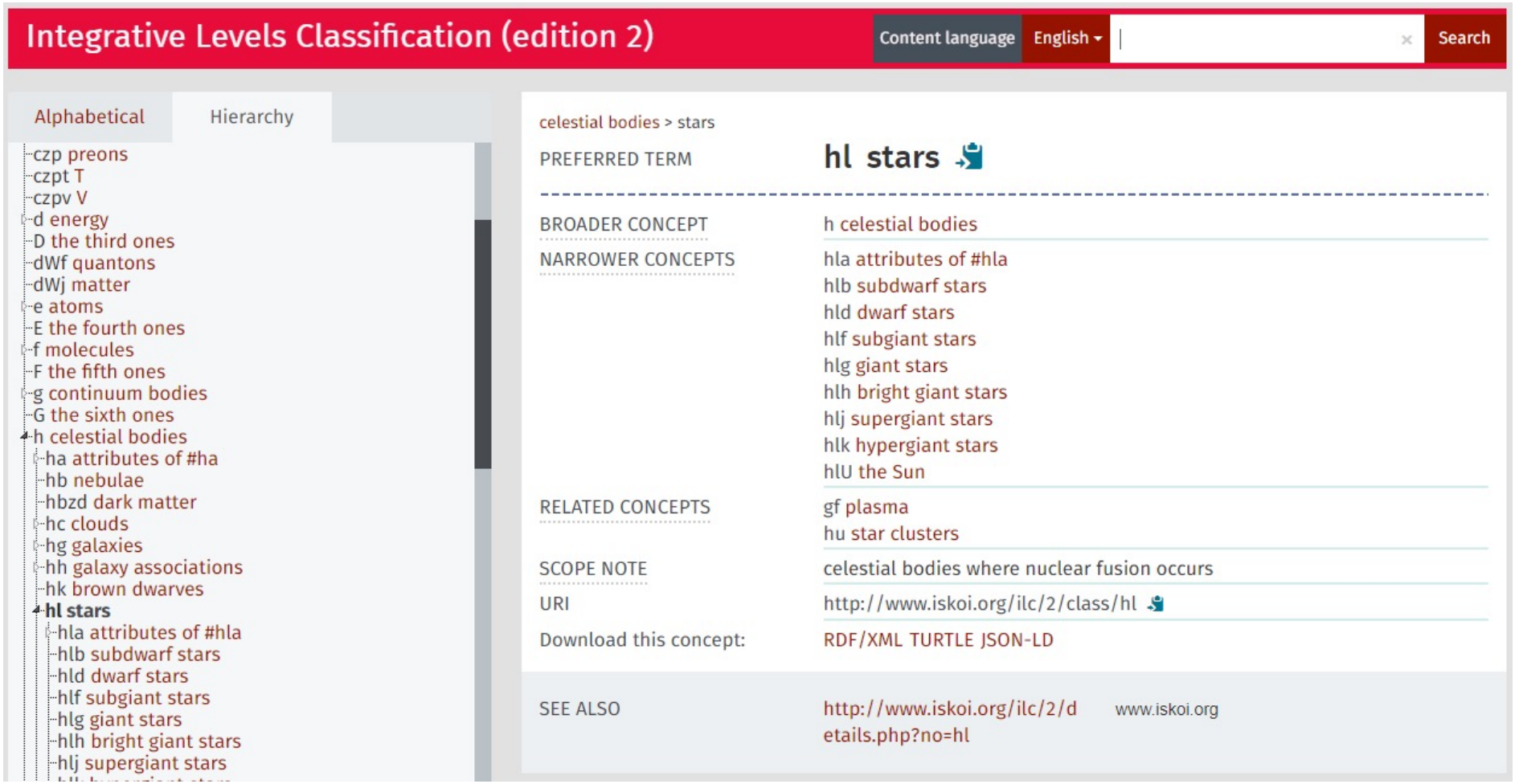
```
@prefix ilc2: <http://www.iskoi.org/ilc/2/class/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

ilc2:h
  skos:notation "h" ;
  rdfs:label "celestial bodies"@en ;
  skos:prefLabel "celestial bodies"@en ;
  a skos:Concept ;
  skos:narrower ilc2:hl .

ilc2:hu
  skos:notation "hu" ;
  rdfs:label "star clusters"@en ;
  skos:prefLabel "star clusters"@en ;
  a skos:Concept ;
  skos:related ilc2:hl .

ilc2:hlb
  skos:notation "hlb" ;
  rdfs:label "subdwarf stars"@en ;
  skos:prefLabel "subdwarf stars"@en ;
  a skos:Concept ;
  skos:broader ilc2:hl .

ilc2:hlg
  skos:notation "hlg" ;
  rdfs:label "giant stars"@en ;
  skos:prefLabel "giant stars"@en ;
  a skos:Concept ;
  skos:broader ilc2:hl .
...
```

```
...
ilc2:hlh
  skos:notation "hlh" ;
  rdfs:label "bright giant stars"@en ;
  skos:prefLabel "bright giant stars"@en ;
  a skos:Concept ;
  skos:broader ilc2:hl .

<http://www.iskoi.org/ilc/2/scheme>
  rdfs:label "Integrative Levels Classification (ILC)"@en ;
  skos:prefLabel "Integrative Levels Classification (ILC)"@en ;
  a skos:ConceptScheme .

ilc2:hla
  skos:notation "hla" ;
  rdfs:label "attributes of #hla"@en ;
  skos:prefLabel "attributes of #hla"@en ;
  a skos:Concept ;
  skos:broader ilc2:hl .

ilc2:hlf
  skos:notation "hlf" ;
  rdfs:label "subgiant stars"@en ;
  skos:prefLabel "subgiant stars"@en ;
  a skos:Concept ;
  skos:broader ilc2:hl .
```

Visualisation of SKOS conversion for Stars



ILC view (Stars) on BARTOC



Modelling ILC Facet Structure in SKOS

ILC fundamental facet categories are represented by notation of single digits *0-9* (eg *2* "in place"). More specific facet indicators have additional digits (eg *214* "in front of"). If the notation consists of a single digit it is modelled as a sub-property of *skos:related;* if more than 1 digit then as a sub-property of the property with the parent notation. Free facet indicators can connect any two classes in the schedules and therefore the domain and range values for these properties is *skos:Concept.*

ILC2 special facets are limited to specific main classes or subclasses and indicated by a numerical notation starting with *9*, with the facet indicator number following the class notation (eg *m981).* They are modelled as instances of *rdf:Property* having specific ILC concepts as their domain and range. Thus for the record m981 ("aged years") the domain is *m* ("organisms") and the range (from the 'foci' field) is *an* ("quantities"). The super property is then *m98* ("developmental stage").

The ILC's fundamental facet relationships are modelled by extensions of the associative relationship - specialized sub-properties of *skos:related,* with facet indicators modelled as *rdf:Property* sub-hierarchies. This approach permits the representation to encompass facets at the RDF level, as an extension to the SKOS standard.

The problem of (SKOS) coordination

Faceted classification schemes, such as ILC, do not enumerate all possible coordinated combinations of concepts, instead providing semantic "building blocks" for creating compound concepts. The issue of coordination was postponed by the SKOS working group and there is no SKOS standard mechanism to support coordination. It was considered out of scope for the ILC conversion exercise. However, we conducted some experiments for the paper, with a view to future faceted applications.

The simplest approach would be to create new coordinated concepts using a combined literal label:

skos:notation "n9mqvtur36wve"^^ilc:notationString ;
skos:prefLabel "populations—of—deer--affected by--road vehicles"@en .

Freely faceted combination of ILC classes *n* ("populations"), *mqvtur* ("deer") and *wve* ("road vehicles"), connected by the fundamental facet category *9* ("of <kind>") and the specific category *36* ("affected by").

Knowledge of the rules for combination are required to create the new coordinated concept but it can be treated like any other concept. The key problem for SKOS representation is that we cannot easily identify or isolate the components of the coordination. From an extract of a message (Leonard Will) in the public archives of the SKOS design discussions at the time:

> *"[…] Any encoding format for a classification scheme needs to be able to represent these compounds, and to display them clearly, showing their structure, both when the scheme is browsed on its own and when it is used to provide headings for a systematic classified catalogue of information resources. It should also be possible to search for any constituent concept and view the compounds in which it occurs. […]"*

MADS/RDF introduced the class *madsrfdf:ComplexSubject* as their approach to the modelling of coordinated concepts. This has as properties a label consisting of a double hyphen separated concatenation of component concept labels, plus a list of the corresponding component concept identifiers. The SKOS Primer suggests a similar property *(coordinationOf*) as a possible extension to list the component members of a coordinated concept.

Coordination in ILC SKOS

In MADS/RDF, the connections between concepts are implied by serial concatenation - but the meaning of these connections is not explicit.

ILC SKOS employs a more nuanced extension property to express coordination. This includes concepts and properties, in order to allow relationships between concepts according to the ILC facet syntax.

Example coordination involving ILC special facet and free facet

```
ilc:n9mqvtur36wve a skos:Concept ;
        skos:notation "n9mqvtur36wve"^^ilc:notationString ;
        skos:prefLabel "populations--of–-deer–-influenced by factor--road vehicles"@en ;
        ex:coordinationOf ( ilc:n ilc:9 ilc:mqvtur ilc:36 ilc:wve )
```

For ILC, the connecting facet indicators (*9* “of kind” and *36* “influenced by factor”) are an integral part of the coordinated concept, conveying the intended semantics more precisely.

See JDoc paper for examples of how this list of coordinated concepts can be queried using SPARQL 1.1.

Complexity in retrieval applications

We assume the main use case for ILC SKOS is retrieval – online search and browsing applications (as with other SKOS vocabularies).

Coordinated indexing with relationships between concepts facilitates very precise description of resources but the effort required to create such indexing and to make use of it in operational systems is high, which hinders implementations.

Some possible avenues for further work to address this complexity:

Auto-suggest functionality in cataloguing tools or services, building on the coordination techniques discussed could assist the creation of complex coordinated concepts for precisely describing resources.

Compound, faceted resource descriptions can be shown in the appropriate (ordered) position within a systematically organised browsing menu.

Search would be facilitated by a menu driven query builder capable of generating coordinated concepts for queries that took account of both ILC concepts and connecting properties (possibly emphasising the 10 fundamental ILC facet categories).

Balancing complexity and standardisation - a low-cost migration path?

Complex classification schemes have complex features extending beyond the boundaries of SKOS – as reflected in the name, *Simple* Knowledge Organization System.

Nonetheless, the SKOS Reference highlights the high degree of commonality of many types of KOS.

"The important point for SKOS is that, in addition to their unique features, each of these families shares much in common, and can often be used in similar ways. … The Simple Knowledge Organization System therefore aims to provide a bridge between different communities of practice within the library and information sciences involved in the design and application of knowledge organization systems. In addition, SKOS aims to provide a bridge between these communities and the Semantic Web, by transferring existing models of knowledge organization to the Semantic Web technology context, and by providing a low-cost migration path for porting existing knowledge organization systems to RDF." (Miles and Bechhofer, 2009).

Balancing complexity and standardisation - a low-cost migration path?

The ILC's top-level fundamental facet relationships are modelled by extensions of the associative relationship - specialized sub-properties of *skos:related.* This is seen as an appropriate compromise for modelling the complexity of the ILC within the constraints of the SKOS model, while facilitating future work that takes account of the syntax of ILC compound statements.

This approach permits the representation to encompass facets at the RDF level, albeit external to SKOS representation. Thus facet indicators (as *rdf:Property)* will be opaque to purely SKOS applications although available to general RDF semantic web applications and any application that is specifically ILC aware.

The ILC conversion follows the aim of SKOS in providing a low-cost migration path to the world of the semantic web, while capturing some of the ILC facet structure by a limited extension to incorporate RDF properties. The consequence of following a simple standardised model (SKOS) is that the output is easily validated, understandable, interoperable and reusable across a range of SKOS-aware applications.

References

Recent paper on this work

Binding, C, Gnoli, C & Tudhope, D 2021, 'Migrating a complex classification scheme to the Semantic Web: expressing the Integrative Levels Classification using SKOS RDF', Journal of Documentation, 77(4), 926-945. https://doi.org/10.1108/JD-10-2020-0167 - see below for OA version.

Open Access version available from https://bit.ly/2ocaHC6

Integrative Levels Classification (ILC) http://www.iskoi.org/ilc/

ILC (edition 2) on BARTOC registry https://bartoc-skosmos.unibas.ch/ilc2/en/

Park, Z, Gnoli, C, Morelli , D. 2020. 'The second edition of the Integrative Levels Classification', Journal of Data and Information Science, 5(1), 39-50

Claudio Gnoli publications https://www.gnoli.eu/worksy.php

KOS related research by Hypermedia Research Group https://hypermedia.research.southwales.ac.uk/kos/

STELETO Open Source code. https://github.com/cbinding/STELETO/

Acknowledgements

Ceri Binding designed and implemented the SKOS transformation of the ILC and created the RDF ILC representation. Andreas Ledl and Jakob Voss have made the ILC and its RDF representation available on BARTOC.

ILC edition 2

http://www.iskoi.org/ilc/2/ilc.php

main classes: *(select + to expand)*

a	forms	+
b	spacetime	+
c	branes	+
d	energy	+
e	atoms	+
f	molecules	+
g	continuum bodies	+
h	celestial bodies	+
i	rocks	+
j	land	+
k	genes	+
l	bacteria	+
m	organisms	+
n	populations	+
o	instincts	+
p	consciousness	+
q	language	+
r	rituals	+
s	communities	+
t	polities	+
u	enterprises	+
v	technologies	+
w	artifacts	+
x	artworks	+
y	knowledge	+

*search by **verbal caption**:*

submit

*browse by **notation**:* submit

*browse **facets** of category:*

Facets key

0	as for aspect	+
1	at time	+
2	in place	+
3	by agent	+
4	affected by disorder	+
5	with transformation	+
6	having property	+
7	with part	+
8	as form	+
9	of kind	+

*browse **special classes**:*

A	those	+
B	the 1st ones	+
C	the 2nd ones	+
D-T	(etc.)	
U	the typical	+
V	the entirety	+
W	together with	+
X	something	+
Y	the actual	+
Z	the mentioned	+