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

Methodological systematization and conceptual stability: the Glossary of Brazilian Educational Censuses

Samuel Santos da Rosa

Federal University of Rio Grande do Sul, Brazil.

Rita do Carmo Ferreira Laipelt

Federal University of Rio Grande do Sul, Brazil.

Guilherme Veiga Rios

National Institute for Educational Studies and Research Anísio Teixeira, Brazil

Geovana Andreia Santos Soares

National Institute for Educational Studies and Research Anísio Teixeira, Brazil

Sonia Elisa Caregnato

Federal University of Rio Grande do Sul, Brazil.

Abstract

The study aims to describe the methodological approach adopted for the development of the Glossary of Brazilian Educational Censuses, with a view to standardizing the concepts and terms used in the production and dissemination of educational data. It analyses the role of terminology and Knowledge Organisation Systems (KOS) in improving information management and retrieval within the context of Brazilian education. The work was structured in four stages: Mapping and selection of terms; Verification of terminological compatibility with other KOS; Construction of a domain tree and conceptual diagram; and Modelling and harmonisation of definitions based on ISO 704:2022, ISO 860:2007 and the Grid Methodology. The results highlight the relevance of intentional terminological definitions for conceptual stability and semantic interoperability. The findings indicate that the glossary represents a strategic tool for strengthening terminological consistency and for supporting the future development of thesauri and ontologies applied to educational statistics.

1.0 Introduction

The management of the growing volume of information requires the use of processes and instruments grounded on Terminology and in technically standardized Knowledge Organisation Systems (KOS) that are conceptually and semantically structured. Aligning information modelling with the specific terminologies of each field enables more accurate retrieval in digital environments. Such terminological rigour allows Information Retrieval

Systems (IRS) to operate with high levels of precision and recall, thereby ensuring the integrity and accessibility of data and information that are organised, indexed, retrieved and shared.

Given the centrality of the concept in terminological work, as explained by Dahlberg (1978), it is important to recognize that concepts and terms evolve differently across languages and communities, influenced by historical, geographical and socio-economic factors. In this context, the standardization and modelling of terminological work aimed at the development and updating of KOS should treat conceptual systems as structured semantic tools that integrate three fundamental elements: terms, definitions and relationships between concepts.

The convergence of these elements enables the reduction of ambiguities, the management of synonyms, and the establishment of semantic relationships between concepts that define a given domain. Terminological definitions clarify the semantic boundaries of concepts and delineate their distinctive features, functioning as structuring tools that enhance precision in information representation. Consequently, such definitions are central to the effective functioning of KOS, such as thesauri, glossaries and ontologies, by promoting the organization and clear communication of scientific and technical knowledge across communities through specialized language.

In Brazil, the Anísio Teixeira National Institute for Educational Studies and Research (Inep) is the government agency responsible for producing studies and research on basic and higher education. Through the implementation and dissemination of educational censuses (School Census and Higher Education Census), Inep provides robust data that support the formulation of public policies, establishing itself as the country's leading institution in educational statistics. In this context, ensuring the effective dissemination, retrieval, and sharing of this extensive dataset requires the implementation of rigorous terminological control. Accordingly, the partnership project involving Inep, the National Education and Research Network (RNP) and the Federal University of Rio Grande do Sul (UFRGS) includes, among its key initiatives, the development of a specialized glossary, with the aim of standardizing census terminology.

From this perspective, this study builds upon and extends the work presented at the 8th Brazilian Congress on Knowledge Organization. Its objective is to describe the methodological approach adopted in compiling the Glossary of Brazilian Educational Censuses, highlighting how the quality and consistency of the definitions directly impact the glossary's effectiveness as a reference tool, a learning resource, and a research aid.

2.0 Glossaries and terminology definitions

In the field of Information Science (IS), glossaries are recognized as structured tools for terminological organization, essential for the organisation and mediation of specialized knowledge. They can be defined as compilations of terms or expressions from a specific domain, usually developed by experts and organized as alphabetically ordered lists for use within a community (Lopes-Nascimento et al. 2024, 3).

Thus, glossaries can be understood both as documentary products and as technical instruments for knowledge organization. This dual characterization highlights their referential, selective, and specialized nature, distinguishing them from general dictionaries by their conceptual and thematic focus, while also conferring methodological and communicative versatility. As KOS, glossaries are not limited to the mere listing of terms; rather, they operate as conceptual structures that make explicit the epistemological boundaries, theoretical choices, and semantic limits of a given domain (Maculan et al. 2023).

Defining is an interpretative and constructive act through which certain aspects of an object are selected according to the 'perspective' of an individual or a discursive community that shares criteria of conceptual relevance and communicative purposes. Krieger and Finatto (2004) emphasize that the act of defining is not neutral, but socially situated, reflecting provisional consensuses constructed within a specialised domain. This approach supports the creation of conceptual subdivisions that express hierarchical, inclusion, and differentiation relationships, thus structuring knowledge in a systematic way. This perspective aligns with Dahlberg's proposal (1978), grounded in concept theory and rooted in the Aristotelian tradition, according to which a definition constitutes a semantic equivalence between the *definiendum* (the symbol to be defined) and the *definiens* (how it is to be defined), intended to precisely circumscribe the understanding of the defined term. A definition, therefore, not only names but also circumscribes the scope of a concept, making explicit the essential characteristics that constitute its meaning.

According to Krieger and Finatto (2004), there are three complementary types of definition in glossaries, employed for different communicative purposes: a) lexicographical; b) encyclopaedic; and c) terminological. Among these, terminological definitions stand out, as they establish clear conceptual boundaries and provide the foundation for formal knowledge modelling. Considered the structural core of specialized glossaries, this type of definition is characterized by the formula 'broad category' (the general category or class to which the defined entity belongs) and 'specific distinction' (an indication of the particularity that distinguishes this entity from others in the same class) (Krieger and Finatto 2004). This formula aligns with the construction of intensional definitions, as it allows the recognition of a concept and its differentiation from others, particularly coordinated concepts.

In practice, intensional definitions are preferable to other types and should be used whenever possible, as they more clearly reveal the characteristics of a concept within a conceptual system. This is reinforced by its definition that states: "a definition that conveys the intension of a concept by stating the immediate generic concept and the delimiting characteristic(s)" (ISO 1087 2019, 7). This definition structure aims to provide concise and consistent information about the concept to which it refers, highlighting it as a unit with an unambiguous intension that reflects a corresponding extension. It thus enables the placement of the concept in a clear and consistent position within a conceptual system (ISO 704 2022; Löckinger et al. 2015; Barité 2025).

Therefore, in order to construct a consistent KOS, such as the glossary used in educational censuses, it is necessary to establish a methodological approach oriented toward

compiling definitions with layered semantic structures. This approach integrates terms, definitions, and relationships between concepts, thereby optimizing precision and recall in IRS systems.

3.0 Methodological systematization and conceptual stability

Inep develops the Brazilian Education Thesaurus (Brased) and uses it, among other functions, as the subject headings system for its Library catalog. It is intended to bring together related terms and concepts based on a predefined conceptual structure for the educational domain. It is currently hosted on a system specifically developed for thesauri, while maintaining its connection with the library's collections.

Brased does not explicitly include terms of a statistical nature; as a result, the definitions used in the censuses are not articulated within a structured conceptual system, and there is no explicit specification or delimitation of relationships between terms. This situation tends to lead to recurring problems of cross-classification and inconsistencies in the organization of information. Therefore, the development of a glossary for Brazilian educational censuses is grounded in the institutional need to ensure terminological coherence and consistency in the modeling and harmonization of statistical and educational terms and their respective definitions, in order to minimize interpretative ambiguities and consolidate essential definitions.

Accordingly, a methodological approach is required to develop terms and definitions with a high degree of conceptual precision and uniformity, ensuring consistency in the description, systematization, and dissemination of specialized knowledge within the institution. Given the complexity involved in compiling the glossary, the construction process is structured into multiple phases, with the first phase being the focus of this study. The methodological process comprises four stages: (1) mapping and selection of terms; (2) terminological compatibility with other KOS; (3) development of a domain tree and conceptual diagram (conceptual system); and (4) modeling and harmonization of definitions.

3.1 Stage I – Mapping and selection of terms (corpus for analysis)

Methodologically, the compilation of the Glossary began with the mapping and extraction of candidate terms from the Educational Censuses (Basic and Higher Education). To this end, internal documents used by the census teams were consulted, as these contain structured metadata employed in Inep's databases. A systematic screening was conducted to identify terms representative of the censuses, followed by their extraction for analysis and terminological processing. At this stage, the analytical corpus was established, comprising 38 terms (24 preferred terms and 14 non-preferred terms).

3.2 Stage II – Terminological consistency with other KOS (for harmonization)

Once the corpus had been established, the next step involved verifying terminological consistency with other controlled vocabularies, thesauri, and glossaries in the fields of education and statistics.

In the field of statistics, KOS such as the IBGE Glossary were selected. This resource provides structured data (metadata) on the statistical and geoscientific information produced by the institution (IBGE 2025). It is particularly relevant for harmonization with the terms used in the Educational Censuses, as IBGE is the leading institution responsible for producing statistical and geographical data in Brazil.

In the field of education, subject classification systems such as the UNESCO Thesaurus (UNESCO Education Thesaurus) were selected. This thesaurus constitutes a systematically organized list of concepts designed to support thematic analysis, as well as the retrieval of documents and publications in the fields of education, culture, natural sciences, social and human sciences, communication, and information, and is continuously updated and expanded. Its relevance to the harmonization and modeling process lies in its multidisciplinary scope, incorporating guidelines and initiatives developed by UNESCO (2026).

This compatibility stage aims to ensure semantic consistency by incorporating nationally and internationally recognized vocabularies into the modeling and harmonization process within their respective domains. Subsequently, during the drafting of definitions, correspondences (analysis of conceptual structures) and compatibilities (comparison of different concepts) among the conceptual systems of the selected KOS are examined (ISO 860 2007). Accordingly, the 24 preferred terms were compared across these vocabularies to verify their presence, establish semantic links, and extract the corresponding terms and definitions.

3.3 Stage III – Domain Tree and Conceptual Diagram (conceptual system)

Once the compatibility process had been completed, the construction of the Domain Tree (conceptual system) began, using the 24 preferred terms to establish hierarchical semantic relationships among them. Such relationships constitute the core of a conceptual system, as they provide the basis for defining its first element (Lopes-Nascimento et al. 2024).

Domain trees consist of hierarchical representations that graphically organize concepts within a field of knowledge through relationships of subordination, progressing from general categories to increasingly specific levels. This structure makes explicit the conceptual architecture of the domain, enabling the visualization of classes, subclasses, and semantic dependencies among terms. From a methodological perspective, it functions as a preliminary tool for conceptual modeling, supporting the definition of thematic scope, the identification of conceptual boundaries, and the clarification of classification criteria. By positioning each concept in relation to a superordinate level (genus) and distinguishing it through specific characteristics, this approach facilitates the formulation of clear intensional definitions, contributing to terminological coherence and reducing semantic overlaps and gaps (Barité 2017).

The construction of the conceptual framework for the Educational Censuses (Figure 1) was based on the existing Brased domain tree, developed by the Inep Terminology Management Team. This process involved determining the position of each term within the corresponding conceptual structure in order to ensure semantic interoperability between Brased and the Glossary of Educational Censuses.

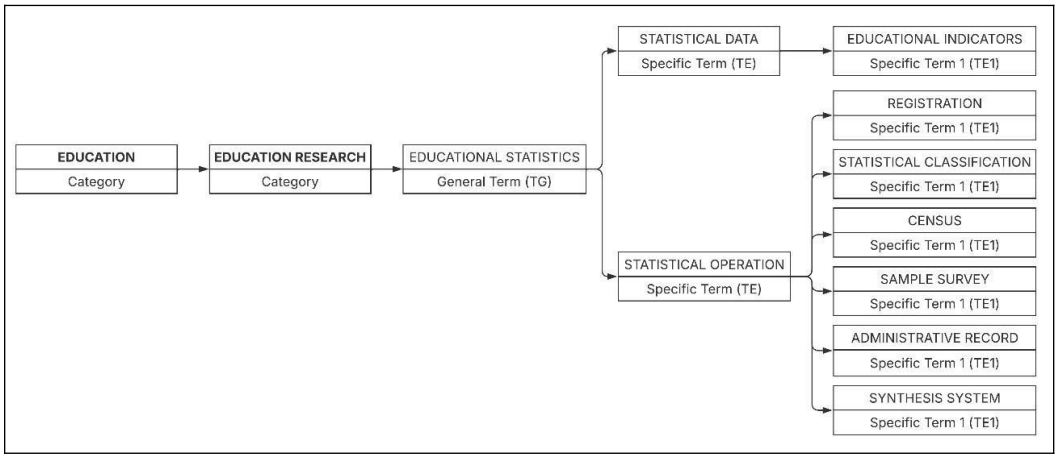

Figure 1. Excerpt from the Domain Tree of the Educational Censuses.

Alongside the domain tree, a conceptual diagram representing the censuses was also drawn up, highlighting the initial associative semantic relationships and enabling their preview. The conceptual diagram, in turn, is a graphical representation of the set of terms in the Censuses that plays a complementary role to the domain tree (showing conceptual interrelationships).

3.4 Stage IV – Modelling and harmonisation of definitions

With the development of the Educational Censuses Domain Tree, it became possible to model and harmonize the 24 preferred terms based on definitions previously extracted from the documents. The terms were modeled and harmonized from the perspective of intensional definitions, using as a normative and theoretical basis the convergence between the guidelines established by ISO 704:2022 and the theoretical framework of the grid method (Barité 2011, 2025).

The guidelines set out in ISO 704:2022 underpin terminological work through a systematic conceptual approach, according to which terms must represent clearly defined concepts within a structured conceptual system. The standard recommends that terminological definitions should primarily be intensional, that is, formulated by specifying the essential characteristics that distinguish a concept from others belonging to the same class (ISO 704 2022).

This approach draws explicitly on the Aristotelian logical tradition of genus and differentia, whereby a definition enables the identification of a concept and its differentiation from other concepts. This same logical foundation is also present in the grid method, which is

likewise based on definition by intension. The method starts from the recognition that drafting definitions involves tensions related to the selection of attributes, precision, and conciseness. For this reason, it proposes a preliminary analytical procedure based on the compilation, segmentation, comparison, and synthesis of pre-existing definitions, whether expressed as statements (essential, accidental, contextual, and informative) or as questions. This decomposition into propositions allows for the isolation of truly distinctive characteristics, bringing terminological practice closer to a systematic and consistent process, in line with the normative requirements of ISO 704 2022 (Barité 2011, 2025).

It can therefore be argued that the grid method empirically operationalizes the theoretical principles of ISO 704:2022. While the standard stipulates that definitions should result from the identification of essential characteristics within an explicit conceptual system, the grid method provides the methodological means to achieve this, by comparatively organizing the attributes of concepts (through statements or questions). The combination of these approaches strengthens the development of standardized terminological definitions that are more coherent, distinctive, and semantically stable. In light of this convergence, the definitions were modeled and harmonized using the following standardized structure (Figure 2):

DEFINITIONAL FORMULA (STANDARDIZED STRUCTURE)

[Essential Statement [General Term + Conceptual Core + Delimiting Characteristics]+ Contextual Statement [Domain]+Accidental Statement [Specific term(s) + Related term(s)]]

Example of term and definition in the educational census glossary

Term: Census

Definition: Type of statistical operation in which all units of analysis in a given population are observed. It involves the collection, recording, compilation, analysis, and dissemination of information at a given point in time or reference period, relating to the Brazilian educational system. It includes the School Census and the Higher Education Census.

Application

STRUCTURE	ELEMENT	EXAMPLE (Term: Census)
Essential Statement	*General Term:* indication of the immediate generic concept (superordinate), which establishes the position of the term within the conceptual system of educational censuses.	Type of statistical operation [...]
	Conceptual Core: establishes the anchor of the definition that highlights whether the term refers, for example, to a principle, a process, a procedure, an instrument, an indicator, an action, a policy, a set of actions/procedures, etc. (represented by a word or phrase that describes the nature of what is being defined).	[...] in which all units of analysis in a given population are observed. [...]
	Delimiting Characteristics: it includes all the essential characteristics of the concept, according to the established structure of the field, as well as non-essential characteristics if necessary for a complete description of the concept.	[...] It involves the collection, recording, compilation, analysis, and dissemination of information at a given point in time or reference period [...]
Contextual Statement	*Domain:* focusing on the domain in which it is inserted, that is, bringing together the dimensions relevant to the specific field.	[...] relating to the Brazilian educational system. [...]
Accidental Statement	*Specific term(s):* indication of concepts subordinate to the respective term, highlighting the branching of the conceptual system (optional element). It assists in structuring the relationships between terms and can serve as a basis for the subsequent development of an ontology.	[...] It includes the School Census and the Higher Education Census.
	Related term(s): concepts related to the respective term in the conceptual system (optional element). It assists in structuring the relationships between terms and can serve as a basis for the subsequent development of an ontology.	Not included in the respective definition as the glossary is still being compiled.

Figure 2. Formula, components and application of the standardized structure for census definitions.

This formula incorporates the essential requirements for constructing intensional definitions, grounded in the principles of ISO 704:2022 and the grid method. It thus enables an understanding of how the conceptual system (domain tree) is structured, highlighting the reflection of the systematic relationships that a concept establishes with other concepts within the same field, as well as its differentiation from similar concepts within the same or across different fields.

The choice of a statement-based approach stems from the role that statements play in the construction of definitions, as it is the statement that describes a notion, process, or

object, constituting a central element in the production and dissemination of specialized, technological, or scientific knowledge, insofar as it materializes and organizes a set of meaning relations specific to a given domain. In this sense, as previously noted, the definition goes beyond a merely explanatory function and assumes a structuring role in the conceptual organization of the field, delimiting meanings and establishing semantic boundaries that contribute to the internal coherence and communication of the discipline (Krieger and Finatto 2004; Barité 2025).

In line with the premise of statements, the guidelines established by ISO 704:2022 are incorporated, which prescribe criteria such as precision, conciseness, objectivity, and the explicit indication of the relevant domain, in order to ensure that each concept is represented by an expression capable of adequately describing it and distinguishing it from related concepts (Bräscher and Café 2010). This methodological orientation supports the systematic organization of conceptual relationships, enabling each terminological unit to be clearly situated within a specific domain and making explicit its logical and semantic connections with other elements of the conceptual system (Bräscher and Café 2010). In this way, the application of such standardization in the modeling and harmonization of definitions in this first phase of the project not only enhances consistency in the construction of intensional definitions and conceptual stability, but also contributes to semantic interoperability among different knowledge organization instruments, such as Brased.

At this stage, terminological records representing each preferred term (structural and semantic characteristics) were also developed. The terminological record is a fundamental instrument for documentation and registration throughout the processes of modeling, harmonization, and validation. It constitutes an essential component in the development of, for example, a glossary or a thesaurus, as it reflects the methodological and modeling decisions underlying a terminographic product. Its fields are defined according to the objectives, functions, users, and subject domain of the product to be developed (Bevilacqua et al. 2023). It should be noted that the construction of these records occurs concurrently with the other stages. The structure of the terminological record can be accessed through the following DOI: 10.5281/zenodo.19026492.

The construction of the present glossary is thus intended to serve as a “backbone” for the subsequent development of a semantic thesaurus and/or an ontology representing the censuses. Glossaries, thesauri, and ontologies should not be viewed as isolated instruments, but rather as complementary stages within a unified methodological trajectory: the glossary stabilizes and standardizes concepts; the thesaurus makes explicit and enhances semantic relationships; and the ontology formalizes these relationships for automated processing. Across all these stages, the terminological definition functions as the core structuring element. Systematically developing definitions therefore strengthens the conceptual coherence of the domain, ensures the consistency of semantic relationships, and enables more precise, interoperable, and cognitively effective knowledge organization systems. As a result, it reduces ambiguities, polysemy, and uncontrolled variation that could otherwise compromise the representation, retrieval, and qualified use of information.

4.0 Conclusions

The methodological approach adopted in this study highlights the relevance of Terminology and KOS for the management and effective retrieval of information in complex institutional contexts, such as Brazilian educational censuses. It is based on the premise that conceptual precision is a central element for scientific and technical communication, as well as for information interoperability, emphasizing the structuring role of terminological definitions in the organization and semantic stabilization of concepts. In this sense, the development of a specialized glossary for educational censuses constitutes a strategic initiative to promote terminological standardization, reduce ambiguities, and ensure greater consistency in the representation and use of the data produced by Inep.

Finally, it is important to emphasize that the construction of a glossary goes beyond its function as a reference tool, serving instead as a foundational structure for the future development of more complex knowledge organization systems, such as thesauri and ontologies applied to educational censuses. By consolidating standardized definitions and explicit conceptual relationships, the glossary is expected to enhance interoperability among different terminological instruments, as well as to improve the analysis, interpretation, and dissemination of educational information. Thus, the continuation of subsequent project stages is expected to expand and consolidate the proposed conceptual system, strengthening knowledge management and information transparency in the context of Brazilian educational statistics.

Statement on the use of Artificial Intelligence (AI)

The authors declare that no artificial intelligence (AI) tools, including generative AI systems, machine learning applications, or AI-assisted writing technologies, were used in the conception, design, data collection, analysis, interpretation of results, writing, editing, or preparation of this manuscript. All aspects of the work were conducted solely by the authors.

References

Barité, Mario. 2011. “La garantia literaria como herramienta de revisión de sistemas de organización del conocimiento: modelo y aplicación.” Ph.D. dissertation. Granada: Universidad de Granada. http://hdl.handle.net/10481/17583.

Barité, Mario. 2017. “Los árboles de dominio.” In Teoría y praxis en terminologia, edited by Mario Barité and Sara Álvarez Catalá, 91–102. Montevideo: Universidad de la República.

Barité, Mario. 2025. “El método de la grilla: nueva herramienta para la elaboración de definiciones originales.” In Proceedings of the Nineteen Ibero-American Terminology Symposium, RITerm, 24. Medellín, Colombia, October 29–31, 2025. https://riterm.org/wp-content/uploads/2025/10/libro_resumenes_RITerm2025.pdf.

Bevilacqua, Cleci R., Denise R. de Sales, and Marcia M. da Silva. 2023. “A ficha terminológica.” In Como elaborar um dicionário especializado?, edited by Cleci R. Bevilacqua, Denise R. de Sales, Márcia M. da Silva, Patrícia C. R. Reuillard, and Sandra D. Loguercio, 90–103. Porto Alegre: Zouk.

Bräscher, Marisa, and Ligia Maria Arruda Café. 2010. “Organização da Informação ou Organização do Conhecimento?” In Proceedings of the Ninth National Meeting of Research and Graduate Studies in Information Science, edited by M. L. G. Lara and J. W. Smit, 85–102. São Paulo: Escola de Comunicações e Artes/USP.

Dahlberg, Ingetraut. 1978. “Teoria do Conceito.” Ciência da Informação 7, no. 2: 101–107. https://doi.org/10.18225/ci.inf.v7i2.115.

Instituto Brasileiro de Geografia e Estatística (IBGE). 2025. “Conceitos e Metadados.” https://metadados.ibge.gov.br/consulta.

ISO 860. 2007. Terminology work: harmonization of concepts and terms. Geneva: International Organization for Standardization.

ISO 1087. 2019. Terminology work and terminology science: vocabulary. Geneva: International Organization for Standardization.

ISO 704. 2022. Terminology work: principles and methods. Geneva: International Organization for Standardization.

Krieger, Maria da Graça, and Maria José Bocorny Finatto. 2004. Introdução à terminologia: teoria e prática. São Paulo: Contexto.

Löckinger, G., H. J. Kockaert, and G. Budin. 2015. “Intensional definitions.” In Handbook of Terminology, vol. 1, edited by H. J. Kockaert and F. Steurs, 60–81. Amsterdam: John Benjamins.

Lopes-Nascimento, Junio, Brenda Vasconcelos Attalla, and Gercina Ângela de Lima. 2024. “Glossário em perspectiva: um estudo intermediário sobre as concepções e características no domínio da Biblioteconomia e Ciência da Informação Brasileira (BCI).” Ciência da Informação Express 5: 1–28. https://doi.org/10.60144/v5i.2024.112.

Maculan, Benildes Coura M. dos Santos, Maristela S. L. Mesquita, and Luander C. de Jesus Falcão. 2023. “A Glossary for Knowledge Organization Systems Terminology.” Knowledge Organization 50, no. 1: 23–33. https://doi.org/10.5771/0943-7444-2023-1-23.

UNESCO. 2026. “About UNESCO Thesaurus.” https://vocabularies.unesco.org/browser/en/about.