

# RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta Tese será disponibilizado somente a partir de 24/02/2025.



Universidade Estadual Paulista “Júlio de Mesquita Filho”  
Faculdade de Filosofia e Ciências  
Campus de Marília

**JACQUELIN TERESA CAMPEROS REYES**

**ADERÊNCIA LEXICAL A DADOS PUBLICADOS PARA  
PRODUTORES RURAIS**

**Marília  
2023**



Universidade Estadual Paulista “Júlio de Mesquita Filho”  
Faculdade de Filosofia e Ciências  
Campus de Marília

**JACQUELIN TERESA CAMPEROS REYES**

**ADERÊNCIA LEXICAL A DADOS PUBLICADOS PARA PRODUTORES RURAIS**

Tese apresentada ao Programa de Pós-graduação em Ciência da Informação do Departamento de Ciência da Informação da Universidade Estadual Paulista Faculdade de Filosofia e Ciências como requisito para a obtenção do título de doutora em Ciência da Informação.

**Linha de Pesquisa:** Informação e Tecnologia

**Orientador:** Dr. Ricardo César Gonçalves Sant’Ana

**Marília  
2023**

C195a Camperos-Reyes, Jacquelin Teresa  
Aderência lexical a dados publicados para  
produtores rurais / Jacquelin Teresa  
Camperos-Reyes. -- Marília, 2023  
144 f. : il., tabs., mapas

Tese (doutorado) - Universidade Estadual Paulista  
(Unesp), Faculdade de Filosofia e Ciências, Marília  
Orientador: Ricardo César Gonçalves Sant'Ana

1. Aderência lexical. 2. Acesso a dados. 3. Dados  
de governo. 4. Necessidades informacionais. 5.  
Pequeno produtor. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp.  
Biblioteca da Faculdade de Filosofia e Ciências, Marília. Dados  
fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

JACQUELIN TERESA CAMPEROS REYES

ADERÊNCIA LEXICAL A DADOS PUBLICADOS PARA PRODUTORES RURAIS

Tese apresentada ao Programa de Pós-graduação em Ciência da Informação do Departamento de Ciência da Informação da Universidade Estadual Paulista Faculdade de Filosofia e Ciências como requisito para a obtenção do título de doutora em Ciência da Informação.

**Linha de Pesquisa:** Informação e Tecnologia

**Orientador:** Dr. Ricardo César Gonçalves Sant'Ana

Marília - SP, 23 de fevereiro de 2023

BANCA EXAMINADORA:

Profa. Dra. Ângela Maria Grossi  
Universidade Estadual Paulista (UNESP)

Profa. Dra. Cristiane Hengler Correa Bernardo  
Universidade Estadual Paulista (UNESP)

Profa. Dra. Luana Farias Sales  
Instituto Brasileiro de Informação em Ciência e  
Tecnologia (IBICT)

Prof. Dr. Guilherme Ataíde Dias  
Universidade Federal da Paraíba (UFPB)

Prof. Dr. Ricardo César Gonçalves Sant'Ana – Orientador  
Universidade Estadual Paulista (UNESP)

*Crear una tesis doctoral conlleva muchos sacrificios, uno de ellos es la obligada distancia con nuestros seres amados. Algunas distancias físicas desaparecen una vez termina el ciclo, otras se mantienen porque algunos de nuestros tesoros ahora viven en la Morada Divina, y quizás, de esa forma, la distancia ya no es una limitante, pues desaparece, y se transforma en una fuerza con poderes indescriptibles. Esa fuerza que me motivó desde el primer minuto de existencia y durante todo el tiempo que tuvo cálido aliento, aún me impulsa a servir a quien pueda y deba hacer.*

*Por eso, este trabajo está dedicado a honrar la memoria de  
Mery Edilia Reyes de Camperos, mi madre.*

## **AGRADECIMIENTOS**

*Con alegría pienso en todas las personas y entidades a las cuales les agradezco por favorecer este recorrido académico y profesional.*

*A César Augusto Gutiérrez Yañez, mi esposo, amigo, compañero y paño de lágrimas, ya sé de qué está hecha esta relación. A María Juliana Gutiérrez Camperos y Juan Diego Gutiérrez Camperos, mis hijos, seres humanos maravillosos que me inspiran, me ayudan, me calman y me engrandecen como madre. A don Rafael Antonio Camperos Higuera, mi padre, fuerte guía y patrón de comportamiento, a doña Mery, por heredarme resiliencia e incentivarme a estudiar. A mis hermanos, Nataly Eliana, Freddy Rafael y Jhonatan Eduardo Camperos Reyes, a mis cuñadas Alexandra Sanchez y Sonia Cabarcas, a mis sobrinos, Salomé, Ángel, Massimiliano, Adrián e Isaac, todos ciertamente me ayudaron. A Doris, Guty, Juan Camilo, Kinino, porque me acompañaron, me enseñaron a servir y a tener paciencia. A mis tíos Celina, Josefa, Stella, Mercedes, Samuel y Wilson, porque siempre desean cosas buenas y oran por nosotros. A mis primos Daniel y Jefry, por ser ejemplos de la familia que nos une.*

*A mi orientador, profesor y consejero, Ricardo César Gonçalves Sant'Ana, brillante profesional y gran ser humano, que brinda a sus alumnos la confianza que precisamos, para ser agentes de soluciones en una sociedad sedienta por profesionales integrales y justos, él se convirtió en un soporte para poder concluir este proceso.*

*A los profesores Luana Farias Sales Marques, Moisés Lima Dutra, Angela Maria Grossi, Cristiane Hengler Corrêa Bernardo, Guilherme Ataíde Dias, por sus sugerencias y valiosos aportes dados en las bancas de cualificación y defensa.*

*A mis amigos Neusa Teixeira dos Santos y familia, Mário Rui Andrade de Moura y familia, Luiz Esteves y Valdercina, Plínio Aparecido Rosetti, Maira Nani França Moura Goulart y familia, Yolanda Perdomo Villalba, Mauricio Zambrano Iglesias, Diana Vilas Boas Souto Aleixo, Elaine Parra Affonso, Fábio Mosso Moreira, Fernando de Assis Rodrigues, Cristian Berrío-Zapata, Ester Ferreira da Silva, Liliana Giusti Serra, Luiza de Menezes Romanetto, Felipe Arakaki, Juan Bernardo Montoya Mogollón y Sandra Milena Roa Martinez, más que palabras, mi cariño por lo mucho que sustentaron este camino. Gracias!*

*A mis colegas del Grupo de Investigación GPTAD por el acogimiento, soporte y las buenas experiencias que continuamente nos hacen crecer juntos; a todos los profesores y funcionarios del programa de posgraduación en Ciencia de la Información de la UNESP, en particular a los profesores Marta Lígia Pomim Valentim, Helen de Castro Silva Casarin y Carlos Cândido de Almeida por su convencimiento de que podemos y debemos participar de iniciativas que la universidad nos brinda.*

*A la profesora Ana Reyes Pacios Lozano y a todos los profesores del programa de doctorado en Documentación de la Universidad Carlos III de Madrid, por acogerme durante mi jornada de capacitación del programa de internacionalización CAPES-PrInt.*

*A la república de Brasil, representada por el Ministério de Educação, y a la Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES, por el apoyo financiero para este estudio doctoral<sup>1</sup>.*

*Dios no está en el último párrafo, Él nos trajo hasta aquí, nos sustentó, nos protegió e hizo esto posible, entonces, siendo principio y fin, con su gracia, cada paso de este recorrido fue posible.*

---

<sup>1</sup> O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## **IMPACTOS RELACIONADOS AOS OBJETIVOS DO DESENVOLVIMENTO SUSTENTÁVEL (ODS)**

Esta tese contribui com o ODS 2, “Acabar com a fome, alcançar a segurança alimentar e melhoria da nutrição e promover a agricultura sustentável”, particularmente na meta interna 2.3 do Instituto de Pesquisa Econômica Aplicada (IPEA), “Até 2030, aumentar a produtividade agrícola e a renda dos pequenos produtores de alimentos, particularmente de mulheres, agricultores familiares, povos e comunidades tradicionais, visando tanto à produção de autoconsumo e garantia da reprodução social dessas populações quanto ao seu desenvolvimento socioeconômico, por meio do acesso seguro e equitativo”. Ainda contribui com o ODS 16, “Paz, Justiça e Instituições Eficazes”, particularmente na meta 16.10: “Assegurar o acesso público à informação e proteger as liberdades fundamentais, em conformidade com a legislação nacional e os acordos internacionais”.

## **IMPACTS RELATED TO THE SUSTAINABLE DEVELOPMENT GOALS (SDG)**

This thesis contributes to SDG 2, “Zero hunger, achieve food security and improved nutrition and promote sustainable agriculture”, particularly in the internal target 2.3 of the Institute for Applied Economic Research (IPEA), “By 2030, increase the agricultural productivity and income of small food producers, particularly women, family farmers, traditional peoples and communities, aiming both at self-consumption production and guaranteeing the social reproduction of these populations and at their socioeconomic development, through safe and equitable access” . It also contributes to SDG 16, “Peace, Justice and Strong Institutions”, particularly in target 16.10: “Ensure public access to information and protect fundamental freedoms, in accordance with national legislation and international agreements” .

## **IMPACTOS RELACIONADOS CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE (ODS)**

Esta tesis contribuye con el ODS 2, “Poner fin al hambre, lograr la seguridad alimentaria y la mejora de la nutrición y promover la agricultura sostenible”, particularmente en la meta interna 2.3 del Instituto de Investigación Económica Aplicada (IPEA), “Para 2030, aumentar la productividad agrícola y los ingresos de pequeños productores de alimentos, particularmente mujeres, agricultores familiares, pueblos y comunidades tradicionales, visando tanto la producción para el autoconsumo como garantizando la reproducción social de estas poblaciones y su desarrollo socioeconómico, a través del acceso seguro y equitativo”. También contribuye al ODS 16, “Paz, Justicia e Instituciones Sólidas”, particularmente en la meta 16.10: “Garantizar el acceso público a la información y proteger las libertades fundamentales, de conformidad con las leyes nacionales y los acuerdos internacionales”.

## RESUMO

A conjuntura contemporânea de abertura de dados governamentais se manifesta em uma ampla gama de opções de recursos disponíveis. O poder público, diante da dinâmica do acesso à informação, publica conjuntos de dados, que ao vir de todas as esferas orgânicas, implica em uma riqueza pelos assuntos tratados. Os usuários dos dados, com as suas características estruturais e contextuais, têm cada vez mais possibilidades de utilização de dados disponibilizados. Este processo de investigação identifica como relevante o usufruto pelo uso de dados por parte de pequenos e médios produtores conforme as suas demandas informacionais, focando na transdução de conteúdos da coleta à recuperação de dados sobre agricultura e desenvolvimento rural, publicados em sites de dados de governo. Parte-se da premissa de que para que os dados sejam proveitosos, é necessário que eles estejam publicados com características que favoreçam a sua interpretação. A pergunta que norteia a pesquisa é: Qual é a Aderência Lexical da fase de recuperação de dados, do detentor governo, com o campo informacional do sujeito alvo, identificado como produtores rurais, considerando as suas necessidades informacionais? Para responder a tal inquietação, foi proposto como objetivo geral Verificar a viabilidade de um modelo que reconheça a Aderência Lexical a dados publicados pelo detentor governo, usando necessidades informacionais de produtores rurais como marco de análise. A abordagem metodológica inclui pesquisa documental, revisão sistemática de literatura, coleta automatizada de dados textuais e análises de similaridade com linguagem R. As fontes de dados escolhidas foram conjuntos de dados de governo e notícias publicadas para grupos de produtores rurais. Os resultados mostram que o modelo de análise da Aderência Lexical proposto é um instrumento útil para identificar o potencial de interpretação de pequenos e médios produtores, pois permite reconhecer a proximidade entre dados publicados pelo detentor governo e os sujeitos alvo considerando as suas necessidades informacionais. É possível apontar que, na amostra escolhida, existe uma comunicação insuficiente, pois, das cinco categorias de necessidades ponderadas, houve proximidade unicamente com uma, Crédito. Desde o ponto de vista da similaridade entre as duas fontes, é entendida como favorável em um primeiro nível na busca de inteligibilidade entre estas, sem deixar de assinalar a existência de palavras que precisam de esforços adicionais para melhor contextualização e possível interpretação por parte dos sujeitos alvo. Estudos futuros objetivam contrastar a proposta com outras medidas de similaridade, reprodução em outras fontes de dados e em outros contextos socioeconômicos.

**Palavras-chave:** Aderência Lexical; Acesso a dados; Dados de governo; Necessidades informacionais; Pequeno produtor.

## ABSTRACT

The contemporary situation of open government data is manifested in a wide range of resource options available. The public power, faced with the dynamics of access to information, publishes sets of data, which, when coming from all organic spheres, imply a richness for the subjects dealt with. Data users, with their structural and contextual characteristics, have more and more possibilities to use the available data. This investigation process identifies as relevant the use of data by small and medium producers according to their informational demands, focusing on the transduction of content from the collection to the recovery of data on agriculture and rural development, published on government data sites. It starts from the premise that for the data to be useful, it is necessary that they are published with characteristics that favor their interpretation. The question that guides the research is: What is the Lexical Adherence of the data recovery phase, of the government holder, with the informational field of the target element, identified as rural producers, considering their informational needs? To respond to this concern, it was proposed as a general objective to verify the viability of a model that recognizes Lexical Adherence to data published by the government holder, using informational needs of rural producers as an analysis framework. The methodological approach includes documentary research, systematic literature review, automated collection of textual data and similarity analysis with R language. The chosen data sources were government datasets and news published for groups of rural producers. The results show that the proposed Lexical Adherence analysis model is a useful instrument to identify the interpretation potential of small and medium-sized producers, as it allows recognizing the proximity between data published by the government holder and the target subjects, considering their informational needs. It is possible to point out that, in the chosen sample, there is insufficient communication, since, of the five weighted categories of needs, there was proximity to only one, Credit. From the point of view of the similarity between the two sources, it is understood as favorable at a first level in the search for intelligibility between them, while noting the existence of words that need additional efforts for better contextualization and possible interpretation by the subjects. target. Future studies aim to contrast the proposal with other measures of similarity, reproduction in other data sources and in other socioeconomic contexts.

**Keywords:** Lexical Adherence; Data access; Government data; Informational needs; Small Producer.

## RESUMEN

La coyuntura de datos abiertos del gobierno se manifiesta en una amplia gama de opciones en recursos disponibles. El poder público ante la dinámica de acceso a la información, publica conjuntos de datos, que, provenientes de todas las esferas orgánicas, implican una riqueza en cuanto a los temas tratados. Los usuarios de datos, con sus características estructurales y contextuales, tienen cada vez más posibilidades de utilizar los datos disponibles. Este proceso de investigación identifica como relevante el uso de datos por parte de pequeños y medianos productores de acuerdo con sus demandas informacionales, enfocándose en la transducción de contenidos desde la recolección hasta la recuperación de datos sobre agricultura y desarrollo rural, publicados en sitios de datos gubernamentales. Parte de la premisa de que para que los datos sean útiles es necesario que se publiquen con características que favorezcan su interpretación. La pregunta que orienta la investigación es: ¿Cuál es la Adherencia Lexical de la fase de recuperación de datos, del detentor gobierno, con el campo informacional del elemento objetivo, identificados como productores rurales, considerando sus necesidades informacionales? Para dar respuesta a esta inquietud, se planteó como objetivo general verificar la viabilidad de un modelo que reconozca la Adherencia Lexical a datos publicados por el detentor gobierno, utilizando como marco de análisis las necesidades informacionales de los productores rurales. El enfoque metodológico incluye investigación documental, revisión sistemática de literatura, recolección automatizada de datos textuales y análisis de similitud con lenguaje R. Las fuentes de datos elegidas fueron conjuntos de datos gubernamentales y noticias publicadas para grupos de productores rurales. Los resultados muestran que el modelo de análisis de Adherencia Lexical propuesto es una herramienta útil para identificar el potencial de interpretación de los pequeños y medianos productores, ya que permite reconocer la proximidad entre los datos publicados por el detentor gobierno y los sujetos informacionales, considerando sus necesidades. Es posible señalar que, en la muestra escogida, hay comunicación insuficiente, ya que, de las cinco categorías ponderadas de necesidades, hubo proximidad solo a una, Crédito. Desde el punto de vista de la similitud entre ambas fuentes, se entiende favorable en un primer nivel en la búsqueda de la inteligibilidad entre ellas, sin olvidar señalar la existencia de palabras que necesitan esfuerzos adicionales para una mejor contextualización y posible interpretación por parte de los sujetos informacionales. Futuros estudios pretenden contrastar la propuesta con otras medidas de similitud, reproducción en otras fuentes de datos y en otros contextos socioeconómicos.

**Palabras-clave:** Adherencia Lexical; Acceso a datos; Datos gubernamentales; Necesidades Informacionales; Pequeño productor.

## ÍNDICE DE FIGURAS

|                                                                                                    |     |
|----------------------------------------------------------------------------------------------------|-----|
| Figura 1 - Localização Cúcuta-Colômbia e relação com localização de Marília-Brasil .....           | 20  |
| Figura 2 - Palavras-chave em trabalhos em eventos, capítulos de livros e artigos.....              | 26  |
| Figura 3 - Aderência lexical entre duas fontes de dados.....                                       | 35  |
| Figura 4 - Diagrama estrutural da tese.....                                                        | 40  |
| Figura 5 - Governança política de dados abertos no Brasil .....                                    | 52  |
| Figura 6 - Modelo de dados do DCAT .....                                                           | 55  |
| Figura 7 - Percorso metodológico para determinar categorias de necessidades<br>informacionais..... | 81  |
| Figura 8 - Distribuição do número de citas versus a quantidade de autores .....                    | 83  |
| Figura 9 - Fluxograma da RSL.....                                                                  | 84  |
| Figura 10 - Diagrama de associação frequências relativas necessidades informacionais. ....         | 91  |
| Figura 11 - Descritores sobre categorias informacionais mais frequentes.....                       | 92  |
| Figura 12 - Diagrama coleta e análise de dados no site da CNA e no dados.gov.br .....              | 109 |
| Figura 13 - Histograma com bigramas mais frequentes no CNA.....                                    | 115 |
| Figura 14 - Histograma com trigramas mais frequentes no CNA .....                                  | 117 |
| Figura 15 - Histograma com bigramas mais frequentes em dados.gov.br.....                           | 120 |
| Figura 16 - Histograma com trigramas mais frequentes em dados.gov.br.....                          | 121 |
| Figura 17 - Aderência Lexical instanciada nesta tese .....                                         | 131 |

## ÍNDICE DE QUADROS

|                                                                                                         |     |
|---------------------------------------------------------------------------------------------------------|-----|
| Quadro 1 - Lista de produções científicas durante o mestrado em Ciência da Informação ..                | 23  |
| Quadro 2 - Publicação de artigos científicos durante o período de doutorado 2018-2022....               | 25  |
| Quadro 3 - Protocolo da RSL.....                                                                        | 42  |
| Quadro 4 - Exemplos do uso de dados disponibilizados em sites de dados abertos de governos.....         | 58  |
| Quadro 5 - Classificação de artigos por dimensões de análise e eixos temáticos.....                     | 68  |
| Quadro 6 - Protocolo da RSL.....                                                                        | 80  |
| Quadro 7 - Estudos compreendidos na RSL.....                                                            | 85  |
| Quadro 8 - Análise de aderência entre necessidades informacionais no acesso a dados na agricultura..... | 88  |
| Quadro 9 - Cálculo de frequências relativas das fontes do estudo.....                                   | 90  |
| Quadro 10 - Instituições gremiais identificadas.....                                                    | 104 |
| Quadro 11 - Datasets recuperados .....                                                                  | 110 |

## LISTA DE ABREVIATURAS E SIGLAS

|         |                                                                             |
|---------|-----------------------------------------------------------------------------|
| 5W1H    | What, Where, Who, When, Why, How                                            |
| ACI     | Aliança Cooperativa Internacional                                           |
| ARIS    | Agricultural Research Information Systems                                   |
| ATER    | Assistência Técnica e Extensão Rural                                        |
| BRAPCI  | Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação |
| BRP     | Bacia do Rio da Prata                                                       |
| BSM     | Brasil sem miséria                                                          |
| CADSOL  | Cadastro Nacional de Empreendimentos Econômicos Solidários                  |
| CAPES   | Coordenação de Aperfeiçoamento de Pessoal de Nível Superior                 |
| CEPEA   | Centro de Estudos Avançados em Economia Aplicada                            |
| CGU     | Controladoria-Geral da União                                                |
| CKAN    | Comprehensive Knowledge Archive Network                                     |
| CNA     | Confederação da Agricultura e Pecuária do Brasil                            |
| CoDAF   | Competências Digitais para Agricultura Familiar                             |
| CSS     | Cascading Style Sheets                                                      |
| DANE    | Departamento Administrativo Nacional de Estadística                         |
| DCAT    | Data Catalog Vocabulary                                                     |
| DCSOL   | Declaração de Empreendimento Econômico Solidário                            |
| DERI    | Digital Enterprise Research Institute                                       |
| DIEESE  | Departamento Intersindical de Estatística e Estudos Socioeconômicos         |
| EES     | Empreendimentos Econômicos Solidários                                       |
| ENANCIB | Encontro Nacional de Pesquisa em Ciência da Informação                      |
| EPA     | Agência de Proteção Ambiental                                               |
| ESALQ   | Escola Superior de Agricultura Luiz de Queiroz                              |
| FAO     | Organização das Nações Unidas para a Alimentação e a Agricultura            |
| GLD     | Government Linked Data do W3C                                               |
| GPNTI   | Novas Tecnologias em Informação                                             |
| GPTAD   | Grupo de Pesquisa Tecnologias de Acesso a Dados                             |
| HLPF    | High-level Political Forum                                                  |
| HRSA    | Administração de Serviços de Recursos de Saúde                              |
| IBGE    | Instituto Brasileiro de Geografia e Estatística                             |
| IBICT   | Instituto Brasileiro de Informação em Ciência e Tecnologia                  |
| INDA    | Infraestrutura Nacional de Dados Abertos                                    |
| IPEA    | Instituto de Pesquisa Econômica Aplicada                                    |
| ODP     | Open Data Platform                                                          |
| ODS     | Objetivos do Desenvolvimento Sustentável                                    |
| ODS     | Objetivos de Desarrollo Sostenible                                          |

|         |                                                                                                                       |
|---------|-----------------------------------------------------------------------------------------------------------------------|
| OGP     | Open Government Partnership                                                                                           |
| OKF     | Open Knowledge Foundation                                                                                             |
| ONESC   | Observatório Nacional da Economia Solidária e do Cooperativismo                                                       |
| ONU     | Organização das Nações Unidas                                                                                         |
| PIB     | Produto Interno Bruto                                                                                                 |
| PICO    | Population, Intervention, Comparison, Outcomes                                                                        |
| PNATER  | Política Nacional de Assistência Técnica e Extensão Rural para a Agricultura Familiar e Reforma Agrária               |
| PRISMA  | Preferred Reporting Items for Systematic reviews and Meta-Analyses                                                    |
| ReCoDAF | Revista Competências Digitais para Agricultura Familiar                                                               |
| RSL     | Revisão Sistemática de Literatura                                                                                     |
| SDG     | Sustainable Development Goals                                                                                         |
| SECIN   | Seminário em Ciência da Informação                                                                                    |
| SENA    | Servicio Nacional de Aprendizaje                                                                                      |
| SENAES  | Secretaria Nacional de Economia Solidária                                                                             |
| SENAR   | Serviço Nacional de Aprendizagem Rural                                                                                |
| SIATER  | Sistema informatizado de ATER                                                                                         |
| SIMOG   | Sistema de monitoramento e Gestão da Secretaria Especial de Agricultura Familiar e do Desenvolvimento Agrário da SEAD |
| TIC     | Tecnologias da informação e da comunicação                                                                            |
| UFPB    | Universidade Federal da Paraíba                                                                                       |
| UFPS    | Universidad Francisco de Paula Santander                                                                              |
| UNESP   | Universidade Estadual Paulista                                                                                        |
| UPFS    | Unidades produtivas familiares                                                                                        |
| USP     | Universidade de São Paulo                                                                                             |
| WoS     | Web of Science                                                                                                        |
| WOUGNET | Women of Uganda Network                                                                                               |

## SUMÁRIO

|                                                                                                    |           |
|----------------------------------------------------------------------------------------------------|-----------|
| <b>PREFÁCIO .....</b>                                                                              | <b>20</b> |
| <b>1. INTRODUÇÃO .....</b>                                                                         | <b>28</b> |
| 1.1 Caracterização do problema.....                                                                | 30        |
| 1.2 Justificativa .....                                                                            | 36        |
| 1.3 Objetivos .....                                                                                | 39        |
| 1.4 Metodologia .....                                                                              | 40        |
| 1.5 Escopo.....                                                                                    | 47        |
| 1.6 Estrutura do documento .....                                                                   | 48        |
| <b>2. CONTEXTO DA PUBLICAÇÃO DE DADOS DE GOVERNO NO BRASIL .....</b>                               | <b>49</b> |
| <b>3. PERSPECTIVA DE PESQUISA SOBRE O ACESSO A DADOS NA ECONOMIA<br/>SOLIDÁRIA NO BRASIL .....</b> | <b>61</b> |
| 3.1 Introdução .....                                                                               | 62        |
| 3.2 Metodologia .....                                                                              | 63        |
| 3.3 Economia Solidária no mundo e no Brasil.....                                                   | 64        |
| 3.4 Dados Coletados E Discussão .....                                                              | 68        |
| 3.5 Considerações .....                                                                            | 71        |
| 3.6 Referências.....                                                                               | 72        |
| APÊNDICE A.....                                                                                    | 76        |
| <b>4. NECESSIDADES INFORMACIONAIS E ACESSO A DADOS NO CONTEXTO DA<br/>AGRICULTURA.....</b>         | <b>78</b> |
| 4.1 Introdução .....                                                                               | 78        |
| 4.2 Descrição do percurso metodológico .....                                                       | 79        |
| 4.3 Resultados da Revisão Sistemática de Literatura.....                                           | 82        |
| 4.4 Resultados e discussão da Meta-Análise .....                                                   | 89        |
| 4.5 Considerações .....                                                                            | 93        |
| 4.6 Referências.....                                                                               | 94        |

|                                                                                                     |            |
|-----------------------------------------------------------------------------------------------------|------------|
| <b>5. USO DA MINERAÇÃO DE TEXTO EM FONTES INFORMACIONAIS PARA GRUPOS DE PRODUTORES RURAIS .....</b> | <b>97</b>  |
| 5.1 Introdução .....                                                                                | 98         |
| 5.2 Metodologia .....                                                                               | 99         |
| 5.3 Mineração textual como técnica de análise.....                                                  | 101        |
| 5.4 Resultados e discussões .....                                                                   | 103        |
| 5.5 Considerações .....                                                                             | 122        |
| 5.6 Referências.....                                                                                | 124        |
| APÊNDICE B.....                                                                                     | 128        |
| <b>6. CONCLUSÕES .....</b>                                                                          | <b>130</b> |
| <b>REFERÊNCIAS .....</b>                                                                            | <b>137</b> |

## 1. INTRODUÇÃO

Estudos do fenômeno da informação e seus regimes no contexto das práticas sociais e públicas, pretendem conciliar com as realidades econômicas, políticas e culturais que caracterizam grupos atingidos por elas, o que implica, entre outros, o aproveitamento dessa informação (FROHMANN, 2008). “A importância que a informação assumiu na atualidade pós-industrial, recoloca para o pensamento questões sobre a sua natureza, seu conceito e os benefícios que pode trazer ao indivíduo e no seu relacionamento com o mundo em que vive.” (BARRETO, 1994, p.1).

Diante disso, particularmente no contexto dos benefícios das relações entre atores dos fluxos de informação, este estudo localiza-se no âmbito da disponibilização de dados e considera aspectos do acesso a esse volume de recursos que a cada instante é criado. Sant’Ana (2016) aponta que

a ciência da informação pode e deve contribuir para que este cenário de acesso e uso intenso de dados se desenvolva da melhor maneira possível, buscando identificar e estudar fatores e características que propiciem ampliação do equilíbrio entre os atores envolvidos no processo e a máxima otimização do uso dos dados (SANT’ANA, 2016, p.119).

Entende-se que o volume de dados disponíveis versus os benefícios obtidos por práticas que os usam, manifesta-se em magnitudes diferentes, densidades de dados que podem ser aproximadas, onde, se bem uma integração plena entre as densidades dos dados seja utópica, ações que contribuam com a conciliação de realidades e o fenômeno informacional, observando demandas dos sujeitos alvos, concebem-se como necessárias.

A desconformidade de agendas entre detentores de dados e os seus usuários, atores imersos em regímenes de informação que são objeto de estudo pela Ciência da Informação, configura uma assimetria que pode implicar na diminuição do valor dos dados que estão sendo disponibilizados.

Sabendo que a intencionalidade dos detentores dos dados publicados pode estar diferenciada na sua essência da intencionalidade no uso dos dados por parte do

grupo alvo<sup>4</sup>, o uso por parte desses interessados é feito na consciência de poder aproveitar esses ativos nas realidades socioeconômicas particulares.

Assim, o foco da pesquisa está no acesso, por parte de determinados grupos, a recursos informacionais importantes para o desenvolvimento socioeconômico de diferentes regiões. Para efeitos desta investigação, consideram-se como recursos informacionais os dados gerados e publicados no âmbito público destinados a produtores rurais do Brasil.

O distanciamento entre atributos de dados disponíveis e os sujeitos alvo, e a necessidade de contribuir com subsídios para o desenvolvimento de estratégias que permitam usar efetivamente esses dados, focando o acesso no âmbito da agricultura particularmente por pequenos e médios produtores, motivam o desenvolvimento dessa pesquisa. Ademais, reflexões promovidas pela pesquisa na área da Ciência da Informação originaram uma sensibilização com aspectos técnicos e sociais que permeiam os fluxos de informação, bem com o aproveitamento de recursos crescidos em variedade e volume.

Certamente, o problema no qual se espera contribuir de forma positiva está localizado nas camadas de abstração a dados publicados em sites de dados de governos, pois de parte dos detentores implicam-se

[...] não só competências técnicas específicas como, ainda, conhecimento sobre os significados envolvidos com os conteúdos tratados e objetivos desejados, criando uma camada de opacidade sobre as operações modeladas, implementadas e realizadas por parte dos detentores e intermediários ao construir e reconstruir as tessituras das denotações e das conotações. (SANT'ANA, 2019, p.121).

Por isso, partindo da afirmação de que as Tecnologias de Informação e Comunicação se configuram como agentes de desenvolvimento em áreas rurais (FERRÁS SEXTO; GARCÍA; POSE, 2011) e que elas adquirem o caráter de usáveis dependendo dos atributos de acesso a elas (GURSTEIN, 2007), esta pesquisa aponta à importância de que conjuntos de dados que estejam sendo publicados pelo detentor governo para grupos de produtores rurais, realmente atinjam esses sujeitos alvo. Embora o pressuposto de que tal disponibilização possa estar impregnada de ruídos,

---

<sup>4</sup> A referência à intencionalidade, extraída da obra de Frohmann (2008), está aqui expressa como característica do cenário tradicional de disseminação de informação onde segundo o autor ela “está ausente na geração de um vasto conjunto de enunciados digitais” (FROHMANN, 2008, p.30).

acredita-se na relevância de discutir oportunidades para que os dados possam ser usados adequadamente pelos sujeitos.

Gurstein (2007) aponta que grupos de usuários com particularidades que os distanciam de um uso efetivo das TIC, os deixam em dissonância de possibilidades de desenvolvimento local, além de melhores condições de justiça social e de empoderamento político.

Este processo de investigação estima relevante a potência do uso de dados e informações por produtores, enquadrada nas demandas informacionais de tais conglomerados, focando a interface em que esses dados são apresentados pelo detentor governo, uma vez que, efetuada a transdução<sup>5</sup> de conteúdos da coleta à recuperação (SANT'ANA, 2019), considera-se importante verificar a proximidade entre esses dados disponíveis e os sujeitos alvos. O esforço está no aprimoramento das condições de acesso a dados significativos para a sustentabilidade de um grupo da sociedade, especificamente, produtores rurais.

### **1.1 Caracterização do problema**

O estudo de metodologias e instrumentos para abordar ações relacionadas, no sentido amplo, com dados como recursos informacionais, configura uma das áreas de interesse da Ciência da Informação.

Segundo o contexto em que eles surgem e as necessidades dos sujeitos que os detém ou quem deles fará uso, poderão ter acontecido as mencionadas ações conforme o fim esperado, seja localizar, armazenar, preservar etc. Assim, na compreensão de que dados inacessíveis podem perder sua relevância e incorrer em uma perda de esforços de quem os produziu, estima-se necessário que eles sejam atingidos com análises de elementos relacionados com seu acesso, e dessa forma servir aos sujeitos que deles precisam.

Na perspectiva desses sujeitos alvos, entende-se que demandam o uso desses recursos para fundamentar processos estratégicos como tomadas de decisões, por isso, considerar características da recuperação desses recursos pode implicar no melhor uso dos dados acessados desde o ponto de vista do sujeito que os precisa.

---

<sup>5</sup> O autor Sant'Ana (2019) conceitua transdução como o processo de transformação de um tipo de sinal em outro, aplicável a distintas energias; o processo, discernido no âmbito desta área de estudo, foi intitulado como transdução informacional, e implica no acompanhamento de processos de persistência, medições, atributos de representações dos conteúdos e suas interfaces, perfiladas como próximas às características dos sujeitos alvo.

Esse estudo localiza-se no contexto de dados gerados no âmbito público, isto é, dados publicados pelo detentor governo. A dinâmica da publicação de dados de governo foi potencializada no contexto da sociedade da informação. Ela tornou-se substancial para que os cidadãos pudessem enxergar-se como agentes ativos das realidades que os circundam. Foi uma possibilidade de ver-se como parte do ciclo informacional enunciado por Barreto (1998): informação - conhecimento - desenvolvimento - informação.

No entendimento de que o acesso à informação implica na melhora da qualidade de vida dos sujeitos, ao quebrar com exclusões sistêmicas que os limitam em opções que possam transformar o entorno ao que pertencem (CARVALHO, 2010; RUEDIGER, 2006), a possibilidade da abertura dos dados de governo veio como início para a construção de uma cidadania digital (CARVALHO, 2010).

Alavancado o acesso a dados de governo no contexto internacional, em 1998 o Parlamento Europeu mediante a *Diretiva 2003/98* (PARLAMENTO EUROPEU, 2003), assinalava um caminho respeito ao fortalecimento da institucionalidade dos seus governos membros, cuidando para que essa força começasse pelo empoderamento da sociedade, agindo assim em sincronia com o pensamento de Saracevic (1996) quanto à busca do equilíbrio da equação homem-tecnologia, visando esse equilíbrio a partir do homem e não da tecnologia em si (CAMPEROS-REYES, 2018).

A diretiva citada<sup>6</sup> indica que a disposição geral dos documentos nos governos, “[...] não só relativos aos atos políticos, mas também aos processos judiciais e administrativos [...]” (PARLAMENTO EUROPEU, 2003), constitui um instrumento fundamental da democracia dos governos. O princípio do livre acesso à informação governamental, deve aplicar-se em todos os níveis e esferas dos governos, assegurando métodos práticos que auxiliem os usuários na utilização e reutilização da informação publicada (CAMPEROS-REYES, 2018).

No mesmo sentido, cabe mencionar que na busca pelo aprimoramento do acesso a ativos informacionais, a iniciativa de caráter mundial nomeada Parceria de Governo Aberto (Open Government Partnership - OGP), busca desde 2011 estabelecer compromissos concretos por parte de governos que estejam interessados em estimular a disseminação aberta de suas ações, promovendo esforços contra a

---

<sup>6</sup> Diretiva revisada com a 2013/37/EU.

corrupção e melhorando condições de acesso aos ativos informacionais gerados por eles (OPEN GOVERNMENT PARTNERSHIP, 2021).

É por isso que a OGP tem criado estratégias com mudanças nos modelos de administração dos governos participantes, onde eles se comprometem de diversas formas com a sociedade que representam, abrindo portas dos órgãos oficiais que os compõem com o intuito de facilitar o acesso a informações públicas, e promover a utilização desses ativos informacionais por parte dos seus cidadãos.

Em um primeiro momento o Brasil presidiu, junto aos Estados Unidos, a OGP, e naquele momento com o México reconheceram-se como países latino-americanos pioneiros na estratégia de abertura de ativos informacionais de governos (OPEN GOVERNMENT PARTNERSHIP, 2021).

Essas iniciativas que buscam o acesso a dados gerados pelos governos e pelas empresas privadas que com eles se relacionam administrando recursos públicos, utilizam o encorpado tecnológico como base para acrescentar as repercussões dessa abertura na esfera global. Verhulst e Young (2016) identificam quatro tópicos nos quais o acesso a dados pode contribuir com o desenvolvimento dos países que decidam somar-se: educação, corrupção governamental, oportunidades econômicas e respostas a desastres.

No contexto dos países em desenvolvimento, se uma opção que se encontra ao interior de cada governo, como os dados governamentais, pode abrir as portas à competitividade e ao desenvolvimento, aliás, dentro de uma conjuntura tecnológica favorável pela abrangência e escopo global, dita opção poderia estar direcionada a áreas que de fato já se encontram contribuindo no corpus econômico e social de cada país. É o caso do domínio da agricultura e do desenvolvimento rural.

Logo, repara-se na conveniência de estudos realizados no âmbito da Ciência da Informação, a fim de examinar as condições de acesso às TIC de modo que estas possam ser usadas e úteis, não somente em países desenvolvidos, mas também em vias do desenvolvimento, na perspectiva de que elas coadjuvem em processos de transformação das condições de vida, bem como na identificação de novas possibilidades que sejam integradas na cotidianidade desses grupos sociais (GURSTEIN, 2007).

Nessa conjuntura de abertura de dados governamentais em um setor determinante como é o da agricultura e o desenvolvimento rural, considera-se necessário focar pesquisas para construir sobre o construído, no que se refere ao

aproveitamento de dados que estão sendo disponibilizados, analisando aspectos do acesso na busca do aprimoramento do ecossistema no qual encontra-se a sociedade latino-americana, e assim, usufruir das condições instauradas com a sociedade da informação. É certa a necessidade de somar na freada ao desmoronamento do universo do conhecimento (WERSIG, 1993) e, neste caso, impera a sensibilização com aspectos técnicos e sociais do acesso a dados.

Gurstein (2007) assevera que além da infraestrutura, ou hardware que em ocasiões origina divisões entre grupos sociais, existem aspectos como são os conhecimentos, habilidades, estruturas organizativas e sociais, que permitem o uso efetivo do acesso às TIC, para alcançar os objetivos sociais e comunitários em determinado contexto, e, até mesmo, atentando às demandas informacionais.

Ainda no acesso a dados, considera-se o Ciclo de Vida dos Dados (CVD) proposto por Sant'Ana (2016), em que são ponderados processos de geração, disseminação e uso de dados, postulado para “estudar fatores e características que propiciam a ampliação do equilíbrio entre os atores envolvidos em um processo e da máxima otimização do uso dos dados” (MOREIRA, 2020), onde os atores são os denominados “Detentor”, quem em determinado domínio possui os dados e os disponibiliza, conforme seus interesses, valores e conhecimento prévio, e o “Usuário” em si, isto é, o sujeito alvo da publicação dos dados (SANTOS; SANT'ANA, 2019).

O CVD estima quatro fases constituintes, a saber, coleta, armazenamento, recuperação e descarte, todas elas realizadas por um sujeito alvo. Em contextos em que interagem vários CVD, e, portanto, vários sujeitos alvo, resulta em um cenário de interação conhecido como Campo Informacional (SANT'ANA, 2016, 2019).

Observando a relação de um sujeito alvo com o detentor governo, na fase de recuperação do CVD, onde ele dispõe dados para estes, existe um vazio diante da fase de coleta desse sujeito que os irá consumir. Conforme apontado, o sujeito pode apresentar características que deixe esse vazio maior ou menor. Trata-se de um distanciamento entre atributos de dados disponibilizados e os sujeitos alvo, o que poderia atenuar as possibilidades de uso dos dados pelos sujeitos e conforme com as suas demandas informacionais. Entende-se ainda, que essas demandas perfilam o desenvolvimento das atividades que os fazem viáveis economicamente, além de possibilidades de desenvolvimento conforme o aproveitamento de recursos de informação disponíveis.

O foco nesta pesquisa encontra-se em analisar aspectos do aproveitamento de

dados disponibilizados para um grupo particular como são produtores rurais no Brasil, observando a proximidade que pode existir entre comunicações particulares desse setor e dados que estão sendo publicados pelo detentor governo. Terá um viés lexical observando, como recomendado por Gray et al. (2009), aspectos que, uma vez tornados acessíveis quanto dados oficiais, determinem o seu uso com base em um formato adequado aos usuários.

Tencionam-se explicações acerca de condições de acesso baseadas em eventos regulares nas fontes de dados escolhidas, visando, mediante o uso de categorias de necessidades informacionais, uma manifestação das proximidades lexicais entre dados das duas pontas da análise, comunicações para produtores rurais e dados publicados pelo detentor governo.

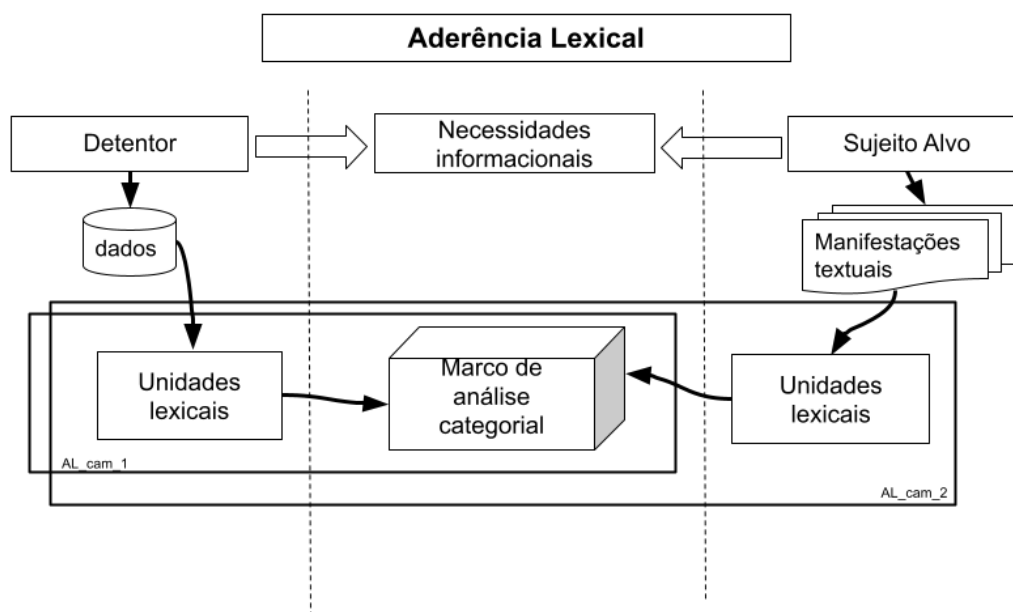
A lente que foi usada para observar um aspecto no acesso a dados é aplicada para analisar a proximidade entre as unidades lexicais geradas, propondo um elemento para propiciar a interpretação de fluxos de informação, como postulado por Borko (1968), e pretendendo um aprimoramento da experiência humana ao participar de fluxos informacionais disponíveis (CAMPEROS-REYES et al., 2020).

Dessa forma, essa pesquisa está norteada pela **questão problema**: Qual é a Aderência Lexical da fase de recuperação de dados, do detentor governo, com o campo informacional do sujeito alvo, identificado como produtores rurais, considerando as suas necessidades informacionais?

Para atender a questão, propõe-se analisar a Aderência Lexical entendida como uma possibilidade de mensurar a similaridade vocabular entre duas fontes de dados, o que permitiria obter um primeiro indício da potencial interpretação de dados publicados por um detentor, e passíveis de serem coletados por sujeitos alvo em determinado contexto.

A Aderência Lexical se alicerça na tríade: dados do detentor + necessidades informacionais + sujeito alvo, que em este estudo está instanciada no contexto de dados de governo, e necessidades informacionais de pequenos e médios produtores (Ver Figura 3). Ela é aplicável em duas camadas: 1) para verificar a proximidade entre dados publicados para o sujeito alvo e as suas necessidades informacionais; 2) para verificar a similaridade vocabular entre duas fontes de dados que mantêm necessidades informacionais em comum.

Figura 3 - Aderência lexical entre duas fontes de dados



Fonte: elaborado pela autora

O **título** da tese “Aderência lexical a dados publicados para produtores rurais”, foi adotado durante esse caminho que pretende observar indícios sobre a proximidade entre dados publicados pelo detentor governo, mediante as palavras usadas em comunicações para produtores rurais, pequenos e médios, considerando este elemento como uma opção para vislumbrar uma interpretação possível a dados publicados pelo governo.

Nesse entorno enunciado, a **tese** desta pesquisa é que a fase de recuperação dos dados disponibilizados pelos governos revela baixa aderência quando se considera seu uso por parte de produtores rurais e suas demandas informacionais.

Não há dúvidas em que pesquisas podem contribuir para aproximar os detentores dos dados aos sujeitos alvo, ainda que um contato ou integração total seja improvável, é possível estudar formas de diminuir esse vazio entendendo esse distanciamento.

## 1.2 Justificativa

Esse estudo se localiza na busca pelo aprimoramento do acesso a dados por pequenos e médios produtores atendendo suas demandas informacionais. Considera o processo de transdução de conteúdos da coleta à recuperação de dados no âmbito da agricultura e do desenvolvimento rural, com dados publicados em sites de dados de governos. O empenho é no âmbito de dados significativos para a sustentabilidade de um grupo da sociedade, especificamente, produtores rurais.

Ao mesmo tempo, observa o papel estratégico que desempenha a Ciência da Informação (SANTOS, 2005), no que tange às dimensões econômicas e sociais de atividades de informação no marco das TIC e suas dimensões sociais (SARACEVIC, 1996), convida a levantar questões dentro da realidade do potencial agrícola da região, como a maior produtora de alimentos do mundo (ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A ALIMENTAÇÃO E A AGRICULTURA, 2023).

Foi certamente indicado na motivação deste estudo o empenho em aproximar a comunidade acadêmica à conjuntura que envolve ativos informacionais, na sua disponibilização e nas condições de acesso, no domínio da agricultura e desenvolvimento rural, uma vez que “a agricultura é o setor que produz mais emprego no mundo, fornecendo a forma de vida de 40% da população mundial e é a maior fonte de ingressos e trabalho nos lares pobres rurais” (ORGANIZAÇÃO DAS NAÇÕES UNIDAS, 2023, tradução própria).

O Brasil é um país com vocação agrícola, que ainda é precursor no contexto latino-americano na promoção e execução de políticas públicas que têm buscado melhorar a percepção de transparência do governo mediante estratégias para disponibilização e uso de dados abertos (BRASIL, 2011). Ademais, possui uma área de 8.510.000 km<sup>2</sup> (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2021b), da qual 2.368.000 km<sup>2</sup> (BANCO MUNDIAL, 2021) correspondem a área agrícola, isto é, um 28% do território nacional para realizar atividades do setor.

Cabe ressaltar que junto às potencialidades territoriais do país, no ano 2020, particularizado pela crise econômica gerada pela pandemia da doença COVID-19, o único setor que obteve resultado positivo no PIB foi a Agricultura: no Brasil cresceu 2% dentro de uma contração no país de 4.1% (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2021).

O setor da agricultura agiu com resiliência durante o período da pandemia; foi

tal que alavancou a queda dos demais setores econômicos, somente superada pelo auge inefável e necessário das tecnologias da informação. Na América Latina e o Caribe, 11 de 16 países experimentaram aumento no Produto Interno Bruto (PIB) do setor agropecuário, e naqueles que não aumentou, contraiu-se em uma taxa menor que o total PIB do país (COMISSÃO ECONÔMICA PARA A AMÉRICA LATINA E O CARIBE, 2021).

Nesse âmbito, os sujeitos informacionais, “pequeno produtor”, constituem um grupo destacado no país, observa-se que no Brasil, do total de estabelecimentos agropecuários (5.073.324), o 67% correspondem a agricultores familiares, sendo um total de 3.897.408 estabelecimentos. (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2021a).

Observa-se que na América Latina está acontecendo uma sensibilização frente ao uso efetivo de dados, mas também, das características do acesso a esses recursos que, contextualizados na realidade de cada país, atributos particulares e a possibilidade de um trabalho mancomunado, referenciam uma alternativa para enfrentar em grupo desafios que os ocupam.

Para Verhulst e Young (2017), seis características demarcam a relevância no acesso a ativos informacionais em economias em desenvolvimento: a) possibilidade de escrutínio pela sociedade; b) igualdade na distribuição da informação e o conhecimento; c) flexibilidade nos formatos de apresentação para a posterior utilização; d) amplo espectro de participação tanto ao interior quanto na tomada de experiências fora de contexto regional; e) aumento da confiança por parte do grupo social no governo que o esteja representando; e, f) amplificação *do valor* dos dados bem como ativos informacionais.

É notório que tais características, de forma implícita, aludem à intencionalidade do uso dos dados e, portanto, a demandas informacionais dos sujeitos alvo, sendo que o seu amparo potencializa um cenário rico no uso desses ativos disponíveis.

Atendendo à infraestrutura tecnologia disponível, aliada às possibilidades de acesso a dados, entende-se que podem permear os mais diversos setores, como a saúde, educação, comércio e, como o caso que nos ocupa, a agricultura, promovendo transformações em prol do desenvolvimento social, econômico e político de forma sustentável. Espera-se que o uso assertivo das TIC seja inserido cada vez mais além de tarefas manuais nas tarefas de análise, sobretudo na tomada de decisões

(INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2019), para o qual devem demarcar-se caminhos de pesquisa que contribuam com elementos que complementem a infraestrutura e os ativos informacionais, junto às competências informacionais dos sujeitos alvos.

Por exemplo, como a agricultura é cada vez mais intensiva em conhecimento, as TICs podem ajudar os agricultores a melhorar os rendimentos das culturas e a produtividade das empresas por meio de um melhor acesso à informação de mercado, previsões meteorológicas, programas de formação e outros conteúdos online adaptados às suas necessidades (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2019, p.35).

Deve considerar-se que, diante a relevância no acesso a ativos informacionais, o ecossistema não pode estar completo focando apenas na tecnologia e nos dados em si. Nossa estrutura tecnológica, pequena como asseverado por Davenport (1998), porém marcante no que diz diante das vantagens competitivas que trazem para aqueles grupos sociais que sabem fazer uso dela, poderia ser observada como pivô na busca do aprimoramento de um Estado, sem desconsiderar que o desafio é fazer que exista equilíbrio na ecologia, percebido mediante o aprimoramento das condições de vida palpável nos grupos sociais atingidos.

Segundo Saracevic (1996), a Ciência da Informação com seu imperativo tecnológico é partícipe da transformação da sociedade, considerando sua forte dimensão social e humana, a qual ultrapassa a tecnologia. Dessa forma, estudos realizados em áreas representativas do desenvolvimento de um país contribuem no robustecimento dela como ciência, gerando contributos que tenham a possibilidade de auxiliar os planos estratégicos em níveis de detentores de dados de governo.

Certamente, estudos que envolvem dados e informações dentro de um setor da economia de importância manifesta, como o da agricultura, revestem-se de importância por estabelecer pontes entre a ciência e o desenvolvimento de uma região ou de um país (CAMPEROS-REYES; SANT'ANA; RAMALHO, 2017). Espera-se impactar promovendo discussões sobre caminhos para o progresso, com a participação de atores que demandam por análises, reflexões, propostas, serviços e processos relacionados com o acesso a dados.

O impacto social da tese está adequado aos Objetivos para o Desenvolvimento Sustentável (ODS) declarados pela Organização das Nações

Unidas (ONU, 2023). O segundo objetivo, ODS 2, dessa agenda é “Acabar com a fome, alcançar a segurança alimentar e melhoria da nutrição e promover a agricultura sustentável”; nesse item, os países latino-americanos destacam-se pelo considerável volume de alimentos produzidos, bem como a expressiva participação dos pequenos e médios produtores. No contexto do Brasil, foi determinada pelo Instituto de Pesquisa Econômica Aplicada (IPEA) a criação de metas internas que levem ao cumprimento dos objetivos, explicitando na meta 2.3,

até 2030, aumentar a produtividade agrícola e a renda dos pequenos produtores de alimentos, particularmente de mulheres, agricultores familiares, povos e comunidades tradicionais, visando tanto à produção de autoconsumo e garantia da reprodução social dessas populações quanto ao seu desenvolvimento socioeconômico, por meio do acesso seguro e equitativo (SILVA, 2018),

Ainda mais, a tese contribui com o ODS 16, “Paz, Justiça e Instituições Eficazes”, particularmente na meta 16.10: “Assegurar o acesso público à informação e proteger as liberdades fundamentais, em conformidade com a legislação nacional e os acordos internacionais” (SILVA, 2018); ambos os objetivos agem como agentes motivadores desta pesquisa pelo impacto social esperado.

Acredita-se que é necessário criar mecanismos que diminuam a distância entre os detentores dos dados e os sujeitos alvo de processos de disponibilização desses recursos e, com isso, propor alternativas para entender o gap entre os atores entende-se como uma ação relevante no avanço dessa relação.

### 1.3 Objetivos

Considerando o delineamento teórico que alicerça esse estudo, a questão problema, as motivações e que o ‘acesso a dados’ é um dos objetos de estudo da Ciência da Informação ao estudar as propriedades e fluxos da informação junto às forças que a governam (BORKO, 1968), apresenta-se como **objetivo geral**: Verificar a viabilidade de um modelo que reconheça a Aderência Lexical a dados publicados pelo detentor governo, usando necessidades informacionais de produtores rurais como marco de análise.

Em decorrência, são propostos os seguintes **objetivos específicos**:

OE1. Identificar categorias de necessidades informacionais, no contexto de acesso a dados, em pequenos produtores rurais relacionados com a agricultura;

OE2. Coletar unidades lexicais contidas em notícias publicadas para grupos de produtores rurais;

OE3. Verificar, nas notícias mineradas, a proximidade com categorias de necessidades informacionais;

OE4. Coletar unidades lexicais contidas em rótulos e descrições de conjuntos de dados publicados no Brasil para grupos de produtores rurais;

OE5. Verificar, nos conjuntos de dados, a proximidade com as categorias de necessidades informacionais;

OE6. Constatar a Aderência Lexical entre palavras advindas de notícias publicadas para grupos de produtores e palavras de conjuntos de dados do detentor governo.

#### **1.4 Metodologia**

Por tratar-se de uma pesquisa descritiva, aplicada em práticas de acesso a dados, cenário enriquecido por fatores complexos e heterogêneos na origem, este estudo desenvolver-se-á utilizando uma abordagem qualitativa com desenho não experimental transeccional (HERNÁNDEZ SAMPIERI; FERNÁNDEZ COLLADO; BAPTISTA LUCIO, 2010). Será aplicada triangulação de procedimentos metodológicos conforme explicitado na Figura 4, usando para tal, Pesquisa Documental, Revisão Sistemática de Literatura, Mineração Textual e Análise de Similaridade.

Figura 4 - Diagrama estrutural da tese

## Título: Aderência Lexical a dados publicados para produtores rurais

**Motivação:** o distanciamento entre atributos de dados disponíveis e os sujeitos alvo, e a necessidade de contribuir com subsídios para o desenvolvimento eficaz de estratégias que permitam usar efetivamente esses dados, focando o acesso no âmbito da agricultura e o desenvolvimento rural, particularmente por pequenos e médios produtores rurais.

**Problema de pesquisa:** distância dentre dados disponibilizados pelo detentor governo (volumen, legislação a favor), e produtores rurais; suporte à tomada de decisão pelos pequenos e médios produtores. Qual a Aderência Lexical da fase de recuperação de dados, do detentor governo, com o campo informacional do elemento alvo identificado como pequenos e médios produtores rurais?

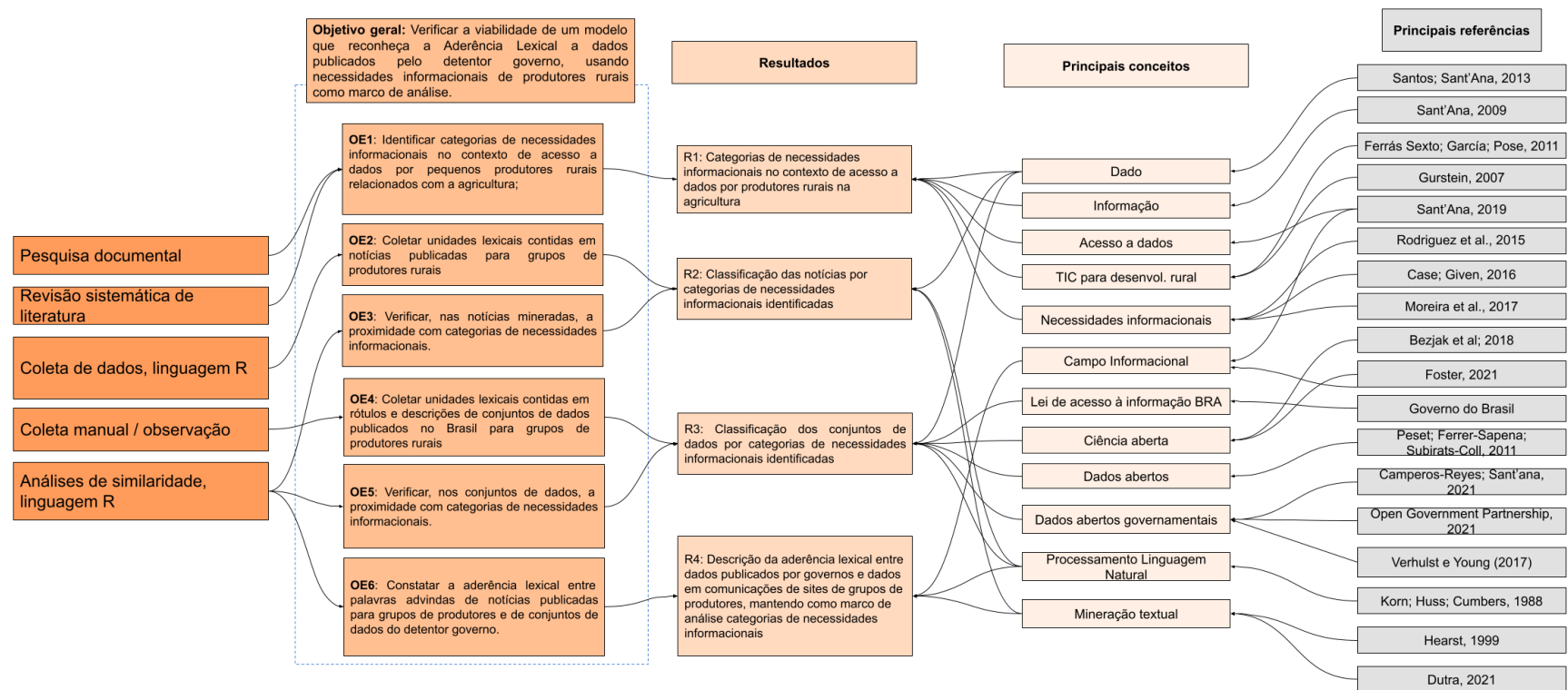
**Justificativa:** a procura pelo aprimoramento do acesso a dados pelos pequenos e médios produtores atendendo suas demandas informacionais, no contexto da transdução de conteúdos da coleta à recuperação de dados, para a agricultura e desenvolvimento rural, publicados em sites de dados de governos. O empenho é no aprimoramento do acesso a dados significativos para a sustentabilidade de um grupo da sociedade, especificamente, os produtores rurais.

**Delimitações:** a análise de aderência foca a camada lexical, por tanto, as técnicas de mineração textual usadas concentram-se na distinção da similaridade lexical entre as fontes de dados escolhidas, não impedindo que estudos futuros abordem camadas com elementos semânticos nas mesmas ou em outras fontes de dados, bem como a interação direta com sujeitos do grupo alvo; os elementos usados como indícios de aderência a dados são dados extraídos a partir de comunicações publicadas para grupos de produtores rurais como instrumento de aproximação aos pequenos e médios produtores. Os dados de governo são um recorte de datasets publicados para grupos de produtores rurais.

**Tese:** A fase de recuperação dos dados disponibilizados pelos governos revela baixa Aderência Lexical quando se considera seu uso por parte de produtores rurais e suas demandas informacionais.

**Impacto Social:** dados e informações por seu papel fundamental na economia tornam-se objeto de esforços comuns em prol do acesso e do aproveitamento, tal o caso do domínio da agricultura, especificamente em regiões em vias de desenvolvimento, e, no caso de dados publicados pelo detentor governo, procurar que o estejam com aspectos intrínsecos que outorgam auxílios adequados aos usuários. ODS 2: Acabar com a fome, alcançar a segurança alimentar e melhoria da nutrição e promover a agricultura sustentável; meta 2.3: "Até 2030, aumentar a produtividade agrícola e a renda dos pequenos produtores de alimentos", e, ODS 16: Paz, Justiça e Instituições Eficazes", meta 16.10 "Assegurar o acesso público à informação e proteger as liberdades fundamentais, em conformidade com a legislação nacional e os acordos internacionais

**Impacto Científico:** incremento de pesquisas em contextos socioeconômicos de reconhecida importância, a partir das possibilidades da Ciência da Informação, contribuindo no desenvolvimento e a competitividade ao fornecer caminhos de pesquisa que tencionam progressão com a participação de atores que demandam por análises, reflexões, propostas, serviços e processos relacionados com o acesso a dados.



Fonte: adaptado de Sant'Ana (2022)

A interseção formada na aplicação dos procedimentos metodológicos remete à referência heurística da Ciência da Informação (GÓMEZ, 2003), que possibilita o descobrimento de fatos durante o percurso da pesquisa.

O foco da Pesquisa Documental serão fontes de dados e informações do detentor governo no Brasil, <https://dados.gov.br/>. Será observada a documentação do processo de acesso e uso aos conjuntos de dados, buscando estruturas de aproximação propostas para o entendimento, por parte de produtores rurais, dos dados publicados.

A Revisão Sistemática de Literatura (RSL) atentou à produção científica em torno do acesso a dados, que permitiu identificar necessidades informacionais de produtores rurais.

Os insumos para a RSL foram autores influentes no contexto de acesso a dados na agricultura, optando para tal propósito por identificá-los em estudos publicados nas bases de dados referenciais Web of Science (WoS) e Scopus. Para a primeira fonte, os autores foram obtidos a partir do artigo de Camperos-Reyes e Sant'Ana (2019), e em segundo momento, realizou-se estudo métrico de informação recuperando dados a partir da Scopus e, ao aplicar técnicas bibliométricas, obtiveram-se os autores mais produtivos e os mais citados na área temática indicada.

Estudos métricos da informação contribuem como insumo para alicerçar o componente teórico em pesquisas, conforme às afirmações dos Hernandez-Sampieri, Collado e Lucio (2010), quando designam à investigação científica como sistemática e empírica: sistemática ao fornecer disciplina e rigurosidade no seu procedimento, e empírica por causa dos exercícios de coleta e análise de dados.

A seguir, procedeu-se a recuperar publicações dos autores identificados previamente, mediante RSL apoiada na recomendação *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA) (MOHER et al., 2009), sendo realizada em abril de 2021 conforme o protocolo indicado no Quadro 3.

Quadro 3 - Protocolo da RSL

|                            |                                                                                                                                                                                                                                                        |
|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Objetivo da RSL            | Identificar produções científicas, de autores previamente identificados, com categorias de necessidades informacionais no acesso a dados pela agricultura                                                                                              |
| Critérios de elegibilidade | <ul style="list-style-type: none"> <li>Abordagem <i>Population, Intervention, Comparison, Outcomes</i> (PICO):</li> </ul> <p>P: produtores pequenos e médios<br/> I: necessidades informacionais<br/> C: categorias de necessidades informacionais</p> |

|                                   |                                                                                                                                                                                                                                                                                                                                              |
|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                   | O: categorias de necessidades informacionais mais significativas <ul style="list-style-type: none"> <li>• Sem delimitação temporária</li> <li>• Idiomas português, inglês, castelhano</li> <li>• Documentos em acesso aberto</li> <li>• Documentos do tipo artigo científico, trabalho apresentado em evento e capítulos de livro</li> </ul> |
| Fonte de informações              | Google Scholar                                                                                                                                                                                                                                                                                                                               |
| Expressão de busca                | Corresponde à "information needs" AND "agriculture" AND "sobrenome" (do autor a pesquisar), mudando "sobrenome" para cada autor identificado previamente.                                                                                                                                                                                    |
| Critérios de exclusão             | <ul style="list-style-type: none"> <li>• Revisões de literatura</li> <li>• Estudos distantes do escopo da pesquisa</li> </ul>                                                                                                                                                                                                                |
| Processamento dos dados coletados | Leitura informativa (MARCONI; LAKATOS, 2003) em todos os documentos recuperados, atentando às fases de reconhecimento, leitura exploratória, seletiva e reflexiva, mantendo coerência com a abordagem PICO determinada.                                                                                                                      |

Fonte: elaborado pelos autores

O protocolo apresentado entregou necessidades informacionais no acesso a dados na agricultura, e para obter um panorama acorde ao contexto da pesquisa, foi incorporado um elemento de análise na realidade científica brasileira. Para tal, procuraram-se os descritores "necessidades informacionais" AND "agricultura\*" no site da Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI), nos anais do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), e nos anais do Seminário em Ciência da Informação (SECIN), buscando por documentos sem limite de temporalidade.

Na continuidade, a fim de relacionar os vários estudos recuperados para sintetizar os resultados, optou-se pela meta-análise, isto é, análise da análise, indicada por Sampaio e Mancini (2007), apelando a procedimentos estatísticos que possam caracterizar a intervenção indicada no protocolo de uma RSL.

Assim sendo, com o objetivo de relacionar os resultados obtidos mediante a meta-análise, optou-se pelo cálculo de frequências relativas para designar peso às categorias identificadas nos estudos recuperados, de um lado, estudos publicados nas bases WoS e Scopus e, do outro, estudos recuperados em fontes de informação do Brasil.

Para a coleta dos dados do lado do produtor rural, ou seja, unidades lexicais contidas em comunicações para eles, foi determinado realizá-la a partir de uma

organização que os agrega; considerou-se factível atingir indícios que caracterizam a pequenos e médios produtores quando eles estão agrupados ou associados.

A linguagem usada nas comunicações, por ser geradas no entorno de agrupações de produtores rurais, é uma linguagem direcionada para alcançar esse público-alvo, portanto, presumem-se escritas com um teor adequado às características dos grupos de produtores.

Entende-se que a linguagem natural ao ser um recurso mediador entre cenários da compreensão humana, é uma linguagem “eficaz na tradução dos fenômenos em pensamento porque é constituída por um elemento primordial para as associações lógica, psicológica, ideológica e interpretativa: o signo” (MARTINES; MOREIRA; ALMEIDA, 2022, p.33)

Abordar documentos escritos na linguagem natural com ferramentas tecnológicas propicia amplas possibilidades de análise. Em um primeiro momento, pelas capacidades *per se* da linguagem natural como sistema de signos expressivo e predileto na comunicação (KORN; HUSS; CUMBERS, 1988), de outro lado, técnicas informáticas permitem revelar informações em grandes corpos textuais, que aproveitando das capacidades da linguagem natural, viabilizam a obtenção de inferências conforme interesses de pesquisa em um contexto determinado.

Assim sendo, a coleta das comunicações foi realizada mediante Mineração de Textos, implementado algoritmos na linguagem R. Outra ferramenta usada durante o processo é o SelectorGadget<sup>7</sup>, extensão de navegador web que facilita o descobrimento de seletores CSS<sup>8</sup> em sites, funcional durante o processo de identificação dos blocos de texto que serão coletados. As duas ferramentas indicadas são software de acesso aberto.

Determinou-se usar a técnica de Mineração de Textos pois permite extrair informação útil das fontes mediante identificação de padrões, tendências, ou índices, conforme interesses de estudos que abordam coleções de textos. Os dados são extraídos do texto fluido que conforma os documentos que são minerados (FELDMAN; SANGER, 2006). A finalidade é extrair informação para a posterior vinculação dos

---

<sup>7</sup> Cf. <https://selectorgadget.com/>

<sup>8</sup> *Cascading Style Sheets* (CSS), em português Folha de Estilo em Cascatas, linguagem usado para estilizar elementos na linguagem de marcado HTML. CSS separa o conteúdo da representação visual de um site. HTML constitui a base do site, enquanto CSS enfatiza no estilo, na estética do site (W3C, 2021a).

elementos discernidos e, assim, conformar novos feitos, encontrando inferências de interesse extraídas a partir de grandes corpus textuais, originalmente não codificados ou estruturados (DUTRA, 2021; HEARST, 2003; KAO; POTEET, 2007).

Em relação à coleta de dados do lado do detentor governo, foi realizada extraíndo rótulos dos recursos de conjuntos de dados e as descrições registradas pelos publicadores, de tal forma que ambas as fontes fornecem dados manifestados na linguagem natural.

Assim sendo, as duas fontes escolhidas provêm de mediadores da informação que estão fora de manifestações explícitas advindas de profissionais da Ciência da Informação, agrupações de produtores rurais e publicadores de dados.

Observaram-se conjuntos de dados, ou datasets, de interesse para produtores rurais, recuperados mediante o uso de descritores no site de dados abertos do Brasil: “Pequeno produtor”, “Desenvolvimento rural”, “Associação agricultura”, “Cooperativa agricultura”. Realizou-se extração manual dos rótulos e descrições justificando-se necessário discriminar a disponibilidade de cada recurso recuperado na busca.

Devido a que a forma de apresentação de informações e os pressupostos em relação ao seu acesso estão mais relacionados com atributos de quem os disponibiliza do que de quem fará uso deles (ALMEIDA, 2008), foi determinado o uso dessa fonte de dados do detentor governo, de modo a ser comparada com os elementos lexicais do lado produtor rural.

As descrições e a rotulagem dos conjuntos de dados foram usados em razão de que eles são referências tangíveis ao conteúdo que têm acesso os usuários de sites de dados abertos. Esses elementos manifestam características dos dados acessíveis pelos usuários e, portanto, descrevem os dados publicados pelo detentor governo, outorgando aos usuários uma aproximação das possibilidades de uso daqueles recursos disponíveis. No conceito de Santos (2013), essas formas de representação são criadas na perspectiva de oferecer condições favoráveis no processo de acesso a recursos informacionais.

Assim, esta pesquisa foi realizada com dados do tipo não estruturados, dados extraídos do site com informações para produtores rurais, e dados estruturados, recuperados dos rótulos e das descrições de conjuntos de dados. Contudo, a mineração central de dados das fontes, independentemente da natureza na origem, foi realizada com procedimentos padronizados.

Para a análise da Aderência Lexical entre comunicações, notícias para grupos

de produtores rurais, e dados publicados pelo detentor governo, em primeira instância foram extraídas as unidades lexicais, palavras das notícias publicadas e dos conjuntos de dados. Logo após, de forma automatizada, foi descoberta a proximidade das duas fontes com as categorias de necessidades informacionais identificadas para agir como marco estrutural da análise da proposta.

Foram usadas as bibliotecas da linguagem R '*tidytext*', '*tm*' e '*philentropy*'; nestas, as funções '*findAssocs*' e '*jaccard*' fizeram trabalho conjunto para a classificação das notícias e conjuntos de dados conforme as necessidades informacionais.

A função '*findAssocs*', permitiu descobrir quais palavras aparecem com maior frequência quando encontrado o nome da categoria de necessidade informacional. Com os resultados da aplicação da função, a lista de palavras foi analisada na ordem decrescente do índice de coocorrência, que na função apresenta-se com valores entre 0 e 1, sendo que valores próximos a 1 indicam maior coocorrência. Como critério de escolha das palavras foram consideradas apenas palavras do tipo substantivo. Com as palavras identificadas foi construído um vetor para cada categoria de necessidade informacional e, mediante a função '*jaccard*', foi calculado com qual categoria, cada dataset e notícia, tinha proximidade, produzindo assim uma classificação por categorias.

Uma vez classificadas tanto as notícias quanto os conjuntos de dados, foi calculado o índice de similaridade entre as palavras usadas pelo governo nos conjuntos de dados e as notícias publicadas para grupos de produtores, mediante a função '*jaccard*' da biblioteca '*philentropy*'. Dita função entrega um índice com valor decimal que está entre 0 e 1, sendo que um valor próximo de zero indica baixa similaridade entre os textos comparados.

Conforme a estrutura adotada para os dados durante o pré-processamento, *jaccard* foi observada como a função mais pertinente para calcular a similitude entre os conjuntos de palavras, de tipo assimétrico, que foram conformados: conjuntos de palavras advindos das comunicações para produtores, conjuntos de palavras advindos de rótulos e descrições de datasets, e conjuntos de palavras que descrevem necessidades informacionais de produtores.

Na mineração textual existem diversas funções que calculam similaridade entre conjuntos de dados; a escolha da função depende tanto dos fins da análise quanto da estrutura em que se encontram os dados. Há funções direcionadas a dados

simétricos, assimétricos, segundo a natureza, dados binários, categóricos, numéricos, ou para combinações de tipos de dados, ainda podendo ser orientadas para análises de dissimilaridade em dados de essas e outras naturezas. Han, Kanker e Pei (2012) indicam *jaccard* como uma opção apropriada para analisar conjuntos de dados como os resultantes da fase de pré-processamento deste estudo. Ainda, considerando o fim dos estudos que usam mineração textual, Huang (2008) indica que, não objetivando análises de significado nos dados coletados, *jaccard* é uma das opções adequadas.

A diferença da análise morfológica, que também considera palavras isoladas em um texto a fim de identificar classes gramaticais, este estudo individualiza as palavras com o intuito de encontrar a proximidade entre o léxico usado nas comunicações orientadas a grupos de produtores rurais e dados publicados por um governo, análise intermediado com um marco estruturante de necessidades informacionais do grupo alvo, produtores rurais.

Assim, o reconhecimento de indícios de aderência lexical a dados publicados pelo governo do Brasil será visto em dois níveis, proximidade dos dados de governo com categorias de necessidades informacionais para pequenos e médios produtores, e proximidade entre unidades lexicais extraídas de dados publicados pelo governo e comunicações publicadas para pequenos e médios produtores.

O intuito é, nas unidades lexicais identificadas nas duas fontes, perceber características da dimensão *input* do campo informacional de grupos de produtores rurais, as quais contribuem como elementos que detalham o acesso a dados desses sujeitos, implicando em uma primeira camada da potencial interpretação dos dados.

## 1.5 Escopo

A análise de aderência foca na camada lexical, portanto, as técnicas de mineração textual usadas concentram-se na distinção da similaridade lexical entre as fontes de dados escolhidas, não impedindo que estudos futuros abordem camadas com elementos semânticos nas mesmas ou em outras fontes de dados, bem como a interação direta com sujeitos do grupo alvo. Esse estudo não foca em uma análise da construção textual dos recursos que serão analisados com fins morfológicos ou sintáticos.

Por outro lado, os elementos usados como indícios de aderência a dados são dados extraídos a partir de notícias publicadas para grupos de produtores rurais, como instrumento de aproximação aos pequenos e médios produtores. Os dados de

governo são um recorte de conjuntos de dados publicados para produtores rurais.

As categorias de necessidades informacionais não constituem uma proposta da tese, elas foram extraídas de categorias segundo resultados de estudos prévios que tiveram contato direto com produtores rurais.

## **1.6 Estrutura do documento**

Apresenta-se o documento da tese como um compêndio de artigos científicos, eles contendo tanto referencial teórico como resultados; exceptuando o capítulo dois, Contexto da publicação de dados abertos de governo no Brasil. Assim, o corpo da tese é como segue:

- Prefácio: trajetória acadêmica da autora;
- Capítulo 1. Introdução: contexto e motivações, caracterização do problema e tese, justificativa e impactos, objetivos, metodologia e escopo;
- Capítulo 2: Contexto da publicação de dados abertos de governo no Brasil;
- Capítulo 3: (Artigo) Perspectiva de pesquisa sobre acesso a dados na Economia Solidária no Brasil;
- Capítulo 4: (Artigo) Necessidades informacionais e acesso a dados no contexto da agricultura;
- Capítulo 5: (Artigo) Uso da Mineração de Texto em fontes informacionais para grupos de produtores rurais;
- Capítulo 6: Conclusões.

## 6. CONCLUSÕES

A questão do acesso a dados na busca pela otimização dos fluxos informacionais direcionou esta pesquisa, entregou as balizas que permitiram focar o projeto e a procura dos resultados esperados. É certo que alguns dados e informações, por seu papel fundamental na economia e na sociedade, devem tornar-se objeto de esforços comuns para aprimorar o seu acesso e aproveitamento, e, no caso de ativos informacionais publicados pelo detentor governo, conforme afirma Gray et al. (2009), dados oficiais devem publicar-se com aspectos intrínsecos que outorguem auxílios para o entendimento por parte dos usuários.

Acredita-se que o incremento de pesquisas em contextos socioeconômicos de notável importância, a partir das possibilidades da Ciência da Informação, deve contribuir com efetividade no desenvolvimento e a competitividade, fornecendo subsídios para que os atores demandantes de análises, reflexões e propostas relacionados com dados como ativos informacionais, possam atender às suas necessidades.

Norteadas pela pergunta problema, qual é a Aderência Lexical da fase de recuperação de dados, do detentor governo, com o campo informacional do sujeito alvo identificado como grupos de produtores rurais, esta tese conduziu um roteiro para abordar elementos teóricos e técnicos que permitiram assentar aspectos tais como relevância do estudo, elementos conceituais, contextos particulares à realidade nacional, e ferramentas tecnológicas que possibilitaram a culminação dos objetivos propostos.

Diante da necessidade de identificar teorias, atores, instituições e demais artefatos contextuais da temática escolhida, foram criados dois capítulos, no capítulo dois, “Contexto da publicação de dados de Governo no Brasil”, partindo das iniciativas globais do livre acesso à informação e o movimento de acesso aberto na ciência, foram identificadas as estruturas de governança e elementos técnicos utilizados para a publicação de dados abertos de governo no país. O capítulo três, “Perspectiva de pesquisa sobre o acesso a dados na Economia Solidária no Brasil” caracterizou as dimensões input e output do Campo Informacional do sujeito alvo Economia Solidária, encontrando como eixos temáticos na dimensão input: a) Uso de dados em redes de colaboração; b) No desenvolvimento de Competências Informacionais para suporte à tomada de decisão e inclusão social; e c) Gestão de informação pelos

Empreendimentos Econômicos Solidários; e na dimensão output: Extração automatizada de dados do setor. A caracterização das dimensões possibilitou uma aproximação a elementos para alicerçar pesquisas na área temática, a qual se percebeu como um campo ainda em construção e determinante, pois, além de que a Economia Solidária constitui uma opção de renda para cada vez mais pessoas, é também uma área de notada importância no desenvolvimento integral de uma sociedade e, portanto, que precisa estudos para minorar desvantagens pelo desequilíbrio entre os dados coletados sobre eles e os dados disponíveis para seus empreendimentos.

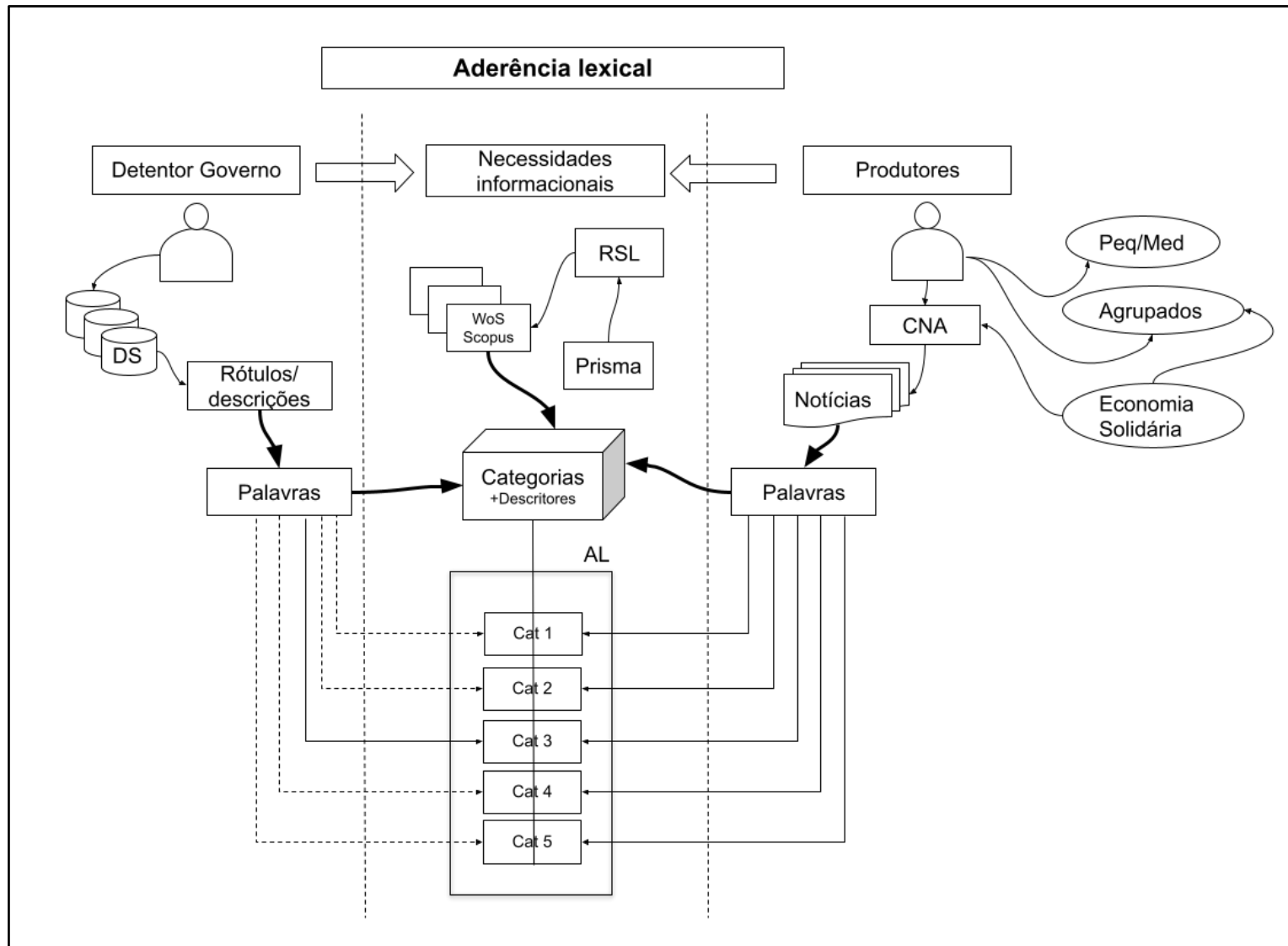
O capítulo quatro, “Necessidades informacionais e acesso a dados no contexto da agricultura” teve como principal função fornecer um marco estruturante para a análise desta pesquisa. Ao delimitar a amostra a pequenos e médios produtores, foram identificadas categorias de necessidades informacionais assim como descritores de cada uma.

Em seguida, implementando técnicas de mineração de texto aplicadas em fontes com dados estruturados e não estruturados, foi analisada a Aderência Lexical entre as duas fontes escolhidas, usando as categorias de necessidades informacionais como marco de análise. Os resultados foram apresentados no capítulo cinco, “Uso da Mineração de Texto em fontes informacionais para grupos de produtores rurais”.

É preciso salientar que Aderência Lexical é entendida como uma possibilidade de mensurar a similaridade vocabular entre duas fontes de dados, o que permitiria obter um primeiro indício da potencial interpretação de dados publicados por um detentor, e passíveis de serem coletados por sujeitos alvo em determinado contexto.

A Aderência Lexical se alicerça na tríade: dados do detentor + necessidades informacionais + sujeito alvo, uma proposta de análise que está manifestada graficamente na Figura 17, que foi instanciada nesta tese no contexto de dados de governo, e as necessidades informacionais de pequenos e médios produtores.

Figura 17 - Aderência Lexical instanciada nesta tese



Fonte: elaborado pela autora

É uma proposta que permite gerar um diagnóstico sobre a proximidade entre as fontes, que não pode ser observado apenas pelo volume entre as unidades lexicais extraídas, e sim deve considerar-se com uma visão holística, no caso deste estudo, pela perspectiva outorgada pelas categorias de necessidades informacionais.

Por esse motivo, foi reconhecida uma Análise Lexical aplicável em duas camadas: 1) para verificar a proximidade entre dados publicados para o sujeito alvo como, por exemplo, neste estudo, notícias e/ou conjuntos de dados de governo para produtores rurais, e as suas necessidades informacionais; 2) para verificar a similaridade vocabular entre duas fontes de dados que mantêm necessidades informacionais em comum.

Acredita-se que, na medida em que dados publicados pelo detentor governo estejam relacionados com as necessidades informacionais dos sujeitos alvos, existe uma maior possibilidade de que eles sejam usufruídos por parte deles. Conforme os desfechos deste estudo, é possível apontar que, na amostra escolhida, existe uma comunicação insuficiente, pois, das categorias de necessidades ponderadas, houve proximidade unicamente com uma, Crédito. Por um lado, a RSL mostrou que as categorias mais relevantes para os pequenos e médios produtores são Mercado e Tratos culturais, nessa ordem, enquanto, nas notícias aqui mineradas, as de maior frequência foram Mercado e Oportunidades.

Em particular, este estudo apontou que tanto para a comunidade científica como para uma instituição representante dos produtores, como a CNA, destacam-se assuntos categorizados como Mercado, tais como preços, tendências de consumo, publicidade, canais de comercialização etc., onde, com os procedimentos aqui realizados, não houve um conjunto de dados que os tratasse. Uma situação da mesma natureza se apresentou com assuntos categorizados como Oportunidades, onde ressaltaram aspectos de associativismo, cooperativismo, negócios, empreendedorismo, e desenvolvimento de carreira, que também não foram identificados na amostra do estudo do lado detentor governo.

A seguir, ao observar a Aderência Lexical entre duas fontes de dados que mantêm necessidades informacionais em comum, aponta que 80.48% das unidades lexicais usadas nos dados de governo, foram similarmente encontradas nas notícias mineradas do site da CNA. Desde o ponto de vista da similaridade vocabular entre as duas fontes, é entendida como favorável em um primeiro nível na busca de inteligibilidade entre estas, como indicado por Sant'Ana (2018), quando prescreve que

as formas de publicação dos dados devem contribuir com as possibilidades de interpretação pelos usuários.

Os dados desta pesquisa assinalam uma oportunidade de melhora na pertinência de dados que estão sendo disponibilizados pelo governo, e revelam que a Aderência Lexical é um instrumento útil para identificar proximidade entre fontes de dados e necessidades informacionais de sujeitos alvo.

Nesse ponto, é importante frisar a existência de unidades lexicais que foram identificadas unicamente do lado governo, que correspondem a siglas como BSM, SIATER, SIMOG, UPFS. Estas são atributos de programas de governo, as quais podem manifestar a necessidade de melhor aproximação por parte dos sujeitos alvo visando o seu potencial uso.

A Aderência Lexical aqui proposta age como um mecanismo de mensuração da integração do contexto de usuário ao contexto de dados (SANTOS; SANT'ANA, 2019), usando as unidades lexicais como pontos de comparação e, portanto, é uma forma de interpretação de fluxos informacionais, matéria de estudo da Ciência da Informação conforme Borko (1968); é um resultado que contribui na busca de soluções a situações permeadas pelo uso das TIC no contexto do acesso a dados e a sua potencial interpretação, como auxílio para o direcionamento e tomada de decisões (SANTOS; SANT'ANA, 2019), neste caso, por pequenos e médios produtores.

Ademais, as unidades lexicais identificadas nas duas fontes indicam características da dimensão input do campo informacional de grupos de produtores rurais, as quais contribuem como elementos que detalham o acesso a dados desses sujeitos, implicando em uma opção de camada de interpretação de dados, conforme Nadkarni, Brandt e Frawley (1998), ao dissertar sobre a dupla atributo-valor e sua implicância nos sistemas de informação.

Esses desfechos permitem já responder à questão norteadora da pesquisa, qual é a Aderência Lexical da fase de recuperação de dados, do detentor governo, com o campo informacional do sujeito alvo, identificado como produtores rurais e suas necessidades informacionais? E ainda, junto aos resultados expostos no capítulo anterior, atingir o objetivo geral definido como “Verificar a viabilidade de um modelo que reconheça a Aderência Lexical a dados publicados pelo detentor governo, usando necessidades informacionais de produtores rurais como marco de análise”.

A Aderência Lexical proposta, que considera as unidades de texto no nível de palavra, diferencia-se de uma análise morfológica, que também considera as palavras isoladas em um texto, porém com o fim de identificar a classe gramatical correspondente, pois este estudo procurou encontrar a proximidade entre o léxico usado nas comunicações orientadas a grupos de produtores rurais e o léxico usado em dados publicados por um governo, em uma análise intermediada com um marco estruturante de necessidades informacionais do grupo alvo produtores rurais.

Como limitações do estudo, aponta-se por um lado que a busca por indícios das possibilidades de interpretação de dados por produtores levou à decisão de usar manifestações indiretas dessas possibilidades, como é o caso das notícias mineradas do site da CNA, o que naturalmente apresenta vieses técnicos, mas, sobretudo, com uma implementação à distância do próprio sujeito informacional. Abre-se a oportunidade de propor novos projetos na busca de uma visão correlata com o que o sujeito manifesta diretamente.

Outro aspecto a ressaltar é a quantidade de dados disponíveis do lado detentor governo, magnitude que aponta para novas estratégias metodológicas que permitam coletar dados em outros contextos, resultando em uma maior quantidade e diversificando manifestações da origem dos dados a analisar.

Esses aspectos assinalados e potencialidades das ferramentas tecnológicas para mineração de dados, particularmente, mineração textual, admitem refletir sobre estudos a serem desenvolvidos com base tanto nos resultados obtidos, quanto de novas camadas que possam chegar a atingir um nível da aderência semântica aos dados publicados.

Além da escolha de outras fontes de informações que tencionam a publicação de dados para os sujeitos alvo, e não sobre eles, propõe-se ampliar o espectro dos dados do detentor governo, usando conjuntos de dados publicados por instituições e não por assuntos, o que entregaria ao menos uma maior quantidade de unidades para serem analisadas, aprofundando nas opções técnicas para análise de aderência lexical entre unidades de texto, até com trabalho interdisciplinar com áreas como da linguística, comunicação e computação, propondo a evolução para um inter-modelo que inclua áreas interessadas no aprimoramento do usufruto de dados disponibilizados.

Outra opção de análise, como o de colocação, foi trazida nesta tese, demonstrando uma alternativa de extração de assuntos a partir do contexto do

usuário, caracterizando-o, e portanto, entregando insumos para futuras pesquisas. É possível combinar a análise de colocação com análise de relevância, opção que evolui no sentido de identificar unidades significativas não unicamente por frequências absolutas, assim como a caracterização de assuntos mediante técnicas de *Topic Modelling*, alternativa para modelagem temático de corpus.

Estas alternativas mencionadas configuram formas de extrair dados a partir de fontes de dados dos sujeitos alvo, porém, é necessário verificar se fornecem indícios acerca do usufruto de dados publicados para eles. Percebe-se que algumas técnicas constituem parte dos percursos metodológicos que podem ser considerados, não como o fim de análises realizadas mediante mineração de textos. Novas propostas precisam coletar direta ou indiretamente dos sujeitos alvo amostras que evidenciem o aproveitamento de dados publicados para eles.

Um outro estudo aponta ao uso do vocabulário controlado AGROVOC, que constitui um sistema de organização do conhecimento, que ao ser mantido por uma comunidade internacional de expertos e coordenado pela FAO, é uma ferramenta que poderia ser usada como alternativa para identificar, logo após a extração automatizada de unidades lexicais, *n-grams*, ou termos advindos de fontes de dados para os sujeitos alvos deste estudo, qual a relação com os termos constituintes dessa ferramenta, avançando na caracterização de relações semânticas entre os dados publicados pelo detentor governo e dados coletados de fontes próprias dos sujeitos alvo.

Em suma, os resultados desta pesquisa e estudos futuros, são ainda oportunidades de reflexão sobre a avaliação da proficuidade do volume de dados disponibilizados para sujeitos alvos como os pequenos e médios produtores, setor certamente relevante para países como o Brasil, onde um dos interesses particulares é a potencialidade de uso outorgada pelo tratamento descritivo dado a esses recursos informacionais.

## REFERÊNCIAS

- ALEIXO, D. V. B. S. et al. A apropriação das Tecnologias de Informação e Comunicação pela agricultura familiar: em foco a Política Nacional de Assistência Técnica e Extensão Rural. *Revista Eletrônica Competências Digitais para Agricultura Familiar*, v. 2, n. 2, p. 81-94, 2016. Disponível em: <http://200.145.54.28:8082/index.php/recodaf/article/view/29/67>
- ALMEIDA, M. A. Mediações da cultura e da informação: perspectivas sociais, políticas e epistemológicas. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 1, n. 1, 2008, p. 1-23., v. 24, n. 2, 2008. Disponível em: <https://brapci.inf.br/index.php/res/download/119743>. Acesso em 03 jan. 2023.
- ARGENTINA. *Encuentro Regional de las Américas de la Alianza para el Gobierno Abierto 2017*. 2022. Disponível em: <https://www.argentina.gob.ar/noticias/encuentro-regional-de-las-americas-de-la-alianza-para-el-gobierno-abierto-2017>. Acesso em 15 jun.2022.
- BANCO MUNDIAL. Tierras agrícolas. 2021. Disponível em: <https://datos.bancomundial.org/indicador/AG.LND.AGRI.K2?end=2018&locations=BR&start=2018&view=map>
- BARRETO, A. de A. A questão da informação. *São Paulo em perspectiva*, v. 8, n. 4, p. 3-8, 1994.
- \_\_\_\_\_. Mudança estrutural no fluxo do conhecimento: a comunicação eletrônica. *Ciência da Informação*, Brasília, v. 27, n.2, p. 122-127, 1998. DOI: <https://doi.org/10.1590/S0100-19651998000200003>.
- BEZJAK, S. et al. Open science training handbook. Tradução. [s.l.] Zenodo, 2018. Disponível em: <https://doi.org/10.5281/zenodo.1212495>.
- BORKO, H. Information Science: What Is It ? *American Documentation*, v. 19, n. 1, p. 3–5, 1968.
- BRASIL. Decreto nº 9.903, de 8 de julho de 2019. Altera o Decreto nº 8.777, de 11 de maio de 2016, que institui a Política de Dados Abertos do Poder Executivo federal, para dispor sobre a gestão e os direitos de uso de dados abertos. 2019. Disponível em [http://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2019/decreto/D9903.htm](http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D9903.htm).
- BRASIL. Ministério do Planejamento, Orçamento e Gestão. Instrução Normativa nº 4, de 12 de abril de 2012. Institui a Infraestrutura Nacional de Dados Abertos - INDA. Disponível em: <http://dados.gov.br/paginas/instrucao-normativa-da-inda>
- BRASIL. Presidência da República. Casa Civil. Lei nº 12.527, de 18 novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm)
- BRASIL. Ministério da Cidadania. Estrutura organizacional. 2021. Disponível em: [http://www.mds.gov.br/webarquivos/cidadania/Organograma%20MC%20-%20Decreto%20jul-2020%20\(3\).pdf](http://www.mds.gov.br/webarquivos/cidadania/Organograma%20MC%20-%20Decreto%20jul-2020%20(3).pdf)

CAMPEROS-REYES, J. T. Metadados nas instruções de governos para publicadores de dados. 2018. Disponível em <http://hdl.handle.net/11449/152838>.

CAMPEROS-REYES, J.T.; BORGES, E. V. E.; MELLO, M. R. G. de; REIS, D. P. dos; NOVAES, F. C. P.; MORAES, I. S. de. CoDAF e a teoria da complexidade: uma relação necessária no contexto da Pós-Modernidade. In: *Aproximando pequenos produtores e o mundo dos dados: Construindo pontes entre necessidades e tecnologias*. 1 ed. Tupã: Faculdade de Ciências e Engenharia UNESP – Câmpus de Tupã, 2019, v.1, p.183-207.

CAMPEROS-REYES, J.T.; ROMANETTO, L.M.; SANT'ANA, R.C.G.; SANTOS, P.L.V.A.C. Intersecção temática de programas de pós-graduação brasileiros: considerações sobre Big Data. In: MARTÍNEZ-ÁVILA, D., SOUZA, E.A., and GONZALEZ, M.E.Q., eds. Informação, conhecimento, ação autônoma e big data: continuidade ou revolução? [online]. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica; FiloCzar, 2019, pp. 67-81. ISBN: 978-85-7249-055-9. Disponível em <https://doi.org/10.36311/2019.978-85-7249-055-9.p67-82>.

CAMPEROS-REYES, J.T.; ROSSI, R. C.; MODENEIS, T. F.; OLIVEIRA, V. H. V. de.; ALEIXO, D. V. B. S. As tecnologias de informação e comunicação e a agricultura familiar: uma revisão de literatura. In: Encontro Competências Digitais para Agricultura Familiar e-CoDAF, 4., Tupã - SP. **Anais** [...] Tupã: Unesp Faculdade de Ciências e Engenharia - Câmpus de Tupã, 2017. Disponível em: [https://dadosabertos.info/events/ecodaf/livecodaf.php/livecodaf\\_proceedings.php?lang=pt\\_BR](https://dadosabertos.info/events/ecodaf/livecodaf.php/livecodaf_proceedings.php?lang=pt_BR).

CAMPEROS-REYES, J. T.; SANT'ANA, R. C. G. Domínio intelectual da literatura de acesso a dados de agricultura na ciência da informação. *Revista Eletrônica Competências Digitais para Agricultura Familiar*, v. 5, n. 1, p. 87-110, 2019. Disponível em <https://owl.tupa.unesp.br/recodaf/index.php/recodaf/article/view/89>

CAMPEROS-REYES, J. T.; SANT'ANA, R. C. G. Abordajes de metadatos para publicadores de datos gubernamentales en el contexto de Iberoamérica. *Revista Interamericana de Bibliotecología*, v. 44, n. 3, e339110, 2021. Disponível em: <https://doi.org/10.17533/udea.rib.v44n3e339110>. Acesso em 13 dez. 2021.

CAMPEROS-REYES, J.T.; ZAFALON, Z. R.; SANTOS, P. L. V. A. da C.; SANTANA, R. C. G. Elementos de modelado para intercambio de información en ciencia de la información e ingeniería de sistemas. *Ciência Da Informação (Online)*, v. 49, n. 1, p 181-195, 2020. Disponível em <http://revista.ibict.br/ciinf/issue/view/284/66>

CAMPEROS-REYES, J.T.; VECHIATO, F. L.; VIDOTTI, S. A. B. G.; SANTANA, R. C. G. Encontrabilidad de la información en sites que promueven Datos Abiertos. *Palabra Clave (La Plata)*, v.10, p.e109, 2020. Disponível em <https://www.palabraclave.fahce.unlp.edu.ar/article/view/PCe109>

CAMPEROS-REYES, J. T.; SANT'ANA, R. C. G.; RAMALHO, R. A. S. A produção científica sobre representação de dados dentro da área temática da agricultura. *Informação & Informação*, v. 22, n. 2, p. 467-480, 2017. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/31465>

CAMPEROS-REYES, J.T.; SANTANA, R. C. G.; RAMALHO, R. A. S. Estruturas de representação de dados agrícolas: trajetória em periódicos e eventos científicos. In: Encontro Internacional Dados, Tecnologia e Informação DTI, 3., 2016. **Anais** [...] Marília: Unesp Faculdade de Filosofia e Ciências – Campus de Marília – SP, 2016.

CAMPEROS-REYES, J.T.; SANTANA, R. C. G.; SANTARÉM SEGUNDO, J. E. Estudo comparativo de datasets governamentais do Brasil e da Colômbia, com Dados de agricultura e desenvolvimento rural. *Anales de Documentación (Internet)*, v.21, p.1 - 14, 2018. Disponível em

<https://revistas.um.es/analesdoc/article/view/302381>

CAMPEROS-REYES, J.T.; SANTANA, R. C. G.; CARVALHO, A. M. G. Tratamento das políticas públicas de informação e tecnologia em periódicos científicos colombianos da ciência da informação. *Revista Interamericana De Bibliotecologia*, v. 41, p. 289-301, 2018. Disponível em

<https://revistas.udea.edu.co/index.php/RIB/article/view/335124>

CARVALHO, A. M. G. de. Apropriação da informação: um olhar sobre as políticas públicas sociais de inclusão digital. 2010. 169 f. Tese (Doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2010. Disponível em:

<http://hdl.handle.net/11449/103358>.

CASTRO, L. N.; FERRARI, D. G. *Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações*. São Paulo: Saraiva, 2016. ISBN 978-85-472-0099-2.

CATALANO, M. M.; VAUGHN, P.; BEEN, J. Using maps to promote data-driven decision-making: one library's experience in data visualization instruction. *Medical reference services quarterly*, v. 36, n. 4, p. 415-422, 2017. Disponível em:

<https://doi.org/10.1080/02763869.2017.1369292>

CKAN. 2021. Disponível em: <https://ckan.org/>

COMISSÃO ECONÔMICA PARA A AMÉRICA LATINA E O CARIBE. La paradoja de la recuperación en América Latina y el Caribe - Crecimiento con persistentes problemas estructurales: desigualdad, pobreza, poca inversión y baja productividad. 2021. Disponível em:

[https://repositorio.cepal.org/bitstream/handle/11362/47043/5/S2100379\\_es.pdf](https://repositorio.cepal.org/bitstream/handle/11362/47043/5/S2100379_es.pdf).

COMISSÃO INTERAMERICANA DE DIREITOS HUMANOS. O direito de acesso à informação no marco jurídico interamericano. 2011. Disponível em:

[https://www.oas.org/pt/centro\\_informacao/default.asp](https://www.oas.org/pt/centro_informacao/default.asp). Acesso em: 15 abr. 2022.

CONFERENCIA REGIONAL DE DATOS ABIERTOS AMÉRICA LATINA Y EL CARIBE. 2017. Disponível em: <https://desdecasa.abrelatam.org/>. Acesso em 20 jul. 2022

CONTROLADORIA-GERAL DA UNIÃO. 2021. Disponível em:

<https://www.gov.br/cgu/pt-br/acesso-a-informacao/dados-abertos/historico>

DAVENPORT, T. H. *Ecologia da informação: porque só a tecnologia não basta para o sucesso na era da informação*. São Paulo: Futura, 1998.

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA. 2021. *Encuesta Nacional Agropecuaria (ENA)*. Disponível em:

[https://www.dane.gov.co/files/investigaciones/agropecuario/enda/ena/2019/boletin\\_e\\_na\\_2019.pdf](https://www.dane.gov.co/files/investigaciones/agropecuario/enda/ena/2019/boletin_e_na_2019.pdf).

DICCIONARIO ETIMOLÓGICO CASTELLANO EN LINEA. Adherencia. 2021.

Disponível em: <http://etimologias.dechile.net/?adherencia>

FECHER, B.; FRIESIKE, S. Open science: one term, five schools of thought. *RatSWD\_WP\_218*, mai. 30, 2013. Disponível em: <http://dx.doi.org/10.2139/ssrn.2272036>. Acesso em: 10 jul. 2022.

FELDMAN, R.; SANGER, J. 2006. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2006. Disponível em: <https://dl.icdst.org/pdfs/files/25a6d982ee80e1db7a4ebf7eeca4e0ec.pdf>. Acesso em 10 dez. 2022.

FERRÁS SEXTO, C.; GARCÍA, Y.; POSE, M. New Ways to Buy and Sell: An Information Management Web System for the Commercialization of Agricultural Products from Family Farms without Intermediaries. In: CRUZ-CUNHA, M. M.; MOREIRA, F. (Eds.). *Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts*. Hershey: Information Science Reference, p. 1182–1198, 2011.

FOSTER. 2021. The Open Science Training Handbook. Disponível em: <https://book.fosteropenscience.eu/en/02OpenScienceBasics/01OpenConceptsAndPrinciples.html>

FROHMANN, B. O caráter social, material e público da Informação. In: FUJITA, M. S. L; MARTELETO, R. M.; LARA, M. L. G. de. (Org.). *A dimensão epistemológica da ciência da informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da informação*. São Paulo: Cultura Acadêmica Editora; Marília: Fundepe, 2008. p. 19-34.

GASCÓ-HERNÁNDEZ, M. et al. Promoting the use of open government data: Cases of training and engagement. *Government Information Quarterly*, v. 35, n. 2, p. 233-242, 2018. Disponível em: <https://doi.org/10.1016/j.giq.2018.01.003>.

GHEPI. 2022. Disponível em: <https://gephi.org/users/download/>

GLOBAL OPEN DATA FOR AGRICULTURE & NUTRITION. *About GODAN*. 2021. Disponível em <https://www.godan.info/www.godan.info/index.html>

GÓMEZ, M. N. G de. As relações entre ciência, Estado e sociedade: um domínio de visibilidade para as questões da informação. *Ciência da Informação*, v. 32, p. 60-76, 2003. Disponível em: <https://www.scielo.br/j/ci/a/KQV7G77RcxK7F6cCH96pVDb/?lang=pt>. Acesso em 02 jan. 2023.

GONÇALEZ, P. R. V. A. Repositórios arquivísticos digitais confiáveis: identificação de requisitos com ênfase no acesso à informação. 2017. Tese doutoral. Disponível em <http://hdl.handle.net/11449/150028>. Acesso em: 14 ago. 2022.

GOOGLE MAPS.2023. Disponível em: <https://www.google.com.br/maps>. Acesso em: 10 jan. 2023.

GRAY, J.; et al. Unlocking the Potential of Aid Information. 2009. Disponível em: <http://www.unlockingaid.info/wp-content/uploads/2010/02/UnlockingAidInformation.pdf>

HAN, J.; KAMBER, M.; PEI, J. *Data mining: concepts and techniques*. 3 ed. Waltham: Morgan Kaufmann. 2012.

HERNÁNDEZ-SAMPIERI, R.; COLLADO, C. F.; LUCIO, M. P. B. *Metodología de la investigación*. 5 ed. México DF: McGraw Hill, 2010.

HUANG, A. Similarity measures for text document clustering. In: PROCEEDINGS OF THE SIXTH NEW ZEALAND COMPUTER SCIENCE RESEARCH STUDENT CONFERENCE (NZCSRSC2008), 6., 2008, Christchurch, New Zealand. Christchurch, 2008. p. 9-56.

INTERNATIONAL COUNCIL FOR SCIENCE. World Data System. 2021. Disponível em <https://www.icsu-wds.org/>

INSTITUTO GEOGRÁFICO AGUSTÍN CODAZZI. 2021. Disponível em: <https://www.igac.gov.co/>

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. 2021a. Disponível em: <https://www.ibge.gov.br>

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. 2021b. Áreas territoriais. Disponível em <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html?=&t=o-que-e>

KAO, A.; POTEET, S. R.. (Ed.). *Natural language processing and text mining*. USA: Springer, 2007.

MARTINES, A. R.; MOREIRA, W.; ALMEIDA, C. C. de. Do signo ao tesouro: contribuições de três correntes da linguagem. *Ciência da Informação*, v. 51, n. 1, 2022. Disponível em: <https://revista.ibict.br/ciinf/article/view/5543>. Acesso em: 10 dez. 2023.

OLIVEIRA, W. (Coord.). *Modelo de referência de abertura de dados*: documento de referência do Marco 5 do Compromisso 2, ecossistema de dados abertos. Rio de Janeiro: FGV DAPP, 2020. Disponível em: [https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/compromisso-2-docs/modelo-de-referencia-de-abertura-de-dados\\_versao-final-2.pdf](https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/compromisso-2-docs/modelo-de-referencia-de-abertura-de-dados_versao-final-2.pdf)

OPEN GOVERNMENT PARTNERSHIP. 2021. Disponível em: <https://www.opengovpartnership.org>

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. *Objetivos de desarrollo sostenible*. 2023. Disponível em: <http://www.un.org/sustainabledevelopment/es/hunger/>. Acesso em: 19 jan. 2023.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A ALIMENTAÇÃO E A AGRICULTURA. 2023. Disponível em: <https://www.fao.org/brasil/pt/>. Acesso em 10 dez. 2022.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. *Government at a Glance 2017*. Paris: OECD Publishing, 2017a. Disponível em: [http://dx.doi.org/10.1787/gov\\_glance-2017-en](http://dx.doi.org/10.1787/gov_glance-2017-en). Acesso em: 20 jul. 2021.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO; ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A ALIMENTAÇÃO E A AGRICULTURA *Perspectivas Agrícolas 2015*. Paris: OECD Publishing, 2015. Disponível em: <https://www.fao.org/3/i4738s/I4738S.pdf>. Acesso em 15 jun. 2022.

PARLAMENTO EUROPEU. Diretiva 2003/98/CE do Parlamento Europeu e do Conselho, 17 nov. 2003. Relativa à reutilização de informações do sector público. *Jornal Oficial*, Bruxelas, n. L 345, p. 90-96, 31 dez. 2003. Disponível em: <http://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32003L0098&rid=1> . Acesso em: 19 jul. 2022.

PEREIRA, T. G. N.; MONTEIRO, E. C. S. A.; CAMPEROS REYES, J.T.; AFFONSO, E. P. Uso das tecnologias da informação e comunicação pela agricultura familiar: um estudo de caso no assentamento Reunidas. *In: Tecnologias de acesso a dados para a gestão de pequenas propriedades rurais*. 1 ed. Tupã: Faculdade de Ciências e Engenharia UNESP, 2019, v.1, p. 55-80.

PESET, F.; FERRER-SAPENA, A.; SUBIRATS-COLL, I. Open data y linked open data: su impacto en el área de bibliotecas y documentación. *Profesional de la Información*, v. 20, n. 2, p. 165-174, 2011. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2011.mar.06/21218>. Acesso em 11 nov. 2022.

PAINEL MONITORAMENTO DE DADOS ABERTOS. 2023. Disponível em: <https://centralpaineis.cgu.gov.br/visualizar/dadosabertos>. Acesso em 14 jan. 2023.

RUEDIGER, M. A. Perspectivas da governança na era da informação: estado e sociedade civil. *In: MARTINS, P. E. M.; PIERANTI, O. P. (orgs.) Estado e gestão pública: visões do Brasil contemporâneo*. Rio de Janeiro: FGV, 2006.

SAMPAIO, R.F.; MANCINI, M.C. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. *Revista Brasileira de Fisioterapia*, São Carlos, v. 11, n. 1, p. 83–89, 2007. Disponível em: <https://www.scielo.br/j/rbfis/a/79nG9Vk3syHhnSgY7VsB6jG/?lang=pt&format=html>. Acesso em 03 jan. 2023.

SANT'ANA, R. C. G. Transdução Informacional: impactos do controle sobre os dados. *In: MARTÍNEZ-ÁVILA, D; SOUZA, E.A.; GONZALEZ, M. E. Q. (orgs) Informação, conhecimento, ação autônoma e big data: continuidade ou revolução? Cultura Acadêmica - FiloCzar - São Paulo*. 2019. p.117-128. ISBN 978-85-7249-054-2

SANT'ANA, R.C.G. Campo Informacional Resultante da Interação de Ciclos de Vida dos Dados. *In: DIAS, G.; FREIRE, B. (orgs). Dados Científicos: perspectivas e desafios*. Editora UFPB - João Pessoa. 2018 p.5-19.

SANT'ANA, R. C. G. Diagrama estrutural para teses e dissertações: uma proposta didática. *In: MOREIRA, F. M. at al. (orgs). Acesso a Dados e a Ciência da Informação: aplicação, tendências e reflexões*. Tupã: Faculdade de Ciências e Engenharia UNESP – Campus de Tupã, 2022. p. 330-345. Disponível em: <https://dadosabertos.info/events/dti/4dti.pdf>

SANTANA, R. C. G. *Tecnologia e gestão pública municipal: mensuração da interação com a sociedade* [online]. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2009. 178 p. ISBN 978-85-7983-010-5. Disponível em: <https://books.scielo.org/id/8v2y2>. Acesso em: 21 mar. 2022.

SANTANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. *Informação & Informação*, v. 21, n. 2, p. 116-142, 2016. DOI: 10.5433/1981-8920.2016v21n2p116. Acesso em: 10 ago. 2021

SANTOS, P. L. V. A. C. Informação e tecnologia para o conhecimento: desafios da ciência da informação. *In: DEL ROIO, M (Org.). A universidade entre o conhecimento e o trabalho: o dilema das Ciências*. Marília: UNESP, 2005. p. 127-139.

SANTOS, P. L. V. A. C. Catalogação, formas de representação e construções mentais. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 6, n. 1,

2013. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/119476>. Acesso em: 15 jan. 2023.

SANTOS, B. S.; RODRÍGUEZ, C. Introdução: para ampliar o cânone da produção. In: SANTOS, B. S. (org.). *Produzir para viver: os caminhos da produção não capitalista*. Rio de Janeiro: Civilização brasileira, v. 3, 2002, p. 23-74.

SANTOS, P. L. A. C.; SANT'ANA, R. C. G. Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela ciência da informação. *Ciência da Informação*, Brasília, v.42, n.2, 2013. Disponível em: [www.revista.ibict.br/ciinf/article/view/1382](http://www.revista.ibict.br/ciinf/article/view/1382). Acesso em: 07 maio 2022.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. *Perspectivas em ciência da informação*, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/22308>. Acesso em: 12 maio 2022

SILVA, E. R. A. da. (coord.). *Agenda 2030: ODS – Metas Nacionais dos Objetivos de Desenvolvimento Sustentável*. Instituto de Pesquisa Econômica Aplicada. Governo Federal Ministério do Planejamento, Desenvolvimento e Gestão. 2018. Disponível em: <http://repositorio.ipea.gov.br/handle/11058/8855>. Acesso em: 04 jun. 2022.

SINGER, P. Prefácio. In: GAIGER, L. I. G. (coord.), KUYVEN, P. S. et al. *A economia solidária no Brasil: uma análise de dados nacionais*. São Leopoldo: Oikos, 2014. p. 11-12.

SOARES, N. F.; CAMPEROS-REYES, J.T. Usabilidade em ambientes informacionais digitais para pequenos produtores. In: *Aproximando pequenos produtores e o mundo dos dados: Construindo pontes entre necessidades e tecnologias*. 1 ed. Tupã: Faculdade de Ciências e Engenharia UNESP – Câmpus de Tupã, 2019, v.1, p. 163-182.

SOCRATA. 2021. Disponível em: <https://www.tylertech.com/products/socrata>. Acesso em: 13 dez. 2021.

THE INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. *DESENVOLVIMENTO E ACESSO À INFORMAÇÃO 2019*. Disponível em <https://da2i.ifla.org/wp-content/uploads/da2i-2019-full-report-pt.pdf>

UNIVERSITY OF TORONTO. G7 Information Centre. 2017. Disponível em: <http://www.g8.utoronto.ca/>. Acesso em 30 dez. 2017.

VERHULST, S. G.; YOUNG, A. Open data in developing economies: Toward building an evidence base on what works and how. *African Minds*, 2017. Disponível em: <http://odimpact.org/files/odimpact-developing-economies.pdf>

VERHULST, S. G.; et al. *The emergence of a third wave of open data*. [SI]: GovLab. 2020. Disponível em <https://opendatapolicylab.org/images/odpl/third-wave-of-opendata.pdf>

VIDOTTI, S. A. B. G.; MOREIRA, F. M.; CAMPEROS REYES, J.T.; RODAS, C. M.; SANTANA, R. C. G. Aplicação da triangulação de métodos para avaliação da usabilidade em ambientes informacionais digitais especializados: um estudo no Portal CoDAF. *Informação & Informação (Online)*, v.23, p.586 - 624, 2018. Disponível em <http://www.uel.br/revistas/uel/index.php/informacao/article/view/29802>

W3C. Cascading Style Sheets. 2021a. Disponível em:

<https://www.w3.org/Style/CSS/Overview.en.html>

W3C. Data Catalog Vocabulary (DCAT) - Version 3. 2021b. Disponível em:

<https://www.w3.org/TR/vocab-dcat-3/>

WERSIG, G. Information science: the study of postmodern knowledge usage.

*Information processing & management*, v. 29, n. 2, p. 229-239, 1993. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/030645739390006Y>. Acesso

em: 02 dez. 2022.

WORLD BANK GROUP. Comparativo plataformas comparativo plataformas de datos abiertos. 2015. Disponível em: <https://encurtador.com.br/syCG0>. Acesso em 13 dez.

2021.