



REDE BIBLIODATA:

estudo bibliométrico da catalogação cooperativa

Camila Ribeiro Teixeira

<https://orcid.org/0009-0009-4112-4089>.

camilateixeira@ibict.br.

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) |

<https://ror.org/006c42y96> | Brasília, Brasil.

Tamiê Moraes de Sousa

<https://orcid.org/0009-0005-3588-8905>.

tamiesousa@ibict.br.

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) |

<https://ror.org/006c42y96> | Brasília, Brasil.

Amanda Nunes Soares

<https://orcid.org/0009-0004-0909-694X>.

amandasoares@ibict.br.

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) |

<https://ror.org/006c42y96> | Brasília, Brasil.

Josina da Silva Vieira

<https://orcid.org/0000-0002-9727-3771>.

josinavieira@ibict.br.

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) |

<https://ror.org/006c42y96> | Brasília, Brasil.

Letícia Santos de Jesus

<https://orcid.org/0009-0008-7556-9660>.

leticiajesus@ibict.br.

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) |

<https://ror.org/006c42y96> | Brasília, Brasil.

Gabrielle Helpis dos Santos Gandra

<https://orcid.org/0009-0002-9964-3710>.

gabrielle.helpis@gmail.com.

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) |

<https://ror.org/006c42y96> | Brasília, Brasil.

Ana Carolina Simionato Arakaki

<https://orcid.org/0000-0002-0140-9110>.

anaarakaki@ibict.br.

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) |

<https://ror.org/006c42y96> | Brasília, Brasil.

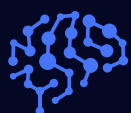
Eixo temático: Produtividade e Colaboração Científica

Modalidade: Resumo expandido

DOI: 10.22477/ix.ebbc.277

Resumo: Este estudo busca aplicar a Lei de Zipf, uma das leis da bibliometria, para identificar padrões nas estruturas temáticas dos dados bibliográficos provenientes da base legada da rede Bibliodata. Para alcançar esse objetivo, a Lei foi adaptada para analisar os termos tópicos desses registros catalogados de livros. As informações coletadas indicam que as Ciências Sociais foram as mais frequentemente registradas pela rede Bibliodata no início do século XXI, fornecendo um suporte vital para os pesquisadores dessa área. Isso ressalta a importância de uma colaboração mais abrangente com outras áreas do conhecimento, a fim de impulsionar a catalogação cooperativa no Brasil.

Palavras-Chave: Bibliometria. Lei de Zipf. Bibliodata. Indexação Temática.



1 INTRODUÇÃO

Originada em 1942 como uma iniciativa pioneira no país, a rede Bibliodata desempenha um papel crucial na criação de uma rede de catalogação cooperativa no Brasil. Esta infraestrutura representa uma base integrada de bibliotecas com o propósito específico de promover a divulgação dos acervos bibliográficos nacionais, os quais estão acessíveis ao público, e compreendem uma vasta coleção de registros (Instituto Brasileiro de Informação em Ciência e Tecnologia, 2023).

Sua responsabilidade primordial consiste em consolidar os registros bibliográficos provenientes das bibliotecas participantes em território brasileiro, estabelecendo assim uma rede de catalogação cooperativa de acesso aberto. A catalogação cooperativa refere-se a cooperação entre bibliotecas e permite que estas compartilhem dados bibliográficos, o que é conduzido por profissionais da informação – catalogadores – com objetivos comuns. Essa prática reduz a redundância na catalogação, economizando tempo e recursos ao evitar o retrabalho na catalogação de registros já disponíveis (Godinho; Faria, 2014, p. 5). Nesse contexto, a rede Bibliodata abrange uma quantidade significativa de dados associados aos registros, o que permite análises estratégicas das informações extraídas de estudos bibliométricos.

O objetivo deste estudo consiste em identificar métricas bibliométricas através da análise dos registros legados do sistema da rede Bibliodata, com o intuito de identificar os temas emergentes na rede, proporcionando uma visão macro da sua atuação e destacando sua importância na difusão da informação científica brasileira.

A pesquisa documental foi adotada como procedimento metodológico, conforme definido por Gerhard *et al.* (2009, p. 71), que a descrevem como “[...] aquela realizada a partir de documentos, contemporâneos ou retrospectivos, considerados cientificamente autênticos [...]”. A abordagem foi quantitativa, destacando a objetividade que “[...] recorre à linguagem matemática para descrever as causas de um fenômeno, as relações entre variáveis, etc.” (Fonseca, 2002, p. 20). Durante a coleta de dados, a Lei de Zipf foi aplicada para formular indicadores bibliométricos com base nos termos de assuntos de livros catalogados na rede Bibliodata, evidenciando os benefícios de uma rede de colaboração catalográfica brasileira e enfatizando a necessidade de integração entre os centros informacionais do país para otimizar a difusão científica nacional.

2 CONTEXTUALIZAÇÃO: BIBLIOMETRIA E A LEI DE ZIPF

A bibliometria, uma subárea da Biblioteconomia e da Ciência da Informação, emprega ferramentas estatísticas para identificar indicadores relacionados à gestão da informação, produção e disseminação do conhecimento. De acordo com Costa *et al.* (2012), a bibliometria é uma técnica quantitativa que mensura índices de produção, disseminação do conhecimento, desenvolvimento das áreas científicas, padrões de autoria e publicações, formando um “[...] conjunto de leis e princípios empíricos que contribuem para estabelecer os fundamentos teóricos da Ciência da Informação” (Guedes; Borschiver, 2005, p. 2). As principais leis bibliométricas são: 1) Lei de Bradford (produtividade de periódicos); 2) Lei de Lotka (produtividade científica de autores); e 3) Lei de Zipf (frequência de palavras).

Este estudo concentra-se na Lei de Zipf, que objetiva “[...] aferir a distribuição de frequências de palavras de um texto” (Rodrigues; Viera, 2016, p. 169), aplicada no contexto da análise de termos tópicos catalogados no banco de dados legado¹ da rede Bibliodata, com o objetivo de consolidar a catalogação cooperativa e a integração das áreas do conhecimento na disseminação científica no Brasil. Para cumprimento dos objetivos desta pesquisa, utilizou-se a primeira Lei de Zipf (número de ocorrência das palavras), e para medir esse número de ocorrências das palavras utiliza-se a seguinte expressão matemática:

$$K = R.F$$

A fórmula explicitada acima simboliza K= constante; R= ordem das palavras; e F= frequência das palavras. Ou seja, multiplica o número da ordem das palavras com o quantitativo da frequência de palavras para obter o resultado da constante. Seguindo o princípio do menor esforço, proposto por Zipf e elucidado por Rodrigues e Viera (2016), existe uma tendência à economia no uso de palavras. Isso implica que, quando se busca utilizar o menor número possível de palavras, estas não se dispersam, mas sim são repetidas, e as palavras mais frequentemente utilizadas indicam o tema predominante do documento.

Aplicando o princípio de Zipf à indexação temática de registros bibliográficos na atividade catalográfica, é viável mensurar quais assuntos foram mais frequentemente indexados nos bancos de dados legados da rede Bibliodata. Isso ocorre porque, durante o processo de catalogação, os profissionais selecionam, por meio de uma análise técnica do documento, as palavras-chave mais relevantes para facilitar a recuperação de informações pelos usuários.

Portanto, adaptou-se a primeira Lei de Zipf, para a mensuração no corpus caracterizado no campo 650 (Assunto tópico) do formato *Machine Readable Cataloging* (MARC 21) dos registros bibliográficos catalogados na rede Bibliodata. O objetivo é observar a predominância de temáticas catalogadas referente aos acervos das bibliotecas integrantes da rede. Dessa forma, é possível inferir os termos mais frequentes, que serão demonstrados no próximo tópico, observando o seguinte: “a Zona I – Informação trivial ou básica define os temas centrais da análise bibliométrica; a Zona II – Informações interessantes [...] mostra ora os temas periféricos, ora a informação potencialmente inovadora” (Rodrigues; Viera, 2016, p. 170). Assim, considerou-se os termos catalogados caracterizando os assuntos mais frequentes dos acervos cooperados entre a rede Bibliodata durante o período 2000 a 2003.

3 ANÁLISE E RESULTADOS OBTIDOS

Considerando que o principal objetivo deste estudo é gerar dados bibliométricos a partir dos registros legados do sistema da rede Bibliodata com o intuito de identificar os temas emergentes mais catalogados no início do século, a primeira etapa envolveu a delimitação dos campos do formato MARC21 pertinentes para análise, com foco na aplicação da primeira Lei de Zipf para o campo 650 (Assunto Tó-

¹ “Diz-se que um banco de dados é legado quando armazena dados que são utilizados por um sistema ou software desenvolvido por uma tecnologia ‘mais antiga’ em comparação com as mais atuais do mercado, dita ‘tecnologia legada’” (Marques, 2022, p. 47).



pico). Após o tratamento dos dados disponíveis da base legada, referente aos registros dos anos 2000 a 2003, foram identificados os seguintes termos mais frequentes na Tabela 1.

Tabela 1 – Assuntos ordenados e mais frequentes (F) no campo 650 do banco de dados legado da rede Bibliodata

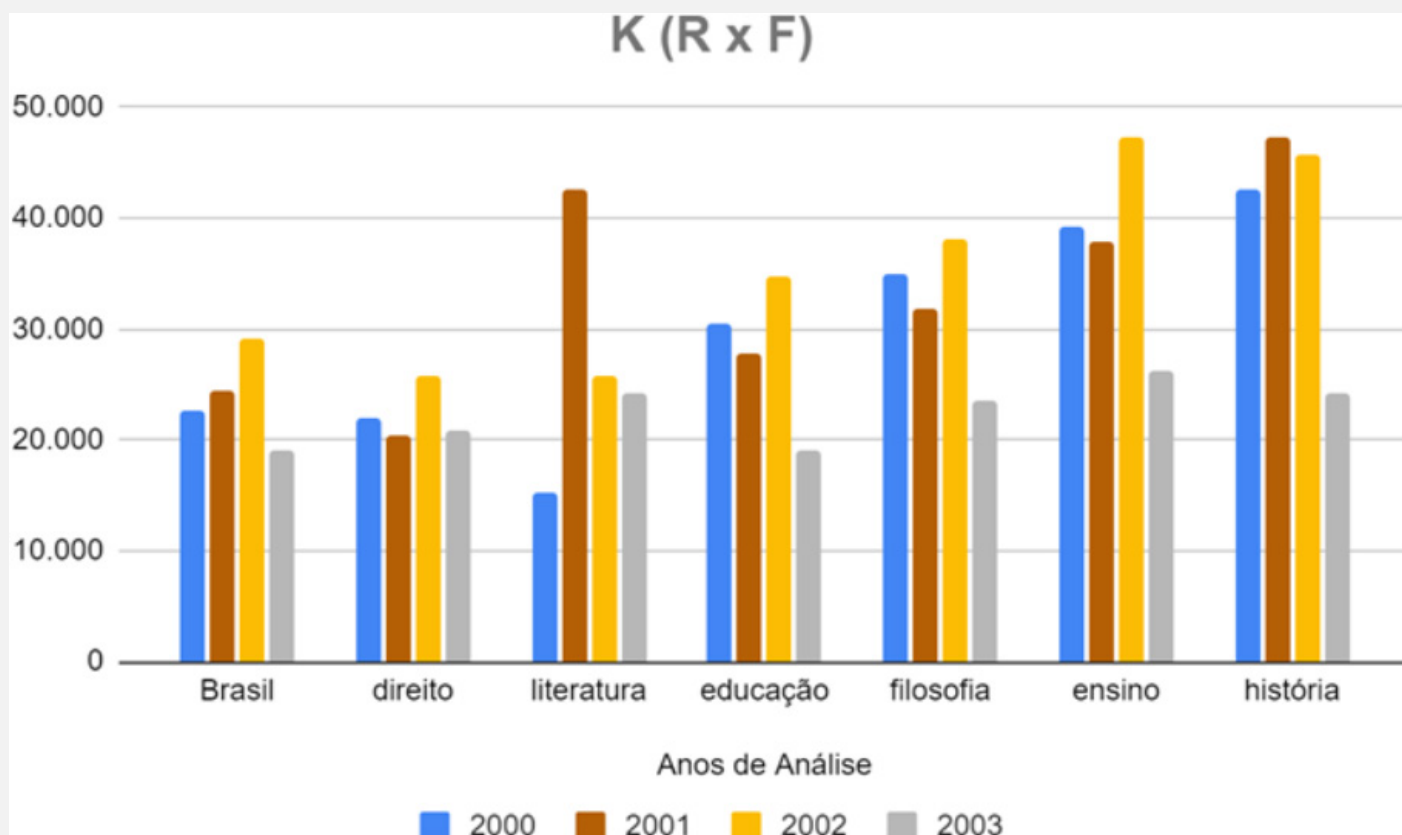
Termo (TF)	Ano	Frequência (F)	Ordem (R)	K (R x F)
Brasil (95.342)	2000	22.683	1	22.683
	2001	24.370		24.370
	2002	29.222		29.222
	2003	19.067		19.067
direito (44.454)	2000	10.953	2	21.906
	2001	10.189		20.378
	2002	12.894		25.788
	2003	10.418		20.836
literatura (35.951)	2000	5.086	3	15.258
	2001	14.184		42.552
	2002	8.620		25.860
	2003	8.601		24.183
educação (28.003)	2000	7.641	4	30.564
	2001	6.929		27.716
	2002	8.666		34.664
	2003	4.767		19.068
filosofia (25.687)	2000	6.974	5	34.870
	2001	6.348		31.740
	2002	7.642		38.210
	2003	4.723		23.615
ensino (25.076)	2000	6.531	6	39.186
	2001	6.298		37.788
	2002	7.886		47.316
	2003	4.361		26.166
história (22.840)	2000	6.083	7	42.581
	2001	6.768		47.376
	2002	6.519		45.633
	2003	3.470		24.290

Fonte: Elaboração própria (2024).

Como é possível observar, os termos mais frequentes foram assuntos que podem ser associados a diversas áreas do conhecimento, apresentando predominância ao serem mais utilizados por áreas relacionadas às Ciências Humanas e Sociais. Abaixo é possível visualizar o Gráfico 1 que apresenta a

incidência de palavras após a aplicação da primeira Lei de Zipf.

Gráfico 1 – Resultados de assuntos mais frequentes no banco de dados legado da rede Bibliodata entre 2000 - 2003



Fonte: Elaboração própria com base na aplicação da Lei de Zipf de palavras mais frequentes de assuntos catalogados na rede Bibliodata (2024).

É importante destacar que o campo 650 (Assunto Tópico) do formato MARC21 apresenta subcampos específicos na rede Bibliodata: a (Termo tópico ou elemento de entrada de nome geográfico); x (Subdivisão geral); y (Subdivisão cronológica); e z (Subdivisão geográfica) (Library of Congress, 2023). Observa-se que o termo “Brasil” é frequentemente encontrado nos assuntos catalogados devido às associações com outras temáticas indexadas em publicações sobre o país, identificadas no subcampo “z” do MARC21.

Para complementar a análise realizada anteriormente, a segunda etapa do estudo envolveu a análise manual dos campos de classificação decimal por área do conhecimento (campos MARC21 080 e 082), a fim de reforçar a relação com os assuntos mais indexados no banco de dados legados, conforme apresentado na Tabela 2.

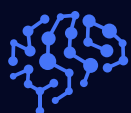


Tabela 2 – Quantitativo de registros de acordo com as classes das classificações decimais.

Áreas do Conhecimento:	Classificação	Ano					
		2000	2001	2002	2003	Total	Total CDU +CDD
CLASSE 0	CDU	2.359	3.367	4.297	2.254	12.277	30.498
	CDD	5.340	4.927	4.873	3.081	18.221	
CLASSE 01	CDU	3.095	2.892	3.937	2.494	12.418	33.104
	CDD	6.837	3.639	6.350	3.860	20.686	
CLASSE 02	CDU	791	484	651	443	2.369	10.254
	CDD	2.822	1.436	1.568	2.059	7.885	
CLASSE 03	CDU	14.381	15.843	23.271	21.518	75.013	155.286
	CDD	24.089	20.605	23.036	12.543	80.273	
CLASSE 04	CDD	2.070	1.543	1.941	844	6.398	6.398
CLASSE 05	CDU	5.239	6.633	8.664	4.592	25.128	56.175
	CDD	8.675	8.902	7.858	5.612	31.047	
CLASSE 06	CDU	10.103	10.549	16.313	10.580	47.545	106.835
	CDD	15.812	16.565	15.424	11.489	59.290	
CLASSE 07	CDU	2.843	3.421	4.623	3.315	14.202	30.097
	CDD	3.263	3.977	4.299	4.356	15.895	
CLASSE 08	CDU	4.705	6.339	7.537	7.509	26.090	49.202
	CDD	5.807	5.942	7.704	3.659	23.112	
CLASSE 09	CDU	2.871	3.504	4.612	3.418	14.405	31.774
	CDD	6.384	3.639	4.694	2.652	17.369	

Fonte: Elaboração própria (2024).

Nota: A classe com mais registros encontrados, em ambas as classificações, foi a classe 3 - Ciências Sociais.

Analisando simultaneamente as Tabelas 1 e 2, nota-se que os termos indexados mais frequentes (Brasil, literatura, educação, filosofia, ensino e história) estão associados à classe 3 da classificação dos registros bibliográficos, pois trazem assuntos relacionados às Ciências Sociais. Isso mostra a relevância das bibliotecas participantes da rede Bibliodata na cooperação catalográfica para os acervos que possuem essa temática.

Além disso, foi conduzida uma análise dos registros no campo 111 (Entrada principal de evento) para mapear a relação entre assunto, classificação e nome do evento. Este campo abrange uma variedade de eventos, como conferências, seminários, congressos, encontros, colóquios, exposições, feiras, entre outros, conforme a Library of Congress (2023). No entanto, na análise da Tabela 3, o foco recai sobre os assuntos e não sobre os tipos de eventos, portanto, esses termos foram excluídos da lista.

Tabela 3 – Quantitativos de registros de Autoridades de Evento (Campo 111) de 2000 a 2003 mais frequentes

Ano	Qtd. de eventos	Categorias (frequência)				
Total de registros:	12.996	Localidade	Educação/ Ensino	Ciência/Pesquisa	Entidade	Áreas do conhecimento
2000	2.341	Nacional (1526)				Engenharia (276)
2001	3.365					Geologia (223)
2002	4.909					Desenvolvimento (201)
2003	2.381	Internacional (947)	Educação (419) Universidade (474) Ensino (274) Graduação (122)	Pesquisa (632) Científica (619) Ciência (152) Iniciação (772) Produção (141)	Sociedade (282) Associação (155)	Tecnologia (188)
		Internacional (857)				Psicologia (182)
		Brasileira (501)				Ambiente (130)
		Brasil (374)				Enfermagem (125)
		Sul (263)				Sistemas (118)
		Americano (257)				
		Latino (216)				
		Regional (213)				

Fonte: Elaboração própria (2024).

A frequência de termos não agrupou os termos semelhantes em questão de variação de língua (português, inglês, espanhol etc.), tendo em vista a ocorrência de eventos estrangeiros que também são indexados, como é o caso dos termos “international” e “internacional”. Por meio da comparação entre os termos mais frequentes da Tabela 1, as classes mais frequentes da Tabela 2 e as palavras mais frequentes dos Eventos da Tabela 3 é possível relacionar que os assuntos mais indexados na rede Bibliodata correspondem às áreas da Classe 3 (Ciências Sociais) e Classe 6 (Ciências Aplicadas) (Tabela 2) e dos assuntos “Brasil, literatura, educação, filosofia, ensino e história” (Tabela 1), destacando os termos mais frequentes: Nacional (1526), Brasileira (501), Brasil (374); Educação (419), Ensino (274); Pesquisa (632), Científica (619); Engenharia (276), Desenvolvimento (201), Tecnologia (188) e Ambiente (130).

A partir dos resultados, destaca-se que a pesquisa desempenha um papel relevante diante da rede Bibliodata e seu relançamento, destacando a importância de compreender o potencial da catalogação cooperativa no Brasil disponibilizada pela rede. Essa compreensão não apenas beneficia pesquisadores, estudantes e bibliotecários, mas também colabora significativamente com a comunidade acadêmica, e instituições fomentadoras, fortalecendo a divulgação científica.

Além disso, concordando com Guedes e Borschiver (2005, p. 15), que consideram a bibliometria como um “[...] instrumento quantitativo, que permite minimizar a subjetividade inerente à indexação e recuperação das informações, produzindo conhecimento, em determinada área de assunto”, a pesquisa contribui assim, “[...] para tomadas de decisão na gestão da informação e do conhecimento, uma vez que auxilia na organização e sistematização de informações científicas e tecnológicas”. Portanto, o



estudo não apenas ressalta a importância da catalogação cooperativa, mas também destaca o impacto positivo que a bibliometria pode ter na gestão eficaz da informação e do conhecimento associada ao controle bibliográfico nacional, um dos papéis da rede Bibliodata no Brasil.

4 CONSIDERAÇÕES FINAIS

A rede Bibliodata está passando por uma reestruturação e é essencial que as bibliotecas contribuam ativamente aumentando os registros em todas as áreas do conhecimento para consolidar sua importância nos centros informacionais em todo o país. Durante o trabalho, foram identificados os termos mais comuns que abrangem diversas áreas do conhecimento, principalmente nas Ciências Humanas e Sociais. Temas emergentes na rede incluem Brasil, literatura, educação, filosofia, ensino e história.

Para estudos futuros, é recomendada a expansão da análise de resultados para incluir a aplicação da 2ª Lei de Zipf e da Lei de Lotka na avaliação da produtividade científica e dos respectivos autores registrados no catálogo de autoridades. Isso permitirá examinar a produtividade e a comunicação científica presentes nos acervos das bibliotecas brasileiras participantes da rede.

Esses dados são de grande relevância para a disseminação e acesso por parte dos pesquisadores em diversas áreas do conhecimento, contribuindo para o avanço da produção científica no Brasil. Além disso, fornecem informações históricas sobre a produção acadêmica de acordo com as áreas do conhecimento, auxiliando na tomada de decisões e no direcionamento de recursos e políticas por parte das entidades responsáveis pelo financiamento de projetos de pesquisa em nível nacional.

REFERÊNCIAS

COSTA, Teresa *et al.* A bibliometria e a avaliação da produção científica: indicadores e ferramentas. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 11., 2012, Lisboa.

Actas [...]. Lisboa: BAD, 2012. Tema: Integração, Acesso e Valor Social. Disponível em: <http://hdl.handle.net/10071/6513>. Acesso em: 29 jan. 2024.

FONSECA, João José Saraiva da. **Metodologia da pesquisa científica**. Fortaleza: UECE, 2002. Apostila.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Bibliodata**. Publicado em: 12 ago. 2022. Atualizado em: 27 jul. 2023. Disponível em: <https://www.gov.br/ibict/pt-br/servicos/informacao-cientifica/bibliodata>. Acesso em: 23 jan. 2024.

GERHARDT, Tatiana Engel *et al.* Estrutura do projeto de pesquisa. In: GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (org.). **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009. Unidade 4, p. 67-90. Disponível em: <https://lume.ufrgs.br/handle/10183/52806>. Acesso em: 4 jun. 2024.

GODINHO, Flavia Martins Alves; FARIA, Francieleide Miguelina dos Santos. Catalogação cooperativa: história, vantagens e desvantagens. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, v. 4, n. 2, 2014. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/16990>. Acesso



em: 28 mar. 2024.

GUEDES, Vânia L. S.; BORSCHIVER, Suzana. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. *In*: CINFORM: ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 6., 2005, Salvador.

Anais [...]. Salvador: ICI/UFBA, 2005. Tema: Informação, Conhecimento e Sociedade Digital. Disponível em: https://cinform-antiores.ufba.br/vi_anais/docs/VaniaLSGuedes.pdf. Acesso em: 7 fev. 2024.

LIBRARY OF CONGRESS. **MARC Standards**. Washington: The Library of Congress, June 15, 2023. Disponível em: <https://www.loc.gov/marc>. Acesso em: 8 fev. 2024.

MARQUES, Glauco Lauria. **Proposta de um framework para apoiar o setor público na adequação dos bancos de dados legados à LGPD**. 2022. Dissertação (Mestrado Stricto Sensu em Governança, Tecnologia e Inovação) – Escola de Educação, Tecnologia e Comunicação, Universidade Católica de Brasília, Brasília, 2022. Disponível em: <https://bdtd.ucb.br:8443/jspui/handle/tede/3141>. Acesso em: 1 fev. 2024.

RODRIGUES, Charles; VIERA, Angel Freddy Godoy. Estudos bibliométricos sobre a produção científica da temática Tecnologias de Informação e Comunicação em bibliotecas. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 7, n. 1, p. 167-180, mar./ago. 2016. DOI: <https://doi.org/10.11606/issn.2178-2075.v7i1p167-180>. Disponível em: <https://www.revistas.usp.br/incid/article/view/98761>. Acesso em: 23 jan. 2024.