TAFNAVEGA as an Interoperable Semantic Vocabulary: From Faceted Taxonomy to Ontological Modeling

Benildes Coura Moreira dos Santos Maculan[1,*,†] and Elisângela Cristina Aganette[2,†]

[1, 2] *Universidade Federal de Minas Gerais, Av. Presid. Antônio Carlos, 6627, Pampulha, BH, MG, CEP 31270-901, Brasil*

Abstract

The navigational faceted taxonomy TAFNAVEGA was originally proposed as a categorical structure for organizing theses and dissertations. With the advancement of Open Science and the Semantic Web, it has become necessary to revisit this proposal to align it with contemporary standards of formal representation and semantic interoperability. This study, which is qualitative, exploratory, and conceptual in nature, presents the reformulation of TAFNAVEGA based on Faceted Classification Theory, semantic standards such as SKOS (Simple Knowledge Organization System) and OWL (Web Ontology Language), and in dialogue with the ontological model CERIF (Common European Research Information Format). The methodology comprised a diagnosis of the original structure (CAFTE), the mapping of semantic relations and logical dependencies, and the alignment with formal representation standards. The results, still at a conceptual stage, indicate the feasibility of structuring CAFTE as an interoperable controlled vocabulary, while highlighting the need for conceptual refinements, computational formalization, and empirical validation in thesis and dissertation repositories. The proposal underscores TAFNAVEGA's contribution to academic knowledge organization and its potential integration into knowledge graphs aimed at Open Science and Artificial Intelligence.

Keywords

Faceted taxonomy; Knowledge organization; Semantic interoperability; Academic ontologies.

1. Introduction

The growing adoption of digital libraries by academic and research institutions has intensified the demand for classificatory structures that enable more precise and semantically oriented information retrieval, consistent with the logic of scientific production. This challenge is particularly evident in repositories of theses and dissertations, given the diversity of topics, methods, and objects of investigation.

In this context, the navigational faceted taxonomy (TAFNAVEGA), proposed by Maculan [1], emerged as a pioneering initiative of categorical organization, structured around ten facets. Its formulation was based on Bardin's thematic categorical analysis [2] and inspired by Ranganathan's Faceted Classification Theory (FCT) [3], providing a strategy for navigable representation of academic content.

More than a decade later, the principles of Open Science, the consolidation of the FAIR standards —Findable, Accessible, Interoperable, and Reusable [4]—and the growing adoption of semantic standards such as SKOS (Simple Knowledge Organization System) [5] and OWL (Web Ontology Language) [6] highlight the need to revisit TAFNAVEGA. In this scenario, models oriented to the formal representation of scientific activity gain relevance, among which CERIF (Common European Research Information Format) [7] stands out, widely used in European Current Research Information Systems (CRIS).

Based on this framework, the general objective of this article is to propose the reformulation of TAFNAVEGA as an interoperable controlled vocabulary, grounded in FCT and formal semantic

[1] *Proceedings of the 18th Seminar on Ontology Research in Brazil (ONTOBRAS 2025) and 9th Doctoral and Masters Consortium on Ontologies (WTDO 2025), São José dos Campos (SP), Brazil, September 29 – October 02, 2025.*

* Corresponding author.

† These authors contributed equally.

benildes@gmail.com (B.C.M.S. Maculan); elisangelaaganette@gmail.com (E. C. Aganette).

0000-0003-4303-9071 (B.C.M.S. Maculan); 0000-0003-4357-8016 (E. C. Aganette).

standards. Specifically, the study seeks to: (i) diagnose the limitations of the original structure; (ii) map its facets in dialogue with the CERIF ontological model; (iii) propose guidelines for formalization in SKOS and OWL; and (iv) outline perspectives for application in academic repositories. The research question guiding the study is: How can TAFNAVEGA be updated to operate as a semantically interoperable component in Open Science ecosystems?

2. Theoretical and methodological foundations

The construction of interoperable classificatory systems in digital environments requires a theoretical-methodological basis that combines consolidated principles of Knowledge Organization (KO) with approaches to formal representation. The Faceted Classification Theory (FCT), originating in Librarianship and consolidated in Information Science, remains a relevant analytical-synthetic model in the face of challenges posed by the Semantic Web and Artificial Intelligence. Its ability to decompose and recombine concepts into multidimensional facets makes it a reference for the development of faceted taxonomies and lightweight ontologies, applicable in interdisciplinary contexts.

This section presents four key axes that underpin the reformulation of TAFNAVEGA: (i) a critical review of FCT (2.1); (ii) description of the categorical matrix CAFTE, derived from empirical analysis of theses and dissertations (2.2); (iii) presentation of the original TAFNAVEGA proposal and its potential for reformulation (2.3); and (iv) discussion of SKOS and OWL standards as technical references for its formalization (2.4).

By articulating these foundations, the study emphasizes the dialogue between KO traditions and Semantic Web models, providing the conceptual basis for the reformulation of TAFNAVEGA and its prospective application in Artificial Intelligence and knowledge representation systems.

2.1. Classification Theory and its application in Information Science

Proposed by Ranganathan in the 1930s, Faceted Classification Theory (FCT) is one of the most influential foundations of Knowledge Organization (KO). By decomposing subjects into five fundamental categories — Personality, Matter, Energy, Space, and Time (PMEST) — it broke with traditional enumerative models, offering a flexible and synthetic structure. Despite its conceptual robustness, FCT lacks formal computational mechanisms, which highlights the importance of integrating it with Semantic Web standards such as SKOS and OWL.

Recent studies confirm its relevance. Silva and Miranda [8] emphasize its precision and adaptability for representing interdisciplinary domains, while Coelho, Lima and Borges [9] point to the effectiveness of faceted taxonomies for semantic retrieval in digital environments, particularly when formalized in interoperable standards. Almeida and Teixeira [10] argue that, in educational contexts, faceted classification also functions as an epistemic instrument, enhancing usability and reusability of digital objects. Lima [11] highlights its compatibility with the epistemological diversity of scientific knowledge, reinforcing its pertinence for thesauri and interoperable taxonomies.

Some proposals have sought to update the theory, such as Martins, Moreira and Santarém Segundo [12], who suggest alternative fundamental categories (thing, cultural object, event, person, place, time, attribute, action, cause, and purpose), thus incorporating dimensions such as causality and intentionality. This effort resonates with the reformulation of TAFNAVEGA, reinforcing the potential of aligning faceted structures with ontological models and semantic standards, thereby expanding their interoperability and applicability in scientific information systems.

2.2. The categorical matrix: construction process

The categorical matrix, named CAFTE (Faceted Analytical Classification of Theses and Dissertations), was developed by Maculan [1] to represent, with granularity and flexibility, the research objects found in theses and dissertations. Its construction was theoretically based on FCT, especially Ranganathan's PMEST model [3], and methodologically on qualitative categorical content analysis [2].

The corpus consisted of titles, abstracts, and keywords from 41 documents in the UFMG Institutional Repository, within the Information Organization research line. The process involved: (i) corpus collection and selection; (ii) coding into emerging thematic categories; (iii) grouping by semantic proximity and discursive function; (iv) definition of facets and subfacets; and (v) experimental validation [1]. This ensured that categories were empirically derived from actual scientific discourse, aligning representation with the internal logic of the texts.

The outcome was the definition of ten functional categories: Theme, Empirical object, Scope, Context, Research type, Data collection, Methods, Theoretical framework, Historical/contextual framework, and Results [1]. Although not identical to PMEST, correspondences with Ranganathan's proposal were established, reaffirming CAFTE's affinity with the faceted tradition of KO.

Originally conceived to support indexing and retrieval in academic repositories, CAFTE also demonstrates potential for more complex computational environments. Its faceted structure can be formalized in SKOS or OWL, becoming an ontologically grounded controlled vocabulary capable of semantic interoperability. As such, it serves as a bridge between empirical conceptual extraction and lightweight ontological modeling, supporting applications in knowledge graphs, recommender systems, and generative AI.

2.3. The original TAFNAVEGA proposal: structure, categories, and implementation

TAFNAVEGA was conceived from the CAFTE matrix [1] as a mechanism for representing and accessing theses and dissertations, aiming to enable faceted navigational exploration rather than simple keyword search, often constrained by ambiguity and low precision.

Its initial implementation took place through a prototype developed in Microsoft Access [16], designed according to the principles of multidimensional classification and faceted navigation. Each CAFTE facet generated a set of terms organized into independent tables, allowing progressive combinations. The interface enabled users to refine queries by selecting multiple categories, such as identifying qualitative research using interviews in the health field, grounded in Activity Theory.

Although it did not incorporate formal representation languages such as RDF, SKOS, or OWL, TAFNAVEGA demonstrated the feasibility of faceted navigation in academic repositories. However, the absence of notations, URIs, and explicit ontological relations limited its interoperability. The reformulation proposed in this study seeks to advance its formalization in SKOS and OWL, expanding its applicability to institutional repositories and broader Open Science ecosystems.

2.4. SKOS and OWL semantic patterns

SKOS and OWL, developed by the W3C, are key standards for representing structured knowledge and enabling semantic interoperability on the Web. In the field of Knowledge Organization, they have been widely adopted for modeling controlled vocabularies, classification schemes, and ontologies.

SKOS provides a lightweight framework for representing concepts and basic semantic relations, such as hierarchy (skos:broader, skos:narrower) and association (skos:related). Its simplicity and machine readability make it especially suitable for expressing taxonomies, thesauri, and faceted schemes in an interoperable way [5].

OWL, in contrast, is a more expressive language grounded in description logic, capable of defining classes, properties, equivalence or disjointness relations, and formal restrictions [6]. These features make it particularly useful in contexts that require automatic inference, consistency checking, and integration with intelligent agents.

In this study, SKOS is considered appropriate for structuring TAFNAVEGA's categories and basic relations, while OWL is proposed as an additional layer to formalize logical dependencies among facets. For instance, the facet C5: Research Type can be represented in SKOS as a skos:Concept, with terms linked through hierarchical relations, while constraints between complementary facets can be expressed in OWL through axioms.

The combined use of SKOS and OWL thus offers a balanced approach between simplicity and expressiveness, enabling the reformulation of TAFNAVEGA as an interoperable controlled vocabulary aligned with the Semantic Web.

3. Methodology for the Reformulation of TAFNAVEGA

The reformulation of TAFNAVEGA followed a qualitative, exploratory, and descriptive approach, based on conceptual modeling and manual semantic analysis. No automated tools were employed; instead, the work relied on the researchers' experience with faceted taxonomies, theoretical frameworks in Knowledge Organization and the Semantic Web, and a critical review of relevant literature. Conceptual decisions were documented in analytical matrices and iteratively revised to ensure coherence and minimize bias.

The proposal builds on the original TAFNAVEGA, developed by Maculan [1] through empirical analysis of theses and dissertations, and focuses on its conceptual and technical update. Rather than collecting new data, we conducted semantic comparisons of categorical structures. Facets were examined against PMEST and CERIF models to identify terminological alignments, logical dependencies, and opportunities for representation in SKOS and OWL. The methodological process comprised four stages:

> **Stage 1** – Diagnosis of the existing structure: critical analysis of CAFTE and the first version of TAFNAVEGA, assessing semantic clarity, coverage, and relevance for representing academic objects. Comparative analyses were made with PMEST and CERIF [7]. CERIF was selected as a reference due to its recognition as an international standard promoted by euroCRIS and its RDF/OWL compatibility [6], though it was not fully adopted to avoid scope limitations.
>
> **Stage 2** – Identification of semantic relations: hierarchical, associative, and equivalence relations among categories were mapped, together with logical dependencies across facets, to support coherent navigation and retrieval.
>
> **Stage 3** – Mapping to formal standards: SKOS schemes were designed to model facets as ConceptSchemes and terms as Concepts, with semantic properties explicitly defined. OWL was also considered for representing more complex dependencies relevant to automated inference.
>
> **Stage 4** – Alignment with external ontologies: equivalences between CAFTE categories and CERIF entities [7] were established to strengthen interoperability in Open Science ecosystems. This alignment supports applications such as content recommendation, semantic enrichment of metadata, and generative AI in academic repositories. ChatGPT 4.0 [13] was used as a heuristic aid for semantic exploration, with results validated by the authors.

As a conceptual exercise, this proposal does not include empirical validation with users. Future work will address this gap through pilot studies in institutional repositories and retrieval experiments. Nonetheless, the methodology provides a consistent conceptual basis—anchored in Information Science and Semantic Web frameworks—for reformulating TAFNAVEGA as an interoperable controlled vocabulary for Open Science environments.

4. Results

The reformulation of TAFNAVEGA focuses on creating a more refined semantic artifact, oriented towards interoperability and advanced computational use. Results are organized according to the four methodological steps.

4.1. Diagnosis and validation of the existing categorical structure

The critical analysis of CAFTE, the matrix that originated TAFNAVEGA, was conducted through a comparison between its ten categories, the classical facets of Ranganathan's Faceted Classification Theory (PMEST) [3], and the components of the CERIF model (Common European Research Information Format) [7]. While Maculan [1, p. 142, table 13] had already established parity between CAFTE and PMEST, this study adds the novel alignment with CERIF.

Correspondences between CAFTE and CERIF were established by conceptual inference, using three criteria: functional equivalence (same structural role with different terms), semantic equivalence (similar meaning with different functions), and mixed equivalence (function + meaning). Table 1 summarizes these mappings.

Table 1
Comparative alignment CAFTE (TAFNAVEGA) × PMEST × CERIF

CAFTE (TAFNAVEGA)	Ranganathan (PMEST)	CERIF (Common European Research Information Format)
1. Theme	Personality	cfResPubl_Class (thematic classification)
2. Object	Personality	cfResPubl / cfResProd / cfResPat (research product/object)
3. Results	Personality	cfResultProduct (scientific output)
4. Scope	Matter	cfResPubl / cfProj (Theme, scope, domain)
5. Historical/contextual foundation	Matter	cfProj / cfEvent (historical context, project, event)
6. Research type	Matter	cfProj / cfResPublClassification
7. Data collection	Energy	cfResPubl / cfMethod (applied methods)
8. Theoretical foundation	Energy	cfTheory / cfResPubl (conceptual base)
9. Methods	Energy	cfMethod / cfTechnique (applied technique)
10. Setting	Espace	cfOrgUnit / cfFacility / cfPlace (institutional or spatial context)

Source: Authors (2025).

Table 2 shows that hierarchical relationships were represented by generalization/specialization structures in SKOS, while associative connections and functional dependencies required greater expressiveness in OWL. This conceptual modeling was defined manually, by inference of the authors, which constitutes a theoretical exercise not yet validated on empirical bases.

In line with CERIF, it was observed that although the model is not classificatory, its relational architecture allows for the incorporation of extensions and external vocabularies. In this sense, the semantic links identified in TAFNAVEGA can be associated with entities such as cfResPubl, cfProj, cfResultProduct, and cfMethod, expanding the potential for semantic interoperability with CRIS systems and knowledge graphs. It should be noted, however, that this alignment is exploratory: it serves as a basis for integration, but does not replace the need for formalization in SKOS/OWL.

The explicitation of semantic relationships strengthens the conceptual consistency of TAFNAVEGA and guides its future computational implementation. In addition to technical robustness, it is recommended that the next steps include interface prototypes in academic repositories in order to empirically test the effectiveness of faceted navigation, as well as the added value to the semantic retrieval of scientific information. Table 2 shows that hierarchical relationships were represented by generalization/specialization structures in SKOS, while associative connections and functional dependencies required greater expressiveness in OWL. This conceptual modeling was defined manually, by inference of the authors, which constitutes a theoretical exercise not yet validated on empirical bases.

In alignment with CERIF, it was observed that although the model is not classificatory, its relational architecture allows for the incorporation of extensions and external vocabularies. In this sense, the semantic links identified in TAFNAVEGA can be associated with entities such as cfResPubl, cfProj, cfResultProduct, and cfMethod, expanding the potential for semantic interoperability with CRIS systems and knowledge graphs. It should be noted, however, that this alignment is exploratory: it serves as a basis for integration, but does not replace the need for formalization in SKOS/OWL.

The explicitation of semantic relationships strengthens the conceptual consistency of TAFNAVEGA and guides its future computational implementation. In addition to technical robustness, it is recommended that the next steps include interface prototypes in academic repositories in order to empirically test the effectiveness of faceted navigation, as well as the added value to the semantic retrieval of scientific information.

4.1.1. Conceptual Convergences

CAFTE categories show strong alignment with TCF principles, particularly the PMEST structure. Facets such as Theme, Object, and Results correspond to the Personality category, central to subject determination in PMEST. Similarly, Methods, Theoretical foundation, and Data collection align with the Energy facet, representing processes applied to research objects.

In the CERIF model, widely used for standardized descriptions of scientific activity, relevant correspondences also emerge:

- Theme ↔ cfResPubl_Class (thematic classification of scientific output);
- Object ↔ cfResPubl / cfResProd / cfResPat (research products or objects);
- Results ↔ cfResultProduct (formal representation of research outcomes);
- Methods and Data collection ↔ cfMethod and cfTechnique;
- Context ↔ cfOrgUnit, cfFacility, or cfPlace.

These mappings support the feasibility of TAFNAVEGA as an interoperable vocabulary, capable of dialoguing with established standards in scientific information. Full adoption of a European model such as CERIF, however, may require contextual adjustments for specific domains, particularly in Brazil, given institutional, cultural, and linguistic differences. Thus, the mapping should be seen as an exploratory reference, open to adaptation and extension.

4.1.2. Gaps and Misalignments

The comparative analysis also revealed some conceptual misalignments. The Results facet, for instance, originally tied to Personality, is better understood as a dimension of Energy, as it denotes the outcome of a methodological process. In CERIF, although cfResultProduct reflects this aspect, it does not capture the causal relationship inherent in the PMEST model.

Other limitations concern more abstract elements, such as Theoretical Foundation and Historical Context, which in CERIF appear only as descriptive fields (e.g., cfResPubl, cfProj), without dedicated entities or explicit hierarchies. Likewise, the Setting facet, central to CAFTE, is fragmented across multiple entities (cfOrgUnit, cfPlace, cfFacility), hindering an integrated and cohesive representation.

These limits show that CAFTE–CERIF compatibility, while strong, is not complete. To bridge these gaps, we recommend: (a) pilot tests in institutional repositories, applying the mapping to real dissertation records to validate accuracy and resolve ambiguities; and (b) complementary ontological extensions in SKOS and OWL, especially for theoretical foundations and historical contexts, which require richer semantic expressivity.

Overall, the analysis confirms the potential for convergence but also highlights the need to complement CERIF with richer formal ontologies to ensure both interoperability and the conceptual expressiveness of TAFNAVEGA.

4.2. Semantic Relations and Logical Dependencies

In the second stage, we developed a model of semantic relationships between the categories of the reformulated TAFNAVEGA, using the standard properties of controlled vocabularies as a reference (broader, narrower, related, and equivalent). The process was documented in categorical comparison spreadsheets, which contained descriptions, examples, and critical comments to ensure transparency and auditability for future external validations. The analysis resulted in the systematization of hierarchical, associative, and functional relationships, which are presented in Table 2 along with implications for their formalization in SKOS and OWL.

Table 2
Semantic relations and logical dependencies of CAFTE categories

Category (CAFTE)	Type of relation	Example of Dependency or Association	Implication for SKOS / OWL
Theme	Hierarchical	‘Education’ > ‘Inclusive education’	skos:broader/narrower
Object	Associative	‘‘Teachers’ related to ‘Teacher training’	skos:related; can be owl:ObjectProperty in complex graphs
Results	Functional dependency with Methods	‘Predictive model’ generated by ‘Statistical analysis’	owl:hasOutput; owl:Restriction
Scope	Hierarchical / Associative	‘School libraries’ narrower than ‘Libraries’	skos:broader; skos:related for adjacent scopes
Historical/contextual foundation	Associative	‘Student movement’ related to ‘Educational reforms’	skos:related; can be integrated with dc:coverage (temporal)
Type of research	Hierarchical	‘Case study’ narrower than ‘Qualitative research’	skos:broader/narrower
Data collection	Dependency with Research type	‘Interviews’ requires ‘field research’	owl:Restriction; conditional logic with owl:someValuesFrom
Theoretical foundation	Associative	‘Freire’ related to ‘Critical pedagogy’	skos:related; owl:equivalentClass between theoretical schools
Methods	Hierarchical / dependency	‘Content analysis’ narrower than ‘Qualitative analysis’	skos:broader; relationship with data type or support via OWL
Setting	Dependency with Population and Theme	‘Public school’ where ‘pedagogical practice’ takes place	owl:hasLocation; or skos:related to Entity and Theme

Source: Authors (2025).

Table 2 shows that hierarchical relationships are represented by generalization/specialization structures in SKOS, while associative links and functional dependencies require the greater expressiveness of OWL. This modeling was defined manually by the authors, as a theoretical exercise not yet empirically validated.

In line with CERIF, although not a classificatory model, its relational architecture supports extensions and external vocabularies. Accordingly, the semantic links identified in TAFNAVEGA can be mapped to entities such as cfResPubl, cfProj, cfResultProduct, and cfMethod, enhancing semantic interoperability with CRIS systems and knowledge graphs. This alignment, however, remains exploratory and does not replace the need for formalization in SKOS/OWL.

Explicit semantic relationships strengthen TAFNAVEGA's conceptual consistency and guide future computational implementation. Next steps should include interface prototypes in academic repositories to empirically test the effectiveness of faceted navigation and its added value for semantic retrieval of scientific information.

4.3. Mapping to Conceptual Representation Standards

Based on the semantic relations and logical dependencies identified in the previous stage, an initial conceptual modeling of TAFNAVEGA was developed according to SKOS principles. This modeling remains at an exploratory level, without implementation in RDF language, and aims to project the future formalization of TAFNAVEGA as an interoperable controlled vocabulary.

Each facet was conceived as a potential skos:ConceptScheme, while the terms were treated as skos:Concept. At the theoretical level, preferred and alternative labels were defined (skos:prefLabel, skos:altLabel), as well as definitions (skos:definition) and hierarchical or associative relations (skos:broader, skos:narrower, skos:related). Figure 1 illustrates an example of modeling for the CAFTE Research Type category [13].

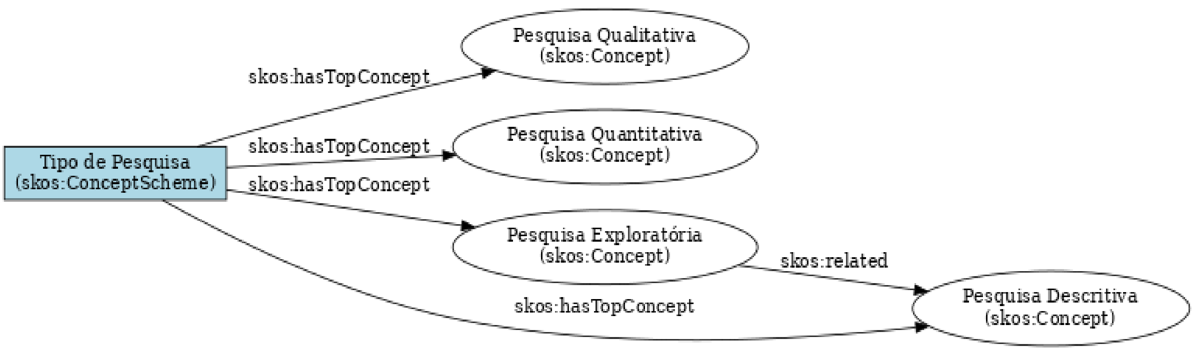

Figure 1 – Example of SKOS modeling for the CAFTE Research Type category [13].

In addition to SKOS, the complementary application of OWL was considered, particularly to express restrictions, functional properties, and logical inferences across facets. Table 3 summarizes the main mapping implications for each CAFTE category.

Table 3
Possibilities for formal representation of CAFTE categories in SKOS and OWL

CAFTE Category	Implication for SKOS / OWL
Theme	skos:broader/narrower
Object	skos:related; may be owl:ObjectProperty in complex graphs
Results	owl:hasOutput; owl:Restriction
Scope	skos:broader; skos:related for adjacent scopes
Historical/contextual foundation	skos:related; may integrate with dc:coverage (temporal)
Research type	skos:broader/narrower
Data collection	owl:Restriction; conditional logic with owl:someValuesFrom
Theoretical foundation	skos:related; owl:equivalentClass between theoretical schools
Methods	skos:broader; related to data type or medium via OWL
Setting	owl:hasLocation; or skos:related to Entity and Theme

Source: the authors (2025).

The choice between SKOS and OWL was guided by the nature of the relations to be represented. Hierarchical, associative, or equivalence relations were modeled in SKOS, given its simplicity and broad interoperability. In turn, functional dependencies, co-occurrence conditions, and logical restrictions (such as “requires,” “generates,” or “occurs in”) were projected in OWL, ensuring semantic expressiveness to support automatic inference and integration with AI systems.

Although still at a conceptual stage, the results of this phase indicate promising paths for the computational implementation of TAFNAVEGA in environments compatible with the Semantic Web and aligned with FAIR principles, reconciling representational simplicity with semantic robustness.

4.4. Alignment with external ontologies and knowledge graphs

Based on the conceptual structure refined in the previous stages, we outlined possibilities for aligning TAFNAVEGA with the CERIF (Common European Research Information Format) model, a widely adopted standard for representing scientific and academic research activities. Although not originally faceted, CERIF has a relational and modular architecture that supports the incorporation of external vocabularies through RDF/OWL extensions, enabling integration with TAFNAVEGA’s categorial logic.

In this mapping exercise, CAFTE categories were associated with core CERIF entities, such as:

- **cfResPubl** – scientific publications, related to Theme, Object, and Results;
- **cfProj** – research projects, aligned with Scope, Research Type, and Theoretical Foundation;
- **cfResultProduct** – research outputs, linked to Results and Methods;
- **cfMethod** – applied methodologies, associated with Data Collection and Procedures;
- as well as other entities addressing infrastructure, historical context, and research agents.

The semantic and logical relations defined in previous stages—hierarchical, associative, and dependency—support this conceptual alignment, projecting TAFNAVEGA as an auxiliary semantic layer in CERIF-based systems. This integration may contribute to: (i) semantic enrichment of metadata for publications and projects; (ii) interoperability across institutional repositories; and (iii) support for recommendation, automated classification, and thematic summarization in open science and AI-driven environments.

Table 4
Correspondence between CERIF entities and CAFTE categories (TAFNAVEGA)

CERIF Entity	CAFTE Category	Notes
cfResPubl	Results	Research outputs (papers, theses, reports)
cfProj	Theme / Object / Scope	Research project as thematic and organizational context
cfResultProduct	Results	Tangible and intangible outputs (models, software, theories)
cfMethod	Methods / Data collection	Scientific methods and applied instruments
cfFund	Theoretical foundation / Historical-contextual foundation	Conceptual and contextual bases of research
cfFacility	Setting	Infrastructure and locations involved in research
cfOrgUnit	Setting / Object	Organizational unit where research takes place
cfPers	Object	Researchers as active agents
cfClassScheme	Research type	Classification scheme representing methodological approaches

Source: the authors (2025).

CERIF, like other widely used standards (e.g., UNESCO Thesaurus, Frascati Manual Taxonomy), supports integration through RDF/OWL mappings. In this scenario, TAFNAVEGA positions itself as a complementary thematic vocabulary, grounded in a multidimensional categorial base, suitable for supporting semantic indexing and retrieval in thesis and dissertation repositories.

As a next step, we propose a pilot study in a real CERIF environment, in partnership with universities or repository consortia. This test will assess the robustness of the alignment, identify potential conceptual adjustments, and validate TAFNAVEGA's applicability in line with open science requirements and FAIR principles.

5. Conclusions

The reformulation of TAFNAVEGA presented in this article constitutes a conceptual sketch, developed through theoretical modeling and intellectual inference. Although the results point to promising directions for the semantic structuring of academic repositories, the current stage remains abstract and lacks practical validation. The semantic relationships between categories, the mappings with the CERIF model, and the preliminary modeling in SKOS/OWL represent a starting point rather than a finalized implementation.

It is acknowledged that the absence of empirical experimentation limits the generalization of the findings. Challenges such as terminological variability across institutions, metadata curation, and the costs of large-scale implementation constitute limitations that should be addressed in future work.

Among the perspectives for continuity, the following stand out: (a) conceptual adjustments in the facets and terms of the taxonomy; (b) application of the reformulated TAFNAVEGA in real subsets of theses and dissertations, aiming to test its effectiveness in information retrieval; (c) integration of the taxonomy into experimental repositories, assessing its performance in categorization, recommendation, and semantic enrichment systems; (d) usability studies with librarians, managers, and end users, to ensure that the structure is intuitive, flexible, and aligned with real needs.

From a theoretical standpoint, this study dialogues with the proposal of fundamental categories presented by Pereira, Moreira, and Santarém Segundo [12], whose emphasis on dimensions such as cause and purpose broadens the repertoire of faceted classification. This approach demonstrates that CAFTE can be refined and extended, consolidating itself as a structure capable of interoperability with contemporary ontological models.

The original contribution of this study thus lies in the articulation between classical faceted taxonomies and interoperable semantic standards, suggesting a hybrid path between classifications and lightweight ontologies applicable to Information Science. Such an approach expands the possibilities of integration with digital systems and strengthens the agenda of Open Science by promoting the standardized semantic description of data and publications.

It is concluded that aligning faceted taxonomies with Semantic Web standards requires continuous and contextualized revisions, but initiatives such as TAFNAVEGA can decisively contribute to interoperability policies, metadata enrichment, and the consolidation of knowledge graphs applied to scientific information.

Acknowledgments

The authors acknowledge the support of the National Council for Scientific and Technological Development (CNPq), Brazil, for funding this research (Processes Author 1 and Author 2).

Declaration on Generative AI

In preparing this work, the authors used ChatGPT-4 as a complementary support tool for the preliminary identification of semantic relations and correspondences between classificatory models, within the context of the conceptual reformulation of TAFNAVEGA. All content was subsequently reviewed and edited by the authors, who take full responsibility for the final version of this publication.

References

[1] B. C. M. S. Maculan, *Taxonomia facetada navegacional: construção a partir de uma matriz categorial para trabalhos acadêmicos*, M.S. thesis, Programa de Pós-Graduação em Ciência da Informação, Univ. Fed. Minas Gerais, Belo Horizonte, Brazil, 2011.

[2] L. Bardin, *Análise de conteúdo*, 4th ed. Lisboa: Edições 70, 2009.

[3] S. R. Ranganathan, *Prolegomena to Library Classification*, 3rd ed. London: Asia Publishing House, 1967. [Online]. Available: http://dlist.sir.arizona.edu/arizona/handle/10150/106370. [Accessed: Jul. 8, 2025].

[4] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, p. 160018, 2016. DOI: 10.1038/sdata.2016.18.

[5] A. Miles and D. Beckett, "SKOS Simple Knowledge Organization System Reference," W3C Recommendation, 2009. [Online]. Available: https://www.w3.org/TR/skos-reference/. [Accessed: Jul. 10, 2025].

[6] M. K. Smith, C. Welty, and D. L. McGuinness, "OWL Web Ontology Language Guide," W3C Recommendation, 2004. [Online]. Available: https://www.w3.org/TR/owl-guide/. [Accessed: Jul. 10, 2025].

[7] euroCRIS, "Common European Research Information Format (CERIF)," 2021. [Online]. Available: https://www.eurocris.org/cerif. [Accessed: Jul. 10, 2025].

[8] M. B. da Silva and Z. D. de Miranda, "Estudo teórico-analítico-sintético sobre a presença de facetas na organização da informação: do físico ao digital," in *Proc. 19th ENANCIB*, Londrina, Brazil, 2018. [Online]. Available: https://proceedings.ci.inf.br/index.php/enancib/article/view/1621. [Accessed: Jul. 8, 2025].

[9] A. G. Coelho, G. Â. de Lima, and M. M. Borges, "As taxonomias navegacionais facetadas e a produção científica da Ciência da Informação: tendências temática e diacrónica (2011–2020)," in *Proc. 6th Encontro Nacional da ISKO Portugal*, Coimbra, Portugal, 2021, pp. 617–633.

[10] M. B. Almeida and L. M. D. Teixeira, "Revisitando os fundamentos da classificação: uma análise crítica sobre teorias do passado e do presente," *Perspect. Ciênc. Inf.*, vol. 25, no. esp., pp. 28–56, Feb. 2020.

[11] G. Â. de Lima, "Organização e representação do conhecimento e da informação na web: teorias e técnicas," *Perspect. Ciênc. Inf.*, vol. 25, no. esp., pp. 57–97, Feb. 2020.

[12] C. M. Pereira, W. Moreira, and J. E. Santarem Segundo, "Classificação facetada: proposta de categorias fundamentais para organizar teses e dissertações em uma biblioteca digital," *Encontros Bibli*, vol. 26, e79427, 2021. DOI: 10.5007/1518-2924.2021.e79427.

[13] OpenAI, *ChatGPT* (versão 4.0), San Francisco, CA, USA: OpenAI, 2023. [Online]. Available: https://chat.openai.com/

[14] N. Oddone and M. Y. F. S. de F. Gomes, "Uma nova taxonomia para a ciência da informação," in *Proc. 5th Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB)*, Belo Horizonte, Brazil, 2003, pp. 1–24.

[15] G. M. Sacco and Y. Tzitzikas, Eds., *Dynamic Taxonomies and Faceted Search: Theory, Practice and Experience.* Berlin, Germany: Springer, 2009.