

**Análise de vocabulários controlados na representação de palavras-chave atribuídas por pesquisadores da Fundação Oswaldo Cruz (FIOCRUZ)**

***Analysis of controlled vocabularies in the representation of keywords assigned by Oswaldo Cruz Foundation (Fiocruz) researchers***

***Análisis de vocabularios controlados en la representación de palabras clave asignadas por investigadores de Fundación Oswaldo Cruz (Fiocruz)***

**Adrianne Oliveira de Andrade da Silva**

Fundação Oswaldo Cruz (FIOCRUZ)  
[adrianne.silva@fiocruz.br](mailto:adrianne.silva@fiocruz.br)

**Débora da Silva Rocha**

Fundação Oswaldo Cruz (FIOCRUZ)  
[debora.srcoha@fiocruz.br](mailto:debora.srcoha@fiocruz.br)

**Julia Dias Mota Magaraia**

Fundação Oswaldo Cruz (FIOCRUZ)  
[julia.mota@fiocruz.br](mailto:julia.mota@fiocruz.br)

**Mariângela Spotti Lopes Fujita**

Universidade Estadual Paulista (UNESP)  
[mariangela.fujita@unesp.br](mailto:mariangela.fujita@unesp.br)

**RESUMO**

**Introdução:** A indexação é fundamental para a organização e representação da informação, especialmente quando há exigência de terminologia adequada. A ausência de um vocabulário compatível com a diversidade temática da Fundação Oswaldo Cruz (Fiocruz) compromete a indexação e a recuperação de conteúdos relevantes.

**Objetivo:** O objetivo é investigar a adequação de vocabulários controlados atualmente utilizados nas bibliotecas da Fiocruz, em especial o DeCS, com foco na identificação de lacunas na representação terminológica dos conteúdos institucionais.

**Metodologia:** Os dados coletados se referem a identificação das palavras-chave atribuídas por 15 (quinze) participantes pesquisadores com pesquisas em andamento, entre mestrandos, doutorandos e pesquisadores vinculados a Fiocruz. Cada pesquisador selecionou de 5 (cinco) a 13 (treze) palavras-chave da linguagem natural referentes as pesquisas que estão desenvolvendo, posteriormente inseridas em Planilha Excel, com o intuito de comparar os termos em linguagem natural com os termos encontrados nos vocabulários controlados. As 84 palavras-chave foram atribuídas durante coletas realizadas entre abril e agosto de 2024 e organizadas em Planilha Excel em 22 de janeiro de 2025.

**Resultados:** Os resultados obtidos revelam que o quantitativo de representações com identidade total e parcial é quantitativamente maior em vocabulários multidisciplinares em relação aos

vocabulários especializados em Ciências da Saúde, assim como o quantitativo de representações sem nenhuma identidade precisa da inclusão de novos descritores.

Conclusões: Conclui-se que é necessária a construção de vocabulário controlado da Fiocruz a partir de registros de autoridade de assunto.

**PALAVRAS-CHAVE:** Vocabulários controlados; Representação da Informação; Palavras-chaves.

## ABSTRACT

Background: Indexing is essential for the organization and representation of information, especially when appropriate terminology is required. The lack of a vocabulary compatible with the thematic diversity of *Oswaldo Cruz Foundation (Fiocruz)* compromises both indexing and the retrieval of relevant content.

Purpose: The objective is to investigate the adequacy of controlled vocabularies currently used in the libraries of Fiocruz, especially DeCS, focusing on identifying gaps in the terminological representation of institutional content.

Methodology: The data collected refers to the identification of keywords assigned by 15 (fifteen) participating researchers with ongoing research, including master's students, doctoral students, and researchers affiliated with Fiocruz. Each researcher selected 5 (five) to 13 (thirteen) keywords from natural language related to the research they are developing, which were subsequently entered into an Excel spreadsheet, in order to compare the terms in natural language with the terms found in the controlled vocabularies. The 84 keywords were assigned during the collection conducted between April and August 2024 and organized in an Excel spreadsheet on January 22, 2025.

Results: The results show that the number of representations with total and partial identity is quantitatively higher in multidisciplinary vocabularies than in specialized Health Sciences vocabularies, and that the number of representations without any identity requires the inclusion of new descriptors.

Conclusion: The conclusion is that it is necessary to build a controlled vocabulary for Fiocruz in order to improve the quality of the information.

**KEYWORDS:** Controlled vocabularies; Information representation; Keywords.

## RESUMEN

Introducción: La indización es fundamental para la organización y representación de la información, especialmente cuando se requiere una terminología adecuada. La ausencia de un vocabulario compatible con la diversidad temática de la *Fundación Oswaldo Cruz (Fiocruz)* compromete la indización y la recuperación de contenidos relevantes.

Objetivo: El objetivo es investigar la idoneidad de los vocabularios controlados utilizados actualmente en las bibliotecas de la Fiocruz, especialmente en el DeCS, centrándose en la identificación de lagunas en la representación terminológica del contenido institucional.

Metodología: Los datos recopilados se refieren a la identificación de palabras clave asignadas por 15 (quince) investigadores participantes con investigación en curso, incluyendo estudiantes de maestría, doctorado e investigadores afiliados a la Fiocruz. Cada investigador seleccionó entre 5 (cinco) y 13 (trece) palabras clave en lenguaje natural relacionadas con la investigación que desarrolla, las cuales se ingresaron posteriormente en una hoja de cálculo de Excel para comparar los términos en lenguaje natural con los términos encontrados en los vocabularios controlados. Las 84 palabras clave se asignaron durante la recolección realizada entre abril y agosto de 2024 y se organizaron en una hoja de cálculo de Excel el 22 de enero de 2025.

Resultados: Los resultados obtenidos revelan que la cantidad de representaciones con identidad total y parcial es cuantitativamente mayor en los vocabularios multidisciplinares en comparación con los especializados en Ciencias de la Salud, así como la cantidad de representaciones sin ninguna identidad precisa, lo que indica la necesidad de inclusión de nuevos descriptores.

Conclusiones: Se concluye que es necesaria la construcción de un vocabulario controlado propio de Fiocruz, basado en registros de autoridad de materia.

**PALABRAS CLAVE:** Vocabularios controlados; Representación de la información; Palabras clave.



## 1 INTRODUÇÃO

A indexação desempenha um papel central nos processos de organização e representação da informação, especialmente em contextos nos quais a produção do conhecimento demanda representatividade terminológica adequada. No caso da Fundação Oswaldo Cruz (FIOCRUZ) a ausência de uma política institucional de indexação resultava em práticas fragmentadas, muitas vezes dependentes de instrumentos vocabulares que não se adaptam às especificidades das produções científicas institucionais. Essa condição motivou o desenvolvimento de pesquisa voltada à proposição de uma política de indexação para a Rede de Bibliotecas, com o objetivo de sistematizar os procedimentos técnicos e desenvolver uma linguagem de indexação institucional.

Ao longo das últimas décadas, a Fiocruz consolidou-se como uma das principais instituições científicas do Brasil, tendo expandido significativamente suas áreas de atuação. Embora tenha mantido seu compromisso histórico com a saúde pública, a Fiocruz aprofundou sua inserção em campos interdisciplinares, como as ciências sociais, ambientais, história, patrimônio cultural, políticas públicas e tecnológicas. Essa pluralidade temática refletiu sua atuação diante dos complexos desafios da sociedade contemporânea e do papel estratégico que vem desempenhando através da produção de conhecimento que vai além das temáticas voltadas estritamente à área de saúde.

Outro fator relevante foi o perfil diversificado dos seus usuários, que inclui pesquisadores, estudantes, profissionais, gestores públicos e técnicos de diversas áreas. Suas necessidades informacionais, igualmente heterogêneas, demandam sistemas que acompanhem a evolução temática e linguística do campo científico. A ausência de um vocabulário que dialogue com essa pluralidade contribui para a exclusão de conteúdos relevantes, impactando não



apenas a pesquisa científica, mas também o desenvolvimento de políticas públicas, ações educativas e estratégias institucionais.

Assim, diante desse cenário, a pesquisa tem como objetivo investigar a adequação dos vocabulários controlados utilizados nas bibliotecas da Fiocruz, especialmente o DeCS, com foco na identificação de lacunas na representação terminológica dos conteúdos institucionais. Para isso, o Subgrupo de Trabalho de Vocabulário Controlado analisou palavras-chave em linguagem natural fornecidas por usuários, comparando-as com termos de vocabulários selecionados como fontes positivas. Os níveis de correspondência (total, parcial ou ausente) permitiram mapear o grau de representação e identificar lacunas, subsidiando a construção de um vocabulário próprio.

Sendo assim, o desenvolvimento de pesquisas sobre o uso de vocabulários controlados mostra-se fundamental para contemplar aspectos relacionados à representação e à recuperação da informação. Esse estudo tomou como base metodológica a pesquisa realizada por Tartarotti (2019) e seus desdobramentos apresentados por Alves, Tartarotti e Fujita (2022, p. 286) que destacam “o vocabulário controlado [como] um instrumento dinâmico e que precisa estar em consonância com o conhecimento científico que vem se modificando e se atualizando constantemente, principalmente no âmbito acadêmico”. No entanto, o estudo em questão tem o foco em repositórios e para fins desta pesquisa, o sistema de informação de investigação será o catálogo da Fiocruz. Isso porque, conforme aponta Silva, Oliveira e Lima (2018, p. 2) “o catálogo expõe informações que são capazes de estabelecer relações entre si e os usuários”.

Dessa forma, considerando a necessidade de representar adequadamente o acervo da Fiocruz e de assegurar maior precisão na recuperação da informação durante as buscas, a presente pesquisa justifica-se para garantir a eficiência da indexação na Rede de Bibliotecas da Fiocruz que se configurou como um desafio estratégico. Isso ocorrerá através da criação de um vocabulário controlado específico para o Catálogo Mourisco, alinhado às



particularidades temáticas da instituição, e que surgiu como solução promissora para aprimorar a busca, qualificar a representação da informação e fortalecer a política de indexação, ampliando o acesso ao acervo e a integração entre sistemas informacionais. Busca-se, assim, produzir conhecimento acerca da recuperação da informação por meio do vocabulário controlado no catálogo on-line da biblioteca, com vistas a garantir comunicação entre o sistema e os usuários.

Considerando que os vocabulários controlados nem sempre conseguem acompanhar a evolução científica nem contemplar as especificidades regionais, podendo ocasionar prejuízos na recuperação da informação, esta pesquisa parte da seguinte problemática: em que medida os vocabulários controlados adotados pelas bibliotecas da Fiocruz, com ênfase no DeCS, são capazes de representar com precisão a diversidade e a especificidade da produção institucional?

O presente estudo organiza-se da seguinte forma: primeiro, aborda o contraste entre a linguagem natural e a controlada na representação temática; em seguida, detalha a metodologia de análise dos vocabulários combinando a abordagem quantitativa e qualitativa. A seção seguinte é dedicada à análise e discussão dos dados coletados junto a 15 pesquisadores participantes. Por fim, o estudo encerra-se com as considerações finais, que retomam os principais pontos da pesquisa.

## **2 VOCABULÁRIOS CONTROLADOS NA REPRESENTAÇÃO DE PALAVRAS-CHAVE**

No campo da organização e representação da informação, dois tipos de linguagem são amplamente utilizados: a linguagem natural (LN) e a linguagem controlada (LC). Ambas desempenham papéis fundamentais no processo de indexação e na busca por documentos em sistemas de



informação, sendo essenciais para garantir que os usuários encontrem o que procuram com precisão e eficiência.

A linguagem natural, também conhecida como “linguagem livre”, refere-se às palavras que ocorrem espontaneamente em textos impressos (Lancaster, 1993, p. 200). Sua flexibilidade e expressividade, embora permitam diversas formas de comunicação, geram desafios como ambiguidade e variação terminológica, dificultando a representação precisa da informação. Na indexação, os termos da linguagem natural são comumente empregados como palavras-chave, extraídos diretamente dos documentos. Elas são úteis para a pesquisa em vocabulários quando estes possuem termos e nomes que compreendem diversas palavras (Harpring, 2016, p. 224).

Por sua vez, a linguagem controlada é um conjunto estruturado e limitado de termos autorizados, denominados descritores. Esses termos são organizados em vocabulários controlados, com o objetivo de padronizar a linguagem usada na indexação e aumentar a precisão na representação de documentos.

O controle vocabular, conforme Svenonious (1976), ocorre em duas fases: a normalização gramatical (ex.: singular/plural, variações morfológicas) e a unificação semântica de sinônimos e termos equivalentes. Essa padronização agrupa termos relacionados, otimizando os resultados de busca ao evitar dispersão. Raitt (1980) adiciona que a atribuição de pesos descritivos e a conexão entre termos são recursos que aprimoram as estratégias de busca. Austin (1986) complementa que a consistência na indexação exige a representação de conceitos por substantivos ou frases substantivadas, além de uma clara definição do uso de formas no singular ou plural. Esses princípios fundamentam as linguagens controladas baseadas em vocabulários controlados, visando à representação uniforme do conteúdo documental e à facilitação de sua recuperação.

Bocato, Fujita e Gil Leiva (2011, p. 56) destacam um desafio crucial: muitos usuários desconhecem ou não têm acesso direto às linguagens



controladas dos sistemas. Essa lacuna vocabular dificulta a formulação de estratégias de busca eficazes, já que a terminologia do usuário pode não coincidir com os descritores do vocabulário controlado empregado na indexação.

Essa falta de articulação entre os vocabulários controlados e a linguagem natural, utilizada tanto por indexadores quanto por usuários, frequentemente resulta na recuperação de grande quantidade de itens irrelevantes.

Em pesquisa qualitativa, Boccatto (2009, p.128) avaliou os princípios de especificidade, exaustividade, revocação e precisão na análise de assunto por catalogadores e usuários da UNESP. Os usuários apontaram a abrangência excessiva dos resultados como fator de revocação elevada, dificultando a recuperação da informação. A autora ponderou que a linguagem adotada apresentou desempenho insatisfatório, evidenciando sua incompatibilidade com a linguagem de busca do usuário. Em estudo quantitativo posterior, Boccato, Fujita e Gil Leiva (2011) confirmaram a baixa precisão dessa linguagem em comparação com outra controlada e com a linguagem natural.

Um vocabulário controlado é um conjunto estruturado e normatizado de termos autorizados, que garante a consistência terminológica na indexação e representação da informação, evitando ambiguidades e controlando sinônimos (Lancaster, 2004; Fujita, 2012; Mazzocchi, 2018). Esses vocabulários integram listas de termos autorizados e incluem, frequentemente, uma estrutura semântica organizada, sendo representados por esquemas de classificação, taxonomias, tesouros e listas de cabeçalhos de assuntos. Lancaster (2004) afirma que são cruciais para a recuperabilidade do conhecimento, otimizando a comunicação entre indexadores e sistemas de busca. Apesar de dialogarem com a linguagem científica ao refletirem os paradigmas e a terminologia padronizada de disciplinas específicas, (Brundage, 1989), sua complexidade exige treinamento para todos os usuários.



O vocabulário controlado apresenta vantagens como padronização terminológica, precisão conceitual por meio de tesouros e notas de escopo, maior confiança do usuário, identificação de conceitos relacionados via relações semânticas e otimização das estratégias de busca. Em contrapartida, envolve altos custos de manutenção, risco de desatualização, distanciamento das demandas dos usuários e necessidade de capacitação (Henzler, 1978; Perez, 1982; Salton, 1986 *apud* Lopes, 2002). Além de converter linguagem natural para representação e recuperação, vocabulários controlados relacionais destacam-se por suas estruturas semânticas complexas, que permitem ao usuário acessar e associar conceitos com base em descritores representativos de sua necessidade informacional (Bocatto, 2011, p. 184).

A estrutura conceitual funciona como um mapa semântico, permitindo que a representação e a busca considerem descritores e termos relacionados, o que amplia a precisão da indexação. Segundo Hjørland (2008, p. 89), o vocabulário controlado segue a regra de Cutter, ao priorizar a expressão mais específica e adequada, tornando os tópicos mais previsíveis. Para Moreiro Gonzalez (2004, p. 51), sua principal vantagem está na consistência da representação e recuperação, garantindo que um conceito seja expresso sempre da mesma forma.

A evolução dos vocabulários controlados, possivelmente influenciada pela indexação social na internet, tem favorecido visualizações mais intuitivas voltadas a usuários não especializados, que demandam suporte terminológico para maior especificidade ou exaustividade nas buscas. O tesouro destaca-se como modelo amplamente adotado por instituições, dada sua estrutura que facilita a compreensão e contextualização de conceitos. A norma ISO 25964 (2011, p. vi, tradução nossa) observa que, 'no passado, os tesouros foram projetados para profissionais de informação treinados em indexação e pesquisa, hoje há uma demanda por Vocabulários que os usuários não treinados acharão intuitivos'.



É baseado nessa premissa que a Norma ISO 25964 sustenta a aplicação do tesauro como vocabulário controlado, também, em situações em que os computadores façam as escolhas, ou seja, “Se o indexador e o pesquisador forem orientados para escolher o mesmo termo para o mesmo conceito, os documentos relevantes serão recuperados” (International Organization for Standardization, 2011, p.vi, tradução nossa).

O tesauro é uma forma especializada e sistematizada de vocabulário controlado. Trata-se de uma linguagem de indexação alfabética, controlada, combinatória e pós-coordenada, composta por descritores semanticamente relacionados, cuja função é representar de maneira precisa os conceitos de documentos em sistemas de informação (Fujita e Moreira, 2021).

O tesauro possui estrutura lógica e hierárquica, conectando termos por relações de hierarquia (mais geral/específico), associação (afinidade temática) e equivalência (sinônimos e variantes). Ao contrário das linguagens pré-coordenadas, como listas de cabeçalhos, os tesauros são voltados a ambientes de pós-coordenação, permitindo combinações flexíveis de descritores na estratégia de busca (Gil Urdiciain, 2004).

Além da função descritiva, o tesauro exerce papel organizacional e pedagógico ao estruturar o conhecimento de uma área, facilitando o acesso para indexadores, pesquisadores e usuários. É especialmente eficaz em domínios técnico-científicos com terminologia complexa, especializada e em constante evolução. A seguir foram detalhados os tesauros e vocabulários controlados utilizados como fontes positivas nesta pesquisa.

### **3 METODOLOGIA PARA ANÁLISE DE VOCABULÁRIOS CONTROLADOS**

Conforme destaca Minayo (2010), a metodologia em um estudo científico compreende o debate de ideias, a definição de opções teóricas e a adoção de práticas coerentes com os objetivos da investigação, sendo a escolha do caminho metodológico fundamental para assegurar a coerência e a



credibilidade do processo investigativo. Nessa perspectiva, o presente estudo trata-se de uma pesquisa exploratória, com uma abordagem metodológica empírica, de natureza mista, combinando procedimentos quantitativos e qualitativos.

Conforme Braga (2007), a pesquisa exploratória concentra-se em temáticas pouco exploradas, visando à coleta de informações que fomentem a criação de novas hipóteses e ideias. O propósito desse método não reside na validação de teorias, mas na detecção de padrões para orientar investigações posteriores. Procedimentos como análise histórica, observação, revisão bibliográfica e estudos de caso são recorrentes nessa modalidade investigativa.

Já a utilização integrada dessas abordagens tem se mostrado uma estratégia eficaz para o aprofundamento da compreensão dos fenômenos investigados, uma vez que, conforme assinala Minayo (2014, p. 57), “cada um tem seu papel, seu lugar e sua adequação”. Para Creswell (2010), a abordagem quantitativa possibilita a apreensão da dimensão objetiva e mensurável dos dados, enquanto a qualitativa permite compreender interpretações, significados e contextos atribuídos pelos sujeitos envolvidos. De modo complementar, Richardson et al. (1999) destaca que a pesquisa quantitativa busca quantificar dados com vistas à identificação de padrões e à testagem de hipóteses, ao passo que a qualitativa se concentra na compreensão das subjetividades e das dinâmicas sociais. A articulação entre essas abordagens, conforme aponta Goldenberg (1997), favorece o cruzamento de diferentes tipos de dados, ampliando a confiabilidade dos resultados e reduzindo vieses metodológicos.

Além disso, os procedimentos metodológicos fundamentaram-se no estudo desenvolvido por Alves, Tartarotti e Fujita (2022), cuja análise e a busca por assuntos teve como foco o Repositório Institucional da Universidade Estadual de Campinas (UNICAMP). A partir desse referencial, o método foi



adaptado ao contexto da Fiocruz, preservando sua estrutura analítica e seus pressupostos teórico-metodológicos.

A pesquisa abrangeu grandes áreas do conhecimento da Fiocruz, conforme apresentado no Quadro 1, contemplando as subáreas das Ciências da Saúde, Ciências Sociais Aplicadas e Ciências Humanas. Participaram da pesquisa: discentes de especialização, discentes de pós-graduação (mestrado e doutorado) e docentes pesquisadores, sendo selecionado um participante por subárea e por categoria, totalizando 15 participantes. Os critérios de seleção consideraram: discentes de especialização com projeto de Trabalho de Conclusão de Curso; discentes de mestrado e doutorado com projetos de pesquisa em andamento; e docentes pesquisadores atuantes como líderes de grupos ou laboratórios de pesquisa, ou docentes permanentes de programas de pós-graduação.

Quadro 1 - Seleção das áreas de conhecimento da Fiocruz

| Grande área do conhecimento                    | Subárea                                   | Categorias de Participantes (*) |
|------------------------------------------------|-------------------------------------------|---------------------------------|
| Ciências da Saúde<br>Ciências Biológicas       | Saúde Coletiva                            | 2 DM; 2 DD; 1 DP                |
|                                                | Saúde Pública                             | 1 DM                            |
|                                                | Ciências                                  | 1 DD                            |
| Ciências Sociais Aplicadas<br>Ciências Humanas | Arquitetura (Patrimônio Cultural)         | 1 DP                            |
|                                                | Divulgação da Ciência, Tecnologia e Saúde | 1 DE; 1 DM                      |
|                                                | História                                  | 1 DD; 1 DP                      |
| Ciências Humanas                               | Sociologia                                | 1 DE; 1 DD; 1 DP                |

Fonte: Elaborado pelas autoras (2025).

(\*) No caso de não existir alguma das categorias de participantes, não considerar para a coleta.

Descrição: O quadro está organizado em três colunas: Grande área do conhecimento, Subárea e Categorias de Participantes. Na grande área Ciências da Saúde/Ciências Biológicas estão incluídas as subáreas Saúde Coletiva, com dois discentes de mestrado (DM), dois discentes de doutorado (DD) e um docente pesquisador (DP); Saúde Pública, com um discente de mestrado; e Ciências, com um discente de doutorado. Na grande área Ciências Sociais Aplicadas/Ciências Humanas estão contempladas as subáreas Arquitetura (Patrimônio Cultural), com um docente pesquisador; Divulgação da Ciência, Tecnologia e Saúde, com um discente de especialização (DE) e um discente de mestrado; e História, com um discente de doutorado e um docente pesquisador. Na área Ciências Humanas consta a subárea Sociologia, com um discente de especialização, um discente de doutorado e um docente pesquisador. O conjunto totaliza 15 participantes distribuídos entre diferentes áreas de conhecimento e categorias acadêmicas.



Na etapa preliminar, os participantes foram convidados por correio eletrônico, ocasião em que receberam esclarecimentos acerca dos objetivos da pesquisa, bem como o Termo de Consentimento Livre e Esclarecido e o Termo de Cessão de Direitos Autorais, assegurando-se o anonimato e a confidencialidade das informações.

Para operacionalização, os procedimentos metodológicos foram organizados em fases sequenciais:

- a) definição de palavras-chave em linguagem natural atribuídas pelos próprios participantes;
- b) pesquisa dos descritores correspondentes em vocabulários controlados (Quadro 2);
- c) construção de planilhas para análise quantitativa dos resultados; categorização dos termos em identidade total, parcial ou ausente; e análise por categoria e por vocabulário, com apoio de representações gráficas.

As coletas de dados foram realizadas de forma remota, por meio da plataforma Microsoft Teams, com gravação das telas e das interações ocorridas durante o processo de busca. Cada participante foi orientado a selecionar de 5 (cinco) a 13 (treze) palavras-chave da linguagem natural, em língua portuguesa, referentes as pesquisas que estão desenvolvendo, posteriormente inseridas em Planilha Excel, com o intuito de comparar os termos em linguagem natural com os termos encontrados nos vocabulários controlados.

No âmbito qualitativo, realizou-se a análise da consistência e da frequência de palavras-chave em linguagem natural fornecidas por usuários da Instituição X, comparando-as com vocabulários controlados.

Isto posto, as palavras-chave de linguagem natural foram pesquisadas nos Vocabulários Controlados descritos no quadro 2 por meio de sua abrangência temática, fonte de desenvolvimento do vocabulário, descrição e



cobertura linguística com base nas informações extraídas de seus respectivos sites onde estão acessíveis.

**Quadro 2 – Descrição dos vocabulários controlados utilizados na pesquisa**

| Vocabulário controlado                                                     | Instituição responsável                                                         | Abrangência temática                                                                                                                                           | Fonte de desenvolvimento                                                                                                                                                                                                                                                | Descrição                                                                                                                                                                                                                                                                                                                                                                                     | Cobertura linguística        |
|----------------------------------------------------------------------------|---------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|
| <i>Descritores em Ciências da Saúde (DeCS)</i>                             | Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME) | Especializado: Ciências da Saúde                                                                                                                               | <i>Medical Subject Headings (MeSH)</i> e categorias exclusivas que abrangem as áreas específicas de Ciência e Saúde, Vigilância em Saúde, Saúde Pública, Homeopatia, e Medicinas Tradicionais, Complementares e Integrativas.                                           | 77 qualificadores e 34.861 descritores; organização hierárquica de conceitos                                                                                                                                                                                                                                                                                                                  | Multilíngue                  |
| <i>Vocabulário Controlado USP (VC-USP)</i>                                 | Universidade de São Paulo (USP)                                                 | Multidisciplinar: Todas as áreas do conhecimento científico                                                                                                    | Vocabulário gerado pela indexação de assuntos do sistema de bibliotecas da USP                                                                                                                                                                                          | Composto por termos autorizados, termos "não-autorizados" (remissivos) e "elos falsos" que agrupam termos mais específicos; macroestrutura com relações lógico-semânticas entre as áreas, subáreas e a terminologia; Lista Alfabética de Assuntos e a Lista Sistemática (Hierárquica), tabelas de Qualificadores de Locais Geográficos e Históricos, Gênero e Forma, e Profissões e Ocupações | Português                    |
| <i>Vocabulário Controlado dos Acervos da Casa de Oswaldo Cruz (VC-COC)</i> | Casa de Oswaldo Cruz                                                            | Especializado: História da saúde pública, História da Medicina, Sociologia e Filosofia da Ciência, Ciências Biomédicas e da Saúde Pública Memória e Patrimônio | Bibliotecas da Casa de Oswaldo Cruz, Biblioteca de História das Ciências e da Saúde, Biblioteca de Educação e Divulgação Científica Iloni Seibel, Departamento de Arquivo e Documentação, Departamento de Patrimônio Histórico e Serviço de Museologia do Museu da Vida | Vocabulário é estruturado com subdivisões específicas — como descritores de assunto, localizações geográficas e autoridades — e organizado de forma hierárquica e relacional, permitindo também a inclusão de equivalências com a linguagem natural dos usuários.                                                                                                                             | Português                    |
| <i>Tesouro Unesp (T-Unesp)</i>                                             | Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)                  | Multidisciplinar: Todas as áreas do conhecimento científico                                                                                                    | <i>Library of Congress Subject Headings (LCSH)</i> , Terminologia da                                                                                                                                                                                                    | O Tesouro Unesp tem vocabulário com termos especializados das áreas de conhecimento de suas atividades de ensino, pesquisa e extensão. Construído a partir da                                                                                                                                                                                                                                 | Bilíngue: português e inglês |



| Vocabulário controlado                             | Instituição responsável                                       | Abrangência temática                                        | Fonte de desenvolvimento                                                                                             | Descrição                                                                                                                                                                                                                                                                                                                                       | Cobertura linguística |
|----------------------------------------------------|---------------------------------------------------------------|-------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
|                                                    |                                                               |                                                             | Biblioteca Nacional do Brasil (BN), <i>Medical Subject Headings</i> (MeSH) e Descritores em Ciências da Saúde (DeCS) | combinação de vocabulários em registros de autoridade MARC21.                                                                                                                                                                                                                                                                                   |                       |
| <u>Thesaurus Brasileiro da Educação</u> – (Brased) | Instituto Nacional de Estudos e Pesquisas Educacionais (INEP) | Especializado: Educação                                     | Vocabulário gerado pela indexação de assuntos da biblioteca do INEP                                                  | Organiza termos e conceitos interligados na área da educação, seguindo uma estrutura conceitual predefinida.                                                                                                                                                                                                                                    | Português             |
| <i>Terminologia da Biblioteca Nacional (BN)</i>    | Biblioteca Nacional do Brasil                                 | Multidisciplinar: Todas as áreas do conhecimento científico | <i>Library of Congress Subject Headings (LCSH)</i>                                                                   | Os assuntos são apresentados em uma lista multidisciplinar, estruturada como um tesouro, onde cada termo de assunto é detalhado com termos gerais (TG), termos específicos (TE) e termos relacionados (TR), além de incluir tópicos, remissivas "ver" e "ver também", e subdivisões gerais, cronológicas e geográficas.                         | Português e inglês    |
| <u>Library of Congress Subject Headings (LCSH)</u> | Biblioteca do Congresso dos EUA                               | Multidisciplinar: Todas as áreas do conhecimento científico | Vocabulário gerado pela Biblioteca do Congresso dos EUA                                                              | Lista alfabética completa de termos a serem usados como vocabulário controlado para conceitos de assunto por catalogadores da Biblioteca do Congresso e de outras bibliotecas, visando fornecer acesso por assunto controlado a registros substitutos. Utilizado desde 1898, o LCSH facilita o acesso por assunto aos recursos em seu catálogo. | Inglês                |
| <u>Medical Subject Headings (MeSH)</u>             | National Library of Medicine (NLM)                            | Especializado: Ciências da Saúde; Ciências Médicas          | Vocabulário produzido pela National Library of Medicine                                                              | Inclui os cabeçalhos de assunto que aparecem na base de dados MEDLINE/PubMed, Catálogo da NLM e em outras bases de dados NLM.                                                                                                                                                                                                                   | Inglês                |

Fonte: Elaborado pelas autoras (2025).

Descrição: O quadro está estruturado em seis colunas: Vocabulário controlado, Instituição responsável, Abrangência temática, Fonte de desenvolvimento, Descrição e Cobertura linguística. Ao todo, o quadro apresenta oito vocabulários controlados utilizados na pesquisa e suas principais características. A primeira linha descreve os Descritores em Ciências da Saúde (DeCS), desenvolvido pela BIREME, com abrangência especializada em Ciências da Saúde. O vocabulário é baseado no MeSH e complementado por categorias específicas relacionadas à Saúde Pública, Vigilância em Saúde, Homeopatia, Ciência e Saúde e Medicinas Tradicionais, Complementares e Integrativas, com cobertura multilíngue. A segunda linha apresenta o Vocabulário Controlado USP (VC-USP), mantido pela Universidade de São Paulo, com abrangência multidisciplinar. O vocabulário foi construído a partir da indexação de assuntos das bibliotecas da USP e é composto por termos autorizados, termos não autorizados e relações semânticas entre áreas e subáreas do conhecimento, com cobertura linguística em português. A terceira linha descreve o Vocabulário Controlado dos Acervos da Casa de Oswaldo Cruz (VC-COC), especializado em história da saúde, patrimônio, memória e áreas correlatas. O vocabulário resulta do trabalho conjunto das bibliotecas, do arquivo e do museu da Casa de Oswaldo Cruz e contempla descritores de assunto, autoridades e localizações geográficas, além de equivalências com a



linguagem utilizada pelos usuários. Está disponível em português. A quarta linha refere-se ao Tesouro Unesp, vocabulário multidisciplinar desenvolvido a partir da combinação de diferentes fontes, como LCSH, Biblioteca Nacional, MeSH e DeCS. Sua descrição destaca a presença de termos especializados voltados às atividades de ensino, pesquisa e extensão universitária, estruturados em registros de autoridade MARC21. O vocabulário possui cobertura bilingue, em português e inglês. A quinta linha apresenta o Thesaurus Brasileiro da Educação (BRASED), produzido pelo INEP e especializado na área de Educação. O vocabulário foi construído a partir da indexação dos acervos da instituição e organiza termos e conceitos educacionais em uma estrutura conceitual predefinida, estando disponível em português. A sexta linha descreve a Terminologia da Biblioteca Nacional (BN), vocabulário multidisciplinar baseado no LCSH. Sua estrutura segue a lógica de um tesouro, com termos gerais, específicos e relacionados, além de remissivas e subdivisões cronológicas e geográficas, possibilitando a navegação entre conceitos. Possui cobertura em português e inglês. A sétima linha corresponde ao Library of Congress Subject Headings (LCSH), desenvolvido pela Biblioteca do Congresso dos Estados Unidos, com abrangência multidisciplinar. O vocabulário consiste em uma extensa lista de cabeçalhos de assunto utilizada para representação temática em catálogos bibliográficos, possibilitando acesso controlado por assunto aos registros documentais. Sua cobertura linguística é exclusivamente em inglês. A oitava linha apresenta o Medical Subject Headings (MeSH), produzido pela National Library of Medicine. Especializado nas áreas da saúde e das ciências médicas, o vocabulário reúne os cabeçalhos de assunto utilizados na indexação da MEDLINE/PubMed e de outras bases da NLM. Sua cobertura linguística é em inglês.

A última etapa correspondeu à análise dos resultados, os quais foram classificados em três categorias: identidade total, quando o termo pesquisado foi encontrado exatamente como formulado na fonte consultada; identidade parcial, quando foram identificados termos alternativos ou sinônimos; e ausência de identidade, quando não foi localizado nenhum termo correspondente. Esse procedimento envolveu a verificação de identidade total, parcial ou ausente entre os termos, por meio de uma análise integrada que considerou tanto aspectos quantitativos — relacionados à frequência de ocorrência dos termos — quanto a aspectos qualitativos, associados aos significados e contextos de uso dos descritores.

Os dados foram organizados em planilhas contendo os termos pesquisados, as fontes consultadas e a respectiva categoria de identificação (Quadro 3 a seguir na seção de resultados). A partir desse material, foram elaborados gráficos que permitiram visualizar os níveis de identificação por fonte de informação, bem como o quantitativo geral de ocorrências de identidade total, parcial e ausente.

## **4 RESULTADOS E DISCUSSÕES**



Ao longo dos anos, a Fiocruz ampliou suas áreas de pesquisa, tornando-se uma instituição cada vez mais multidisciplinar, ainda que com foco no campo da Saúde. As áreas de conhecimento científicas prevalentes na Fiocruz são Ciências da Saúde e Ciências Biológicas, seguidas por Ciências Humanas, Ciências Sociais Aplicadas e Química, menos prevalentes. No entanto, os debates e questões sociais nos quais a Fiocruz está inserida não estão plenamente representados no atual vocabulário utilizado pelas suas bibliotecas, os Descritores em Ciências da Saúde (DeCS).

Conforme dito anteriormente, os dados coletados consistiram na identificação de palavras-chaves atribuídas por uma amostra de 15 (quinze) participantes. A seleção da participação procurou abranger pesquisadores, mestrandos e doutorandos da Fiocruz.

Na pesquisa, foi solicitado que os participantes selecionassem de 5 (cinco) a 13 (treze) palavras-chave de linguagem natural e depois as buscassem nos vocabulários controlados selecionados conforme quadro 2. A orientação era que utilizassem termos de suas pesquisas em andamento.

Como resultado, obteve-se a atribuição de 84 palavras-chaves durante a realização das coletas. As palavras-chave de linguagem natural foram então pesquisadas nos Vocabulários Controlados (quadro 2), considerando a abrangência temática, fonte de desenvolvimento do vocabulário, descrição e cobertura linguística com base nas informações extraídas de seus respectivos sites.

Em análise comparada dos 8 vocabulários controlados observa-se que todos possuem cobertura temática compatível com as áreas de conhecimento prevalentes nas bibliotecas da Fiocruz e apenas o Thesaurus Brasileiro de Educação (BRASED) é específico da área de Educação. A escolha dos vocabulários controlados levou em consideração a proposta de verificar coberturas temáticas multidisciplinares (LCSH, BN, T-Unesp, VC-USP), especializadas em Ciências da Saúde (DeCS, MeSH, VC-COC) e em Educação



(BRASED). Essa diversidade de abrangência temática teve como objetivo analisar as possibilidades de representação de assuntos multidisciplinares que não tem alcance nos vocabulários controlados mais específicos da área de ciências da saúde.

Na planilha demonstrada como exemplo abaixo (Quadro 3) referente ao participante 1, foram registrados os termos pesquisados, as respectivas fontes de informação consultadas e a categoria de identificação atribuída a cada termo (identificação total, parcial ou nenhuma). Para esse estudo, considerou-se como “Identidade total” (IT) quando a palavra-chave pesquisada no vocabulário controlado é exatamente igual ao descritor existente na fonte positiva; a “identidade parcial” (IP), por sua vez, foi caracterizada pelo encontro de descritores alternativos ou sinônimos quando pesquisada a palavra-chave; e, por último, classificou-se como “nenhuma identidade” (NI) quando nenhum resultado correspondente foi encontrado entre a palavra-chave e o descritor.

**Quadro 3** - Planilha de pesquisa das palavras-chaves atribuídas pelo Participante 1 nos vocabulários controlados (fontes de informação)

| Vocabulário<br>s<br>controlados | Palavras-chave      | Keywords              | Resultado das buscas nos<br>vocabulários controlados | Identidade |
|---------------------------------|---------------------|-----------------------|------------------------------------------------------|------------|
| LCSH                            | Pensamento social   | Social thought        | Não foi encontrado                                   | NI         |
|                                 | Sociologia rural    | Rural sociology       | Rural sociology                                      | IT         |
|                                 | Antropologia urbana | Urban<br>anthropology | Urban anthropology                                   | IT         |
|                                 | Sertão              | Hinterland            | Hinterland                                           | IT         |
|                                 | Medicina rural      | Rural medicine        | Medicine, rural                                      | IT         |
| MeSH                            | Pensamento social   | Social thought        | Thinking                                             | IP         |
|                                 | Sociologia rural    | Rural sociology       | sociology                                            | IP         |
|                                 | Antropologia urbana | Urban<br>anthropology | anthropology                                         | IP         |
|                                 | Sertão              | Hinterland            | Não foi encontrado                                   | NI         |
|                                 | Medicina rural      | Rural medicine        | medicine                                             | IP         |
| DeCS                            | Pensamento social   | Social thought        | Não foi encontrado                                   | NI         |
|                                 | Sociologia rural    | Rural sociology       | Não foi encontrado                                   | NI         |
|                                 | Antropologia urbana | Urban<br>anthropology | Não foi encontrado                                   | NI         |
|                                 | Sertão              | Hinterland            | Não foi encontrado                                   | NI         |
|                                 | Medicina rural      | Rural medicine        | Saúde da População Rural                             | IP         |
| VC-USP                          | Pensamento social   | Social thought        | Não foi encontrado                                   | NI         |
|                                 | Sociologia rural    | Rural sociology       | Sociologia rural                                     | IT         |
|                                 | Antropologia urbana | Urban<br>anthropology | Antropologia urbana                                  | IT         |
|                                 | Sertão              | Hinterland            | Não foi encontrado                                   | NI         |
|                                 | Medicina rural      | Rural medicine        | Não foi encontrado                                   | NI         |



| Vocabulário<br>s<br>controlados | Palavras-chave      | Keywords              | Resultado das buscas nos<br>vocabulários controlados | Identidade |
|---------------------------------|---------------------|-----------------------|------------------------------------------------------|------------|
| Tesauro<br>UNESP                | Pensamento social   | Social thought        | Não foi encontrado                                   | NI         |
|                                 | Sociologia rural    | Rural sociology       | Sociologia rural                                     | IT         |
|                                 | Antropologia urbana | Urban<br>anthropology | Antropologia urbana                                  | IT         |
|                                 | Sertão              | Hinterland            | Não foi encontrado                                   | NI         |
|                                 | Medicina rural      | Rural medicine        | Medicina rural                                       | NI         |
| VC-COC                          | Pensamento social   | Social thought        | Não foi encontrado                                   | NI         |
|                                 | Sociologia rural    | Rural sociology       | Não foi encontrado                                   | NI         |
|                                 | Antropologia urbana | Urban<br>anthropology | Não foi encontrado                                   | NI         |
|                                 | Sertão              | Hinterland            | Sertões                                              | IT         |
|                                 | Medicina rural      | Rural medicine        | Não foi encontrado                                   | NI         |
| BRASED                          | Pensamento social   | Social thought        | Não foi encontrado                                   | NI         |
|                                 | Sociologia rural    | Rural sociology       | Sociologia rural                                     | IT         |
|                                 | Antropologia urbana | Urban<br>anthropology | Não foi encontrado                                   | IT         |
|                                 | Sertão              | Hinterland            | Não foi encontrado                                   | NI         |
|                                 | Medicina rural      | Rural medicine        | Não foi encontrado                                   | NI         |
| BN                              | Pensamento social   | Social thought        | Não foi encontrado                                   | NI         |
|                                 | Sociologia rural    | Rural sociology       | Sociologia rural                                     | IT         |
|                                 | Antropologia urbana | Urban<br>anthropology | Antropologia urbana                                  | IT         |
|                                 | Sertão              | Hinterland            | Não foi encontrado                                   | NI         |
|                                 | Medicina rural      | Rural medicine        | Medicina rural                                       | IT         |

Fonte: Elaborado pelas autoras (2025).

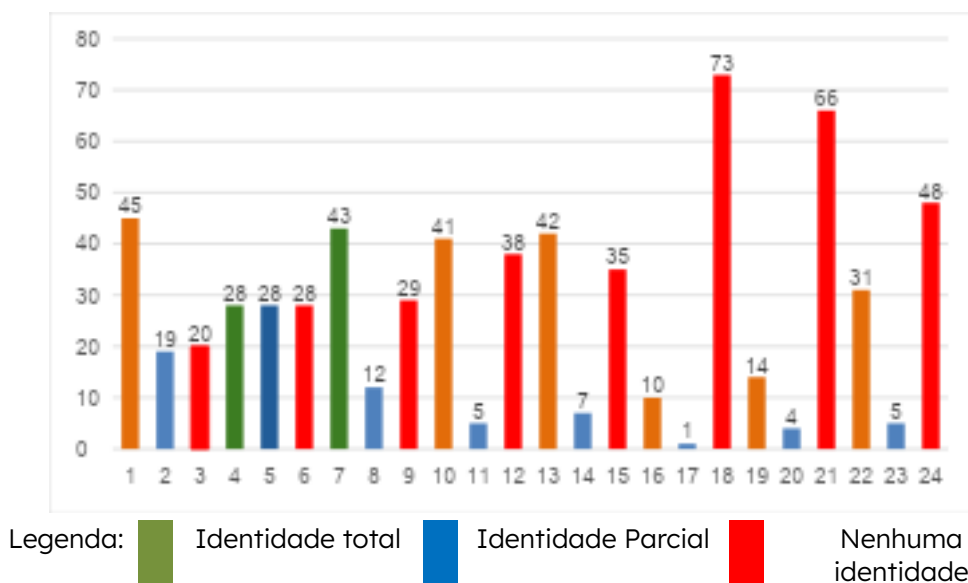
Descrição: O quadro está organizado em cinco colunas e quarenta linhas de dados, distribuídas por oito vocabulários controlados: LCSH, MeSH, DeCS, VC-USP, Tesauro Unesp, VC-COC, BRASED e Biblioteca Nacional. A primeira coluna identifica o vocabulário controlado analisado. Como cada vocabulário foi utilizado para pesquisar cinco palavras-chave, seu nome se repete ao longo de cinco linhas consecutivas. A segunda coluna apresenta as palavras-chave atribuídas pelo Participante 1: Pensamento social, Sociologia rural, Antropologia urbana, Sertão e Medicina rural. A terceira coluna apresenta as respectivas traduções para o inglês: Social thought, Rural sociology, Urban anthropology, Hinterland e Rural medicine. A quarta coluna registra o resultado encontrado em cada vocabulário controlado, podendo corresponder ao mesmo termo pesquisado, a um termo relacionado ou à ausência de resultados. A quinta coluna apresenta a classificação da correspondência encontrada, utilizando as categorias Identidade Total (IT), Identidade Parcial (IP) e Nenhuma Identidade (NI). No bloco referente ao LCSH, a palavra-chave Pensamento social não apresentou correspondência. Já Sociologia rural, Antropologia urbana, Sertão e Medicina rural foram localizadas respectivamente como Rural sociology, Urban anthropology, Hinterland e Medicine, rural, sendo todas classificadas como identidade total (IT). No MeSH, nenhuma das cinco palavras-chave foi recuperada exatamente na forma pesquisada. Pensamento social foi relacionado ao termo Thinking; Sociologia rural ao termo Sociology; Antropologia urbana ao termo Anthropology; e Medicina rural ao termo Medicine. Esses resultados receberam classificação de identidade parcial (IP). A palavra-chave Sertão não foi encontrada, sendo classificada como nenhuma identidade (NI). No DeCS, Pensamento social, Sociologia rural, Antropologia urbana e Sertão não apresentaram resultados correspondentes, sendo classificados como nenhuma identidade. Apenas Medicina rural apresentou recuperação parcial por meio do descritor Saúde da População Rural, sendo classificada como identidade parcial. No VC-USP, Sociologia rural e Antropologia urbana foram recuperadas exatamente com os mesmos termos utilizados na busca, recebendo identidade total. As palavras-chave Pensamento social, Sertão e Medicina rural não foram encontradas, sendo classificadas como nenhuma identidade. No Tesauro Unesp, observou-se comportamento semelhante ao do VC-USP. As palavras-chave Sociologia rural e Antropologia urbana apresentaram identidade total, enquanto Pensamento social e Sertão não foram localizadas. O termo Medicina rural foi



recuperado exatamente com a mesma denominação, embora o quadro o apresente na coluna de classificação conforme o registro original dos dados. No VC-COC, apenas a palavra-chave Sertão apresentou correspondência, por meio do descritor Sertões, registrado como identidade total. As demais palavras-chave não foram encontradas e foram classificadas como nenhuma identidade. No BRASED, somente Sociologia rural foi encontrada exatamente como descritor do vocabulário. As palavras-chave Pensamento social, Sertão e Medicina rural não apresentaram correspondência. O termo Antropologia urbana também não foi localizado, conforme registrado no quadro. No vocabulário da Biblioteca Nacional, Sociologia rural, Antropologia urbana e Medicina rural foram recuperadas exatamente como descritores existentes na base, correspondendo a identidades totais. As palavras-chave Pensamento social e Sertão não foram encontradas.

Após a sistematização, ocorreu a análise quantitativa a partir da geração do Gráfico 1 tanto para melhor visualizar os níveis de identificação por fonte de informação, quanto para demonstrar um panorama geral da distribuição entre “identificações totais, parciais e nenhuma”. Essa análise possibilitou a identificação de quais fontes apresentaram maior incidência aos termos descritores utilizados na pesquisa.

**Gráfico 1 - Relação do quantitativo de representações por vocabulários controlados**



Fonte: Elaborado pelas autoras (2025).

Descrição: O gráfico apresenta a quantidade de ocorrências de identidade total (T), representada na cor verde; identidade parcial (P), representada na cor azul; e nenhuma identidade (N), representada na cor vermelha; obtidas para cada um dos oito vocabulários controlados analisados: Library of Congress Subject Headings (LCSH), Medical Subject Headings (MeSH), Descritores em Ciências da Saúde (DeCS), Vocabulário Controlado USP (VC-USP), Tesouro Unesp, Vocabulário Controlado dos Acervos da Casa de Oswaldo Cruz (VC-COC), Thesaurus Brasileiro da Educação (BRASED) e Terminologia da Biblioteca Nacional (BN). O eixo horizontal apresenta os vocabulários controlados e o eixo



vertical apresenta o quantitativo de ocorrências registradas para cada categoria de identidade. No LCSH foram registradas 45 ocorrências de identidade total, 19 de identidade parcial e 20 de nenhuma identidade. No MeSH foram registradas 28 identidades totais, 28 identidades parciais e 28 ocorrências sem identidade. No DeCS foram observadas 43 identidades totais, 12 identidades parciais e 29 ocorrências sem identidade. No VC-USP foram registradas 41 identidades totais, 5 identidades parciais e 38 ocorrências sem identidade. O Tesouro Unesp apresentou 42 identidades totais, 7 identidades parciais e 35 ocorrências sem identidade. O VC-COC registrou 10 identidades totais, 1 identidade parcial e 73 ocorrências sem identidade. O BRASED apresentou 14 identidades totais, 4 identidades parciais e 66 ocorrências sem identidade. Por fim, a Terminologia da Biblioteca Nacional registrou 31 identidades totais, 5 identidades parciais e 48 ocorrências sem identidade.

O vocabulário que representou mais palavras-chave com Identidade Total foi a LCSH (=45) seguido de DeCS =43, Tesouro UNESP (=42), VC-USP (=41), Terminologia da Biblioteca Nacional (BN) (=31), e MeSH (=28). Este resultado demonstra que os 4 vocabulários, LCSH, DeCS, Tesouro Unesp e VC-USP estão alinhados no índice quantitativo de representação de assuntos em linguagem natural atribuídos pelos pesquisadores da Fiocruz. Entretanto, observa-se que o maior índice quantitativo de identidade parcial é o do MeSH entre todos os vocabulários e entre os 4 vocabulários com mais identidade total. A quantidade de nenhuma identidade do grupo de 4 vocabulários é apresentada pelo VC-USP e a menor pela LCSH.

O VC-COC apresentou desempenho inferior ao DeCS e aos demais, refletindo sua ênfase em áreas específicas como Patrimônio Cultural, Memória, História, Arquitetura, Conservação, Comunicação e Divulgação Científica, ainda que os termos dessas áreas apareçam com menor frequência nas buscas dos pesquisadores da amostra analisada. O BRASED e a BN apresentaram um desempenho inferior em relação às demais fontes, indicando que poderiam ser menos adequadas para a pesquisa de termos/descriptores específicos das áreas de conhecimento específicas da Fiocruz, ainda assim interessantes como fontes positivas.

A área da saúde apresenta uma constante evolução em relação ao surgimento de novos termos, patologias e abordagens. Dessa forma, o uso do DeCS apresenta também o aspecto negativo no que diz respeito aos outros termos relacionados às demais áreas que também fazem parte dos estudos e



pesquisas da Fiocruz, ocasionando conforme gráfico 1, nenhuma identidade e identidade parcial para algumas palavras-chaves indicadas pelos participantes da pesquisa, que não foram contempladas utilizando o DeCS.

## 5 CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo investigar a adequação dos vocabulários controlados utilizados na Fiocruz para a representação de palavras-chave em linguagem natural atribuídas por seus pesquisadores, considerando o caráter progressivamente multidisciplinar da produção científica institucional. Esse objetivo foi alcançado por meio da comparação sistemática entre termos indicados por pesquisadores com pesquisas em andamento e diferentes vocabulários controlados, de caráter multidisciplinar e especializados, permitindo identificar níveis de correspondência e lacunas terminológicas.

De forma sintética, a pesquisa demonstrou que, embora a Fiocruz mantenha forte predominância nas áreas de Ciências da Saúde e Ciências Biológicas, sua produção científica incorpora, de maneira crescente, temas oriundos das Ciências Humanas, Ciências Sociais Aplicadas e de debates sociais mais amplos. Os resultados evidenciaram que vocabulários controlados de caráter multidisciplinar apresentaram maior aderência quantitativa às palavras-chave utilizadas pelos pesquisadores, sobretudo no que se refere às ocorrências de identidade total e parcial, quando comparados aos vocabulários especializados em Ciências da Saúde e Educação. Esse achado confirma que a diversidade temática da produção institucional não é plenamente contemplada pelo uso exclusivo do DeCS, atualmente adotado como principal instrumento de representação temática realizado pela Rede de Bibliotecas da Fiocruz em seu catálogo online.

Do ponto de vista crítico, os resultados indicam que, embora o DeCS seja um vocabulário consolidado e altamente eficaz para a padronização terminológica no campo biomédico, sua aplicação isolada mostra-se limitada



diante da complexidade e da interdisciplinaridade que caracterizam a produção científica contemporânea da Fiocruz. Nesse aspecto, é preciso aumentar a precisão com o aumento da representação de modo a evitar os documentos inúteis e alta revocação para assim recuperar os documentos úteis (Lancaster, 2004). Isso tendo em vista que vocabulários controlados buscam auxiliar o indexador e usuário na identificação dos conceitos relacionados, promovendo assim a consistência/coerência na indexação. Essas lacunas impactam diretamente a visibilidade, o acesso e a circulação do conhecimento produzido institucionalmente, sobretudo em um contexto marcado pela valorização da ciência aberta, da interoperabilidade entre sistemas e do acesso amplo à informação.

No que se refere às limitações da pesquisa, destaca-se o tamanho da amostra qualitativa, restrita a 15 (quinze) participantes, o que não permite generalizações absolutas sobre o comportamento informacional de toda a comunidade científica da Fiocruz. Além disso, a análise concentrou-se em um conjunto específico de vocabulários controlados e em palavras-chave fornecidas em determinado recorte temporal, não contemplando variações terminológicas ao longo do tempo nem diferenças mais profundas entre subáreas do conhecimento. Também não foram analisados aspectos relacionados à prática efetiva de indexação realizada pelos profissionais da informação, o que poderia enriquecer a compreensão do processo de mediação terminológica.

Como perspectivas futuras, a pesquisa aponta para a necessidade de aprofundar os estudos sobre a construção de instrumentos de representação temática mais adequados ao perfil multidisciplinar da Fiocruz. Recomenda-se a elaboração de um vocabulário controlado institucional, fundamentado em registros de autoridade de assunto, que integre descritores provenientes de múltiplos vocabulários, bem como a inclusão sistemática de novos termos identificados a partir da linguagem natural dos pesquisadores. Esse processo deve ser orientado por métodos consolidados no campo da Terminologia,



garantindo padronização, coerência semântica e interoperabilidade com outros sistemas de organização da informação. Dessa forma, a pesquisa contribui para as discussões institucionais sobre representação temática envolvendo campos multidisciplinares e sugere um passo seguinte para o caminho de aprimoramento das práticas de indexação e recuperação da informação da Fiocruz.

## REFERÊNCIAS

ALVES, L. S. S.; TARTAROTTI, R. C. Dal'Evedove; FUJITA, M. S. L. Avaliação de vocabulário controlado para a representação e recuperação de teses e dissertações em repositório institucional RICI: **Revista Ibero Americana de Ciência da Informação**, Brasília, DF, v. 15, n.1, jan. abr. 2022. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/download/42496/33461>. 07 set. 2026.

AUSTIN, D. **Vocabulary control and information technology**. Aslib Proceedings, London, v. 38, n. 1, p. 1-15, jan. 1986.

BOCCATO, V. R. C. A linguagem de indexação vista pelo conteúdo, forma e uso na perspectiva de catalogadores e usuários. *In*: Fujita, M. S. L. (org.). **A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias. Um estudo de observação do contexto sociocognitivo com protocolos verbais**. São Paulo: Cultura Acadêmica, 2009. p. 119-35.

BOCCATO, V. R. C. Os sistemas de organização do conhecimento nas perspectivas atuais das normas internacionais de construção. **InCID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 2, n. 1, p. 165-192, jan./jun. 2011. Disponível em: <https://revistas.usp.br/incid/article/download/42340/46011/50538>. 07 set. 2026.

BOCCATO, V. R. C.; FUJITA, M. S. L.; GIL LEIVA, I. Avaliação comparada do uso de linguagens de indexação em catálogos coletivos de bibliotecas universitárias para recuperação por assunto. **Scire**, Zaragoza, v.17, n.2, p.55-64, jul.-dic. 2011.

BRAGA, Kátia Soares. Aspectos relevantes para a seleção de metodologia adequada à pesquisa social em Ciência da Informação. *In*: MUELLER, Suzana Pinheiro Machado *et al.* (org.) **Métodos para a pesquisa em Ciência da Informação**. Brasília: Thesaurus, 2007. p. 17-38.



BRUNDAGE, C. A. Teaching controlled vocabulary and natural language to end-users. **Science & Technology Libraries**, Philadelphia, v. 10, n. 1, p. 3-13, Fall 1989.

CRESWELL, J. W. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto**. 3. ed. Porto Alegre: Artmed, 2010.

DESCRITORES EM CIÊNCIAS DA SAÚDE. **Sobre o DeCS/MeSH**. [Online]: Bireme, 2024. Disponível em: <https://decs.bvsalud.org/ths/treeView>. Acesso em: 20 fev. 2025.

FUJITA, M. S. L. A política de indexação e recuperação da informação. In: GIL LEIVA, I.; FUJITA, M. S. L. (org.). **Política de indexação**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2012. p. 17-28. DOI: <https://doi.org/10.36311/2012.978-85-7983-199-7>. p.17-28.

FUJITA, M. S. L.; MOREIRA, W. **Manual do planejamento, construção e manutenção do Tesauro Unesp para bibliotecas: do conceitual a prática**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2021. DOI: <https://doi.org/10.36311/2021.978-65-5954-069-3>.

FUNDAÇÃO OSWALDO CRUZ. Casa de Oswaldo Cruz. **Política de indexação dos acervos da Casa de Oswaldo Cruz**. Rio de Janeiro: Fiocruz/COC, 2018. 24 p.

GIL URDICIÁIN, B. **Manual de languages documentales**. 2. ed. rev. ampl. Gijón: Trea, 2004. 280 p.

GOLDENBERG, M. **A arte de pesquisar: como fazer pesquisa qualitativa em Ciências Sociais**. Rio de Janeiro: Record, 1997.

HARPRING, P. **Introdução aos vocabulários controlados: terminologia para arte, arquitetura e outras obras culturais**. São Paulo: Secretaria da Cultura do Estado; Pinoteca de São Paulo; ACAM Portinari, 2016. Disponível em: <https://www.sisemsp.org.br/wp-content/uploads/2023/03/vocabularios-controlados-digital.pdf>. Acesso em: 20 fev. 2025.

HENZLER, R. G. Free or controlled vocabularies. **International Classification**, München, v. 5, n. 1, p. 21-26, 32, mar. 1978.

HJØRLAND, B. What is Knowledge Organization (KO)? **Knowl. Org.**, [S.l.], v. 35, n. 2/3, p. 86-101, 2008.



INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964-1:2011 Information and documentation -- **Thesauri and interoperability with other vocabularies** -- Part 1: Thesauri for information retrieval. Geneva: International Organization for Standardization, 2011.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Brasília, DF: Briquet de Lemos, 1993.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília, DF: Briquet de Lemos, 2004.

LOPES, I. L. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ci. Inf.**, Brasília, v. 31, n. 1, p. 41-52, jan./abr, 2002.

MAZZOCCHI, F. Knowledge Organization Systems (KOS). **Knowledge Organization**, Baden-Baden, v. 45, p. 54-78, 2018. DOI:10.5771/0943-7444-2018-1-54.

MINAYO, M. C. de S. **O desafio do conhecimento: pesquisa qualitativa em saúde**. 14. ed. São Paulo: Hucitec Editora, 2014.

MOREIRO GONZÁLEZ, J. A. **El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural**. Gijón: Trea, 2004.

PEREZ, E. Text enhancement: controlled vocabularies vs free text. **Special Libraries**, Mount Laurel, NJ, v. 72, n. 3/4, p. 183-192, 1982.

RAITT, D. J. Recall and precision devices in interactive bibliographic search and retrieval systems. **Aslib Proceedings**, London, v. 32, n. 7/8, p. 281- 301, july/aug. 1980. Disponível em: <https://www.emerald.com/ajim/article-abstract/32/7/281/62895/Recall-and-precision-devices-in-interactive?redirectedFrom=fulltext>. 07 set. 2026.

RICHARDSON, R. J. *et al.* **Pesquisa social: métodos e técnicas**. 3. ed. São Paulo: Atlas, 1999.

SALTON, G. On the use of term associations in automatic information retrieval. *In: Coling 1986 (Volume 1: The 11th International Conference on Computational Linguistics, 1986)*.

SILVA, L. S. O.; OLIVEIRA, L. R. S.; LIMA, S. I. A. Estudo de uso do catálogo online da Universidade Federal de Alagoas. **Múltiplos Olhares em Ciência da Informação**, v.8, n. 2 esp, 2018. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/16903>. 07 set. 2026.



SVENONIOUS, E. Natural language vs-controlled vocabulary. *In: CANADIAN CONFERENCE ON INFORMATION SCIENCE*, 1976, Ontario. **Proceedings...** [S. l. : s. n.], 1976. p. 141-150.

## NOTAS

### **Adrienne Oliveira de Andrade da Silva** **Fundação Oswaldo Cruz (FIOCRUZ)**

**Minicurriculo:** Possui graduação em Biblioteconomia e Documentação pela Universidade Federal Fluminense (2016) e mestrado em Ciência da Informação pela Universidade Federal Fluminense (2019). Atua como bibliotecária na Biblioteca de História das Ciências e da Saúde da Casa de Oswaldo Cruz - Fundação Oswaldo Cruz desde 2018. É membro do GT de Indexação da Rede de Bibliotecas da Fiocruz, foi um dos membros responsáveis pela elaboração da Política de Indexação da Rede de Bibliotecas da Fiocruz e é coordenadora do SubGT de Vocabulário Controlado da Fiocruz. É gestora da Comunidade COC no Repositório Institucional Arca desenvolvendo ações de Ciência Aberta em sua unidade. Atua também como apoio à gestão da biblioteca.

**ORCID:** <https://orcid.org/0000-0002-9676-0412>

**Lattes:** <http://lattes.cnpq.br/1686104039491486>

**Email:** [adrienne.silva@fiocruz.br](mailto:adrienne.silva@fiocruz.br)

### **Débora da Silva Rocha** **Fundação Oswaldo Cruz (FIOCRUZ)**

**Minicurriculo:** Graduada em Biblioteconomia pela Universidade Federal do Amazonas e Pós-Graduada em Informática na Educação pelo Instituto Federal do Amazonas (IFAM). Bibliotecária Analista no Instituto Leônidas & Maria Deane (Fiocruz Amazônia) e atua na área de gestão da informação científica em contextos de saúde e pesquisa. É integrante do grupo de pesquisa GICA (Gestão da Informação e do Conhecimento na Amazônia).

**ORCID:** <https://orcid.org/0009-0000-3971-392X>

**Lattes:** <http://lattes.cnpq.br/2687438548642753>

**Email:** [debora.srocha@fiocruz.br](mailto:debora.srocha@fiocruz.br)

### **Julia Dias Mota Magaraia** **Fundação Oswaldo Cruz (FIOCRUZ)**

**Minicurriculo:** Possui graduação em Engenharia Civil pelo Centro Universitário de Patos de Minas - UNIPAM (2018). Pós-graduada em Engenharia de Segurança no Trabalho pela Universidade Federal do Rio de Janeiro - UFRJ (2019). Graduada em Biblioteconomia pela Universidade Federal do Estado do Rio de Janeiro - UNIRIO (2023). Atua como bolsista em biblioteconomia no Projeto Preservo, da Fiocruz. É membro do GT de Indexação da Rede de Bibliotecas da Fiocruz, foi um dos membros responsáveis pela Política de Indexação da Rede de Bibliotecas da Fiocruz.



**ORCID:** <https://orcid.org/0009-0007-4094-0450>

**Lattes:** <http://lattes.cnpq.br/1064490054622153>

**Email:** [julia.mota@fiocruz.br](mailto:julia.mota@fiocruz.br)

### **Mariângela Spotti Lopes Fujita**

#### **Universidade Estadual Paulista (UNESP)**

**Minicurrículo:** Docente permanente do Programa de Pós-Graduação em Ciência da Informação na Unesp, Brasil, onde realiza pesquisa e orienta alunos de mestrado e doutorado e supervisiona pesquisadores de pós-doutorado. É pesquisadora nível 1-B do Conselho Nacional de Ciência e Tecnologia (CNPq) na área de Ciência da Informação e seus temas de pesquisa estão em Indexação, Política de Indexação, Tesouro e Controle de vocabulário em sistemas de informação acadêmicos. Publicou numerosos artigos de pesquisa em revistas nacionais e internacionais e é parecerista de periódicos científicos e eventos acadêmicos nacionais e internacionais. Líder do Grupo de Pesquisa “Representação Temática da Informação”. Presidente da Comissão Permanente de Publicações e do Conselho de Editores de Periódicos Científicos da Faculdade de Filosofia e Ciências da Unesp.

**ORCID:** <https://orcid.org/0000-0002-8239-7114>

**Lattes:** <http://lattes.cnpq.br/6530346906709462>

**Email:** [mariangela.fujita@unesp.br](mailto:mariangela.fujita@unesp.br)

#### **TAXONOMIA CREDIT: N/A**

#### **LICENÇA DE USO**

CC BY-NC-ND.

#### **ENTIDADE EDITORA**

Associação Catarinense de Bibliotecários.

#### **EDITORADO POR:**

Beatriz Moraes Borges, Débora Crystina Dias Reis, Kariane Regina Laurindo, Marcelo Werneck de Souza Saraiva

#### **HISTÓRICO**

Recebido em: 22-07-2025 - Aprovado em: 27-07-2026

