



Article

# From Faceted Taxonomy to Explainability: Operationalizing TAFNAVEGA With BPM-lite in Thesis and Dissertation Repositories

Benildes Coura Moreira dos Santos Maculan<sup>1,\*</sup>, Gercina Ângela de Lima<sup>1</sup>,  
Elisângela Cristina Aganette<sup>1</sup>, Frederico Cesar Mafra Pereira<sup>1</sup>, Francis Bento Marques<sup>2</sup>,  
Yuri Bento Marques<sup>3</sup>

<sup>1</sup>Graduate Program in Knowledge Management and Organization, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil

<sup>2</sup>Instituto de Ciência, Engenharia e Tecnologia, Universidade Federal dos Vales Jequitinhonha e Mucuri, 39803-371 Teófilo Otoni, MG, Brazil

<sup>3</sup>Departamento de Informática, Instituto Federal do Norte de Minas Gerais, 39800-430 Teófilo Otoni, MG, Brazil

\*Correspondence: [benildes@gmail.com](mailto:benildes@gmail.com) (Benildes Coura Moreira dos Santos Maculan)

Academic Editor: Natália Tognoli

Submitted: 26 January 2026 Revised: 26 March 2026 Accepted: 14 April 2026 Published: 29 June 2026



**Benildes Coura Moreira dos Santos Maculan**

PhD and Master's degree in Information Science from the Federal University of Minas Gerais, where she also earned her degree in Library Science. She is an Associate Professor at the School of Information Science at UFMG and a researcher at the Graduate Program in Knowledge Management and Organization (PPGGOC/UFMG). She is a member of the MHTX, RECRI, and Digital Humanities research groups (University of Coimbra, Portugal), as well as the Scientific and Technical Advisory Council (STAC) of ISKO International. Her research focuses on Knowledge Organization Systems, knowledge representation and information retrieval, data integration, language interoperability, information services, and information security.



**Gercina Ângela de Lima**

Full Professor at the School of Information Science of the Federal University of Minas Gerais, Brazil. She holds a PhD in Information Science from UFMG and a Master's in Library Science from Clark Atlanta University-USA. She is a professor and researcher in the Graduate Program in Knowledge Management and Organization. Serves as Vice-Chair of the Scientific and Technical Advisory Council of ISKO and coordinates the MHTX Research Group. She has been a Productivity Research Fellow of the Brazilian National Council for Scientific and Technological Development since 2010, focusing on knowledge organization and representation, hypertext, digital libraries, and the semantic web.



**Elisângela Cristina Aganette**

As a professor and researcher at the Federal University of Minas Gerais, I work in the fields of Information Science, university administration, digital transformation, and artificial intelligence applied to information organization. I currently serve as Vice-Rector for Administration at UFMG, developing projects focused on administrative innovation, institutional governance, and the modernization of public processes. My academic and professional background involves research on knowledge representation, automatic indexing, information retrieval, and the use of Large Language Models in educational and scientific contexts. I also work at the intersection of technology, management, and institutional policies, with a focus on improving university public services.



**Frederico César Mafra Pereira**

Postdoctoral Fellow (UFSCar) and Specialist in Innovation in Information Units at UFSCar. Ph.D. and M.A. in Information Science, Economist, and Marketing Specialist from UFMG. Professor at UFMG and Coordinator of the Graduate Program in Knowledge Management and Organization. Researcher in Information and Knowledge Management, Competitive Intelligence, Digital Transformation, Innovation in Information Units, and Strategy. Leader of the IKI Research Group and member of the ITI/UFSCar and Academic BPM 2.0/UFMG projects. Mentor, speaker, member of the Editorial Review Board of The Bottom Line (Emerald), board member of i2i, and associate of ISKO-Brasil.



**Copyright:** © 2026 The Author(s). Published by IMR Press.  
This is an open access article under the [CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/).

**Publisher's Note:** IMR Press stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---

Francis Bento Marques



He holds a bachelor's degree in Information Systems from Presidente Antônio Carlos University (2008), a specialization in Systems Engineering from the Open University of Brazil (2015), a Professional Master's degree in Technology, Environment, and Society from the Federal University of the Jequitinhonha and Mucuri Valleys (2018), and a Ph.D. in Knowledge Management (PPGGOC) from the School of Information Science at the Federal University of Minas Gerais (UFMG), with a sandwich period at the University of Vienna through the PDSE—Sandwich Doctoral Program Abroad. He is currently a computer lab technician at the Federal University of the Jequitinhonha and Mucuri Valleys - Mucuri Campus

---

Yuri Bento Marques



He holds a bachelor's degree in INFORMATION SYSTEMS from Presidente Antônio Carlos University (2009), a master's degree in Computer Science from the Federal University of Viçosa (2015), and a Ph.D. in Knowledge Management from the Federal University of Minas Gerais (2021). Since 2010, he has been a tenured professor at the Federal Institute of Education, Science, and Technology of Northern Minas Gerais. He has been the leader of the NsoluTI research group since 2020 and a member of the PAEBCOMP research group since 2016.

---

## Abstract

This paper presents a socio-technical redesign of TAFNAVEGA to support its operation in thesis and dissertation repositories, guided by two principles: explainable information retrieval and lightweight vocabulary governance. The study adopts a qualitative, design-oriented approach structured in five phases: discovery and planning; terminological curation and modeling; operation of a lightweight Business Process Management approach (BPM-lite) workflow for indexing and controlled term inclusion; linking question to be answered (QR) to faceted combinations with core facet categories (CAFTE) to test explainability; and synthesis of the resulting artifacts. The results indicate reduced terminological ambiguity, improved retrieval explainability (i.e., “why an item was retrieved”), and the generation of minimal decision traces sufficient to support institutional curation without increasing bureaucratic overhead. The study demonstrates that repositories can function as knowledge organization ecosystems when they combine an operational faceted vocabulary, minimal curatorial roles, and a modular infrastructure.

**Keywords:** knowledge organization; faceted taxonomy; thesis and dissertation repositories; lightweight governance; BPM-lite

## 1. Introduction

Within Knowledge Organization, the faceted taxonomy known as TAFNAVEGA (Maculan, 2011) is understood as an operational Knowledge Organization System (KOS), aligned with the principles of faceted classification formulated by Ranganathan (1967). Based on this premise, this paper establishes two verifiable objectives: (i) to demonstrate the explainability of information retrieval through question to be answered (QR), materialized in the interface as facet chips (visible filters); and (ii) to demonstrate the validity of lightweight governance of the vocabulary, supported by a minimal and sufficient trace (decision and change logs) capable of ensuring auditability and incremental maintenance in thesis and dissertation (T&D) repositories (Maculan, 2011; Santos and Flores, 2020).

T&D repositories constitute critical infrastructures for scholarly communication in Brazil (Santos and Flores, 2020; Sganzerla, 2024). Despite this central role, these environments face two persistent challenges: retrieval explainability—that is, users' ability to understand why a given item was retrieved—and lightweight governance, understood as the definition of who decides, based

on which rules, and how changes are recorded. Current practices in such environments tend to prioritize keyword-based search mechanisms, while devoting limited attention to institutional roles, curatorial routines, and decision transparency (Sganzerla, 2024). In response to this scenario, this study proposes repositioning the repository as a socio-technical Knowledge Organization Ecosystem (KOE), with TAFNAVEGA as its faceted core, structured into ten basic categories (C1–C10), operated through human curation and supported by a short workflow of states (Proposed → Under review → Approved → Published), accompanied by concise decision and change records.

Methodologically, the study adopts a qualitative socio-technical design approach aimed at making the relationships among people, processes, and technologies explicit and auditable (Gregoriades and Sutcliffe, 2008). Rather than comparing algorithms, the focus lies on explicating roles, curatorial rules, facets, and artifacts that sustain informational mediation within the repository. The investigation starts from a prototype developed in Microsoft Access, in which theses and dissertations were already indexed using faceted structures. In preparing the migration of this prototype to an ontology-oriented web arrange-

ment, systematic terminological curation was conducted to eliminate homographs, qualify labels, and stabilize facet-specific Uniform Resource Identifiers (URIs). This process enables retrieval explanations to be materialized both in the interface—through chips representing active facets—and in vignettes, understood as short query scenarios that illustrate use contexts.

Following this introduction, Section 2 presents the theoretical framework underpinning the proposal, covering TAFNAVEGA, the socio-technical KOE, faceted structures and explainability, and information needs. Section 3 describes the methodology, including the corpus, terminological inputs, and analytical procedures. Section 4 presents the results according to the methodological phases (curation, workflow operation, QR linking, and synthesis). Finally, Section 5 summarizes the main conclusions and outlines future developments.

## 2. Theoretical and Methodological Framework

This section aligns the theoretical and methodological foundations underpinning the proposed approach. Section 2.1 presents TAFNAVEGA in its original formulation (Maculan, 2011), establishing the conceptual basis of the faceted vocabulary. Section 2.2 situates the proposal within a socio-technical KOE, understanding repositories as configurations of people, processes, and technologies. Section 2.3 addresses how facets (C1–C10) support retrieval explainability. Section 2.4 introduces the lightweight Business Process Management approach (BPM-lite), which enables consistent vocabulary use through a short state-based workflow (Proposed → Under review → Approved → Published) and lightweight decision records. Section 2.5 discusses the repository as an institutional infrastructure. Finally, Section 2.6 frames information needs by linking user profiles and tasks to faceted pathways.

The original taxonomy adopted in this study (named TAFNAVEGA) is organized into ten facets (C1–C10), distinguishing conceptual roles such as theme (C1) and empirical object (C2), and enabling controlled combination of research dimensions.

### 2.1 A TAFNAVEGA: Original Overview

TAFNAVEGA is a faceted navigational taxonomy for theses and dissertations composed of ten recurrent research-oriented categories: Topic, Empirical Object, Scope, Setting, Type of Research, Data Collection, Methods, Theoretical Foundation, Historical/Contextual Foundation, and Results (Maculan, 2011). Its objective is to structure T&D content to support search, progressive refinement, and the visibility of scientific elements such as topics, methods, theories, instruments, and results.

The taxonomy was developed through a combination of Domain Analysis, the Documentary Reading Model (Fujita and Rubi, 2006), Content Analysis, Ranganathan's faceted principles, and contributions from the Classification

Research Group (CRG). Terminological control was supported by the Taxonomy of Information Science (Oddone and Gomes, 2003).

This process resulted in a set of ten categories (named core facet categories, CAFTE) and a corpus-derived repertoire of terms to support faceted navigation. The categories: C1 (Topic), C2 (Empirical Object), C3 (Scope), C4 (Setting), C5 (Type of Research), C6 (Data Collection), C7 (Methods), C8 (Theoretical Foundation), C9 (Historical/Contextual Foundation), and C10 (Results), represent complementary dimensions of research and are applied in faceted indexing. In its original formulation, these ten basic categories were progressively populated with taxonomic term structures as theses and dissertations were indexed. This process drew on a prior taxonomy developed by Oddone and Gomes (2003), providing an initial terminological basis that was subsequently refined through curatorial practice.

In this model, users begin with a single facet and progressively refine results through combinations, without relying on explicit Boolean logic. This mechanism reduces ambiguity and supports exploratory search.

In the original prototype, facets functioned as semantic menus, enabling users to combine values and refine results iteratively. The system, implemented in Microsoft Access, supported navigation by exposing distributions across facets and privileging access to scientific components of research rather than only descriptive metadata. In this sense, TAFNAVEGA operates as a navigational retrieval model, in which users construct their search path through successive facet selections rather than formulating explicit queries.

Having outlined its structure and purpose, TAFNAVEGA is now framed as part of a socio-technical KOE, in which its use depends on the articulation of people, processes, and technologies.

### 2.2 Socio-Technical KOE

In this study, the socio-technical KOE is understood as an integrated configuration of people, processes, and technologies organized to perform institutional functions such as describing, representing, retrieving, and preserving information, with transparency and accountability. Rather than reducing the repository to software or search algorithms, this arrangement is conceived as a system of interdependencies among actors, tasks, infrastructure, and informational artifacts.

This perspective aligns with approaches that emphasize the need to make explicit the interactions between human and technological components and to analyze usage and decision-making scenarios prior to system redesign (Gregoriades and Sutcliffe, 2008). Although qualitative in nature, such analysis is fundamental for understanding the organizational effects of design interventions.

A study on usability and accessibility in institutional repositories shows that barriers persist when technical de-

sign is not accompanied by continuous review routines and clearly defined roles (Marcelino and Siebra, 2015). This reinforces the socio-technical nature of repository performance and supports the emphasis, in this study, on lightweight governance structures, iterative improvement cycles, and documented curatorial decisions.

In line with this perspective, the service blueprint is organized into three levels: (i) institutional (purpose, roles, and lightweight policies), (ii) curatorial (facet application rules, citation order, and scope notes), and (iii) operational (interface elements such as filters, fields, and logs) (Joseph et al., 2022). This layered view connects design decisions to observable system behavior.

Socio-technical configurations are not always directly observable, as many rules remain tacit; however, they can be inferred from practices, documents, and narratives, allowing the reconstruction of how technologies and institutional arrangements evolve over time (Miörner et al., 2022). This perspective informs the analysis of logs, curatorial decisions, and usage vignettes as evidence of KOE dynamics.

From a methodological standpoint, this study adopts a qualitative, socio-technical design-oriented approach. It draws on Socio-Technical Grounded Theory (STGT) to ensure alignment between data, analysis, and explanation in contexts where social and technical elements are intertwined. The approach prioritizes theoretical variation, constant comparison, and adequacy in terms of clarity, usefulness, and plausibility (Hoda, 2021).

Within this framework, the KOE organizes TAFNAVEGA as an explainable faceted core supported by minimal roles and a BPM-lite workflow (Proposed → Under review → Approved → Published), along with provenance logs and lightweight policy artifacts. The focus is not on optimizing performance metrics, but on ensuring the traceability, justifiability, and auditability of decisions in the operation of the repository.

### 2.3 Facets and Explainability

Faceted analysis constitutes an analytic–synthetic method in which subjects are decomposed into constituent categories (facets) and recombined through a controlled syntax. In Ranganathan’s formulation (1958), this approach is expressed across the planes of ideas, verbal expression, and notation—that is, what a concept is, how it is named, and how it is encoded—as well as through the principle of hospitality to new subjects, according to which a system may grow without losing coherence or representational stability. These foundations provide a solid basis for the explainability of information retrieval, as they allow users to visualize which dimensions of a subject have been activated—through facet chips—and to understand why a given item has been retrieved. At the same time, curatorial practice ensures consistency of decisions over time (Ranganathan, 1958, 1963).

Within the postulational approach, Ranganathan (1958) makes explicit a set of postulates and stages that move from basic classes to isolates and facets, while also establishing a separation between distinct functions. Ordering for shelving or browsing is not equivalent to optimizing representation for specific retrieval purposes. Operationally, this implies: (i) determining the conceptual role of each term, that is, identifying the facet to which it belongs; (ii) ensuring labels and URIs that are compatible with this role; and (iii) guaranteeing that the structure implements, at the operational level, the decision taken at the ideational level, in accordance with criteria such as helpful order, uniqueness, and the possibility of synthesis (Broughton, 2012; Ranganathan, 1958).

Historically, the development of faceted classification also draws on the contributions of the CRG, which proposed starting from concepts and their roles—such as entity, process, agent, space, and time—and making explicit the relations that form a network of nodes, rather than relying on enumerated universes. This orientation underpins interfaces in which facet combinations are understandable and manipulable, forming the basis of the chips that make visible to users the reasons for retrieval outcomes (Wilson, 1972).

While the classical faceted approach, as developed by Ranganathan and the CRG, focuses on the decomposition of the subject itself—typically in terms of categories such as entity, process, agent, space, and time—the present study adopts a different operational perspective. TAFNAVEGA does not aim to produce a full faceted classification of subject aboutness in the strict Ranganathan and CRG sense. Instead, it organizes facets around recurrent components of academic research reports, such as research focus (topic), empirical object, methods, data collection, theoretical foundation, and results.

In this sense, the model draws on the analytical–synthetic logic of faceted classification, but reorients it toward the structuring of epistemic and methodological dimensions for navigational retrieval. Rather than decomposing subject content, TAFNAVEGA treats these components as independent yet combinable facets, enabling exploratory, explanatory, and reuse-oriented access to research outputs.

Additionally, the theme-rheme distinction provides a useful lens for interpreting the Results facet (C10). While the theme refers to what a document is about, the rheme concerns what is asserted or produced in relation to that theme. Accordingly, the Results facet does not duplicate subject content, but captures outcome-oriented elements—such as models, profiles, guidelines, or evaluative findings—that are particularly relevant for retrieval and reuse. This interpretation reinforces the role of faceted structures not only in representation, but also in explainability and action-oriented navigation (Gnoli, 2018; Weinberg, 1988).

In the contemporary literature, Broughton (2017) emphasizes that faceting has evolved from a construction technique into an integrated model of knowledge organization and information retrieval, in which categories, ordering, and intra- and inter-facet relations become explicit and machine-explorable. This evolution explains its adoption in classifications, thesauri, and web-based applications, including faceted revisions of the Universal Decimal Classification. From this perspective, explainability emerges from syntactic regularity associated with preferred labels and clear scope notes.

In navigational faceted taxonomies, classical principles—such as the selection and ordering of facets, citation rules, and criteria of admissibility—remain operational in the design of chip-based filtering experiences (Maculan, 2011; Medeiros, 2013). In such instruments, the interface does not function as an appendage to representation, but as its pragmatic expression. In the proposal presented here, this logic is materialized in active facet chips, accompanied by qualifiers and C1–C10 codes, as well as in vignettes (scenarios) that explicitly explain why a particular item appeared in the results.

Applications of faceted classification beyond the bibliographic domain—in business, engineering, and manufacturing contexts—demonstrate the portability of the method, albeit with variations in formulation and technological support (Pestana, 2017). These uses reinforce faceting as a mode of structured reasoning that cuts across domains and supports explainability, provided that the combination of dimensions remains transparent to the user.

In practical terms, explainability in faceted systems depends on the explicit separation of conceptual roles across facets. Typical ambiguities include: (i) the conflation of phenomenon, method and product within a single term; (ii) the use of data collection techniques without distinguishing between procedures and artefacts; and (iii) the overlap between analytical focus, empirical object and contextual framing. These types of ambiguity are addressed through facet differentiation and are illustrated with empirical examples in Section 4.

#### 2.4 *Lightweight Governance and BPM for Academic Repositories*

The literature on public administration and process management highlights the emergence of lightweight governance, oriented toward coordination rather than control. Unlike hierarchical or performance-driven models, this approach emphasizes procedural instruments and collaborative arrangements among autonomous actors (Nederhand et al., 2019). In this context, governance operates through interaction mechanisms rather than centralized authority.

Applied to institutional repositories, these principles support a BPM-lite, focused on roles, workflows, and minimal control points sufficient to ensure consistency, traceability, and transparency. Quality emerges from the clar-

ity of procedures rather than from the proliferation of indicators, making this approach suitable for scientific information environments characterized by autonomy and diversity.

At the same time, repositories must be understood as infrastructures for preservation and organizational memory. In this regard, Silva (2023) shows that the integration of Data Management Body of Knowledge (DMBOK) domains with preservation models such as Open Archival Information System (OAIS) and the Digital Curation Centre (DCC) lifecycle provides a framework for ensuring authenticity, integrity, and long-term accessibility. This perspective situates governance within the broader context of digital curation.

From this standpoint, lightweight governance is grounded in distributed responsibility and verifiable provenance. BPM-lite operates as a mediation mechanism that institutionalizes the curation lifecycle, from submission to dissemination, with each workflow step generating provenance records (W3C PROV). Traceability thus emerges from the process itself, rather than from external control structures.

Empirical evidence supports this approach. Lake and Regenauer (2024) demonstrate that mediated and automated workflows—such as metadata transformation, permission checking, and batch ingestion—reduce friction, increase traceability, and expand operational capacity in repository management.

Together, these perspectives indicate that lightweight governance, supported by BPM-lite and aligned with preservation frameworks, provides a viable approach for thesis and dissertation repositories. This arrangement balances academic autonomy with institutional responsibility, ensuring transparency and auditability without introducing unnecessary procedural overhead.

#### 2.5 *The Repository as Infrastructure*

In this study, the institutional repository is understood as a socio-technical infrastructure supporting ingestion, curation, preservation, and access. This perspective emphasizes structural capacities, clear responsibilities, and the alignment between technology and academic purposes (Sganzerla, 2024).

A functional repository must ensure: (i) stable identification of items and terms (IDs/URIs); (ii) metadata standards with a clear application profile (e.g., Dublin Core and persistent identifiers); (iii) observability (logs and metrics); (iv) security and privacy; and (v) interoperability for integration and audit. These requirements align with recommendations for information infrastructures in the Brazilian context (Sganzerla, 2024).

As an integration node, the repository interacts with institutional systems and external services (e.g., ORCID, DOI, and harvesting protocols). Interoperability depends less on adopting new tools and more on explicit applica-

tion profiles, schema mappings, and consistent Applications Programming Interface (APIs) (Matos, 2021; Santos and Flores, 2020).

Within this framework, TAFNAVEGA operates as a semantic infrastructure: each term is assigned a URI and, when necessary, a facet-specific scope note. This organization prevents ambiguity and supports faceted navigation with traceable curatorial decisions.

From an architectural perspective, it is recommended to separate the domain layer (repository rules and TAFNAVEGA) from the infrastructure layer (ingestion, storage, workflow), enabling modularity and technological evolution without loss of consistency (Microsoft, 2023, 2025). Observability and event logging support auditability as a natural outcome of system operation.

Conceived as modular, observable, and testable, the repository can evolve incrementally while maintaining consistency and reliability, even in resource-constrained environments (Santos and Flores, 2020; Sganzerla, 2024). In this configuration, the repository functions as an institutional infrastructure that supports retrieval explainability and auditability, while TAFNAVEGA ensures semantic consistency by assigning each term a defined conceptual role within the faceted structure.

### 2.6 Information Needs

In this study, information needs are understood as the gap between a user's current cognitive state and a task requiring evidence in the context of T&D. In institutional repositories, such needs are not expressed solely through keywords, but emerge from user profiles and tasks—such as teaching, research, and management—and are operationalized through faceted pathways (e.g., Theme, Empirical Object, and Type of Research), treated here as components of the research report rather than as a full representation of subject aboutness. Addressing these needs requires both explainability of retrieval and reduction of interaction frictions.

Empirical studies indicate distinct patterns of repository use and persistent gaps between needs and system performance. In Brazilian federal universities, repository use is significantly higher among graduate students than undergraduates, with general search engines remaining the primary entry point (Ávila et al., 2017). Evaluations of institutional repositories also identify recurrent problems related to navigation, search, and accessibility, suggesting a structural mismatch between information needs and system design (Custódio and Vechiato, 2017; Marcelino and Siebra, 2015). Additionally, studies involving researchers point to usability issues and limited awareness of repository functions, indicating that many information needs are organizational and procedural rather than purely technological (Ramos et al., 2022).

In the current context, the increasing use of generative AI tools has begun to reshape search behavior. While

widely adopted, such tools are not yet considered reliable sources of scientific evidence, reinforcing the continued relevance of curated repositories as trustworthy infrastructures (Amoah et al., 2025; Deschenes and McMahon, 2024; Silvestre et al., 2023).

Within this proposal, information needs are translated into QR and mapped onto combinations of core facet categories (CAFTE), materialized as chips in the interface. This mapping enables retrieval explainability by making explicit the criteria for item selection (“why did this item appear?”) and supports the identification of friction points for continuous improvement. The operationalization of this approach is detailed in Sections 3 and 4.

## 3. Materials and Methods

This present study adopts a qualitative, design-oriented approach, understood as one that is not limited to the description or analysis of existing phenomena, but is oriented toward the proposition and construction of artifacts capable of addressing concrete problems (Matta, Silva, and Boaventura, 2015). It is therefore aligned with the field of Design Science Research (DSR), in which the central objective is to conceive, structure, and validate applicable models, frameworks, or processes based on explicit criteria of relevance, utility, and theoretical grounding (Hevner et al., 2004; Santana et al., 2023; Santana et al., 2024). Within this framework, the research assumes an interventive and constructive character, aiming to produce a solution that is both operationalizable and institutionally viable.

Specifically, the present research proposes, justifies, and documents a socio-technical arrangement of lightweight governance (with lightweight BPM-lite) for the faceted curation of TAFNAVEGA in T&D repositories. The artifact consists of a socio-technical BPM-lite model oriented toward provenance capture (W3C PROV) and the structured generation of editorial logs throughout the curation lifecycle. This methodological orientation does not diminish scientific rigor; rather, it repositions rigor in criteria such as conceptual coherence, problem adherence, procedural transparency, and the possibility of controlled analytical generalization. In this sense, intervention is understood as a legitimate form of knowledge production within Information Science.

The methodological procedure was organized into five integrated phases, guided by principles of lightweight governance (minimal roles, short states, and succinct decision records), as presented in Table 1.

The five-phase structure makes explicit a controlled, iterative, and decision-oriented pathway in which terminological curation and the BPM-lite workflow are integrated from the outset as operational components of the design. By adopting lightweight governance—with minimal roles, short states, and succinct records—the methodology seeks to ensure traceability without compromising process fluidity. This design creates conditions for evolutionary con-

**Table 1. Integrated methodological phases (curation + BPM-lite + lightweight governance).**

Phase	Objective
Phase 1. Discovery and Planning	Frame the case and inventory the repository/Access prototype; identify QR ; information needs); plan curation with a focus on priority ambiguities and conceptual gaps.
Phase 2. Terminological Curation and Modeling	Eliminate homographs; normalize labels; resolve overlaps by facet; ensure univocity within each facet; consolidate term and authority lists; create scope notes and qualifiers; assign URIs by facet.
Phase 3. Workflow Operation (with BPM-lite)	Execute the short workflow (Proposed → Under review → Approved → Published) with logs; record term proposals and adjustments; apply CAFTE guidelines; keep the vocabulary state and annotations updated; ensure stable IDs/URIs.
Phase 4. QR Linking and Interface	Map QR to facet combinations; test explainability; adjust terms, notes, and relations when necessary; ensure curatorial consistency.
Phase 5. Review and Synthesis	Consolidate lessons learned; prepare a replicable package (checklists, logs, QR matrix); minimize operational frictions.

Source: Authors' own research data (2026). Note: BPM-lite, lightweight Business Process Management approach; QR, question to be answered; CAFTE, core facet categories; IDs, identifiers; URIs, Uniform Resource Identifiers.

sistency, auditability, and subsequent institutionalization of the model, while preserving the flexibility required for adaptive adjustments in real-world contexts.

The results are presented in direct correspondence with the five methodological phases described, demonstrating how each design decision materialized in the proposed socio-technical arrangement. Section 4 presents the results and discussions derived from the application of this methodological pathway.

## 4. Results

This section presents the main results obtained, highlighting how the phases of terminological curation and the BPM-lite workflow materialized from the attempt to convert the Microsoft Access prototype into a web-based arrangement within the context of the analyzed T&D repository, as illustrated in Fig. 1.

The presentation of the findings is accompanied by a critical discussion, in which the observed effects are described and interpreted, as well as their implications for information governance, for the explainability of faceted structures, and for the feasibility of institutionalizing the model in real-world environments.

### 4.1 Results of Phase 1: Discovery and Planning

The starting point was the TAFNAVEGA prototype implemented in Microsoft Access, in which a corpus of 41 theses and dissertations had already been described using the ten basic categories (C1–C10). The prototype was designed for faceted navigation without empty result sets: each record is explicitly associated with the ten facets through structured table relationships (record ↔ facet). This modeling ensures filtering consistency and retrieval predictability.

At this stage, the methodological framing of the study was defined as a socio-technical arrangement of lightweight governance (with lightweight, BPM-lite). The purposes of the evidence artifacts to be employed were also aligned, namely: vignettes (short usage scenarios), log excerpts, and “before/after” tables. These artifacts were adopted to support explainability, consistency, and auditability of the process, without resorting to comparative metrics.

Inspection of the prototype and preparation for migration to a web environment revealed a critical issue: the occurrence of label overlap across facets, that is, the same term appearing with different meanings in different categories. This finding indicated that the vocabulary was not yet sufficiently disambiguated. It therefore became necessary to conduct a systematic curation process of TAFNAVEGA prior to migration. This insight guided the formulation of the QR, which subsequently oriented both the curation process and the interface design.

For each QR, the user's informational intent was recorded, along with a draft CAFTE mapping (C1–C10), sufficient to guide subsequent phases. In parallel, the minimal roles activated in the BPM-lite workflow were made explicit: Proposer (term or adjustment suggestion), Curator (application of CAFTE guidelines, verification of homonymy/polysemy, and drafting of scope notes), Approver (inter-facet validation), and IT/Maintainer (publication and versioning, ensuring stable IDs/URIs and exportability). The operational states of the workflow (Proposed → Under review → Approved → Published) were also defined, with concise decision records.

As planning criteria, priority was given to ambiguities that directly affect the QR and to recurrent conflicts between facets of different natures, especially C1×C2 (Theme × Empirical Object) and C6×C7 (Data Collection Instru-

Navegue pelas facetas de classificação para descobrir documentos e produções científicas organizadas.  
Utilize a árvore abaixo para explorar os temas.



**Fig. 1. TAFNAVEGA web interface.** Source: Authors' own research data (2026).

ment  $\times$  Method/Procedure). Limits and risks associated with the work were also acknowledged, such as uneven terminological coverage in specific subdomains, modeling legacies inherited from the prototype that may require fine-tuning, and the need to keep logs anonymized and restricted to a minimum of personal data.

Upon completion of the discovery and planning phase, the study baseline was established (prototype, QR, and BPM-lite), along with the roadmap for Phase 2, dedicated to terminological curation and modeling (label normalization, scope notes, and facet qualification), from which the remaining results were derived.

At this stage, an initial set of ten QR was developed, later refined and exercised in Phase 4 through interface testing. Two types of QR were considered: (1) collection overview queries (P), aimed at curatorial and thematic diagnosis of the T&D “map”, with the ability to reveal ambiguities, asymmetries, and gaps; and (2) specific faceted queries (S), representing real-use information needs and evidencing the potential of content indexing through CAFTE combinations. Table 2 summarizes this initial matrix.

This set of ten QR constitutes the initial matrix of informational intentions and needs, which was refined and exercised in Phase 4 (scenarios and explainability tests). The prioritized QR directly guided the curatorial work in Phase 2, particularly with regard to conflicts between basic categories prone to ambiguity, such as C1 $\times$ C2 and C6 $\times$ C7.

#### 4.2 Results of Phase 2: Terminological Curation and Modeling

With the baseline established, Phase 2 focused on vocabulary normalization and on resolving overlaps between facets, thereby preparing the ground for QR-based testing. The process started from the complete list of terms in the

Access prototype and proceeded along two complementary curation tracks, both anchored in terminological authorities: (i) the taxonomy by Oddone and Gomes (2003), used to recognize the thematic universe and identify likely zones of polysemy; and (ii) the Brazilian Thesaurus of Information Science (Pinheiro and Ferrez, 2014), adopted as the source of preferred labels and writing standards, also serving as a basis for scope notes in Information Science domains. For the methodological facets (C5–C7), methodological works used in the original project (Maculan, 2011) were consulted in order to stabilize distinctions among research approach/type, data collection instruments, and procedures/methods.

The first step was lexical normalization: revising synonyms and orthographic variants, adopting preferred labels, and creating local-to-authority mappings when the term used in the prototype did not coincide with the thesaurus preferred term (while preserving the mapping for future export). For each validated label, univocity within the facet was checked to ensure that the term retained a single operational meaning across C1–C10. In parallel, the authority source (taxonomy, thesaurus, or methodological work) was recorded for each entry as a reference field to support consistent use in subsequent phases.

Next, inter-facet homographs were addressed. Whenever the same label appeared with different roles — for example, Repository as a theme (C1) and as an empirical object (C2), or Interview as a data collection instrument (C6), as a research method when applicable (C7), and as a research product (C10)—the rule of facet-specific individualization was applied. This involved adding qualifiers to labels (e.g., “Repository [theme]” / “Repository [object of study]”), assigning distinct URIs, and, when necessary, drafting concise scope notes to clarify the intended use

**Table 2. Initial matrix of QR and CAFTE mapping.**

ID	QR (formulation)	Minimum facets (CAFTE)	Expected evidence (how to verify)	Example chips (filters)
P1	Which themes predominate in the collection?	C1	List/view by C1 without noise from other facets; non-generic labels	Knowledge organization (C1) · Library services management (C1)
P2	Which objects are most frequently studied?	C2	Items with concrete instances (systems/collections)	Data repository — object of study (C2) · Thesaurus X (C2)
P3	Distribution of methods and data collection	C7 (+C6, if applicable)	Methods separated from instruments; no C6×C7 overlap	Content analysis (C7) · Semi-structured interview (C6)
P4	Main regulatory/historical contexts	C9	Presence of “— context” qualifiers; no intrusion of C1	Open access — context (C9) · Linked Data — context (C9)
P5	Which products/artifacts were generated?	C10	Reusable products clearly identified	Interview-based data collection protocol (C10) · Metadata profile (C10)
S1	Research on service management (theme) addressing automation (scope)	C1 + C3	Domain-specific C1 + explicit objective/intent in C3	Library services management (C1) · Process automation (C3)
S2	Research on indexing (theme) using verbal protocol (method)	C1 + C7	C1 focused on indexing; C7 with “verbal protocol” procedure	Indexing — phenomenon (C1) · Verbal protocol (C7)
S3	Research using a statistical method and delivering a metadata standard	C7 + C10	Quantitative method present; product is a profile/standard (not only discussion)	Mathematical and statistical method (C7) · Metadata profile (C10)
S4	Research on information retrieval (theme) studying a specific IR system (object)	C1 + C2	Research on information retrieval (theme) + concrete repository instance	Information retrieval (C1) Institutional repository — object (C2)
S5	Research with interviews (collection) and case study (method), with an attached protocol (product)	C6 + C7 (+C10)	C6 and C7 present; C10 when an artifact exists	Semi-structured interview (C6) · Case study (C7) · Data collection protocol (C10)

Source: Authors’ own research data (2026). Note: QR, question to be answered; CAFTE, core facet categories.

within each facet. In such cases, conceptual splitting was performed to avoid level mixing. Table 3 presents representative before/after examples of these curation decisions. The interview example in Table 3 focuses specifically on the distinction between C6 (Data Collection) and C10 (Research Product). C7 (Research Method) remains an independent facet and is assigned whenever the methodological characteristics of the study require it. Thus, Table 3 illustrates representative curation cases rather than the complete set of possible facet combinations.

The interview example in Table 3 focuses specifically on the distinction between C6 (Data Collection) and C10 (Research Product). C7 (Research Method) remains an independent facet and is assigned whenever the methodological characteristics of the study require it. Thus, Table 3 illustrates representative curation cases rather than the complete set of possible facet combinations. For each candidate term, three simple rules guided curation: (1) a univocity test within the facet (“does the label always describe the same thing within this category?”; if not, qualify the label or relocate it to another facet); (2) drafting a short scope note to clarify the intended use (e.g., in C6 Data Collection, Interview as an instrument, not as a product/deliverable); and (3)

confrontation with terminological authority, checking synonyms and variants to adopt the Brazilian Institute of Science and Technology Information (IBICT) Thesaurus preferred term when applicable. When the domain required a local term, the local label was retained as preferred and mapped to the closest authority term for future export.

In cases of redundancy—when labels added no semantic value or duplicated existing terms—terms were marked as obsolete (deprecated), with the decision recorded in a historical note and a reference term indicated. The “obsolete/deprecated” marker is internal to the Master Term List (MTL) and does not correspond to a BPM-lite workflow state. URIs followed a facet-based pattern, ensuring stability and enabling resolution of deprecated terms via redirection.

As a result, an MTL was consolidated containing, per entry: preferred label, facet, facet-generated URI, status (valid, obsolete, under review), scope note, “see also” references (cross-facet, when applicable), authority source, and date/author of the last decision. When no direct authority equivalent existed, the local label was retained and equivalences were recorded for future export.

**Table 3. Before/After curation (labels, facet, decision).**

Case	Previous situation (Access)	Identified issue	Curatorial decision (After)	Scope note	Impact
1	Repository (C1 Theme and C2 Object)	Homograph between Theme and Object	Split into Repository and digital library [theme] and Data repository [object of study].	C1: repositories as a phenomenon or concept; C2: the system or concrete instance under study.	Improved explainability in filters; C1×C2 combinations no longer collide.
2	Interview (C6 Data collection and C10 Product)	Potential C6/C10 overlap	Split into Semi-structured interview → C6 (Instrument/Data collection); Interview data collection guide → C10 (Product).	C6: data collection instrument; C10: data collection guide.	Reduced ambiguity in QR; “Why did this item appear?” becomes clear.
3	Indexing (C1 Theme; C7 Methods; C10 Product)	Potential C1/C7/C10 overlap	Split into Indexing [phenomenon] → C1; Indexing [method] → C7; Criteria for automatic indexing [product] → C10 (artifact/guide).	C1: indexing as a phenomenon; C7: indexing techniques or procedures; C10: indexing criteria (deliverables).	Reduces ambiguity in QR and enhances explainability in filters; separates macro-level from operational content.

Source: Authors’ own research data (2026). Note: QR, question to be answered.

As a structural outcome of this phase, a concise CAFTE guide per category was consolidated, including operational rules (“what it is”/“what it is not”), examples, and counterexamples. This guide oriented recurring decisions and served as a reference for the next phase. All curatorial decisions were logged (who, what, why, when, and based on which authority), including version-to-version changes, thereby ensuring provenance and auditability. At the end of Phase 2, the controlled vocabulary was normalized and disambiguated by facet, identified by URIs, and ready for Phase 3 (operation of the BPM-lite workflow) and for QR chaining in subsequent phases.

#### 4.3 Results of Phase 3: Operation of the Workflow with BPM-lite

With curation completed and the web prototype updated with the revised Master Term List (MTL), operation of the short cycle began, with each transition recorded in editorial logs. This phase focused primarily on faceted indexing of the 41 T&D, based on the revised TAFNAVEGA (CAFTE C1–C10). The short cycle (Proposed → Under review → Approved → Published) was applied to item annotations, ensuring a minimal trace and inter-facet consistency.

The minimal roles activated in the workflow were:

I. Proposer (Indexer): performs facet annotations (chips C1–C10) according to CAFTE and scope notes.

II. Curator (Indexing reviewer): checks adherence, with special attention to critical pairs (C1×C2 and C6×C7×C10), applies “see also” references, and requests adjustments when necessary.

III. Approver (Responsible editor): validates inter-facet coherence and compatibility with the QR.

IV. IT/Maintainer: publishes the approved annotation, preserves IDs/URIs, and records the decision.

Proposals for new terms were exceptional, as prior vocabulary reformulation reduced the need for expansion during this cycle. When they occurred, they followed the same workflow with a specific decision log, without interrupting indexing cadence.

Logs (who, what, when, and under which rule/authority) revealed expected decision patterns: (i) inter-facet qualification (e.g., “Repository [theme]” vs. “Repository [object of study]”); (ii) separation among C6/C7/C10 (instrument × method × product); and (iii) use of cross-references to guide indexing and search. The MTL thus functioned as a stability contract: punctual annotation adjustments generated concise change logs (from/to) without altering URIs.

From the user perspective, explainability improved: displayed chips more accurately reflected faceted intent (e.g., “Indexing [method]”; “Data collection guide [product]”), and the justification “Why did this item appear?” became transparent in QR scenarios. Table 4 presents evidence of the lightweight governance cycle in operation.

Overall, BPM-lite operation consolidated a short, predictable, and auditable cycle in which each indexing decision leaves a minimal trace (decision/change logs) while preserving ID/URI stability. As illustrated in Fig. 2, the evidence shows recurring patterns (C1×C2; C6×C7×C10), the effectiveness of MTL cross-references, and the usefulness of coordination artifacts. For users, explainability advanced through faceted chips and scenarios that explicitly

**Table 4. Evidence from the operation of the workflow operation (with BPM-lite).**

Aspect observed	Evidence (where)	Example (ID/description)	Perceived impact	Next action (if applicable)
Application of CAFTE in indexing	Indexing spreadsheet + record screenshot	IDX-012: record with C1, C2, C3...C10 filled	Consistency across records; reduction of gaps in facet registration	Sample double-check round (5%)
Separation of instrument × method × product (C6/C7/C10)	Decision log + T&D record	LG-031: “Semi-structured interview” in C6; “Interview guide” in C10	Explainability in QR (“how data were collected?” vs. “what can be reused?”)	Require a short scope note when C10 is present
Theme × Object (C1/C2)	Decision log + MTL (cross-references)	LG-023: “Repository [theme]” vs. “Institutional repository [object of study]”	C1×C2 filters no longer collide; improved precision	Add cross “see also” references in the LMT
Use of scope notes per facet	MTL (scope note field)	C7 “Content analysis” – scope delimits technique, not domain	Less noise in combinations; guidance for indexers	Standardize scope note template (1–2 lines)
Publication with stable URI per facet	Change log + URI policy	TAFNAVEGA/C7/content-analysis (timestamped versions)	Technical traceability; references remain resolvable	Document versioning policy
Visible handoffs in the short workflow	Kanban/worksheet snapshot	Columns: Proposed → Under review → Approved → Published	Less rework; clarity of queue and responsibilities	Automate notification upon state change

Source: Authors’ own research data (2026). Note: BPM-lite, lightweight Business Process Management approach; CAFTE, core facet categories; IDX, indexing identifier; T&D, theses and dissertations; LG, log; MTL, Master Term List.

answer “why did this item appear?”. With this arrangement stabilized, Phase 4 proceeds to chain QR to facets for qualitative adequacy checks and fine-tuning of scope notes and qualifiers.

To make the curatorial operation more concrete, Table 5 presents real examples of theses and dissertations indexed with the original taxonomy and subsequently re-indexed using the curated TAFNAVEGA. The examples illustrate recurrent ambiguities identified during indexing and the corresponding facet-based resolutions.

The examples can be analytically grouped into three types of ambiguity: (i) functional ambiguity, (ii) procedural vs. artefactual ambiguity, and (iii) semantic role collision. The first example (indexing) illustrates a recurrent problem in knowledge organization: the conflation of conceptual roles within a single label. In the original taxonomy, “indexing” simultaneously referred to the phenomenon under study, the method applied, and the resulting artefact. The curated version separates these roles across facets (C1, C7, C10), aligning with the analytic-synthetic principle of facet analysis (Ranganathan, 1958) and with Szostak’s (2013) argument for distinguishing methods as an independent dimension of classification. This separation enables the system to explain retrieval outcomes in terms of clearly defined conceptual functions.

The second example (interviews) highlights the distinction between data collection procedures and their material outputs. The original use of “interview” as a generic descriptor obscured whether the reference concerned the act

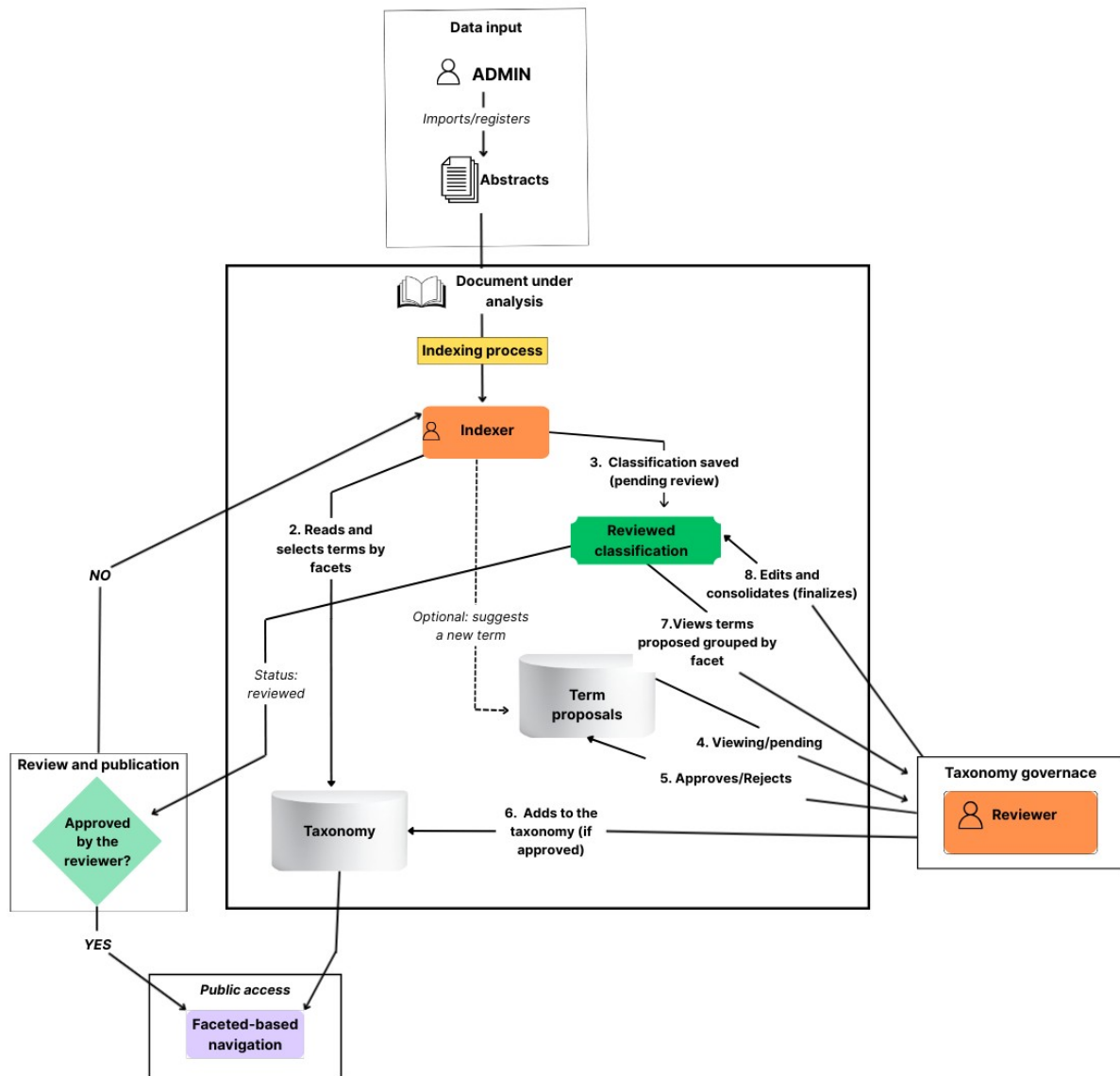
of data collection or the reusable instrument. By separating *interview [data collection]* (C6) from *interview guide [product]* (C10), the curated taxonomy makes explicit the operational role of each element. This distinction is essential for retrieval scenarios focused either on methodological replication or on access to research artefacts.

The third example (digital repositories) demonstrates the importance of distinguishing analytical focus, empirical object, and contextual framing. In the original taxonomy, “digital repository” was used across these dimensions without differentiation. The curated structure distributes these meanings across C1 (phenomenon), C2 (object), and C9 (context), which is consistent with the CRG’s emphasis on role-based categorization (Wilson, 1972). This separation prevents semantic overlap and supports more precise and explainable retrieval.

Across the three cases, the curatorial intervention does not merely refine terminology but restructures the semantic roles of descriptors. Explainability emerges from this restructuring: users can identify which dimension of a subject is being activated and why a given item has been retrieved. In this sense, facet differentiation operates as an epistemic mechanism that makes retrieval decisions transparent and inspectable.

#### 4.4 Results of Phase 4: QR Chaining and Explainability Tests

With the MTL stabilized and BPM-lite in operation, each QR was chained to its corresponding CAFTE combi-



**Fig. 2. Conceptual BPM-lite workflow for faceted curation of TAFNAVEGA in thesis and dissertation repositories.** Source: Authors' own research data (2026). Note: BPM-lite, lightweight Business Process Management approach.

nations and faceted search was exercised in the web prototype. Preference was given to validated MTL terms; when necessary, new-term proposals were logged for later workflow treatment without interrupting tests.

For each QR, the following were defined: (i) semantic intent; (ii) minimal facet combinations (C1–C10); and (iii) any needed adjustments to scope notes, qualifiers, or “see also” references. Queries were then executed and documented through vignettes, including applied filters, the explanation “Why did this item appear?” (via facet chips), and any curatorial adjustments triggered.

Vignette analysis revealed three recurring patterns. First, the distinction among instrument × method × product (C6/C7/C10): QR framed as “How were data collected?” consistently led to C6; “How were data analyzed?” to C7; and “Is there reusable material?” to C10. Residual generic

labels were resolved by scope reinforcement (e.g., interview vs. interview guide).

Second, separation between theme × object (C1/C2): phenomenon-oriented QR retrieved C1, while those focused on concrete systems or instances retrieved C2. Collisions were resolved via cross-references and refined qualifiers/scope notes.

Third, the distinction between macro-context (C9) and practice/implementation (C1/C7): panorama/regulatory QR retrieved C9 without contaminating results with technical studies (C1/C7). Qualifier adjustments (e.g., “Open access [context]”) proved sufficient to maintain analytical level separation.

At the interface level, vignettes showed clear gains in explainability: users quickly recognized inclusion reasons (readable chips) and refined navigation by adding/removing

**Table 5. Real examples of thesis indexing before and after TAFNAVEGA curation.**

T&D (title – abbreviated)	Before curation (original taxonomy)	After curation (revised TAFNAVEGA)	Faceted path (C1–C10)	Effect on explainability
Automatic indexing in digital repositories	“Automatic indexing” used as a single term (combining process, method, and product)	Separation into: Indexing [phenomenon] (C1), Automatic indexing criteria [product] (C10), and Indexing method (C7)	C1 (Indexing) + C7 (Method of indexing) + C10 (Criteria for automatic indexing)	Clarifies whether retrieval refers to the phenomenon, the procedure, or the deliverable
The use of interviews in user studies	“Interview” used interchangeably as a general technique	Differentiation into: structured interview collection] (C6) and Interview guide [product] (C10)	C1 (User studies) + C6 (Interview) + C10 (Interview protocol)	Enables distinction between how data were collected and what can be reused
Digital repositories in the context of open science	“Digital repository” appears as a theme, subject, and context without distinction	Disambiguation into: repository [phenomenon] (C1), Repository system [object] (C2), and Open science [context] (C9)	C1 (Repository – phenomenon) + C2 (Repository system) + C9 (Open science context)	Separates analytical focus from empirical object and contextual framing

Source: Authors’ own research data (2026). Note: T&D, thesis and dissertation.

facets. The QR→facet matrix supported fine scope-note revisions. Whenever noise appeared, Phase 3 was re-entered for minimal corrections or new-term proposals for future evaluation. Each QR ultimately mapped to 1–3 stable, documented, and replicable faceted paths, confirming term-set sufficiency for the analyzed corpus and indicating few candidates for controlled future extension.

#### 4.5 Results of Phase 5: Review and Synthesis

This phase assessed whether the executed path: (1) attempted conversion from Access to a web environment; (2) identification of conceptual limitations in the original TAFNAVEGA; (3) authority-based terminological curation; (4) reconfiguration of the web prototype with the revised vocabulary; (5) reindexing of the 41 T&D through the short cycle (Propose → Review → Approve → Publish); and (6) QR fulfillment testing via faceted combinations—produced an arrangement coherent with the theoretical-methodological framework (Section 2).

Five main points stand out. First, web conversion revealed that the original TAFNAVEGA allowed facet overlaps—especially C1×C2 and C6×C7×C10—directly compromising interface explainability. This was interpreted as a conceptual modeling issue rather than a technological one: loosely qualified labels tolerable in a closed Access environment become problematic with visible filters.

Second, curation directly addressed this bottleneck. Label normalization, facet-specific individualization, scope notes, and URI assignment were necessary conditions for web operation and end-to-end explainability.

Third, with the curated vocabulary, full web conversion and reindexing of the 41 T&D with stabilized C1–C10 became possible. This confirmed an operational hypothe-

sis: after an intensive initial curation round, the short indexing workflow tends to receive few new-term proposals, concentrating effort upfront rather than in daily operation.

Fourth, QR testing verified: (i) whether panorama QR yield clean facet distributions; (ii) whether specific QR (e.g., theme + method + product) produce clear user explanations; and (iii) whether any result still requires scope-note or cross-reference adjustment. Only minor refinements were needed, indicating curatorial sufficiency for the analyzed corpus.

Fifth, validation enabled assembling an institutionalizable artifact set: (a) revised TAFNAVEGA with facet-qualified labels; (b) CAFTE checklists per category; (c) decision/change log model aligned with the short cycle; (d) QR→facet→chip matrix; and (e) an indexing workflow applicable to repository ingestion. This package is the study’s main product, enabling replication in other repositories without conceptual rebuilding.

In sum, Phase 5 closes the cycle where it began: the need to render TAFNAVEGA explicit enough to operate in a web environment with lightweight governance and interface-level explainability.

## 5. Conclusions

The revision of TAFNAVEGA demonstrated that a faceted vocabulary, when subjected to rigorous terminological curation, can operate as an explainable KOS. Migration to a web environment made latent conceptual limitations of the original taxonomy visible, showing that retrieval explainability depends primarily on the semantic stability of labels, consistent assignment by facet, and the explicit delimitation of conceptual scopes through scope notes.

From a methodological perspective, the study confirmed that a BPM-lite workflow is sufficient to record curatorial decisions in a traceable manner without introducing additional layers of bureaucracy. The combination of min-

imal institutional roles, short states, and a four-step cycle (Propose → Review → Approve → Publish) showed that the logic of “few rules, but every rule recorded” is a viable strategy for operating terminological control with light governance. In this arrangement, each inclusion or adjustment of a term remains associated with who decided, when the decision was made, on what basis, and for which facet, providing the provenance required to sustain auditability and explainability.

The articulation between retrieval questions and CAFTE combinations further evidenced that faceted indexing of content—rather than indexing limited to descriptive metadata—supports real usage paths by clearly combining theme, method, scope, and product for the user. Making these paths explicit reinforces the mediating role of the KOS and helps reduce the opacity typical of searches based exclusively on keywords.

At the applied level, the results indicate that the model is operationally viable for Brazilian academic repositories, particularly when these are understood as socio-technical infrastructures rather than isolated software solutions. The incorporation of roles, minimal logs, lightweight versioning, and controlled terminology proved compatible with contexts of limited resources, in line with national studies on repositories. Some limitations remain—such as the restricted corpus, partial simulation of the approval routine for new terms, and the absence of quantitative comparison with text-based searches—but these are consistent with the adopted socio-technical design and the qualitative, design-oriented nature of the research.

From a socio-technical standpoint, the most relevant outcome lies in treating the repository as institutional infrastructure rather than merely as software: roles are defined, decisions are recorded, controlled terminology is in place, and a lightweight versioning policy is adopted. This configuration proved operable with reduced resources, aligning with findings from Brazilian studies on repositories.

As a future development, the same short workflow is proposed for the deposit and ingestion processes of theses and dissertations, ensuring symmetry between vocabulary curation and record curation. Overall, the revised TAFNAVEGA, operated through BPM-lite, consolidates itself as a transferable operational KOS capable of balancing terminological control and institutional lightness, sustaining explainability, provenance, and controlled evolution, in direct alignment with the objective stated in the introduction.

### **Availability of Data and Materials**

The dataset supporting the conclusions of this study is publicly available in Zenodo at: <https://zenodo.org/records/19497457>; DOI 10.5281/zenodo.19497457. The dataset includes the conceptual structure of the TAFNAVEGA taxonomy, its iterative development, empirical materials derived from a prototype implementation, scripts supporting

data structuring, and a README file describing the contents of the deposited materials.

### **Author Contributions**

BCMSM: Project administration; Supervision; Conceptualization; Methodology; Data curation; Formal analysis; Writing-original draft; Writing-review & editing. GÂL: Conceptualization; Methodology; Writing-review & editing. ECA: Conceptualization; Methodology; Formal analysis; Validation; Writing-review & editing. FCMP: Conceptualization; Methodology; Formal analysis; Validation; Data curation; Writing-review & editing. FBM: Software; Data curation; Validation; Visualization; Writing-review & editing. YBM: Software; Data curation; Validation; Writing-review & editing. All authors read and approved the final manuscript. All authors participated sufficiently in the work and agreed to be accountable for all aspects of the study.

### **Acknowledgment**

The authors would like to acknowledge the institutional support provided by the Brazilian National Council for Scientific and Technological Development (CNPq), which enabled the development of this research.

### **Funding**

This research was partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) through Productivity Fellowships held by three of the authors (grant numbers 307765/2023-7, 313645/2020-5, and 306713/2025-0).

### **Conflicts of Interest**

The authors declare no conflicts of interest. Benildes Coura Moreira dos Santos Maculan, Gercina Ângela de Lima, and Elisângela Cristina Aganette are serving as one of the Guest editors of this journal. We declare that Benildes Coura Moreira dos Santos Maculan, Gercina Ângela de Lima, and Elisângela Cristina Aganette had no involvement in the peer review of this article and have no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to Natália Tognoli.

### **Declaration of AI and AI-Assisted Technologies in the Writing Process**

During the preparation of this work, the authors used AI-assisted tools (including large language models) to assist in the organization of data, structuring of tables, and refinement of textual expression, including improvements to the English language, as the authors' native language is Portuguese. The conceptual foundations of this study derive from prior research conducted in 2011, developed independently of AI-based tools and prior to their widespread use in

research contexts. The intellectual design of the study, including the conceptualization of the TAFNAVEGA model, the definition of categories, the indexing processes, and the analytical interpretation within a qualitative socio-technical design approach, was entirely carried out by the authors. AI tools were not used to generate the conceptual model, define categories, or conduct the analytical interpretation. All outputs generated with the support of AI tools were critically reviewed, validated, and, when necessary, revised by the authors, who take full responsibility for the content of the manuscript.

## References

- Amoah A, Asiana RK, Kwablah E. Chatgpt early usage among students: a global evidence of determinants. *Development and Sustainability in Economics and Finance*. 2025; 7: 100065. <https://doi.org/10.1016/j.dsef.2025.100065>
- Ávila BT, Silva M, Cavalcante L. Uso de Repositórios Digitais como Fonte de Informação por Membros das Universidades Federais Brasileiras. *Informação & Sociedade: Estudos*. 2017; 27: 97–120. <https://doi.org/10.22478/ufpb.1809-4783.2017v27n3.31514> (In Portuguese)
- Broughton V. Brian Vickery and the Classification Research Group: the legacy of faceted classification. In Gilchrist A, Vernau J (eds.) *Facets of Knowledge Organization: Proceedings of the Second ISKO UK Conference, London, 4–5 July 2011*. Emerald Group Publishing: London. 2012.
- Broughton VD. Faceted classification as the basis of all information retrieval. In Slavic A, Gnoli C (eds.) *Faceted classification today: theory, technology and end users: proceedings of the International UDC Seminar, 14–15 September 2017, London, United Kingdom* (pp. 149–162). Ergon Verlag: London, UK. 2017.
- Custódio NC, Vechiato FL. Encontrabilidade da informação em repositórios institucionais: uma proposta de instrumento de avaliação. *Revista Informação Na Sociedade Contemporânea*. 2017; 1: 1–17. <https://doi.org/10.21680/2447-0198.2017v1n0id12284>
- Deschenes A, McMahon M. A Survey on Student Use of Generative AI Chatbots for Academic Research. *Evidence Based Library and Information Practice*. 2024; 19: 2–22. <https://doi.org/10.18438/ebliip30512>
- Fujita MSL, Rubi MP. Um modelo de leitura documentária para a indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores. *DataGramaZero*. 2006; 7. <https://cip.brapci.inf.br/download/44949> (Accessed: 13 October 2025)(In Portuguese)
- Gnoli C. Classifying phenomena. Part 4: Themes and rhemes. *Knowledge Organization*. 2018; 45: 43–53.
- Gregoriades A, Sutcliffe A. A socio-technical approach to business process simulation. *Decision Support Systems*. 2008; 45: 1017–1030. <https://doi.org/10.1016/j.dss.2008.04.003>
- Hevner AR, March ST, Park J, Ram S. Design science in information systems research I. *MIS Quarterly*. 2004; 28: 75–106. <https://doi.org/10.2307/25148625>
- Hoda R. Socio-Technical Grounded Theory for Software Engineering. *IEEE Transactions on Software Engineering*. 2021; 48: 3808–3832. <https://doi.org/10.1109/tse.2021.3106280>
- Joseph K, Chen HYW, Ionescu S, Du Y, Sankhe P, Hannak A, et al. A qualitative, network-centric method for modeling socio-technical systems, with applications to evaluating interventions on social media platforms to increase social equality. *Applied Network Science*. 2022; 7: 49. <https://doi.org/10.1007/s41109-022-00486-8>
- Lake S, Regenauer S. Growing an Institutional Repository: Leveraging a Citation Database as a Tool for Sourcing Deposits and Conducting Outreach. *Library Resources & Technical Services*. 2024; 68. <https://doi.org/10.5860/lrts.68n1.8217>
- Maculan BCM. Taxonomia facetada navegacional: construção a partir de uma matriz categorial para trabalhos acadêmicos [Master's thesis]. Universidade Federal de Minas Gerais. 2011. (In Portuguese)
- Marcelino CS, Siebra SA. Análise da interação de usuários com repositórios institucionais de instituições federais de ensino superior brasileiras. *GESTÃO. Org*. 2015; 13: 300–309. (In Portuguese)
- Matos ML. Relatório técnico de proposta de reestruturação da infraestrutura da Universidade de Brasília baseada no conceito de smart campus [Master's thesis]. Universidade de Brasília: Brasília (DF). 2021. (In Portuguese)
- Matta AER, Silva FPS, Boaventura EM. Design-based research ou pesquisa de desenvolvimento: pesquisa aplicada para educação a distância. In: *Anais do 21º Congresso Internacional ABED de Educação a Distância; 2015; Bento Gonçalves (RS)*. São Paulo: ABED; 2015. p. 1–10. Available at: [https://www.abed.org.br/congresso2015/anais/pdf/BD\\_313.pdf](https://www.abed.org.br/congresso2015/anais/pdf/BD_313.pdf) (Accessed: 23 January 2024).
- Medeiros JW. Taxonomia navegacional facetada: análise à luz dos princípios teóricos da classificação facetada [master's dissertation]. Florianópolis: Universidade Federal de Santa Catarina; 2013. (In Portuguese)
- Microsoft. Designing the infrastructure persistence layer: .NET microservices (DDD/CQRS). 2023. Available at: <https://learn.microsoft.com/en-us/dotnet/architecture/microservices/microservice-ddd-cqrs-patterns/infrastructure-persistence-layer-design> (Accessed: 23 March 2024).
- Microsoft. Implementing the infrastructure persistence layer with EF Core. 2025. Available at: <https://learn.microsoft.com/en-us/dotnet/architecture/microservices/microservice-ddd-cqrs-patterns/infrastructure-persistence-layer-implementation-entity-framework-core> (Accessed: 23 January 2024).
- Miörner J, Truffer B, Binz C, Heiberg J, Yap XS. Guidebook for applying the socio-technical configuration analysis method. *GEIST Working Paper Series*. 2022; 1: 1–19. Available from: [https://www.geist-wp.com/papers/geist\\_wp\\_2201.pdf](https://www.geist-wp.com/papers/geist_wp_2201.pdf) (Accessed: 12 July 2025)
- Nederhand J, Klijin EH, van der Steen M, van Twist M. The governance of self-organization: Which governance strategy do

- policy officials and citizens prefer? *Policy Sciences*. 2019; 52: 233–253. <https://doi.org/10.1007/s11077-018-9342-4>
- Oddone N, Gomes MYFSF. Uma nova taxonomia para a ciência da informação. In: *Anais do 5º Encontro Nacional de Pesquisa em Ciência da Informação*. Belo Horizonte. Belo Horizonte: ECI/UFGM. 2003. (In Portuguese)
- Pestana O. Organização da informação e do conhecimento em contexto empresarial: a aplicação da análise por facetas. In: Simões MG, Borges MM, editors. *Tendências atuais e perspectivas futuras em organização do conhecimento: atas do III Congresso ISKO Espanha-Portugal, XIII Congresso ISKO Espanha*; 2017 Nov 23–24; Universidade de Coimbra. Coimbra: Centro de Estudos Interdisciplinares do Século XX (CEIS20), Universidade de Coimbra; 2017. Available from: <https://repositorio-aberto.up.pt/handle/10216/109913> (Accessed: 25 August 2025) (In Portuguese)
- Pinheiro LVR, Ferrez HD. Tesauro brasileiro de ciência da informação. 2014. Available at: <https://www.gov.br/ibict/p t-br/central-de-conteudos/publicacoes/tesauro-brasileiro-d e-ciencia-da-informacao> (Accessed: 16 October 2025) (In Portuguese)
- Ramos C, Rockembach M, Jerónimo P. Usabilidade dos Repositórios Institucionais de Universidades e Politécnicos: a perspectiva dos Professores e Investigadores em Portugal. *Perspectivas Em Ciência Da Informação*. 2022; 27: 68–88. <https://doi.org/10.1590/1981-5344/39979> (In Portuguese)
- Ranganathan SR. *Colon classification: basic classification*. 6th edn. Asia Publishing House: Bombay. 1963.
- Ranganathan SR. Postulational approach to faceted classification. *Annals of Library Science*. 1958; 5: 33–51.
- Ranganathan SR. *Prolegomena to Library Classification*. 3rd ed. The Library Association: London. 1967.
- Santana LD, Pereira FCM, Mattos MCD. Design Science Research na prática: ideação de um repositório institucional baseado em periódico científico. *RDBCI Revista Digital De Biblioteconomia E Ciência Da Informação*. 2023; 21: e023011. <https://doi.org/10.20396/rdbci.v21i00.8672896> (In Portuguese)
- Santana LD, Silva RAPPG, Pereira FFCM, Maculan BCMDS. Rigor e relevância na Ciência da Informação: pesquisa-ação, Design Science Research e Action Design Research. *Informação & Informação*. 2024; 28: 273–297. <https://doi.org/10.5433/1981-8920.2023v28n2p273> (In Portuguese)
- Santos HMD, Flores D. Infraestrutura de segurança para gestão de riscos do Repositório Arquivístico Digital Confiável: um diálogo com a ISO 16363. *Revista Brasileira de Biblioteconomia e Documentação*. 2020; 59–83. <https://doi.org/10.21747/16463153/41a5> (In Portuguese)
- Sganzerla PRP. *Desenvolvimento de projetos de investimento em infraestrutura: um framework orientado à construção de indicadores* [Master's thesis]. Brasília (DF): ENAP. 2024. (In Portuguese)
- Silva FP. *Gestão de dados em organizações: uma análise da aplicação da preservação e curadoria digital no corpo de conhecimento do DMBOK* [Master's thesis]. Porto Alegre: Universidade Federal do Rio Grande do Sul. 2023. (In Portuguese)
- Silvestre ASS, de Moura Amaral EL, Holanda ME, Canedo ED. Students' perception about ChatGPT's impact on their academic education. In: *Proceedings of the 34th Brazilian Symposium on Computers in Education (SBIE)* (pp. 1260–1270). SBC. 2023. Available at: <https://sol.sbc.org.br/index.php/sbie/article/download/26752/26571/> (Accessed: 12 August 2025)
- Szostak R. Classification in terms of basic concepts. *Advances in Classification Research Online*. 2013; 23: 33–34.
- Weinberg BH. Why indexing fails the researcher. *The Indexer*. 1988; 16: 3-6.
- Wilson TD. The work of the British Classification Research Group. In Wellisch H, Wilson TD (eds.) *Subject retrieval in the seventies* (pp. 62–71). Greenwood: Westport (CN). 1972.