Proposal for NKOS Consolidated Workshop, September 2020
Theme: Terminology Development and KOS Mapping

Faceted classifications as linked data: a logical analysis

Claudio Gnoli
Science and Technology Library, University of Pavia, Italy
claudio.gnoli@unipv.it

Faceted classifications are an information-rich and relatively complex kind of knowledge organization system (KOS) (Broughton 2011). This makes their formal representation as linked data an important but demanding task. The Integrative Levels Classification (ILC) is a faceted classification based on phenomena, instead of disciplines as is usual in the bibliographic tradition. It has been developed since 2004 by an international team. Its second edition (ILC2) was published in 2019 (Park et al. 2020) and consists of more than ten thousands classes and facets.

In the same year, ILC2 was also converted and published in the SKOS format. Work to produce the SKOS version has built on previous analyses of the structural elements of a freely faceted classification (Gnoli et al. 2011; de Santis and Gnoli 2016) and implied various choices concerning how elements could be best represented (for a preliminary discussion see Binding et al. 2020). As a result, basic classes (e.g. `xf` "paintings"), facet indicators (`x99` "of *style*") and their possible foci (e.g. "baroque") have been covered in the SKOS version, while their combinations, such as `xf99o` "paintings, baroque" have been excluded. This paper focuses on how facets can be modeled as linked data in any faceted classification, including ILC.

A facet basically expresses a relationship, so in the logic of linked data is equivalent to an RDF property that connects the subject and the object of a triple (Trzmielewski and Gnoli 2019, p. 9). Each RDF property is qualified by a *domain*, specifying which classes in a scheme can be the subject of the triple, and a *range*, specifying which classes can be the object of the triple. In the terminology of faceted classification, the subject is often called a *basic class*, the property is called a *facet indicator* and the object is called a *focus*. The following example shows the correspondence between this terminology and its representation in SKOS Turtle syntax:

BasicClass	facetIndicator	Focus
`xf`	`29`	`f`
painting	in	Italy

```
<29> a rdf:Property;
      skos:notation "29"^^xsd:string;
      skos:prefLabel "in country"@en;
      rdfs:label "in country"@en;
      rdfs:domain skos:Concept;
      rdfs:range <tt>;
      rdfs:subPropertyOf <2>.
```

In ILC and other faceted KOSs, it is possible to attach more than one facet to one and the same basic class. For example, the basic class "paintings" in the previous example can be qualified by both a quality facet and a place facet, to give `xf99o29f` "paintings, baroque, in Italy". Such

multi-faceted classes are equivalent to several RDF triples connected by a relation of intersection. A combination with two facets could then be translated into two triples sharing the same subject:

`xf99o29f = xf99o ∩ xf29f`

Specification of the possible subjects (domain) and objects (range) of each facet as linked data actually depends on what exactly one is meaning by “facet”. Indeed, the term *facet* can be used to denote many different ways in which concepts are combined (Vickery 1975; Gnoli 2017). “When reference is made to their nature or essence, the categorization process consists in allocating objects and concepts to a facet” as defined by a limited number of fundamental categories, “e.g. living being, physical object, attribute, activity, space, time” (Hudon 2019, § 4.2). In this sense, a concept can be a subclass of a more general concept (`skos:broader`) or a property can be a sub-property or another (`rdfs:subPropertyOf`): for example, “style” may be a subproperty of “attribute”. In ILC, fundamental categories are expressed by the digits used as facet indicators, and multi-digit facets reflect their parent facets and categories: `29` “in *country*” is a subproperty of `2` “in *situation/place*”, as reflected in the SKOS example above.

“But the facet can also specify the role played by the concept” (Hudon 2019, § 4.2) and be used as a role operator, with a syntactic rather than semantic function. Such ambiguity between semantic nature and syntactic role, that is found in many classical faceted classifications (Maniez 1999), needs to be better analyzed in order to represent facets in formalized ways. The very need to specify domain and range in linked data can act as a guide to such disambiguation. Indeed, each of them can be either unrestricted or restricted, which gives rise to four logical possibilities (plus two variations). We will illustrate them by using examples from the Dewey Decimal Classification (DDC), as this KOS is better known to a large public, although its original structure is not faceted, which is reflected in various inconsistencies in its notation. Being designed more recently, ILC has formal ways and terminology to distinguish between such different facet kinds. To make the syntactic structure clear, we will translate example notation into the verbal form “basic class, facet indicator: focus” and artificially separate these three elements with blank spaces.

The kind of facets most familiar in bibliographic classifications, known as “common subdivisions”, “common auxiliaries” or (in UDC and ILC) “**common facets**”, have unrestricted domain and restricted range:

`X 09 45` "any subject, in: Italy"

These facets work as suffixes, like `-0945`, that can be appended to any basic class to specify that its meaning is limited to a given region, historical period, document form etc. In linked data, they can be represented as a property having any `skos:Concept` (or, more generally, any `rdfs:Class`) as its domain but only a specific class as its range, e.g. `tt` “countries” in the SKOS example above. These express both the nature of the introduced focus (the fact that Italy is a place) and its role as a specification of the basic class, while the nature of the basic class remains unspecified.

The second possibility is that both domain and range are restricted. These are most of **special facets** typical of such classical faceted classifications as Colon and Bliss 2. In DDC they only occur in classes that have been recently restructured in a faceted way, such as music:

786.2 1 83 "piano, musical form: sonata"

Special facets specify a syntactic role plus the nature of both the basic class (as this facet can only be used with subclasses of music) and the focus (as the facet can only introduce concepts of a musical nature). Gnoli (2006) has described the latter as *context-defined foci (CDF)*, as their very meaning is defined only inasmuch as facets of music; these facets we can call ***bound*** *special facets*.

A variation of this is that foci are taken from other, still restricted parts of the schedules that are external to music. These, called *extra-defined foci (EDF)* in papers on ILC, can be taken from a specific class, to which the present one is then said to have “parallel divisions” (we can then call these ***parallel*** *special facets*). A DDC example is in vocal music for specific non-Christian religions, where foci must be taken from the subdivisions of 290 “non-Christian religions” such as 294.5 “Hinduism”:

782 3 45 "vocal music for service, of religion: Hinduism"

If we now keep the domain restricted to music, but leave the range to cover any class from the scheme, we get yet another type of special facets. Such ***free*** *special facets* take their foci from the generality of subjects in the KOS (001/999 in DDC notation), thus still have a restricted domain (e.g. music, or library science) but this time an unrestricted range:

78 00 61 "music, in relation with: medicine"
02 6 34 "libraries, specializing in: law"

In this case, the facet expresses a syntactic function plus the nature of the basic class, while the nature of foci is not expressed. This situation is typical of such classes as those of libraries, documents, artworks, philosophies or languages, that can have any concept as their object.

The last possibility is to have both domain and range unrestricted. These can be called ***free*** *common facets* as they connect any pair of classes. They correspond to “operators” of such other KOSs as PRECIS or to “phase relationships” in Colon and UDC:

X 015 X "any subject, principles: any science"

Again, we are using a very particular example from DDC, as in this KOS free facets are only available to connect any basic class with a class from 500 “pure sciences”, to give -0151 “mathematical principles”, -0154 “chemical principles” etc. (so they could be described more properly as parallel common facets). However, this example shows how there would be no obstacles to introduce free facets in all DDC, e.g. by defining (like in music) a “common subdivision” 00 that can be followed by 001/999 to give X 00 X "any subject, in relation with: any subject".

Free common facets only express a syntactic role, as they do not inform about the nature of either the basic class or the focus. The inverse logical possibility, expressing nature but not roles, occurs in thesauri, where a term is said to belong to a given “facet” such as activities, places etc.; here its combination with other terms is left to post-coordination.

To summarize, we suggest that in a faceted classifications six different syntactic kinds of facets should be distinguished, each with a different notation (a modified, lighter notation for future ILC3 is also hypothesized):

	domain	**range**	**ILC2**	**ILC3 ?**
common facets	unrestricted	restricted:		
- bound		to domain itself: CDF	-	- (occur in DDC)
- parallel		to other class: EDF	09, 19, 29...	0-2, 09, 39, 49
- free		unrestricted	0-8	3-4, 10, 20
special facets	restricted	restricted:	91-99	
- bound		to domain itself: CDF		5-7, 86, 96
- parallel		to other class: EDF		8-9, 59, 69, 79
- free		unrestricted	91V-99V	50, 60, 70, 80, 90
semantic-only facets	-	-	(occur in thesauri)	

Our analysis is meant to be a contribution to making the meaning of facets in specific KOSs and KOS classes more clear. This can have effects on a rationalization of notation in the developing version of ILC. We also wish that, in light of it, the appropriate varieties of facets can be presented in a more systematic way in future introductions to faceted KOSs and their representation as linked data.

References

Binding, Ceri, Claudio Gnoli, Gabriele Merli, Marcin Trzmielewski and Douglas Tudhope. 2020. “Integrative Levels Classification as a networked KOS: a SKOS representation of ILC2”. In Daniel Martinez Avila and Marianne Lykke eds., *Knowledge organization at the interface*, Baden Baden: Ergon, in press.

Broughton, Vanda. 2011. “Facet analysis as a tool for modelling subject domains and terminologies”. In Aida Slavic and Edgardo Civallero eds., *Classification and ontology: formal approaches and access to knowledge: proceedings of the International UDC Seminar 2011*, Würzburg: Ergon, p. 207-227.

de Santis, Rodrigo and Claudio Gnoli. 2016. “Expressing dependence relationships in the Integrative Levels Classification using OWL”. In José Augusto Chaves Guimarães, Suellen Oliveira Milani and Vera Dodebei eds., *Knowledge organization for a sustainable world*: proceedings of the Fourteenth International ISKO Conference, Rio de Janeiro, September 27-29 2016. Würzburg: Ergon, p. 368-375.

Gnoli, Claudio. 2006. “The meaning of facets in non-disciplinary classifications”. In Gerhard Budin, Christian Swertz and Konstantin Mitgutsch eds., *Knowledge organization for a global learning society: proceedings of the Ninth International ISKO Conference, 4-7 July 2006, Vienna,* Würzburg: Ergon, p. 11-18.

Gnoli, Claudio. 2017. “Syntax of facets and sources of foci: a review of alternatives”. In Aida Slavic and Claudio Gnoli eds., *Faceted classification today: proceedings of the International UDC Seminar 2017*, *London*, Würzburg: Ergon, p. 243-256.

Gnoli, Claudio, Philippe Cousson, Tom Pullman, Gabriele Merli and Rick Szostak. 2011. “Representing the structural elements of a freely faceted classification”. In Aida Slavic and Edgardo Civallero eds., *Classification and ontology: proceedings of the International UDC Seminar 2011, The Hague*, Würzburg: Ergon, p. 193-206.

Hudon. Michèle. 2019. “Facet”. In Birger Hjørland and Claudio Gnoli eds., *ISKO Encyclopedia of Knowledge Organization*, https://www.isko.org/cyclo/facet.

Maniez, Jacques. 1999. “Des classifications aux thésaurus: du bon usage des facettes”. *Documentaliste - Sciences de l'information* 36: 249-262.

Park, Ziyoung, Claudio Gnoli and Daniele P. Morelli. 2020. “The second edition of the Integrative Levels Classification: evolution of a KOS”. *Journal of Data and Information Science* 5, n. 1, p. 39-50, http://manu47.magtech.com.cn/Jwk3_jdis/EN/10.2478/jdis-2020-0004 .

Trzmielewski, Marcin and Claudio Gnoli. 2019. “Une classification interdisciplinaire pour l’échange et la médiation des données ouvertes de la recherche”. *HAL*, https://hal.archives-ouvertes.fr/hal-02307108 .

Vickery, Brian C. 1975. *Classification and indexing in science*. 3rd ed. London: Butterworths.