ADVANCES IN KNOWLEDGE ORGANIZATION
Proceedings of International ISKO Conference - ISSN: 0938-5495 - 2026 - São Paulo

Taxonomy and Information Indexing for Institutional Content

Luciana Kramer Pereira

Universidade Federal do Rio Grande do Sul (UFRGS), Brazil

Thiago Bragato Barros

Universidade Federal do Rio Grande do Sul (UFRGS), Brazil

Abstract

This paper reports on the implementation of a domain-specific taxonomy of legal subjects within a collaborative digital hub for document management, content storage, and team collaboration at the Procuradoria Geral do Estado do Rio Grande do Sul (PGE-RS). By documenting this applied case, the paper contributes to the field of Knowledge Organization by illustrating how taxonomic principles can be operationalized in public-sector legal institutions where terminological control, operational efficiency, and reliable access to information are critical.

1.0 Introduction

The Procuradoria Geral do Estado of Rio Grande do Sul[1] (PGE-RS) is the legal body responsible for representing the state in judicial and extrajudicial matters. It provides legal advice to the state government, defends the state's interests in court, and issues legal opinions on public administration matters. The PGE-RS also supports the drafting and review of legislation, contracts, and internal regulations, playing a central role in the legal and administrative governance of the state. Because of the volume and complexity of legal documentation it produces and manages—such as case files, legal opinions, pleading templates, and internal normative acts—the PGE-RS faces significant challenges in information and knowledge management, especially in a context of distributed and remote work.

Therefore, this paper reports on the implementation of a taxonomy within a centralized hub for document management, content storage, and team collaboration in business and government organizations, drawing specifically on the experience of the Procuradoria Geral do Estado of Rio Grande do Sul (PGE-RS). In this case, an interdisciplinary team developed a domain-specific taxonomy of legal subjects and integrated it into a collaborative environment that supports the management of diverse information assets, such as templates for legal pleadings and internal regulations. While the literature on taxonomies and knowledge organization has extensively addressed libraries, digital repositories, and corporate environments, there are still relatively few documented cases in public-sector

[1] That would be equivalent to an Office of the Attorney General of the State of Rio Grande do Sul.

legal institutions. By presenting the PGE-RS experience, this paper provides an applied example of how taxonomic principles and controlled vocabularies can be operationalized within a government legal office to improve information indexing and retrieval and, consequently, enhance information and knowledge management.

2.0 Literature review

This section presents a brief literature review of key topics relevant to this paper, including knowledge organization, knowledge organization systems, taxonomies, and collaborative content hubs, as well as their relationship to the research topic.

2.1 Knowledge organization

According to Guimarães (2014), knowledge organization (KO) concerns knowledge that has been recorded in written form and has achieved a certain degree of consensus, which constitutes the object of KO. From this perspective, Hjørland (2016) argues that traditional approaches to KO research can be divided into human-based and machine-based approaches. However, they may also be classified into several other types, such as practicalist and intuitivist, consensus-based, facet-analytic, user-based, and domain-analytic or epistemological approaches. In addition, there are approaches developed outside KO, such as bibliometrics, information retrieval, and others.

Considering Hjørland's definitions presented above, the approach adopted in this article can be understood as a combination of consensus-based and user-based approaches. Consensus-based approaches view knowledge organization as a reflection of shared agreement within a scholarly community. By contrast, user-based approaches argue that knowledge organization systems should be informed by users or by empirical studies of user behavior and information needs. Their central idea is that the structure of a knowledge organization system should reflect how people search for, interpret, and use information. In this study, however, the users are also subject specialists; therefore, their behavior reflects the knowledge structure of the field. Additionally, Dahlberg affirms:

> [...] the most essential item in the theoretical background of knowledge organization is the fact that any organization of knowledge must be based on knowledge units – which are nothing else but concepts. Concepts consist of concept elements, also called concept characteristics and exactly these arc the factors by which concept systems – and classification systems are such concept systems – can be constructed. (1993, 211).

Hjørland adds (2003, 101) "The importance of concepts has been recognized in the KO community.". In other words, this is not in question, however, the author argues that the concepts of concept has been much better studied in other fields like philosophy, cognitive

and developmental psychology, artificial intelligence and linguistics. (Hjørland 2003). On the other hand, Barité and Petroccelli (2024) state:

> A concept is nothing more than an idea or notion defined with the greatest precision that has been possible to achieve. [...] A concept is, in a sense, a concentrated unit of knowledge.
>
> [...]
>
> Each concept is represented by one or more terms, which, in the vocabulary of knowledge organization, and more specifically of indexing, receive names such as headings, subject headings, descriptors, non-descriptors, categories, subjects, verbal representations, labels, subject authorities, facets, or keywords, depending on the context from which they come or in which they are used. (23-24).

Taken together, these perspectives suggest that knowledge organization is fundamentally anchored in concepts as the basic units of knowledge and in the terms used to represent them. If KO concerns recorded knowledge that has achieved a degree of consensus within a community, then its systems must not only reflect the conceptual structure shared by a domain, but also account for the ways users seek, interpret, and apply information. In this sense, concepts are not isolated entities, but relational knowledge units whose meaning depends on their characteristics and connections with other concepts.

2.2 Knowledge Organization Systems (KOS)

According to Hjørland (2016, 475) “Knowledge organization systems (KOS) are selection of concepts with indication of selected semantic relations, for example, classification systems, lists of subject headings, thesauri, ontologies and other systems of metadata.” Ortega (2023) points out that the term was proposed by Hodge (2000). According to Hodge, KOS encompass everything from simple instruments, such as dictionaries and glossaries, to more complex systems, such as thesauri and ontologies. The author highlights that, in the referenced text, Hodge includes everything from dictionaries and glossaries to thesauri and ontologies, with the latter being of greater complexity. The author further adds:

> If Hodge does not emphasize the function of documentary instruments, distinct from that exercised by terminological instruments, the appropriation of his text in Brazil also did not prioritize this distinction, adopting the term KOS in a generic manner. This dispersive movement has fostered the erasure of studies on the foundations of documentary languages and their construction and update methodologies. (21).

Thus, Schiessl and Shintaku (2012) define KOS as: “[...] representations of knowledge, which in turn are models of abstraction of the real world, constructed for a specific purpose.” (50). The authors exemplify the types of KOS, highlighting: Classification systems; Taxonomies; Semantic networks; Thesauri; Ontologies.

Taxonomies can be described as classification schemes; however, Lambe (2014) emphasizes that they are not only that. For the author, taxonomies are primarily semantic: “That is, they provide a fixed vocabulary to describe their knowledge and information assets, and this vocabulary needs to be meaningful and transparent to ordinary users.” (6).

Lambe also highlights that a taxonomy can provide a controlled vocabulary and that, when it evolves to describe synonyms, related-term relationships, and hierarchical positions of terms within the domain, we can say we have a thesaurus. A taxonomy, however, can be a simpler form of controlled vocabulary, in which we can see the hierarchy between terms, but not necessarily other types of relationships.

While taxonomies also rely on hierarchical relationships, they differ from thesauri because they do not typically include equivalence or associative relationships. Nevertheless, our experience shows that equivalence relationships can be used; thus, it is important to review the definitions of thesauri. According to Chaumier (1988, 70–71), a thesaurus is '[...] a controlled language, constituted of descriptors (words or expressions) capable of being combined with one another at the time of indexing, to express complex notions'.

Complementarily, Sales and Café (2009, online) state that thesauri are "[...] controlled vocabularies formed by descriptor terms that are semantically related, and they act as instruments of terminological control." Authors such as Pickler (2007) and Duarte (2007) emphasize that this terminological control promotes a uniformization of language, avoiding polysemy and ambiguities. Fujita (2011, 36) adds: "[...] the use of a documentary language, by the indexer and by the user of the information retrieval system, is pragmatically intended for the selection of the correct descriptor term, in order to reduce the diversity and ambiguity of vocabulary”.

Its essential structure can be defined as a collection of concepts represented by terms and relationships between these terms, which are hierarchical, of equivalence, and associative. (Dextre Clarke 2019). The same can happen to taxonomies, using the definitions provided by Chaumier (1988):

> a) Equivalence relationship: allows referring descriptor terms to terms considered synonyms or quasi-synonyms. In a "thesaurus," one thus distinguishes between "descriptor terms" and "equivalent terms." Thus, in a "thesaurus," a term will be referred to its synonym or quasi-synonym elected as the descriptor. (71).
>
> [...]
>
> b) Hierarchical relationship: allows placing a term in relation to another, from the general to the particular. One thus distinguishes between the generic relationship and the specific relationship. (72).

Equivalence relationships, according to Dodebei (2002), establish that different forms of linguistic expression—whether terms, expressions, or symbols—can be considered, in each context, as carriers of the same content of meaning. Such relationships allow

organizing and systematizing the variation inherent in natural language into distinct dispersions:

a) lexical, when dealing with synonyms and quasi-synonyms;

b) symbolic, when linking distinct signs to the same conceptual referent, as occurs in abbreviations, different spellings, and translations;

c) syntactic, when recognizing that different structures can perform the same function or meaning, for example, in the use of gender or number.

Thus, descriptors can receive equivalent terms, which will aid in information retrieval. As Laipelt and Krebs explain:

> [...] the remaining terms (variants) are inserted into thesauri as non-descriptors, which cannot be used to represent the subjects of documents in bibliographic records. Although they cannot be used in bibliographic records, non-descriptors play an important role in indicating to the user that the concept they are seeking is represented by another term, the descriptor. This occurs because thesauri are metalanguages, whose function is to control the terms that will be used to represent the subject of documents. (Laipelt and Krebs 2021, online).

Regarding hierarchical relationships, Fujita (2011) teaches that, since they express subordination between terms, there will be reciprocity between the more generic and more specific terms. The author further reflects that the relationships can be:

a) Generic: existing between a class or category and its members;

b) Whole/part: expressing a relationship where one term is contained within another, for example: organs of the body;

c) Poly-hierarchical: a term can be subordinated to more than one concept simultaneously.

Thus, the perspective of developing a taxonomy will require that the domain be mapped and exhaustively analyzed. Laipelt and Krebs (2021, online) further highlight that "Knowing the language used by users during the information retrieval process is fundamental both for the practice of information representation (indexing), authority catalog management, and for the planning and creation of Knowledge Organization Systems (KOS)."

For the work of developing and maintaining the taxonomy structure, it is quite important and more practical for the documentalist to insert new terms and establish semantic relationships between descriptors based on a previously defined structure of categories. (Fujita 2011).

By constituting controlled vocabularies through equivalence, hierarchical, these instruments contribute to reducing ambiguities, uniformizing indexing language, and aligning user language with system language. From a methodological viewpoint, the construction of a taxonomy requires both mastery of conceptual approaches and understanding of the language effectively used by users in real contexts of information

production and use, in the case of PGE-RS we already established that the users are also subject-matter experts.

2.3 Collaborative digital hubs

According to Leggett (2012), "a collaborative content hub is more than traditional knowledge management. It is a synthesis of knowledge management, discussion forums, collaborative authoring, search, gamification, reporting, and text analytics that no vendor is tying together right now." Although its purpose extends beyond the mere management of organizational knowledge, it is clear that this concept is important to a collaborative content hub. Another definition says:

> A content hub is a centralized digital space where an organization stores, organizes, and shares all its content. Its main purpose is to arrange the content so it's easy to find and use, helping the whole team to access and interact with the information they need. Also, with permission and roles within the content hub, organizations can get control over the content. (Stojović 2025).

The definition above is appropriate because it highlights several essential aspects of a content hub: content is maintained in a centralized digital environment; it is readily accessible to team members; and it supports interaction with information, thereby fostering collaboration.

Additionally, Tukkanen (2025) highlights the difference between a hub and a traditional office: a traditional office is generally organized around fixed desks, hierarchical seating arrangements, and mostly synchronous work schedules, whereas a collaboration hub is intentionally flexible, using digital platforms to maintain continuity across locations, and cultural rituals that support both synchronous and asynchronous work.

Therefore, collaborative digital hub is essentially a content hub that combines digital technology with active collaboration, where teams exchange knowledge, co-create digital solutions, and access resources for innovation. It can facilitate constant, dynamic, and transdisciplinary communication, promoting the creation, sharing, and dissemination of digital knowledge among the team.

3.0 Methodological approach

This paper adopts a practice-oriented case-study approach, drawing on the authors' direct involvement in the design and implementation of the collaborative information environment at PGE-RS. The case description draws on internal project documentation, configuration records, and observations of user interactions with the system from 2023 to 2025.

4.0 The PGE-RS experience

PGE-RS handles a large volume of legal information assets, including case files, templates for legal pleadings, internal regulations, and the procedures that guide the workflows of each team. Prior to the initiative described in this paper, these materials were often dispersed across different systems and individual storage practices, hindering efficient retrieval and reuse. In addition, many employees now work remotely, from different towns and via online access. This made it urgent to establish a collaborative work environment capable of supporting teams working from home or from the office, while, above all, preserving and organizing all this information and knowledge.

The solution was to look for collaborative tools that enable the organization to create sites for storing and sharing files and a wide range of institutional content. Librarians and information professionals usually think about information processing and the use of metadata in the context of libraries, digital libraries, and repositories; however, information management and retrieval in organizational settings can also benefit significantly from these practices. At PGE-RS, our experience has shown that well-structured metadata can make a substantial difference in day-to-day work processes.

Our goal is to improve information and knowledge management for our peers at PGE-RS. To this end, we propose and have implemented a collaborative workspace hub in which users can upload documents and other informational assets, each of which is associated with a set of previously planned and implemented metadata appropriate to its type. The planning and implementation of this structure are carried out by the Office of Data and Information Management; however, any user can upload and manage their own documents and information within this environment. The collaborative environment is designed to store and manage different types of information assets, including templates for legal pleadings, internal regulations, and other institutional documents.

The main contribution of this process is the implementation of a taxonomy of legal subjects, incorporating both hierarchical and equivalent relationships, which has been developed by an interdisciplinary team consisting of a librarian, two lawyers and a designer. In this way, documents can be easily accessed and edited by users, while simultaneously enabling controlled indexing, without requiring the more complex customizations often needed in software such as DSpace. At the moment, the taxonomy includes more than 700 terms that can be used to index any document or asset in the environment by any user. However, only the librarian is authorized to maintain and expand the taxonomy, and each new term is added only after careful research into its meaning and actual use in the PGE-RS context.

The structuring of this environment includes configuring metadata with a wide range of customization options. One such option is managed metadata, in which the values entered in each field are controlled by the system's term store. The term store supports the inclusion of hierarchical terms, the definition of synonyms, and the specification of advanced properties that can store additional information, such as associative

relationships. In other words, it enables the design of a domain-specific taxonomy that aligns with the organization's informational needs and practices.

According to Khaga (2025), the adoption of controlled vocabularies in the metadata fields of a corporate information management platform ensures consistency across the organization, while still allowing for the natural evolution of terminology over time. This approach is particularly valuable in specialized areas of practice, where terminological precision directly impacts the retrievability of documents.

As such, we argue that features such as managed metadata, the term store, and Document Libraries can be valuable allies for knowledge organization on organizations' internal sites, especially when managed by librarians, specialists, and other information professionals.

5.0 Final remark

The experience reported in this paper demonstrates that taxonomies can play a decisive role in strengthening information and knowledge management in public-sector legal institutions. In the case of PGE-RS, the implementation of a domain-specific taxonomy within a collaborative digital environment improved the organization, indexing, retrieval, and reuse of legal and administrative documents, while also supporting more consistent practices in a context marked by documentary complexity and distributed work. The case further highlights the strategic contribution of librarians and information professionals in the design and governance of metadata structures suited to institutional needs. By documenting this applied experience, the paper seeks to contribute to the field of Knowledge Organization by showing how taxonomic principles may be translated into concrete solutions for government organizations, particularly in environments where terminological control, operational efficiency, and reliable access to information are critical. Future developments may include the continuous refinement of the taxonomy, the analysis of user interaction with the system, and the assessment of its impact on organizational workflows and decision-making processes.

AI was used for spelling and grammar proofreading.

References

Barité, Mario and Patricia D. Petroccelli. 2024. *Manual de introducción a la organización del conocimiento*. Montevideo: Universidad de la República, Comisión Sectorial de Enseñanza.

Bräscher, M., and L. Café. "Organização da Informação ou Organização do Conhecimento?" In *Anais do IX Encontro Nacional de Pesquisa em Ciência da Informação* (ENANCIB). São Paulo: ANCIB, 2008.

Chaumier, Jacques. 1988. “Indexação: Conceito, Etapas e Instrumentos.” *Revista Brasileira de Biblioteconomia e Documentação* 21, nos. 1–2: 63–79.

Dahlberg, Ingetraut. 1993. “Knowledge Organization: Its Scope and Possibilities.” *Knowledge Organization* 20, no. 4: 211–222. https://doi.org/10.5771/0943-7444-1993-4-211.

Dextre Clarke, Stella G. 2019. “The Information Retrieval Thesaurus.” *Knowledge Organization* 46, no. 6: 439–459. https://doi.org/10.5771/0943-7444-2019-6-439.

Dodebei, Vera Lúcia Doyle. 2002. *Tesauro: Linguagem de Representação da Memória Documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência.

Fujita, Mariângela Spotti Lopes. 2011. “Modelos de Categorização para a Construção de Tesauros: Metodologia de Ensino.” In *Estudos de linguagem em ciência da informação*, edited by Vera Regina Casari Boccato and Luciana de Souza Gracioso. Campinas: Alínea.

Hjørland, Birger. 2003. “Fundamentals of Knowledge Organization.” *Knowledge Organization* 30, no. 2: 87–111. https://doi.org/10.5771/0943-7444-2003-2-87.

Hjørland, Birger. 2016. “Knowledge Organization.” *Knowledge Organization* 43, no. 6: 475–484. https://doi.org/10.5771/0943-7444-2016-6-475.

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

Khaga, Sivaprasad Yerneni. 2025. “The Strategic Implementation of SharePoint in Modern Legal Practices: Enhancing Document Management, Security, and AI Integration for Law Firms.” *European Journal of Computer Science and Information Technology* 13, no. 25: 127–139. https://doi.org/10.37745/ejcsit.2013/vol13n25127139.

Laipelt, Rita do Carmo Ferreira and Luciana Monteiro Krebs. 2021. *Termos sob a Superfície: Elementos Teóricos, Metodológicos e Terminológicos para a Representação do Conhecimento*. 1st ed. Rio de Janeiro: Interciência.

Lambe, Patrick. 2014. *Organizing Knowledge: Taxonomies, Knowledge and Organizational Effectiveness*. 2nd ed. Oxford: Chandos Publishing.

Leggett, Kate. 2012. “What Is a Collaborative Content Hub for Customer Service?” *Forrester*. https://www.forrester.com/blogs/12-05-29-what_is_a_collaborative_content_hub_for_customer_service/.

Ortega, Cristina Dotta. 2023. “Categorias fundamentais da Organização da Informação: do conceitual, metodológico e pragmático ao normativo e tecnológico.” In *Organização e Representação do Conhecimento em diferentes contextos: desafios e perspectivas na era da datificação*, edited by Natália Bolfarini Tognoli, Ana Cristina de Albuquerque, and Brígida Maria Nogueira Cervantes. Londrina: ISKO Brasil.

Pickler, Maria Elisa Valentim. 2007. “Web Semântica: Ontologias como Ferramentas de Representação do Conhecimento.” *Perspectivas em Ciência da Informação* 12, no. 1: 65–83. https://doi.org/10.1590/S1413-99362007000100006.

Sales, Rodrigo de and Lígia Café. 2009. “Diferenças entre Tesauros e Ontologias.” *Perspectivas em Ciência da Informação* 14, no. 1: 99–116. https://doi.org/10.1590/S1413-99362009000100008.

Schiessl, Marcelo and Milton Shintaku. 2012. “Sistemas de Organização do Conhecimento.” In *Organização da informação e do conhecimento: conceitos, subsídios e aplicações*, edited by A. M. M. Cintra and M. F. G. M. Tálamo. São Paulo: Fundamentos Editoriais.

Stojović, Arso. 2025. “Content Hub Guide: Structuring Content for Better Team Collaboration.” *BCMS*. https://thebcms.com/blog/content-hub-guide.

Tukkanen, Julia. 2025. “What Is a Collaboration Hub?” *Monteco Coworking*. https://montecoworking.me/coworking-fundamentals/what-is-a-collaboration-hub/.