# Faceted Classifications as Linked Data: A Logical Analysis†

Claudio Gnoli

University of Pavia Science and Technology Library, via Ferrata 1, Pavia 27100, Italy,
<claudio.gnoli@unipv.it>

Claudio Gnoli is librarian at the University of Pavia Science and Technology Library. His interests include knowledge organization in general, theory of classification, their philosophical grounds and their application to online knowledge bases. He has authored papers and books in the KO field, including *Introduction to Knowledge Organization* (Facet, 2020). He is a former vice-president and the current webmaster of international ISKO and a member of the UDC Consortium Advisory Board. He is a member of the editorial boards of *Knowledge Organization*, *Axiomathes* and *AIDA Informazioni*. His website is at gnoli.eu.

Gnoli, Claudio. 2021. "Faceted Classifications as Linked Data: A Logical Analysis." *Knowledge Organization* 48(3): 213-218. 16 references. DOI:10.5771/0943-7444-2021-3-213.

**Abstract:** Faceted knowledge organization systems have sophisticated logical structures, making their representation as linked data a demanding task. The term *facet* is often used in ambiguous ways: while in thesauri facets only work as semantic categories, in classification schemes they also have syntactic functions. The need to convert the Integrative Levels Classification (ILC) into SKOS stimulated a more general analysis of the different kinds of syntactic facets, as can be represented in terms of RDF properties and their respective domain and range. A nomenclature is proposed, distinguishing between common facets, which can be appended to any class, that is, have an unrestricted domain; and special facets, which are exclusive to some class, that is, have a restricted domain. In both cases, foci can be taken from any other class (unrestricted range: free facets), or only from subclasses of an existing class (parallel facets), or be defined specifically for the present class (bound facets). Examples are given of such cases in ILC and in the Dewey Decimal Classification (DDC).

Received: 6 January 2021; Revised 9 September 2021; Accepted 10 September 2021

Keywords: facet domain, facet range, faceted classification, Resource Description Framework (RDF)

† This paper was originally presented at the NKOS Consolidated Workshop 2020, September 9-10, 2020 (Virtual). https://nkos-eu.github.io/2020/content/Coladangelo-NKOS2020-Slides-min.pdf

## 1.0 Introduction

In the spectra where different kinds of knowledge organization systems (KOS) are listed, classification schemes are often placed at a medium level of semantic complexity, before thesauri and ontologies (e.g. Hodge 2000). However, the principles on which such spectra are based have been questioned (Souza et al. 2012). As for classification schemes, authors of such models may have old enumerative schemes in mind, such as the Library of Congress Classification. Actually, modern classification schemes, especially faceted ones, are richer systems which encode for a variety of relationships, similar to what thesauri and ontologies do, although in different forms (Broughton 2011). Even such originally enumerative schemes as the Dewey Decimal Classification (DDC) and, to a greater extent, the Universal Decimal Classification (UDC) have gradually incorporated syntactic mechanisms that are also described as facets and allow for a greater, though not complete, flexibility in the construction of compound classmarks.

Today, KOSs are increasingly published as linked data expressed in such formats as SKOS (Simple Knowledge Organization System) and RDF (Resource Description Framework). These formats allow encoding of the KOS conceptual structures in ways that can be shared digitally among different users.

The formal representation of faceted structures as RDF and SKOS data is thus an important though demanding task, in order to leverage the power of faceted classifications in the global cloud of linked data. This paper focuses on how facets in any classification scheme can be modeled as RDF triples, using examples from a natively faceted scheme, Integrative Levels Classification (ILC), and a partially facetized one, DDC.

## 2.0 Modelling facets in ILC

The Integrative Levels Classification is a general 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 led by the present author. Its second edition (ILC2) was published in 2019 (Park et al. 2020) and consists of more than ten thousand 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 has implied various choices concerning how such elements could be best represented (Binding et al. 2020; 2021). 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 and left to specific applications of the scheme. Work for the SKOS conversion of ILC implied more general questions concerning how facets should be modelled in the logic of RDF. Basically, a facet expresses a relationship; in the logic of linked data, this is equivalent to an RDF property that connects the subject and the object of a triple (Trzmielewski and Gnoli 2019, 9; Binding et al. 2021).

Each RDF property is qualified by a *domain*, specifying which classes in a scheme can take the subject role in the triple, and a *range*, specifying which classes can take the object role in 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, taken by the real ILC SKOS linked data (http://www.iskoi.org/ilc/skos.php):

| BasicClass | facetIndicator | Focus |
|---|---|---|
| xf | 29 | f |
| painting | in | Italy |

```
<29> a rdf:Property;
     skos:notation
     "29"^^xsd:string;
     skos:prefLabel   "in   coun-
     try"@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. Therefore, a combination with two facets will have to be translated into two triples sharing the same subject:

```
xf99o29f  =  xf99o    xf29f
```

## 3.0 Facet types

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 listed and combined (Vickery 1975; Gnoli 2017a).

A first distinction is made clear by Hudon (2019, referring to Maniez 1999), between facets as essential categories and facets as role operators. "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, "country" may be a subproperty of "space". In ILC, fundamental categories are expressed by the digits used as facet indicators, and multi-digit facets reflect their parent facets and categories: thus 29 "in *country*" is a subproperty of 2 "in *situation/place*", as reflected in the last line of 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 classical faceted classifications (Maniez 1999), as well as the ambiguity among different types of syntactic facets, needs to be better analyzed before facets can be appropriately represented in formalized ways.

The very need to specify domain and range in linked data can act as a guide to such analysis. Indeed, both domain and range can be either unrestricted or restricted, which gives rise to four logical possibilities (plus two variations, as we will see). We will illustrate them by using examples from the Dewey Decimal Classification (DDC), as this KOS is better known to a broad public, although its original structure is not faceted, which is reflected in certain inconsistencies in its notation. Being designed more recently, ILC has formal ways as well as special terminology to distinguish between such dif-

ferent facet kinds, although some of its classes are not developed in depth yet. In order to make the syntactic structure clear, we will translate example notation into the verbal form "basic class, facet indicator: focus" and will separate these three elements within DDC classmarks by blank spaces.

### 3.1 Common facets

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"
```

Such 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, these 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 ILC SKOS example above or, in DDC, Table 2 for geographical areas. Common facets thus 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.

Common facets in the example above are *parallel* to Table 2 and to such basic classes as 914.5 "geography of Italy". However, other common facets exist, such as -07 "education", that only *bound* to their specific suffix, with no reference to any basic class (cfr. 3.2). The notions of bound (see 3.2), parallel (see 3.3) and free (see 3.4) are discussed in more detail in the next sections which provide examples from special facets.

### 3.2 Bound special facets

The second possibility is that both domain and range are restricted. These are most of the special facets typical of such classical faceted classifications as Colon Classification (CC) and the second edition of Bliss Bibliographic Classification (BC2). 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 a specific context, like music; these facets we can call *bound special facets*.

An additional issue concerning bound special facets is whether a facet defined for a given domain can be inherited by subclasses of that domain. The simplest solution seems to be that it is indeed inherited, which was described as "cascading facets" in ILC terminology. For example, the special facet "with *organ*" can be defined as having the class of animals as its domain, and "with stomach" can be one of its foci ("animals, with: stomach"). Subclasses of animals, like mammals or bovids, can also take the same facet and foci to give such meaningful combinations as "mammals, with: stomach" etc. (In this perspective, even common facets are nothing but facets whose domain is inherited by all classes in the scheme.) However, there are also foci like "with: rumen" whose domain is only bovids, not their parent classes mammals and animals, as not all of these have a rumen. Notational solutions are needed to express such situations.

### 3.3 Parallel special facets

A variation of the music case above 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 them *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"
```

### 3.4 Free special facets

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 facet. 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 such as those of libraries, documents, artworks, philosophies or languages, that can have any concept as their object. (In the case of artworks, DDC does not reuse notation for 001/999 but weirdly provides a whole new mini-classification of all possible subjects listed in a different order, that is an array of

bound special facets which can only be appended to 700.4 "artworks on specific themes and subjects" and some related classes; cfr. Beall 2020.)

### 3.5 Free facets

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 Preserved Context Index System (PRECIS) or to "phase relationships" in CC 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 introducing free facets throughout 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" (in case one is afraid that a series of noughts may generate confusion, a different notation currently unused could instead be adopted, e.g. -013).

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.; its combination with other terms is there left to post-coordination.

## 4.0 Proposed syntax for ILC3

In ILC2, facets are basically defined as free facets, which are represented by indicators 0/9; special facets are expressed as 91/99 (Park et al. 2020). This corresponds to the interdisciplinary character of the scheme, that allows free combination of any pair of phenomenon concepts. On the other hand, when documents in specific domains are indexed, special facets occur very often and lead to lengthy, repetitive classmarks, such as xf99o98g95q94r "paintings, baroque, allegorical, reproduction, damaged by break". Common facets in ILC2 are also free by default, so that -2 "in *situation/place*" introduces any other class, like -2wvlh "in ships", and much more common occurrences such as "in Italy" are also long (-2ttf) or need to be shortened by defining an additional parallelizing digit (-29f as in our initial example).

Previous work (Gnoli 2017b) has also made clear that, out of the ten ILC fundamental categories, those with lesser ordinal value (0 perspective, 1 time, 2 place, 3 agent) most often express relationships of the considered system to its external environment: these are typically expressed by common facets. On the other hand, the next categories (4 disorder, 5 process, 6 property, 7 part...) most often refer to specifications internal to the system itself: these are typically expressed by special facets.

These considerations, in light of the present analysis on special and common facets, all together suggest that notation for a next edition of ILC (ILC3, which is currently in its initial developing stage) can be improved by adopting some default syntax concerning domains and ranges. After various experiments, it seems that a convenient solution is defining 0, 10, 20... as free common facets, 1-9 and 09 ... 99 as bound special facets and all other multiple-digit facets as parallel special facets. This would mean that e.g. -27 "in *region*" would now imply [tU] "country" as its default range so that "in Italy" would be shortened to -27f. While this requires indexers to learn some additional rules, it offers a predictable grammar that is consistent throughout the whole scheme.

To keep the freedom of combination which is a typical feature of ILC, any facet can always be made free by just appending a -0, so that "in ships" can still be expressed as -20wvlh; or made bound by appending a -9. Experimentation with this renewed, "lighter" syntax is currently being carried out.

To summarize, we suggest that in faceted classifications six different syntactic kinds of facets should be distinguished, each with a different notation (Table 1). Semantic-only facets do not occur in classification schemes but they do in thesauri.

## 5.0 Conclusion

Having to define the domains and ranges of ILC facets more formally has stimulated more general considerations on kinds of facets in any classification scheme. These could also be extended to other KOS kinds, such as subject headings, thesauri and ontologies.

Current experimentation with ILC shows that different facet kinds are more or less convenient according to how a scheme is meant to be used. Free facets are more useful for the classification of interdisciplinary collections, where any pair of concepts can happen to be associated; this has been done, for example, by role operators of such general verbal KOSs as PRECIS and Syntol. Bound facets, on the other hand, are useful to provide a compact notation for indexing domain-specific and specialized collections; classical faceted classifications such as CC and BC2 mainly work in this way, as every main class of them is like a separate domain-specific classification. Being conceived as a flexible system for many potential uses, ILC is hoped to offer a consistent syntax that can be used for all options, at the cost of learning some additional rules.

| | domain | range | ILC2 | ILC3 draft |
|---|---|---|---|---|
| **common facets** | unrestricted | restricted: | | |
| - bound | | to domain itself: CDF | - | - (occur in DDC) |
| - parallel | | to other class: EDF | 09, 19, 29... | 01–08 |
| - free | | unrestricted | 0-8 | 0, 10, 20... |
| **special facets** | restricted | restricted | 91-99 | |
| - bound | | to domain itself: CDF | | 1–9, 19, 29... |
| - parallel | | to other class: EDF | | 11–18, 21–28... |
| - free | | unrestricted | 91-99V | |

*Table 1.*

Our analysis is meant to be a contribution to making the meaning of facets in specific KOSs and KOS classes more clear. We also wish that, in this light, the variety of facets can be presented in more systematic ways in future introductions to faceted KOSs and can be represented as linked data in appropriate ways.

**Acknowledgments**

Other members of the ILC research project contributed elements useful to the present analysis. In particular, modeling of facets as RDF properties was conceived independently by Ceri Binding, in order to implement it in SKOS syntax, and by Marcin Trzmielewski in general terms. Application to BARTOC of free vs. faceted ILC classmarks was discussed and tested with Ziyoung Park. Both Marcin and Douglas Tudhope have provided useful suggestions for improving this paper.

**References**

Beall, Julianne. 2020. "Table of Mappings: DDC 001-999 to Table 3C—3 Arts and literature dealing with specific themes and subjects." *025.431: the Dewey blog*, https://ddc.typepad.com/025431/2020/12/.

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 *Knowledge Organization at the Interface*, edited by Marianne Lykke, Tanja Svarre, Mette Skov and Daniel Martínez-Ávila. Advances in Knowledge Organization 17. Baden-Baden: Ergon, 49-58.

Binding, Ceri, Claudio Gnoli, and Douglas Tudhope. 2021. "Migrating a Complex Classification Scheme to the Semantic Web: Expressing the Integrative Levels Classification using SKOS RDF." *Journal of Documentation*, in press.

Broughton, Vanda. 2011. "Facet Analysis as a Tool for Modelling Subject Domains and Terminologies." In *Classification and Ontology: Formal Approaches and Access to Knowledge: Proceedings of the International UDC Seminar 2011*, edited by Aida Slavic and Edgardo Civallero. Würzburg: Ergon, 207-27.

De Santis, Rodrigo and Claudio Gnoli. 2016. "Expressing Dependence Relationships in the Integrative Levels Classification using OWL." In *Knowledge Organization for a Sustainable World*: *Proceedings of the Fourteenth International ISKO Conference, Rio de Janeiro, September 27-29 2016*, edited by José Augusto Chaves Guimarães, Suellen Oliveira Milani and Vera Dodebei. Würzburg: Ergon, 368-75.

Gnoli, Claudio. 2006. "The Meaning of Facets in Non-disciplinary Classifications." In *Knowledge Organization for a Global Learning Society: Proceedings of the Ninth International ISKO Conference, 4-7 July 2006, Vienna,* edited by Gerhard Budin, Christian Swertz and Konstantin Mitgutsch. Würzburg: Ergon, 11-18.

Gnoli, Claudio. 2017a. "Syntax of Facets and Sources of Foci: a Review of Alternatives." In *Faceted Classification Today: Proceedings of the International UDC Seminar 2017*, edited by Aida Slavic and Claudio Gnoli. London, Würzburg: Ergon, 243-56.

Gnoli, Claudio. 2017b. "Classifying Phenomena, Part 3: Facets." In *Dimensions of Knowledge: Facets for Knowledge Organization*, edited by Richard Smiraglia and Hur-Li Lee. Würzburg: Ergon, 55-67.

Gnoli, Claudio, Philippe Cousson, Tom Pullman, Gabriele Merli and Rick Szostak. 2011. "Representing the Structural Elements of a Freely Faceted Classification." In *Classification and Ontology: Proceedings of the International UDC Seminar 2011, The Hague*, edited by Aida Slavic and Edgardo Civallero. Würzburg: Ergon, 193-206.

Hodge, Gail. 2000. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington, DC: Council on Library and Information Resources. Available at http://www.clir.org/pubs/reports/pub91/contents.html.

Hudon. Michèle. 2019. "Facet." In *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjørland and Claudio Gnoli. https://www.isko.org/cyclo/facet. Republished in *Knowledge Organization* (2020), 47(4): 320-33.

Maniez, Jacques. 1999. "Des Classifications aux Thésaurus: du Bon Usage des Facettes." *Documentaliste - Sciences de l'Information* 36: 249-62.

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(1): 39-50, http://manu47.magtech.com.cn/Jwk3_jdis/EN/10.2478/jdis-2020-0004

Souza, Renato Rocha, Douglas Tudhope and Maurício Barcellos Almeida. 2012. "Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems." *Knowledge Organization* 39(3): 179-92.

Trzmielewski, Marcin and Claudio Gnoli. 2019. "Une Classification Interdisciplinaire pour l'Échange et la Médiation des Données Ouvertes de la Recherche." 12th French ISKO Conference, Montpellier, 9-11 October 2019. *HAL*, https://hal.archives-ouvertes.fr/hal-02307108.

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