Wikidata and the Thesaurus Nuovo Soggettario: Together to Build a Domain Ontology in the Field of Photography

Silvia Bruni, Alida Daniele, Fabrizio Nunnari, Elena Cencetti, Elisabetta Viti, Francesca Palareti

Abstract: The project aims to build a domain ontology on photography in Wikidata, in dialogue with the Nuovo soggettario of the National Central Library of Florence, the official Italian subject indexing tool for resources of various kinds. In the Nuovo soggettario, terms are organized in a hierarchical structure (from the most general to the most specific), while in Wikidata, entities are defined through their properties and relationships. Aligning conceptual hierarchies and enriching both databases with new terms and links to other Knowledge Organization Systems (KOSs) highlights the value of participatory semantic modeling to enhance interoperability between tools and cultural resources.

Keywords: Nuovo soggettario, ontology, searchability, accessibility, interoperability, reusability of tools and information resources.

Since 2018, the Social Sciences Library has set up a Wikistation to promote Wikimedia projects within the University of Florence. A Wikidata working group has also been formed with the aim of deepening the study of ontology and its application in the field of research (Bruni 2025).

In 2024, the UNIFI Wikidata group began collaborating with the National Central Library in Florence to develop a discussion on methods and tools for organising knowledge, with a particular focus on thesauri, classifications and ontologies (Kless et al. 2012). A project was launched to refine an ontology in the domain of photography in Wikidata, starting from the terminology of the *Thesaurus Nuovo Soggettario* (ThNS). Both thesauri and ontologies are structured formal representations designed to organize information and data on a given subject, defining concepts expressed by terms or entities, their semantic relations, reference categories, and links. Both use a standardized, machine-

Silvia Bruni, University of Florence, Italy, silvia.bruni@unifi.it, 0000-0001-5461-8226
Alida Daniele, University of Florence, Italy, alida.daniele@unifi.it,
Fabrizio Nunnari, University of Florence, Italy, fabrizio.nunnari1@unifi.it,
Elena Cencetti, Ales S.p.A., Italy, elena.cencetti@unifi.it,
Elisabetta Viti, Central National Library of Florence, Italy, elisabetta.viti@cultura.gov.it, 0000-0002-7627-4062
Francesca Palareti, University of Florence, Italy, francesca.palareti@unifi.it

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Silvia Bruni, Alida Daniele, Fabrizio Nunnari, Elena Cencetti, Elisabetta Viti, Francesca Palareti, *Wikidata and the Thesaurus Nuovo Soggettario: Together to Build a Domain Ontology in the Field of Photography,* © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-1010-2.39, in Elena Marangoni, Camillo Carlo Pellizzari di San Girolamo (edited by), *Wikidata e la ricerca. Condividere esperienze. Atti del Convegno, Firenze, 5-6 giugno 2025 / Wikidata and Research. Sharing Experiences. Conference Proceedings, Florence, 5-6 giugno 2025*, pp. 323-329, 2026, published by Firenze University Press, ISBN 979-12-215-1010-2, DOI 10.36253/979-12-215-1010-2

readable language. While a thesaurus is created to index resources of various types (such as texts, images, websites, etc.) within catalogs and databases, ontologies aim to schematically represent knowledge domains for computational processing (ISO 2013; 2011; Studer et al. 1998, 184–88).

To make this description more concrete, let us take an example from the field of photography, which we have identified as the focus of the project: a photograph is a digital or physical object, taken with a photographic technique (albumen print, digital photograph, daguerreotype, etc.) by a given device, and may serve a particular function (advertising photograph, travel photograph, etc.).

Defining in a consistent and unambiguous way the terms representing these concepts, the categories to which they belong, and their relationships is fundamental for describing and querying data available in a database. In this case, the ThNS, compliant with ISO 25964, uses hierarchical, associative, and equivalence relationships.

For example, the ThNS models the concept of "advertising photography" (in Italian "*fotografia pubblicitaria*" as follows:

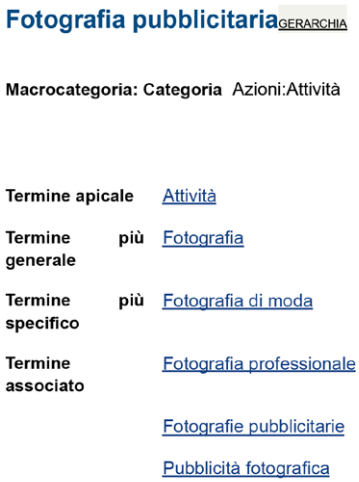

Figure 1 – Structure of the item, in Italian, "fotografia pubblicitaria" ("advertising photography") in the Nuovo soggettario, from the broader term down to the most specific term, along with the associated terms.

In an ontology these relationships are represented by functions and relations, logical rules that constrain meanings. Here is a very simplified representation:

The instances and relations of the ThNS are verified and monitored over time. For this reason, it is a valuable source for controlling and enriching instances, classes, attributes, and logical relations in Wikidata (Cencetti et al. 2025).

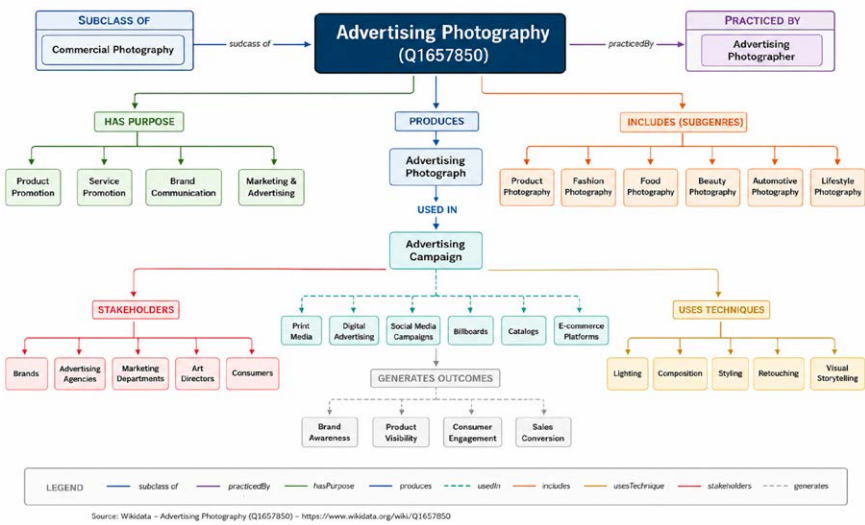

Figure 2 – Ontological model of Advertising Photography (Wikidata Q1657850). The diagram represents the conceptual structure of Advertising Photography as a subclass of Commercial Photography, highlighting its purposes, agents, outputs, application contexts, specialized subgenres, and technical components. Source: adapted from Wikidata Q1657850. Image created by the authors using ChatGPT 5.3 mini.

Figure 3 – Diagram of an ontology of "fotografia pubblicitaria" ("advertising photography"), showing the possible relationships of this item (subclass, context, objectives, equipment, connections).

1. The project

The project focused specifically on the semantic and technical dialogue between Wikidata and ThNS, with the following goals:

1) experimentation with the creation and refinement in Wikidata of a domain ontology on photography, in semantic and technical dialogue with the ThNS,
2) comparative study of methods and tools for importing data into Wikidata,
3) mutual enrichment of both databases with new terminology,
4) mutual enrichment of both databases with links to other KOSs,
5) compliance with FAIR principles (guide lines to make data findable, Accessible, Interoperable and Reusable).

The experimentation had a very practical character and was mainly based on the analysis and comparison, using an inductive method, of data structures, with particular attention to meanings and hierarchical relationships of a significant and representative set of terminology in the domain of photography (about 160 terms selected from the ThNS and and repositories in specific domain (Cencetti 2025; Corti and Gioffredi Superbi 2004).

We verified the correspondence in form and meaning of the terms in the ThNS and in Wikidata entities, in particular adding synonyms (UF/aliases), clarifying definition of concepts (through scope and definition notes; descriptions), reviewing the hierarchical structure; introducing new entities in Wikidata and new terms inTHNS, and finally increasing reciprocal links and links to other foreign national library thesauri (LCSH, RAMEAU, GND, EMBNE), to specialized thesauri (AAT), to archive and museum databases and to Wikipedia articles (Lanza 2022). One example is the ThNS term "*Procedimento all'albumina* (75362)", meaning albumen process, which corresponds in meaning, but not in form, to the Wikidata entity "*stampa all'albume* (Q107387614)", albumen print. In this case, we did not standardize the two entries because the Wikidata entity derives directly from the Wikipedia entry and, as a rule, the label should reflect the entry of the encyclopedia.

In other cases, we eliminated some aliases in Wikidata that seemed irrelevant or misleading. For example, "*ferrotipo*" (tintype) did not belong among the aliases of "*ferrotipia* (Q913381)," nor did "*macro obiettivo*" (macro lens) among those of "*macrofotografia (*Q272444)", macrophotography.

We also corrected some incorrect or redundant statements in Wikidata. For example, in the case of "*procedimento carbro (*Q1035568)", carbro process, which refers to a specific technique, we removed the subclass relationship with "*stampa fotografica* (56055236)", photographic print, because a hierarchical link between a process and an object could not exist.

Instead, in the case of "*ologramma (*Q3139490)", hologram, which was a subclass of both "*positivo* (Q1413646)", positive, and "*fotografia* (Q125191)", photograph, we removed the second link because it was redundant, since "positive" is already a subclass of "photograph."

Figure 3 – Poster submitted at "Wikidata and Research". Conference Proceedings, Florence, 5-6 giugno 2025.

To further enrich the network of connections, we leveraged the associative relations of the ThNS. For example, we linked activities to the person and profession practicing them: "*fotografia scientifica* (Q2586218)", scientific photography, and "*fotografo scientifico* (Q21550957)", scientific photographer, were connected through the statement "practiced by (P3095)" and "field of this occupation (P425)."

2. Conclusions

This analysis and comparison of the two tools has brought many benefits on several aspects. Firstly, it enabled the mutual enrichment of both KOSs with new terminology and links to other knowledge organisation systems (Hodge 2000, 1–22). Secondly, it also highlighted the need for analysis and revision of critical issues that emerged during the process, due to semantic misalignment, in order to optimize the interoperability of the two tools and to foster integrated access to information.

Intervening in the ontology of Wikidata not only means defining more precisely the concepts represented by individual items, but also embedding them in a broader relational system, structured according to principles of semantic coherence and internal logic (Pellizzari di San Girolamo 2024). In this way, individual elements do not remain isolated entities, but become nodes in a dynamic network of connections that enables interoperability between different domains, fostering the construction of an informational fabric capable of organically and effectively linking areas of knowledge that may be very distant from one another (Gruber 1993; Morin 2012, 41–42).

References

Bruni, Silvia. 2025. "Realizzare una Wikistazione e vivere più felici." *Bibelot: notizie dalle biblioteche toscane* 31 (1).

Cencetti, Elena. 2025. "Thesaurus Nuovo Soggettario e Wikidata: analisi di un corpus terminologico relativo all'ambito della fotografia." Tesi di master in Organizzazione e gestione degli archivi, catalogazione e metadatazione di risorse manoscritte, stampate e digitali.

Cencetti, Elena, Camillo Carlo Pellizzari di San Girolamo, and Elisabetta Viti. 2025. "Termini, dati e collegamenti: conversazioni tra il Thesaurus del Nuovo Soggettario e Wikidata." *Imagines* 12: 85–94.

Corti, Laura, and Fiorella Gioffredi Superbi. 2004. *Glossario di terminologia fotografica.* Roma: AIB. https://www.aib.it/aib/lis/lpi13eg.htm.

Gruber, Thomas R. 1993. "A Translation Approach to Portable Ontology Specifications." *Knowledge Acquisition* 5 (2): 199–220. https://doi.org/10.1006/knac.1993.1008.

Hodge, Gail. 2000. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.* Washington, DC: Digital Library Federation, Council on Library and Information Resources.

International Organization for Standardization. 2011. *ISO 25964-1:2011 – Information and Documentation: Thesauri and Interoperability with Other Vocabularies. Part 1: Thesauri for Information Retrieval.* Ginevra: ISO.

International Organization for Standardization. 2013. *ISO 25964-2:2013 – Information and Documentation: Thesauri and Interoperability with Other Vocabularies. Part 2: Interoperability with Other Vocabularies.* Ginevra: ISO.

Kless, Daniel, Ludger Jansen, Jutta Lindenthal, and Jens Wiebensohn. 2012. "A Method for Re-Engineering a Thesaurus into an Ontology." In *Frontiers in Artificial Intelligence and Applications,* edited by Giancarlo Guizzardi and Maureen Donnelly. Amsterdam: IOS Press.

Lanza, Claudia. 2022. *Semantic Control for the Cybersecurity Domain: Investigation on the Representativeness of a Domain-Specific Terminology Referring to Lexical Variation.* Boca Raton: CRC Press.

Morin, Edgar. 2012. *I sette saperi necessari all'educazione del futuro.* 14a rist. Milano: Cortina (Minima/Cortina; 59).

Pellizzari di San Girolamo, Camillo Carlo. 2024. "Conflations and Duplications in Wikidata Items: Causes, Detection, Solutions, and Issues." In *Proceedings of the Wikidata Workshop 2023 Co-Located with 22nd International Semantic Web Conference (ISWC 2023).* Athens: CEUR.

Studer, Rudi, V. Richard Benjamins, and Dieter Fensel. 1998. "Knowledge Engineering: Principles and Methods." *Data & Knowledge Engineering* 25 (1–2): 161–97. https://doi.org/10.1016/s0169-023x(97)00056-6.