From health professionals' discourse to KOS:
representing the facets of allergy in the Integrative Levels Classification

Marcin Trzmielewski* and Claudio Gnoli**

* Laboratoire d'Etudes et de Recherches Appliquées en Sciences Sociales (LERASS), Université Paul-Valéry
Montpellier 3, France
** Biblioteca della Scienza e della Tecnica, Università di Pavia, Italy

Allergy is a major health issue in our society in terms of care and prevention. In France, allergy or allergology, a domain that studies and treats allergies, was only recognized as a specialty in its own right in 2017 (Demoly 2017), and there is no KOS that might be used by professionals in this domain for their activities of organization, classification and search for information (Trzmielewski *et al.* 2019). This situation led Trzmielewski (2022) to cooperate with the Allergy Unit of the University Hospital of Montpellier to develop the AllergiDoc project, aiming to create an ontology to represent and organize allergy knowledge and support work activities of allergy professionals. The project started with a study of information practices, which was carried out in 2020-2021 in the Allergy Unit by Trzmielewski. He elaborated 16 participants' observations of 8 journal club meetings, devoted to the presentation and critical analysis of scientific articles and conference presentations, and 8 clinical meetings focused on the presentation and analysis of patient records. He also conducted 20 interviews with professionals investigating their practices.

He gathered 497 terms from the corpus of data on practices, based on the reports of the observations and on the transcripts of the interviews. The thematic analysis of these data brought up 17 facets, further validated by allergy professionals, by checking whether they are useful for document indexing and characterization of allergic cases. The validation led to the identification of several facets of *phenomena* (Gnoli 2016), used by professionals to search for information: "Allergen", "Healthcare circuit", "Comorbidity", "Diagnostic methods", "Disease", "Mechanism", "Person", "Prevention", "Quality of life", "Risk factor", "Symptom", "Treatment".

Facets of other knowledge dimensions (*Ibid.*) were also validated, in particular the *perspective* dimension, which allow to express particular points of view on the contents: "Data source", "Discipline", "Methodology"; and the *document* dimension, which expresses

features of the production, form and function of document instances: "Author", "Document". While in a general classification a consistent framework is needed, special concepts represented in it should also be identified from the actual usage contexts of particular domains, such as allergy.

Discrimination between facets of phenomena and facets of perspective proved to be a significant issue, as both options were considered for such facets as "Mechanism" or "Discipline". A paper entitled "Clinical relevance of cross-reactivity in food allergy" could be indexed as "cross-allergy, food allergy, mechanism", where mechanism is considered as a perspective facet, or as "food allergy, cross reactivity (mechanism)" where mechanism is considered as a phenomenon facet; after discussion with healthcare professionals, the latter option has been chosen, as mechanism was meant to express various kinds of allergic reactions such as "crossed", "IgE-mediated" or "non-IgE-mediated". The "Discipline" facet has emerged in the AllergiDoc study to express the different disciplinary perspectives adopted in *e.g.* the allergological approach versus the pneumological approach in the treatment of asthma, as literature is indeed discussing the recent development of allergy as a field on its own (Demoly 2017).

Facets and terms as identified in the AllergiDoc study are then applied in the Integrative Levels Classification (ILC) developing version, the ongoing evolution of the existing second edition (ILC2) of this faceted KOS (Binding *et al.* 2021). One research question is how may we transpose and apply the strongly contextual, bottom-up data about representation of the allergy domain (facets, concepts, terms) into a general classification with main classes mostly developed in a top-down way. Indeed, the representation of medical concepts seems to be an issue in the development of detailed subdivisions and facets of any general KOS, as indexing of medical documents raises the challenge of standardization versus personalization (Morquin & Ologeanu-Taddei 2018).

According to the theory of facet analysis (Vickery 1960), the facets of any domain will belong to a limited set of general categories. ILC provides for the following ten basic categories:

0 perspective
1 position; time
2 situation; place
3 agent
4 opposition
5 change
6 property
7 part

8 quantity
9 quality

Thus, in ILC class *rh* "healthcare", facet *rh3* "by medical equipment" belongs to the general category of agents. Further category numerals allow to express more specific facets, for example *rh33* will mean the agent-of-agent in healthcare, that is, health worker. Numerals also rule the syntactical properties of a facet, including what in ontology terms are called their domain and range (Gnoli 2021).

ILC classes and their facets are stored in a MySQL database including different tables for the different editions of the KOS (ILC1, ILC2, ILC developing version). One feature of ILC is its expressive notation, in which the number and kind of symbols (literals, numerals etc.) reflect the structure of the system and are designed in such a way to produce meaningful orders. This is exploited in the web interface by which ILC schedules may be browsed and searched (http://www.iskoi.org/ilc/ilc.php), as classes, subclasses and their facets are automatically displayed in sorted ways and in different colors.

As allergy is part of healthcare, existing facets for ILC class *rh* "healthcare" have been considered as a starting point. Healthcare is a domain particularly rich in facets and these were drafted in ILC previous editions by Gnoli and Tom Pullman. For example, facets for diseases (listed under *rh9-* in the developing version) have to cover not only a list for types of disease – like "hypersensitivity" including allergy, "inflammation" and so on – but also one for treated parts, one for symptoms, one for complications, one for severity, etc. Such special facets may be expressed by further numerals:

rh   "healthcare"
rh59   "treated by *therapy*"
rh94   "healing *condition*"
rh942   "with *concurrent condition*"
rh943   "caused by *organic pathogen*"
rh946   "showing *symptom*"
rh947   "complicated by *complication*"
rh96   "for *patient*"
rh97   "of *treated part*"
rh98   "*severity*"

So for example the subject "immunotherapy in children food allergy" can be expressed as *rh96i94iplf59m* "healthcare, for children, healing food allergy, treated by immunotherapy".

Comparison between these draft ILC facets for healthcare and the facets identified by the bottom-up methods of the AllergiDoc project proves to be a very useful exercise. Indeed,

draft facets are tested by the examination of actual cases in allergy documentation, and may be validated, refined or even reorganized as a result. Also, allergy concepts may show the need for additional special facets to be represented as specifications of existing facets by additional numerals. The representation of allergy knowledge will soon be completed by further terms coming from different kinds of textual documents used by allergy professionals: titles and abstracts of scientific articles, messages from a general-public health forum, and clinical documents redacted in the Allergy Unit. As such documents are produced by different actors – researchers and allergists, patients and the wider public, professionals from the Allergy Unit – it will be interesting to see how terminological, conceptual and discursive varieties resulting from such complexity may be represented in a general faceted classification. Comparison with the classes for allergies in the International Classification of Diseases or the ontologies available in the BioPortal may also be useful.

References

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* 77, n. 4: 926-945.

Demoly, Pascal 2017. "L'allergologie ? Désormais une spécialité médicale universitaire". *Info Respiration*, n. 139: 15-16.

Gnoli, Claudio. 2016. "Classifying phenomena, part 1: Dimensions", *Knowledge Organization* 43, n. 6: 403-415.

Gnoli, Claudio. 2021. "Faceted classification as linked data: a logical analysis". *Knowledge Organization* 48, n. 3: 213-218.

Morquin, David and Roxana Ologeanu-Taddei. 2018. "Le dossier patient informatisé: enjeux de la standardisation et personnalisation de l'information pour les professionnels de santé". In: Céline Paganelli (ed.). *Confiance et légitimité dans le champ de la santé*. London: ISTE.

Trzmielewski, Marcin. 2022. Vers l'élaboration d'un système de représentation des connaissances en allergologie: analyse des documents et des pratiques informationnelles des acteurs. Doctorales SFSIC. Dijon, 23-24 June 2022.

Trzmielewski, Marcin, Pascal Demoly, Luciana Kase Tanno, Davide Caimmi and Claudio Gnoli. 2019. "Information and Data in Allergology: Theoretical and Methodological Proposals to Build a Knowledge Organization Model dedicated to the Design of Info-Communication Devices". In: *ISKO Low Countries Conference. Morsels of Knowledge*, Bruxelles, 20-21 June 2019.

Vickery, Brian C. 1960. *Faceted classification: a guide to construction and use of special schemes*. London: Aslib.