

**DANIEL STRAUCH RIBEIRO**

Evolução do escopo temático da ciência aberta:  
um estudo exploratório

**Dissertação de Mestrado**  
**Abril de 2020**



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO – UFRJ  
ESCOLA DE COMUNICAÇÃO – ECO  
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA –  
IBICT  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – PPGCI

**DANIEL STRAUCH RIBEIRO**

Evolução do escopo temático da ciência aberta: um estudo exploratório

RIO DE JANEIRO  
2020

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO – UFRJ  
ESCOLA DE COMUNICAÇÃO – ECO  
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA –  
IBICT  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – PPGCI

**DANIEL STRAUCH RIBEIRO**

Evolução do escopo temático da ciência aberta: um estudo exploratório

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação do Instituto Brasileiro de Informação em Ciência e Tecnologia e Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Ciência da Informação

Orientadora: Profa. Dra. Sarita Albagli

RIO DE JANEIRO

2020

### Catálogo da Publicação na Fonte

Ribeiro, Daniel Ribeiro

R484e Evolução do escopo temático da ciência aberta: um estudo exploratório / Daniel Strauch Ribeiro. – Rio de Janeiro, 2020.  
137 f.

Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio de Janeiro, Escola de Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós- Graduação em Ciência da Informação, Rio de Janeiro, 2020.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Sarita Albagli

1. Ciência Aberta. 2. Ciência da Informação. 3. Análise de Domínio.  
4. Campo Científico. 5. Organização do Conhecimento. I. Albagli, Sarita (Orient.) II. Universidade Federal do Rio de Janeiro. III. Título.

CDD 020

**DANIEL STRAUCH RIBEIRO**

Evolução do escopo temático da ciência aberta: um estudo exploratório

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação do Instituto Brasileiro de Informação em Ciência e Tecnologia e Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Ciência da Informação

Aprovada em 29 de abril de 2020.

---

Profa. Dra. Sarita Albagli – Orientadora  
PPGCI/IBICT – UFRJ/ECO

---

Dr. Andre Luiz Appel  
IBICT-DF

---

Prof. Dr. Fabio Castro Gouveia  
PPGCI/IBICT – UFRJ-ECO

Dedico esse trabalho a todos que contribuem para o movimento da Ciência Aberta, bem como aqueles que tiveram a paciência de ouvir minhas ponderações sobre o trabalho e, especialmente, a todos os meus professores, amigos e parentes que me apoiaram e incentivaram nessa jornada.

## AGRADECIMENTOS

A palavra “agradecer” tem a raiz no termo latim “gratus”, que significa ser acolhido ou acolher com favor, de maneira cortês. E como, na atual conjuntura do país, não reconhecer todo o abrigo e proteção necessários para fazer e concluir o curso de pós-graduação *stricto sensu*?

Por isso, começo manifestando minha enorme gratidão aos meus pais que, apesar da correria constante e das diversas atribuições da vida acadêmica e profissional, sempre foram um porto seguro e estiveram dispostos a ouvir e aconselhar. No período final do trabalho, eu saí da casa deles e ambos me apoiaram bastante, amenizando pressões e deixando a vida mais leve a cada bate-papo.

Também sou imensamente grato a minha consorte Daniele Gullo e a sua família. Dividir minha vida com ela não só colore meu dia a dia, como também realça todas as perspectivas dessa nossa biografia compartilhada. Toda angústia, obstáculo, confusão ou imperfeição foram suavizados ao seu lado, não só por ter com quem compartilhar, mas sim por poder ponderar e decidir ao lado de quem amo. Agradeço muito por toda atenção, ajuda, carinho e amor.

Tenho profunda gratidão aos meus amigos, que entenderam meus afastamentos, sumiços e preocupações ao longo de todo esse período. Com um destaque especial para Iacy Storino e Luiz Henrique Pereira, por me apoiarem sempre, e Talita Loureiro e Alex Barbarioli, por ouvirem minhas reclamações da vida sem reclamar de volta.

Agradeço também aos meus professores do curso de Biblioteconomia e Gestão de Unidades de Informação da Universidade Federal do Rio de Janeiro, bem como dos professores do Programa de Pós-Graduação do Instituto Brasileiro de Informação em Ciência e Tecnologia. O mestrando que fui e a pesquisa que elaborei são frutos de cada aula que assisti, textos que li e exercícios que fiz.

Sou muitíssimo grato a minha estimada orientadora, Sarita Albagli. A sua experiência e sabedoria, aliada aos seus pontos de vistas, contribuições construtivas e engenhosas considerações foram fundamentais para dar seguimento à pesquisa. Apreciei todas as conversas que tivemos.

E, em derradeiro, mas não menos importante, agradeço a você que leu essa seção de agradecimentos. Ler todo seu conteúdo, e até o fim, é apurar um pouquinho mais da intimidade do autor.

*Ler atentamente a complexa história das transformações que criaram o ethos da ciência aberta é o primeiro passo para uma apreciação mais completa de uma construção cultural bastante frágil, na medida em que uma constelação de distintas instituições apoia e molda a orientação da pesquisa científica e técnica. Estamos lidando com o legado de um processo histórico extenso, intrincado e fortuito, que não pode ser entendido como a produção de um sistema fundamental de equilíbrio e regeneração automáticos que responde somente aos imperativos da técnica científica moderna e que, portanto, não requer preservação social nem proteções políticas. É certo que, como é o caso de muitas composições formais, as infraestruturas institucionais da "ciência aberta" possuem uma certa plasticidade. Mas a elasticidade tem limites (DAVID, 2008, p. 6, tradução nossa).*

## RESUMO

RIBEIRO, Daniel Strauch. **Evolução do escopo temático da ciência aberta**: um estudo exploratório. 2020. 137 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicação, Universidade Federal do Rio de Janeiro, Programa de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2020.

Diante da complexidade conceitual da Ciência Aberta, a falta de um consenso sobre a definição da expressão “*open science*” agrava a disputa semântica da produção acadêmica. Esta pesquisa busca alcançar uma operacionalização teórica que caracterize a temática e a terminologia da Ciência Aberta na literatura formal, destacando as repercussões de algumas de suas vertentes em torno de seu campo semântico. Por meio de uma pesquisa exploratória, fez-se um constructo teórico associado a metodologias de estudos métricos com o objetivo de validar uma sistematização de Análise de Domínio que pudesse comprovar o Campo Científico da Ciência Aberta. Para isso, reuniu-se e confrontou-se os resultados encontrados em pesquisas pelo termo “*open science*” em três bases de dados: *Web of Science*, *Scopus* e *DOAJ*. Os resultados combinados dessa amostragem permitiram traçar recortes sobre o gênero do material documental coletado, além de especificidades sobre os principais locais de publicação, idiomas preferidos pelos autores, as associações entre os coautores com maior número de publicações, as palavras-chaves mais indexadas e as classificações de áreas de conhecimento que mais se destacaram. Essa análise também evidenciou a preponderância da Ciência da Informação para a Ciência Aberta, bem como para algumas de suas vertentes. O desenlace do trabalho visa apontar como a instrumentalização das bases de dados pode ser utilizada para pesquisar uma amostragem da produtividade científica, apontando a pormenorização temática e cronológica do conceito e da representação da expressão “*open science*” na literatura formal.

**Palavras-chave:** Ciência aberta. Ciência da Informação. Análise de domínio. Campo científico. Organização do conhecimento.

## ABSTRACT

RIBEIRO, Daniel Strauch. **Evolução do escopo temático da ciência aberta**: um estudo exploratório. 2020. 137 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicação, Universidade Federal do Rio de Janeiro, Programa de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2020.

In view of the conceptual complexity of Open Science, the lack of consensus to the definition of the term “open science” aggravates the semantic dispute of academic production. This research seeks to achieve a theoretical operationalization that characterizes an Open Science theme and terminology in the formal literature, highlighting the repercussions of some of its aspects around its semantic field. Through exploratory research, it created a theoretical concept associated with methods of metric studies in order to validate a systematization of Domain Analysis that could prove the Scientific Field of Open Science. For this, it gathers and confront the results found in research by the term “open science” in three databases: Web of Science, Scopus and DOAJ. The combined results of this sampling allow to trace clippings about the gender of document material collected, in addition to specificities about the main places of publication, preferred languages by the authors, the associations between the co-authors with the largest number of publications, most indexed keywords and the ranking of knowledge areas that most. This analysis also showed the preponderance of Information Science for Open Science, as well as for some of their thematic trends. The outcome of the work aims to point out how the instrumentalization of the databases can be used to research a sample of scientific productivity, pointing out the thematic and chronological details of the concept and representation of the expression “open science” in formal literature.

**Keywords:** Open science. Information Science. Domain analysis. Scientific field. Knowledge organization.

## LISTA DE FIGURAS

Figura 1 –	Metodologia para análise bibliométrica do termo de Ciência Aberta.....	49
Figura 2 –	Coordenadas da pesquisa na <i>Web of Science</i> .....	60
Figura 3 –	Evolução das publicações sobre “ <i>open science</i> ” na <i>Web of Science</i> no período de 1970-2018.....	61
Figura 4 –	Coordenadas da pesquisa na <i>Scopus</i> .....	62
Figura 5 –	Evolução de publicações sobre “ <i>open science</i> ” na <i>Scopus</i> no período de 1979-2018.....	62
Figura 6 –	Coordenadas da pesquisa no DOAJ.....	63
Figura 7 –	Evolução das publicações sobre “ <i>open science</i> ” no DOAJ no período de 2006-2018.....	64
Figura 8 –	Conjunto de documentos sobre Ciência Aberta com a expressão “ <i>open science</i> ” encontrados na <i>Web of Science</i> , <i>Scopus</i> e DOAJ sem duplicidades (1970-2018).....	67
Figura 9 –	Evolução das publicações sobre “ <i>open science</i> ” nas bases de dados <i>Web of Science</i> , <i>Scopus</i> e DOAJ no período de 1970-2018.....	69
Figura 10 –	Publicações sobre ciência aberta nas bases de dados <i>Web of Science</i> (WOS), <i>Scopus</i> e DOAJ no período de 2006-2018.....	69
Figura 11 –	Publicações sobre ciência aberta nas bases de dados <i>Web of Science</i> (WOS), <i>Scopus</i> e DOAJ no período de 2006-2018, especificando o volume de documentos cronologicamente.....	70
Figura 12 –	Países de editoras que publicaram sobre Ciência Aberta na <i>Web of Science</i> , <i>Scopus</i> e DOAJ no período de 1970 a 2018.....	71
Figura 13 –	Publicações sobre “ <i>open science</i> ” na língua inglesa nas bases <i>Web of Science</i> , <i>Scopus</i> e DOAJ de 1970 a 2018.....	74
Figura 14 –	Publicações sobre “ <i>open science</i> ” em idiomas diferentes da língua inglesa nas bases <i>Web of Science</i> , <i>Scopus</i> e DOAJ de 1970 a	75

	2018.....	
Figura 15 –	As relações entre autores de documentos com a expressão “ <i>open science</i> ” na <i>Web of Science</i> (de 1970 a 2018).....	77
Figura 16 –	As relações entre autores de documentos com a expressão “ <i>open science</i> ” na <i>Scopus</i> (de 1979 a 2018).....	78
Figura 17 –	As relações entre autores de documentos com a expressão “ <i>open science</i> ” na <i>Web of Science</i> (WOS), <i>Scopus</i> e DOAJ (de 1970 a 2018).....	81
Figura 18 –	Nacionalidades de coautores de documentos sobre Ciência Aberta no levantamento feito pela <i>Web of Science</i> (1970-2018).....	83
Figura 19 –	Nacionalidades de coautores de documentos sobre Ciência Aberta no levantamento feito pela <i>Scopus</i> (1979-2018).....	84
Figura 20 –	Ligações entre as ocorrências das palavras-chave de autores em documentos com a expressão “ <i>open science</i> ” na <i>Web of Science</i> (1970-2018).....	89
Figura 21 –	Ligações entre as ocorrências das palavras-chave da <i>Web of Science</i> em documentos com a expressão “ <i>open science</i> ”, indexados pela própria base (1970-2018).....	90
Figura 22 –	Ligações entre as ocorrências das palavras-chave de autores em documentos com a expressão “ <i>open science</i> ” na <i>Scopus</i> (1979-2018).....	91
Figura 23 –	Ligações entre as ocorrências das palavras-chave da <i>Scopus</i> em documentos com a expressão “ <i>open science</i> ”, indexados pela própria base (1979-2018).....	92
Figura 24 –	A relação de palavras-chave de autores em documentos com a expressão Ciência Aberta, mais relevantes sobre o tema, pelo viés da <i>Web of Science</i> (1970-2018).....	101
Figura 25 –	A relação de palavras-chave de autores em documentos com a expressão Ciência Aberta, mais relevantes sobre o tema, pelo viés da <i>Scopus</i> (1979-2018).....	102
Figura 26 –	Artigos sobre Ciência Aberta registrados simultaneamente nas bases de dados <i>Web of Science</i> , <i>Scopus</i> e DOAJ no período de 2007-2018.....	125

## LISTA DE QUADROS

Quadro 1 –	Comparação entre os metadados das bases de dados utilizadas.....	65
Quadro 2 –	Tipos de Documentos e Categorias de Publicação de documentos com a expressão “ <i>open science</i> ” sobre Ciência Aberta na <i>Web of Science</i> , <i>Scopus</i> e DOAJ (1970-2018).....	66
Quadro 3 –	Tipos de Documentos com a expressão “ <i>open science</i> ” sobre Ciência Aberta na <i>Web of Science</i> (WOS), <i>Scopus</i> e DOAJ, com dupla contagem (1970-2018).....	66
Quadro 4 –	Total de documentos sobre Ciência Aberta com a expressão “ <i>open science</i> ” encontrados na <i>Web of Science</i> , <i>Scopus</i> e DOAJ com duplicidades (1970-2018).....	68
Quadro 5 –	<i>Clusters</i> de países de coautores nas Figura 18 e 19, separados por cor.....	85
Quadro 6 –	<i>Clusters</i> identificados em cada representação gráfica de palavras-chave nas bases Web of Science e Scopus (Figuras 20, 21, 22 e 23).....	93
Quadro 7 –	<i>Clusters</i> identificados em cada representação gráfica com as palavras-chave de autores nas bases <i>Web of Science</i> e <i>Scopus</i> (Figuras 24 e 25).....	103
Quadro 8 –	Coincidências de temas no campo “Categorias da <i>Web of Science</i> ” entre o recorte de documentos da base de dados sobre “Ciência Aberta” e suas vertentes (1970-2018).....	115
Quadro 9 –	Coincidências de temas no campo “Área de pesquisa” entre o recorte de documentos da base de dados sobre “Ciência Aberta” e suas vertentes (1970-2018).....	116
Quadro 10 –	Aproximação dos temas de “Categoria da <i>Web of Science</i> ” com o tema “Ciência da computação” em “Área de pesquisa” da base de dados.....	117
Quadro 11 –	Correlação dos temas coincidentes em “Categoria da <i>Web of Science</i> ” e “Área de pesquisa” nos grupos temáticos desse campo, no recorte em documentos da base de dados sobre “Ciência Aberta” e suas vertentes (1970-2018).....	118
Quadro 12 –	Resultados da Análise de Domínio de Ciência Aberta da amostragem.....	121-122

## LISTA DE TABELAS

Tabela 1 –	Países de editoras que publicaram documentos com a expressão “ <i>open science</i> ” na <i>Web of Science</i> , <i>Scopus</i> e DOAJ (de 1970 a 2018).....	72-73
Tabela 2 –	Idiomas de editoras que publicaram documentos com a expressão “ <i>open science</i> ” na <i>Web of Science</i> , <i>Scopus</i> e DOAJ (de 1970 a 2018).....	76
Tabela 3 –	Os 20 autores que mais publicaram documentos com a expressão “ <i>open science</i> ” na <i>Web of Science</i> , <i>Scopus</i> e DOAJ (de 1970 a 2018).....	80
Tabela 4 –	Ranking com os cinco países com mais documentos sobre Ciência Aberta com mais coautores, segundo o VOSviewer.....	86
Tabela 5 –	Levantamento de palavras-chave nos documentos sobre Ciência Aberta na <i>Web of Science</i> , <i>Scopus</i> e DOAJ (1970-2018).....	87
Tabela 6 –	Coincidências de palavras-chave na <i>Web of Science</i> e na <i>Scopus</i> no levantamento do VOSviewer.....	94
Tabela 7 –	Palavras-chave mais registradas nos documentos sobre Ciência Aberta da <i>Web of Science</i> e <i>Scopus</i> , com maior proximidade com o tema e indexadas pelas bases de dados (1970-2018).....	96-97
Tabela 8 –	Palavras-chave mais registradas nos documentos sobre Ciência Aberta da <i>Web of Science</i> , <i>Scopus</i> e DOAJ, com maior proximidade com o tema e indexadas pelos próprios autores (1970-2018).....	97-98
Tabela 9 –	Termos coincidentes na análise das 100 palavras-chave mais registradas nos documentos sobre Ciência Aberta nas bases de dados <i>Web of Science</i> , <i>Scopus</i> e DOAJ (1970-2018).....	99
Tabela 10 –	Áreas de pesquisa e Categorias da <i>Web of Science</i> de documentos da base de dados com a expressão “ <i>open science</i> ” (Ciência Aberta) (1970-2018).	108
Tabela 11 –	Áreas de pesquisa e Categorias da <i>Web of Science</i> de documentos da base de dados com a expressão “ <i>open access</i> ” (Acesso Aberto) (1990-2018)...	109
Tabela 12 –	Áreas de pesquisa e Categorias da <i>Web of Science</i> de documentos da base de dados com a expressão “ <i>open source</i> ” (Fontes Abertas) (1973-	110

2018)...

Tabela 13 – Áreas de pesquisa e Categorias da <i>Web of Science</i> de documentos da base de dados com a expressão “ <i>open data</i> ” (Dados Abertos) (1978-2018).....	111
Tabela 14 – Áreas de pesquisa e Categorias da <i>Web of Science</i> de documentos da base de dados com a expressão “ <i>citizen science</i> ” (Ciência Cidadã) (1997-2018).....	112
Tabela 15 – Áreas de pesquisa e Categorias da <i>Web of Science</i> de documentos da base de dados com a expressão “ <i>open education</i> ” (Educação Aberta) (1971-2018).....	113
Tabela 16 – Incidência do campo “Ciência da Informação” em Áreas de Pesquisa e Categorias da <i>Web of Science</i> de documentos da base de dados sobre “Ciência Aberta” e suas vertentes (1970-2018).....	114
Tabela 17 – Países em que foram publicados os artigos sobre ciência aberta registrados simultaneamente nas bases de dados <i>Web of Science</i> , <i>Scopus</i> e DOAJ no período de 2007-2018.....	125
Tabela 18 – Periódicos que mais se destacaram por publicar os artigos sobre ciência aberta registrados simultaneamente nas bases de dados <i>Web of Science</i> , <i>Scopus</i> e DOAJ no período de 2007-2018.....	126

## LISTA DE ABREVIATURAS E SIGLAS

.csv	<i>Comma-Separated Values</i>
.txt	<i>Plain text</i>
.xls	<i>Excel Binary File Format</i>
.xml	<i>Extensible Markup Language</i>
AACR 2	<i>Anglo-American Cataloguing Rules – 2ª Edition</i>
ALA	<i>American Library Association</i>
BOAI	<i>Budapest Open Access Initiative</i>
DCMI	<i>Dublin Core Metadata Initiative</i>
DOAJ	<i>Directory of Open Access Journals</i>
DOI	<i>Digital Object Identifier</i>
ESI	<i>Essential Science Indicators</i>
HEFCE	<i>Higher Education Funding Council for England</i>
ICAR	<i>Indian Council of Agricultural Research</i>
IS4OA	<i>Infrastructure Services for Open Access</i>
ISBN	<i>Information Serial Book Number</i>
JCI	<i>Journal of Clinical Investigation</i>
MARC	<i>Machine Readable Cataloging</i>
MDS	<i>Multidimensional Scaling</i>
NISO	<i>National Information Standard Organization</i>
OKF	<i>Open Knowledge Foundation</i>
SRI	<i>Sistema de Recuperação da Informação</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>18</b>
1.1	CONTEXTO E COLOCAÇÃO DO PROBLEMA.....	18
1.2	QUESTÕES DE PESQUISA.....	21
1.3	OBJETIVOS GERAL E ESPECÍFICO.....	22
1.4	DAS PRÁTICAS À REPERCURSSÃO DA ANÁLISE.....	22
<b>2</b>	<b>OS CONCEITOS DE CIÊNCIA ABERTA, CAMPO E ANÁLISE DE DOMÍNIO.....</b>	<b>24</b>
2.1	UM BREVE HISTÓRICO SOBRE CIÊNCIA ABERTA.....	25
2.2	O CONCEITO DE ANÁLISE DE DOMÍNIO.....	27
2.3	ACEPÇÕES SOBRE A NOÇÃO DE CAMPO.....	33
<b>2.3.1</b>	<b>O Campo Científico.....</b>	<b>34</b>
2.3.1.1	O Capital Científico.....	36
2.3.1.2	A definição de Autonomia e de Heteronomia.....	37
2.3.1.3	Considerações sobre as “Revoluções” Científicas.....	38
<b>3</b>	<b>PROCEDIMENTOS METODOLÓGICOS E INSTRUMENTOS DA PESQUISA EMPÍRICA.....</b>	<b>41</b>
3.1	INFORMAÇÃO SOBRE OS DADOS: DA REPRESENTAÇÃO DESCRITIVA AOS METADADOS.....	42
<b>3.1.1</b>	<b>Representação Descritiva.....</b>	<b>42</b>
<b>3.1.2</b>	<b>Metadados.....</b>	<b>44</b>
3.2	ETAPAS METODOLÓGICAS.....	47

3.3	BASES	DE	DADOS	
	ANALISADAS.....			50
3.3.1	<i>Web of Science</i> .....			51
3.3.2	<i>Scopus</i> .....			53
3.3.3	DOAJ.....			53
3.4	SOFTWARES UTILIZADOS.....			55
3.4.1	VOSviewer.....			55
3.4.2	Gephi.....			57
4	APRESENTAÇÃO	E	ANÁLISE	DOS
	RESULTADOS.....			59
4.1	RECORTE	HISTÓRICO	SOBRE	CIÊNCIA
	ABERTA.....			59
4.1.1	<i>Web of Science</i> .....			60
4.1.2	<i>Scopus</i> .....			62
4.1.3	DOAJ.....			63
4.2	RECORTE	MATERIAL	DOS	DOCUMENTOS
	ABERTA			SOBRE
				CIÊNCIA
				65
4.3	RELAÇÕES	E	ESPECIFICIDADES	DA
	ANÁLISE			SOBRE
	ABERTA.....			CIÊNCIA
				70
4.3.1	Local			de
	publicação.....			71
4.3.2	Idioma.....			74
4.3.3	Autores.....			76
4.3.3.1	Nacionalidade de coautores.....			83
4.3.4	Palavras-chaves.....			86
4.3.4.1	Análise			pele
	VOSviewer.....			88
4.3.4.2	Análise empírica.....			95
4.3.4.3	Problematização	sobre	as	diferentes
	Aberta.....			definições
				de
				Ciência
				104
4.4	ANALOGIAS	ENTRE	A	CIÊNCIA
	VERTENTES.....			ABERTA
				E
				SUAS
				106
4.4.1	Discussão sobre as áreas de conhecimento da Ciência Aberta e suas vertentes.....			114

4.5	O CONCEITO DE CAMPO CIENTÍFICO COM UMA ANÁLISE DE DOMÍNIO NOS RESULTADOS.....	118
4.5.1	Uma visão geral do Domínio na Ciência Aberta.....	119
4.5.2	Os contornos de um Campo Científico na Ciência Aberta.....	122
4.5.3	Uma breve análise sobre a intercessão entre as bases de dados.....	124
5	CONSIDERAÇÕES FINAIS.....	127
	REFERÊNCIAS.....	131
	APÊNDICE A – Documentos sobre Ciência Aberta publicados antes de 1998.....	139
	APÊNDICE B – 100 palavras-chave mais registradas nos documentos sobre Ciência Aberta da <i>Web of Science</i> e <i>Scopus</i> , indexadas pelas bases de dados (1970-2018).....	153
	APÊNDICE C – 100 palavras-chave mais registradas nos documentos sobre Ciência Aberta da <i>Web of Science</i> , <i>Scopus</i> e <i>DOAJ</i> , indexadas pelos próprios autores (1970-2018).....	155

## 1 INTRODUÇÃO

Tradicionalmente, o âmbito da comunidade científica segue ritos e práticas que a caracterizam como um meio hermético, conservador e reservado. No entanto, nos últimos anos, a Ciência Aberta tem exteriorizado um estandarte de transformações, não só no que tange a divulgação de publicações acadêmicas, e do material levantado para produzi-las, mas, principalmente, na transparência e socialização das práticas científicas.

O progresso tecnológico e a intensificação em inovações de informação e comunicação fizeram surgir uma mobilização espontânea pela Ciência Aberta por parte da comunidade científica. Nesse sentido, tal desenvolvimento foi sendo paulatinamente coberto pela literatura formal, com artigos de periódicos realizando praticamente uma metanálise sobre o fazer científico. É neste contexto que se procura compreender a extensão do conceito de Ciência Aberta por meio das próprias publicações sobre a temática.

### 1.1 CONTEXTO E COLOCAÇÃO DO PROBLEMA

A expansão e popularização da internet, aliada à explosão da informação e do interesse pela disponibilidade de periódicos de acesso aberto, no final da década de 1990, foram alguns dos motivos que estimularam a disseminação do conceito de “Ciência Aberta” (DAVID, 2004, p. 10). O movimento da Ciência Aberta pretende tornar a pesquisa científica, incluindo o conteúdo publicado, os dados de pesquisa, as amostras físicas e as fontes abertas de softwares, e sua disseminação acessíveis a todos os cidadãos da sociedade, sejam curiosos, amadores ou profissionais. Tais práticas não se limitam às publicações de pesquisas abertas, abrangendo também campanhas de acesso aberto que vão desde o encorajamento da divulgação de cadernos de pesquisa (transcrições de anotações dos pesquisadores) até a facilitação da difusão e da comunicação de conhecimentos científicos.

No âmago deste movimento, percebe-se, de um lado, uma tensão social e política entre a socialização do conhecimento, e sua privatização de outro (ALBAGLI; APPEL; MACIEL, 2014, p. 8). Tal conflito tem como consequência fundamental a dificuldade de traçar um percurso bem delimitado do conceito de Ciência Aberta, permitindo assim apontar os avanços e retrocessos em sua trajetória histórica, bem como os contornos semânticos de sua definição terminológica.

Uma vez que a Ciência Aberta se define como um termo guarda-chuva, o amplo repertório derivado de suas produções acadêmicas abriga descrições e posicionamentos bastante variados. A problemática quanto à formação do campo como área de conhecimento acaba refletindo também em sua definição terminológica e na exposição de sua análise de domínio – ou seja, a representação do entendimento das áreas do conhecimento sobre o seu papel e o discurso da comunidade científica, bem como no desenvolvimento de novas vertentes do próprio movimento da Ciência Aberta.

Desde os exórdios do uso da expressão “Ciência Aberta” até a sua consolidação, com o consequente surgimento de novos desdobramentos desse movimento, o confronto sociopolítico entre a abertura e o cerceamento intelectual causou distensões e retrocessos semânticos que dificultam um consenso na definição de Campo Científico<sup>1</sup> e a delimitação da Análise de Domínio<sup>2</sup>. Tal tensão possibilitou a emergência e a especialização das atividades da Ciência Aberta, pois, uma vez que tal abertura passou a ser ameaçada, foi necessário também haver um recrudescimento do movimento, ultrapassando o livre acesso às publicações científicas e incluindo outras frentes, tais como: dados científicos abertos, ferramentas científicas abertas, fontes científicas abertas, hardware científico aberto, software científico aberto, cadernos científicos abertos, *wikipesquisa*, ciência cidadã e educação aberta. O conceito, às vezes, foi representado por definições momentaneamente em desacordo, similares e, muitas vezes, provisoriamente, paralelas.

O próprio termo “*Open Science*”, em si, já é problemático. Uma vez que a palavra “Ciência” aglutina uma série de práticas do meio acadêmico, há expressões que se aproximam da relação que ela engloba, como “*Open Scholarship*”, “*Open Research*”, “*Open Standards*”, “*Open Infrastructure*” etc. No entanto, optou-se por examinar o vocábulo “Ciência Aberta” justamente pela prevalência das publicações deste em relação aos demais. Todavia, vale ressaltar que essa hegemonia não é excludente, de forma que os citados termos, bem como alguns outros, apareceram como sinônimos e “quase sinônimos”.

Cabe destacar ainda a dificuldade de compreender a definição terminológica de “Ciência Aberta”, assim como a maneira com que o movimento se transforma. Uma vez que suas práticas buscam promover a eliminação de filtros e barreiras, além de rechaçar

---

<sup>1</sup> Discussão sobre a luta concorrencial de agentes que detêm o argumento de autoridade sobre determinada área do conhecimento.

<sup>2</sup> Paradigma da Organização do Conhecimento em que se discute a forma de visualizar a emergência e coerência de um domínio, bem como maneira de dominar os parâmetros do universo semântico em que este domínio opera.

preconceitos, a produção científica tem alcançado uma composição cada vez mais igualitária. Isto tem resultado em uma realidade em que indivíduos de diferentes culturas, regionalidades, gêneros e níveis sociais têm participado de publicações acadêmicas.

É certo ainda que a Ciência Aberta possibilita um constante aumento do volume de conhecimento público, assegurando a ampliação das taxas gerais de produtividade científica e de inovação, da mesma forma que também proporciona um considerável retorno social dos níveis de investimento em ciência e tecnologia (ALBAGLI, 2015, p. 14).

No entanto, os conceitos e definições sobre a Ciência Aberta não derivam da literatura formal, registrada em base de dados. Muitas das principais definições da área têm origem em iniciativas<sup>3</sup>, manifestos e declarações<sup>4</sup> e livros abertos. Porém, essas fontes e recursos informacionais não serão abordadas neste trabalho justamente por conta de seu escopo investigativo, feito exclusivamente pelo viés da produção publicada pela comunidade acadêmica.

Atualmente, não há um consenso semântico quanto à extensão, ao significado e *modus operandi* do que venha a ser Ciência Aberta (ALBAGLI, 2015, p. 14). Contudo, há definições e desdobramentos de conceitos em publicações científicas que acabam delimitando sua atuação e seu significado, podendo inclusive agravar enfrentamentos e contradições terminológicas.

O objetivo deste trabalho consiste em demonstrar como a disputa semântica do conceito de Ciência Aberta acabou por resultar em uma expansão do movimento para outras questões, sobrecedendo o mote inicial de definição de acesso aberto. Através desse argumento, propõe-se comprovar como o repertório de palavras-chave e vocábulos em títulos e resumos de artigos sobre tais temáticas acabam por realçar o valor de sua rede conceitual,

---

<sup>3</sup> Por exemplo, a *Open Knowledge Foundation* (OKF), uma rede sem fins lucrativos fundada em 2004 que promove e divulga gratuitamente informações, incluindo conteúdo e dados; o *Polymath Project*, empreendimento entre matemáticos de 2009, criado para resolver coletivamente questões importantes e difíceis da área; o *Center for Open Science*, organização de 2013, também sem fins lucrativos, que tem como objetivo aumentar a abertura, a integridade e a reprodutibilidade da pesquisa científica; e a *The Monarch Initiative*, plataforma on-line concebida em 2015, agregando uma coleção de dados e ferramentas que tornam a exploração do conhecimento biomédico mais eficiente e eficaz.

<sup>4</sup> Como a Declaração de Budapeste de 2002, que apresentou, como formas para enfrentar o monopólio das editoras sobre as publicações científicas, a Via Verde (com o lançamento de publicações em repositórios) e a Via Dourada (incentivando a publicação em periódicos de Acesso Aberto); a Declaração de Bethesda de 2003, que trouxe a definição de “publicação de Acesso Aberto”, na qual a autoria que cede o acesso gratuito de suas publicações permite sua ampla divulgação digital ou analógica, mas destacando a importância do depósito em repositórios on-line e por instituições ligadas ao Movimento de Acesso Aberto; ou a Declaração de Berlim, também de 2003, em que o acesso aberto foi destacado como fonte de conhecimento científico e patrimônio cultura, apontando a necessidade de repositórios abertos em bibliotecas, arquivos e museus, importantes produtores de conhecimentos, e imprescindibilidade da transparência da internet e dos softwares abertos.

podendo auxiliar na definição de subdivisões terminológicas e destacar estudos sobre o desenvolvimento conceitual, a formação semântica e a demarcação temática do termo “Ciência Aberta”.

Desta forma, busca-se traçar a relevância da definição de Campo Científico da Ciência Aberta por meio de uma Análise de Domínio estruturada na produção da literatura formal. Essa apreciação também busca justificar a hegemonia do conceito frente a vertentes que se desenvolveram sob sua égide, marcando um liame que justifique a classificação de todos na área do conhecimento da Ciência da Informação.

## 1.2 QUESTÕES DE PESQUISA

O presente trabalho efetua uma investigação sobre a transformação temática e terminológica do conceito de Ciência Aberta, evidenciando os desdobramentos na ampliação da delimitação do campo semântico desse entendimento por meio da análise da produção acadêmica.

Tem como fonte plataformas digitais que indexam as publicações de trabalhos acadêmicos e catalogam periódicos científicos. Assim, com dados inter-relacionados e organizados de forma a permitir a recuperação da informação e de suas referências, permitiu levantar as publicações científicas indexadas com o termo “Ciência Aberta” entre o período de 1970 a 2018. Isso possibilitou traçar o desenvolvimento do campo semântico e os seus desdobramentos.

A partir do cruzamento de informações de metadados extraídos de pesquisas em três bases de dados diferentes, fez-se um exame da evolução das correlações em torno do tema e do conceito de Ciência Aberta. Este levantamento buscou observar a evolução da categorização terminológica e semântica do tema dentro das áreas de conhecimento, bem como destacar o surgimento e consagração de suas vertentes como novas frentes de pesquisa.

Desta forma, procuramos investigar como o termo “Ciência Aberta” se desenvolveu e foi tematicamente categorizado ao longo das últimas décadas. Para isso, procuramos responder as seguintes questões:

- Quais marcos bibliométricos são relevantes para as publicações acadêmicas identificadas?
- Quais desdobramentos semânticos podem ser identificados?
- Como se delimita o seu campo semântico e os desdobramentos de suas vertentes?

- Como se estrutura a coocorrência e as correlações entre esses termos?
- É possível identificar especificidades segundo países, idiomas e áreas de conhecimento?
- Quais são as relações entre os autores que mais publicaram sobre Ciência Aberta ao longo dos períodos pesquisados?
- Há um Campo Científico da Ciência Aberta?

### 1.3 OBJETIVOS GERAL E ESPECÍFICOS

Este trabalho tem por objetivo geral analisar a evolução do escopo temático e terminológico em torno do conceito de Ciência Aberta, evidenciando o desdobramento de suas vertentes na delimitação do campo semântico.

Destarte, para alcançar este objetivo geral, o trabalho tem como objetivos específicos:

- a) situar o debate sobre os conceitos de Campo Científico e Análise de Domínio quanto à sua pertinência teórico-metodológica para os objetivos da pesquisa;
- b) estipular uma estratégia metodológica para trabalhar com a pesquisa em mais de uma base de dados, fazendo recortes históricos, materiais e específicos, de acordo com suas classificações; e
- c) apurar as correspondências encontradas entre a Ciência Aberta e algumas de suas vertentes.

### 1.4 DAS PRÁTICAS À REPERCURSSÃO DA ANÁLISE

Neste trabalho, objetivou-se elaborar um constructo teórico do Campo Científico da Ciência Aberta, por meio dos resultados encontrados ao se fazer uma Análise de Domínio, também sobre Ciência Aberta. Vale ainda ressaltar que esse estudo instrumental foi arquitetado mediante as respostas de pesquisas sobre Ciência Aberta encontradas em buscas feitas em três bases de dados: *Web of Science*, *Scopus* e *Directory of Open Access Journals* (DOAJ).

Foram estipulados limites para a amostragem recolhida, de forma a configurar um campo de análise com interseções cronológicas, qualitativas e quantitativas. Os resultados reunidos, em um primeiro momento, foram analisados segundo os critérios de descrição de

cada plataforma, para então serem observados de maneira integrada, eliminando possíveis redundâncias, falhas no registro e equívocos na classificação.

De forma geral, os resultados trouxeram perspectivas cronológicas quanto ao volume de documentos indexados em cada base de dados, tal como a materialidade do quociente da pesquisa e especificidades da representação descritiva de cada plataforma; a saber: o local de publicação de editoras de periódicos científicas, o principal idioma dos documentos, os coautores que mais publicaram e as palavras-chave mais registradas.

Ainda quanto aos resultados, aproveitou-se para fazer uma análise das áreas de conhecimentos mais classificadas para os documentos de Ciência Aberta, tal qual de matérias identificadas sob o abrigo desse “tema guarda-chuva”: Acesso Aberto, Fonte Aberta, Dados Abertos, Ciência Cidadã e Educação Aberta. Tal investigação procurou reforçar o vínculo dessas vertentes com a Ciência Aberta, mas também demonstrar as aproximações possíveis mediante as comparações das áreas de conhecimento classificadas.

De forma geral, os resultados forneceram substratos para análises bibliométricas da literatura formal nas três bases de dados escolhidas. Tal repercussão permitiu edificar uma Análise de Domínio da Ciência Aberta, instrumentalizando a investigação proposta para aprofundar o diagnóstico do Campo Científico.

Esse estudo exploratório se debruça sobre uma breve acepção sobre o arcabouço teórico da Análise de Domínio e Campo Científico, para então descrever os processos metodológicos, bases de dados e softwares utilizados. Em seguida, apresenta-se as análises dos resultados pesquisados, que se encerram com um exame dos resultados segundo os objetivos propostos. Feita essa sinalização prévia do conteúdo literal desta pesquisa, passa-se então para a sua composição em si.

## 2 OS CONCEITOS DE CIÊNCIA ABERTA, CAMPO E ANÁLISE DE DOMÍNIO

A investigação e categorização do escopo temático da Ciência Aberta foi feita por meio de métodos empíricos, porém os resultados são melhor interpretados através de uma afinação de teorias, conceitos e conhecimentos científicos. Assim, optou-se por particularizar o exame por meio de arcabouços da noção de Campo Científico, sistematizados por uma das teorias de Análise do Domínio.

Ao se discutir a utilização do termo Ciência Aberta, faz-se necessário distinguir os seus antecedentes, bem como a evolução do reconhecimento de seu potencial como movimento científico. Por isso, esse capítulo se inicia com a consolidação do termo “*Open Science*”, indo dos seus primórdios até a sua estabilização como estandarte de resistência aos problemas do cerceamento da propriedade intelectual.

Em seguida, discute-se a Análise de Domínio como um paradigma da Organização do Conhecimento. Tal questão engloba a noção de visualização de um Domínio até a sua maneira de dominar os parâmetros do universo em que esse opera. Segundo Smiraglia (2012, p. 1), a forma como a Análise de Domínio é abordada interfere nos resultados dessas pesquisas. Por exemplo, a Análise de Domínio, quando usada para auxiliar no desenvolvimento de tesouros e linguagens documentárias, tem seus conceitos usados e aplicados metodologicamente na área do conhecimento da pesquisa. Já quando usada para fins epistemológicos e teóricos, seus conceitos são aprofundados, gerando arcabouço teórico para aquele estudo. No caso, a Análise de Domínio será utilizada com fins para aprofundar a discussão teórica do escopo temático da Ciência Aberta, principalmente como um Campo Científico.

O capítulo encerra discutindo Campo Científico. Ao definir um lugar de luta concorrencial de agentes que detêm o argumento de autoridade sobre determinada área do conhecimento, a questão do Campo Científico reproduz elucidaciones sobre muitas das questões sobre o monopólio da autoridade científica e o acúmulo de capital científico. Uma vez que as práticas científicas não podem ser entendidas como práticas desinteressadas, produzindo e supondo formas determinadas de interesse, a noção de Campo Científico é fundamental para analisar a questão do escopo temático da Ciência Aberta.

## 2.1 UM BREVE HISTÓRICO SOBRE CIÊNCIA ABERTA

No final da década de 1990, com o agravamento do boom informacional, a internet passou a ser instrumentalizada como um recurso para aumentar a disponibilidade de informação. Uma vez que o custo da produção de conteúdo digital é fixo e se distancia do custo marginal mínimo da distribuição on-line, os periódicos digitais poderiam também ser acessíveis instantaneamente e em qualquer local do mundo. Tal questionamento se alinhou com as indagações de pesquisadores pelo acesso aberto a publicações científicas, bem como o surgimento do movimento da Ciência Aberta.

A Ciência Aberta não tem uma definição clara quanto à cunhagem de seu termo. Alguns autores defendem um marco com o trabalho de Steven Mann (2015, p. 4) sobre a importância da “Open Science” para as publicações científicas no estudo da interação de dispositivos móveis. No entanto, a Ciência Aberta passou a ser um movimento social de fato não por conta da sua criação ou definição terminológica na literatura formal, mas, sim, pela publicação de Manifestos, pela promoção de Iniciativas e pela publicação de Livros Abertos. Reunidos em torno de discussões sobre o Acesso Aberto, pesquisadores passaram a discutir sobre o cerceamento do conhecimento frente a maior disponibilidade de informação, a questão da cobrança de taxas exorbitantes para a assinatura de periódicos ou para a publicação de artigos e a padronização mercadológica da produção científica

Atendendo tais anseios do momento, alguns periódicos digitais já haviam disponibilizado o acesso livre a suas publicações. É o caso do *The Public-Access Computer Systems Review* (de 1989 até 1996), o *Bryn Mawr Classical Review* (de 1990 até hoje em dia), o *Postmodern Culture* (também de 1990 e os dias atuais) e o *Psycoloquy* (de 1990 a 2002). Todas essas revistas científicas utilizavam infraestruturas pré-existentes da internet, tais como e-mail ou grupos de notícias, e até trabalho voluntário, sem qualquer intenção de gerar lucro (JACOBS, 2006, p. 24).

Ainda que tivessem ocorrido antecedentes do movimento da Ciência Aberta por diversos autores, como Nielsen (2011, p. 172) faz referência, ao discorrer sobre as personalidades de destaque nos primórdios da produção científica, como Galileu Galilei, Johannes Kepler e Isaac Newton, essas personalidades históricas escreviam suas descobertas em documentos codificados para garantir a posse do conhecimento desenvolvido, evitar cópias e reivindicar quaisquer possíveis constatações de autoria. Outrossim, David (2004, p.

10; 2008, p. 28) sugere que o regime de patronato individual da produção científica foi abandonado em detrimento do surgimento das academias de ciências, tais como a *Académie Royale des Sciences* e a *Royal Society*, incentivando a disseminação do conhecimento especializado produzido por seus membros e estimulando a discussão entre pares.

A primeira grande declaração internacional sobre Acesso Aberto foi a Iniciativa de Acesso Aberto de Budapeste, em fevereiro de 2002, lançada pelo *Open Society Institute*. Seguiram-se duas outras declarações: a Declaração de Bethesda sobre Publicação de Acesso Aberto, em junho de 2003; e a Declaração de Berlim sobre Acesso Aberto ao Conhecimento em Ciências e Humanidades, em outubro de 2003. Também em 2003, a Cúpula Mundial sobre a Sociedade da Informação incluiu acesso aberto em sua Declaração de Princípios e Plano de Ação.

Identificado como um movimento guarda-chuva, a Ciência Aberta (*Open Science*) engloba diferentes conceitos, atividades e iniciativas; ou ainda pontos de vista, conjeturas e implicações. Os seus desdobramentos acabam ilustrando diferentes vertentes, tais como “Acesso Aberto” (*Open Access*)<sup>5</sup>, “Educação Aberta” (*Open Education*), “Recursos Educacionais Abertos” (*Open Educational Resources*), “Dados Abertos” (*Open Data*), “Código Aberto” (*Open Source*), “Softwares Livres” (*Open Software*)<sup>6</sup>, “Hardware Aberto” (*Open Hardware*), “Ciência Cidadã” (*Citizen Science*)<sup>7</sup>, entre outros.

Diante de tal fato, a produção e comunicação científicas constituem processos indissociáveis, dado que a comunicação é uma forma produtiva, o que realça a importância das novas formas de registro, de documentação e de abertura de etapas da pesquisa, pois o

---

<sup>5</sup> O *open access* está relacionado com o acesso aberto às publicações científicas, que ganha força em diversos países a partir da Iniciativa de Budapeste (*Budapest Open Access Initiative – BOAI*), em 2002, seguida das declarações de Bethesda e de Berlim, em 2010. Este pode ser considerado um dos primeiros movimentos da ciência aberta. O acesso aberto é uma prática bastante difundida no mundo e cresce anualmente no campo científico, através das políticas de acesso aberto às informações e dos repositórios institucionais. Vários estudos apontam que uma das principais vantagens das políticas de acesso aberto é dar maior visibilidade às pesquisas e pesquisadores (KURAMOTO, 2008, p. 157).

<sup>6</sup> As licenças adotadas para o software livre e de código aberto (*open source*), “como as licenças *copyleft*, e também as licenças *Creative Commons* tem garantido a manutenção do acesso aberto dos produtos e licenças derivados da cópia, adaptação e redistribuição do que foi originalmente licenciado” (ALBAGLI; APPEL; MACIEL, 2013, p. 15).

<sup>7</sup> A ciência cidadã (*citizen science*) é um tipo de colaboração externa ou fora de instituições de pesquisa (DELFANTI, 2013, p. 36). Está relacionada com a participação social nos projetos científicos, como é o caso dos grandes volumes de dados produzidos pelos big data, conjunto de soluções tecnológicas que produzem, captam e processam dados digitais em grande volume, variedade e alta velocidade, que desafiam as análises científicas tradicionais. Hoje, a ciência busca ajuda de não cientistas (muitas vezes voluntários) para processar esses dados (SOARES; SANTOS, 2011, p. 39).

processo de produção do conhecimento torna-se tão ou mais importante do que seu resultado (JORGE; ALBAGLI, 2017, p. 233).

As práticas da ciência aberta ganharam espaço como alternativa à estratégia de proteção de direitos da propriedade intelectual. Esse movimento parte do pressuposto de que a privatização de dados e informações tende a dificultar e atrasar o avanço científico, bem como o progresso econômico e social. [...] Em suma, este é um jogo que envolve diversos atores, como empresas, universidades, órgãos públicos e novas instituições, um jogo que está em plena disputa no campo científico a partir de suas interações com os diversos campos da sociedade.

A evolução da Ciência Aberta se mostrou heterogênea não só por conta dos múltiplos atores e iniciativas, mas também por suas diferentes definições terminológicas<sup>8</sup>. Uma arquitetura conceitual com base na literatura formal poderia estabelecer os contornos sobre o desenvolvimento conceitual da Ciência Aberta e seus desdobramentos. Visto que observar a produção da construção social da ciência pode dirimir discrepâncias conceituais, alargar significados e reunir atores e interesses em análise, determinar o escopo terminológico da Ciência Aberta ajudaria a identificar as suas áreas de conhecimento, bem como o seu domínio e o seu campo.

Pode-se dizer que a Ciência Aberta é uma resistência aos problemas do cerceamento da propriedade intelectual, caracterizando-se como uma apropriação coletiva da informação e dos conhecimentos produzidos. Por isso, apesar das relevantes discrepâncias entre as diversas perspectivas, atores e iniciativas que formam a base social do movimento, é preciso delimitar as definições de cada vertente ou desdobramento da Ciência Aberta em prol da organização do conhecimento e da divulgação das produções e atividades científicas, bem como amplo acesso a esses conteúdos.

## 2.2 O CONCEITO DE ANÁLISE DE DOMÍNIO

Para organizar uma noção com os propósitos temáticos do termo Ciência Aberta, pensou-se em trabalhar com o conceito de Análise de Domínio, pois, além de ser um paradigma da Organização do Conhecimento, também é uma linha teórica da Ciência da Informação – área do conhecimento também vinculada à Ciência Aberta, conforme os resultados apontados nos capítulos a seguir.

---

<sup>8</sup> Tais como “*open research*”, “*scholarship*”, “*standards*” etc. A problematização de tais conceitos será discutida nos resultados desse trabalho, conforme as decorrências do levantamento feito.

A expressão “Análise de Domínio” propriamente dita, foi cunhada por Prieto-Díaz (1990) para a disciplina de engenharia de software. O termo foi relacionado diretamente à Biblioteconomia e à Ciência da Informação, por conta da análise de facetas<sup>9</sup>. No entanto, diferente dessa, o termo descreveria a atividade de identificar os objetos e operações de uma classe de sistemas similares em um domínio com problemas e linguagens específicas (PRIETO-DÍAZ, 1990, p. 45).

No entanto, dentro do âmbito da Ciência da Informação, o conceito de “Análise de Domínio” passou a ser adotado como uma abordagem de caracterização e avaliação da ciência, na medida em que permite identificar as condições pelas quais o conhecimento científico se constrói e se socializa. Uma das principais aplicações teóricas da “Análise de Domínio” foi criada por Hjørland e Albrechtsen (1995, p. 400), sendo aprofundado conceitualmente em trabalhos desse último no decorrer dos anos.

A Análise de Domínio de Hjørland (2002, p. 422) é uma tentativa de aplicar uma teoria abrangente à Ciência da Informação. Sua principal premissa é um conjunto de ferramentas analíticas que podem ser empregadas em subconjuntos ou como um todo, a fim de estudar os efeitos de diferentes campos sociais, epistemológicos e culturais nas teorias da Ciência da Informação. Assim, a organização é aplicada em relação aos diferentes domínios dos quais emana e é consumida. Em vez de estudar o usuário nesse caso, o domínio ou o ambiente é o foco principal da teoria. Os indivíduos são vistos como membros de culturas ou domínios distintos e seus sistemas de documentos subsequentes.

Semelhante à Análise de Domínio, a teoria conceitual visa criar uma teoria de conceitos que possa ser aplicada aos sistemas de organização do conhecimento através da análise de relações semânticas conceituais (HJØRLAND, 2009, p. 1523). Diferentes teorias de conceitos, ou seja, como os usuários interagem com a informação, moldam a interação com a informação em três aspectos específicos: bibliometria, competência informacional e organização do conhecimento. Essa teoria indica que conceitos são definições do mundo

---

<sup>9</sup> O esquema de classificação resultante das técnicas de processamento de informações na Biblioteconomia se tornou um modelo conceitual para o domínio de uma coleção. Segundo Vickery (1960, p. 14) a aplicação de Análise de Domínio na Biblioteconomia permite a classificação de esquemas especializados, de forma que as “classificações facetadas são derivados de um processo manual que consiste em agrupar termos relacionados de uma amostra de títulos selecionados, definindo nomes de facetas desses grupos, ordenando os termos em cada faceta e especificando regras para sintetizar classes compostas”. O agrupamento de termos de títulos é equivalente a encontrar objetos e operações em um domínio de aplicativo. “A nomeação de facetas e a definição de regras de classificação são equivalentes da derivação de um modelo de domínio ou criação de uma linguagem de domínio” (PRIETO-DÍAZ 1990, p.50, tradução nossa).

criadas de acordo com as inclinações e ideias de uma pessoa; eles não podem existir isoladamente dos interesses que os criaram.

Hjørland e Albrechtsen (1995, p. 420) desenvolveram os estudos sobre a análise de domínio como uma nova abordagem para a área da Ciência da Informação. Destacando o conteúdo do conhecimento e o horizonte de estudos em domínios do conhecimento, podendo ser aproveitados para classificar, na ciência, comunidades de pensamento ou de discursos (HJØRLAND; ALBRECHTSEN, 1995, p. 408).

Sete anos depois, Hjørland (2002, p. 450) sugeriu 11 modos pelas quais a Ciência da Informação poderia abordar um determinado domínio de uma forma relativamente específica. São eles:

1. Produção e avaliação de guias de literatura e gateways de assuntos;
2. Produção e avaliação de classificações especiais e tesouros;
3. Pesquisa de competências em indexação e recuperação de informações em especialidades;
4. Conhecimento de estudos empíricos de usuários em áreas temáticas;
5. Produção e interpretação de estudos bibliométricos;
6. Estudos históricos de estruturas e serviços de informação em domínios;
7. Estudos de documentos e gêneros em domínios do conhecimento;
8. Estudos epistemológicos e críticos de diferentes paradigmas, suposições e interesses em domínios;
9. Conhecimento de estudos terminológicos, linguagens para fins especiais e análise de discurso em campos do conhecimento;
10. Estudos de estruturas e instituições em comunicação científica e profissional em um domínio;
11. Conhecimento de métodos e resultados de estudos analíticos de domínio sobre cognição profissional, representação de conhecimento em ciência da computação e inteligência artificial.

Tais pontos enfatizam que os objetos de estudo de pesquisa seriam as entidades sociais e teóricas, estudando as infraestruturas do que se entende por Ciência da Informação (HJØRLAND, 2002, p. 451). Porém, ainda segundo o autor, os 11 pontos não devem necessariamente ser combinados em um mesmo estudo, mas, no mínimo, duas abordagens sobre um domínio poderiam ser combinadas para auxiliar em sua compreensão.

Se considerarmos comunidades discursivas como uma atividade, uma área de interesse ou um grupo de pessoas, desde que haja limites definidos para a atuação, o método pragmático da análise de domínio sugerido por Hjørland (2002, p. 444) poderia se basear na determinação de objetivos e utilidades dos conhecimentos produzidos. Assim, o uso da análise de domínio pode ser vinculado aos aspectos sociocognitivos das relações entre a sociedade e o conhecimento produzidos por ela.

Juntas, essas 11 abordagens formam uma perspectiva que é exclusiva para informações científicas. Deveria também ser óbvio como esta perspectiva oferece investigações práticas e teoricamente relevantes. Pesquisar em ciência da informação combinando várias das abordagens mencionadas acima, no meu ponto de vista, fortalece a identidade da Ciência da Informação e fortalece a relação entre pesquisa e prática em Ciência da Informação. Porque, como implícito em trabalhos anteriores, esta abordagem deve tornar o campo mais coerente com a sua própria história e proporcionar um estoque de conhecimento mais profundo e satisfatório. Também poderia proporcionar um melhor contato interdisciplinar e intercâmbio com áreas como sociologia, linguística e filosofia. Seria, na minha opinião, um objetivo apropriado para a Ciência da Informação desenvolver estas 11 abordagens (e possivelmente outras novas), fortalecendo significativamente a ciência da informação em geral (HJØRLAND, 2002, p. 451, tradução nossa).

A análise de domínio, como método, focalizaria o estudo da estrutura pelo entendimento dos padrões de cooperação e comunicação, as linguagens especializadas, os sistemas de informação etc. Concebe que cada domínio tem uma linguagem própria<sup>10</sup>, pois a natureza da informação a ser compartilhada acaba por definir a característica da linguagem que se quer falar.

Cada área do conhecimento reúne suas próprias formas documentais mais ou menos específicas, como partituras, mapas e bibliotecas especializadas, por exemplo, bibliotecas de gene, de imagens e outras. No campo científico, os indivíduos epistêmicos formam as comunidades discursivas por meio das práticas científicas que se ajustam à produção de conhecimento, adequadas ao arranjo social do domínio.

O homem da ciência se engaja na aquisição de *habitus*<sup>11</sup> científico e, para ser “reconhecido”, incorpora um conjunto de disposições, incluindo a linguagem especializada, os ritos – como a avaliação por pares –, em busca de reconhecimento e autoridade científica, outorgados a ele (homem da ciência) por sua competência

---

<sup>10</sup> No caso, pegando emprestado um dos jargões de Bourdieu (1983, p. 60), seria um *habitus* linguístico, mas visto dentro de um mesmo campo científico, como, por exemplo, a linguagem de divulgação, a comunicação em pesquisadores ou a linguagem científica.

<sup>11</sup> O conceito de *habitus*, cunhado por Bourdieu (ORTIZ, 2003, p. 43), faz um vínculo direto entre a socialização (ou trajetória social) e as ações dos indivíduos, estruturando todas as disposições, padrões de ação ou percepções, adquiridas por sua experiência social. Assim, os agentes desenvolveriam estratégias baseadas em um pequeno número de disposições adquiridas através da socialização que, embora inconscientes, são adaptadas às necessidades do mundo social.

técnica. À medida que se capitaliza cientificamente, pode converter capital científico em capital social, especialmente ao estabelecer contatos com outros atores. As agências de fomento passam a financiá-lo, alunos o elegem para orientações de tese de doutoramento, recebe convites para participação em pesquisas e projetos etc. (CARVALHO, 2017, p. 203).

Portanto, a análise de domínio vê o indivíduo no espaço social e não o indivíduo e o espaço social. A informação é vista como um produto construído socialmente por meio da interação entre o indivíduo e o contexto da informação. Dias (2015, p. 8) argumenta que a análise de domínio é um importante método de identificação dos objetos e dos processos, e de suas relações na construção de um vocábulo de um determinado domínio, podendo ser utilizados principalmente para a organização e recuperação da informação e do conhecimento.

De acordo com Hjørland e Albrechtsen (1995, p. 8), trata-se também de uma nova abordagem na Ciência da Informação, definida como sociocognitivista, que se contrapõe ao paradigma cognitivo com o indivíduo isolado, possuidor de modelos mentais do mundo exterior e com suas transformações mentais geradas pelo processo informacional.

Assim, os ambientes onde ocorrem estas interações, com suas regras, formalizações, padrões e convenções, acabam por formar uma comunidade discursiva, culminando, de acordo com a atividade realizada, na necessidade de abranger um maior e mais complexo grau de informação. Ao considerar o campo da Ciência da Informação, vemos que a análise de domínio dá ênfase também ao tratamento do assunto, alcançando relevância também nas representações e referências sobre o tema estudado, bem como indexações, bibliografias e outras ferramentas de recursos informacionais.

Pode-se avaliar a importância da conceituação existente quando as palavras e frases são utilizadas como ponto de acesso e desprovidas de contexto: tais expressões ficariam completamente soltas, sem significado e sem sentido. Para Hjørland:

[...] a Ciência da Informação deve estudar como as palavras e símbolos dão sentido em seu contexto específico, a forma como as disciplinas constroem seus significados, como a composição de documentos fornece significado às palavras e símbolos e como diferentes vocabulários controlados constroem significados (HJØRLAND, 2003, p. 97, tradução nossa).

Temos aqui um marco na Organização do Conhecimento que, na visão de Hjørland (2003, p. 97), antes apresentava abordagens universais, reducionistas ou individualistas. A abordagem da análise de domínio “está apoiada na visão histórica e pragmática por considerar os aspectos culturais, sociais e históricos tanto do contexto de produção do documento como do usuário deste documento” (HJØRLAND, 2005, p. 131).

Assim, para fazer uma Análise de Domínio da Ciência Aberta por meio das bases de dados, optou-se por utilizar três dos 11 modos sugeridos por Hjørland (2002, p. 450): o terceiro, indexação e recuperação nas especialidades; o quinto, estudos bibliométricos; e o sétimo, estudos documentais e de gênero.

Sobre a indexação e recuperação de especialidades, Hjørland (2002, p. 450) defende que, apesar do enorme volume de indexações diárias em centenas de bases de dados, ainda há poucas pesquisas sobre os métodos utilizados, o que valoriza a indexação por conta do objetivo de recuperar e dar visibilidade dos “potenciais epistemológicos” dos documentos. De maneira que a análise de domínio acaba sendo favorecida pela produção de classificações especializadas, como aquelas utilizadas por duas das bases de dados escolhidas – *Web of Science* e *Scopus*.

Quanto aos estudos bibliométricos, Hjørland (2002, p. 432) ressalta que a bibliometria pode ser usada como uma ferramenta e um método na análise de domínio de diversas maneiras, de forma que “uma forte abordagem mostra conexões detalhadas e reais entre os documentos”. Todavia, o autor adverte sobre a análise dos diferentes vieses que podem ser aplicados a esse tipo de pesquisa, bem como a interpretação desses exames, que podem contribuir em estudos históricos, epistemológicos e críticos. Dessa forma, fez-se estudos bibliométricos com os resultados de pesquisas coletadas em base de dados.

Em derradeiro, os estudos documentais e de gênero se relacionam diretamente aos diferentes tipos de documentos de cada domínio, bem como a relação daqueles mais importantes para defini-los. Podem ser publicações formais ou não, como, por exemplo, os mapas na Geografia, as partituras em Música ou os textos normativos no Direito. Para os estudos de análise de domínio, essa abordagem deve se apoiar em pesquisas sobre indexação e recuperação em áreas especializadas, no caso, a base de dados para a literatura formal da produção científica sobre Ciência Aberta.

Uma vez que os domínios não seriam divisões prontas do mundo, mas representações dinâmicas e com uma trajetória histórica bem desenvolvida, a sua análise requer o cruzamento dos resultados analisados para estabelecer as relações que os estruturam. Uma análise documental pode utilizar registros de palavras e expressões em títulos, resumos e palavras-chave para fundamentar um dos instrumentos de representação que mais se adequa à instrumentalização dessa dinâmica. De modo que, através da investigação dos resultados obtidos em bases de dados, objetiva-se alcançar uma análise de domínio que ofereça os

subsídios necessários para conceituar uma concepção de Campo Científico para o movimento da Ciência Aberta.

### 2.3 ACEPÇÕES SOBRE O CONCEITO DE CAMPO

Segundo Bourdieu (1996, p. 27), o mundo social seria constituído por campos, como espécies de ambientes com relações objetivas e suas próprias regras. A estrutura do campo tanto teria a força de constranger os agentes nele envolvidos, como estimularia o conflito hierárquico entre os agentes nele – de forma a conservar ou transformar suas estruturas (BOURDIEU, 1996, p. 50).

Os campos resultam de processos de diferenciação social, da forma de ser e do conhecimento do mundo. Como tal, cada campo cria o seu próprio objeto (artístico, educacional, político etc.) e o seu princípio de compreensão. São “espaços estruturados de posições” em um determinado momento. Podem ser analisados independentemente das características dos seus ocupantes, isto é, como estrutura objetiva. São microcosmos sociais, com valores (capitais, cabedais), objetos e interesses específicos. [...] Os campos são mundos, no sentido em que falamos no mundo literário, artístico, político, religioso, científico. São microcosmos autônomos no interior do mundo social. Todo campo se caracteriza por agentes dotados de um mesmo *habitus*. O campo estrutura o *habitus* e o *habitus* constitui o campo. (THIRY-CHERQUES, 2006, p. 36).

As interações entre os diferentes campos seriam estruturadas de acordo com os recursos que cada agente – inserido em cada um destes contextos – possui e mobiliza. Tais artifícios são sintetizados pela definição metafórica de “capital”<sup>12</sup>, seja ele econômico, cultural ou simbólico. Assim, o “campo” seria um espaço de posicionamento em que os participantes compartilhariam interesses e buscariam absorver parte do “capital específico” do campo – por exemplo, capital político no campo político, capital intelectual no campo intelectual etc. –, enquanto a forma como o capital é repartido dispõe as relações internas ao campo, isto é, equivaleria a sua estrutura interna (BOURDIEU, 1983, p. 114).

Da mesma forma, as oportunidades de capturar esses capitais específicos dependeriam diretamente do volume de capitais do agente individual, tidos como “básicos”, e de acordo com uma taxa de conversibilidade particular. Por exemplo, um sujeito com capital cultural

---

<sup>12</sup> Segundo Ortiz (2003, p. 15), Bourdieu tem vocabulário e sintaxe complexos, utilizando muitos jargões e neologismos importantes. No caso, por se considerar herdeiro da sociologia clássica, o autor sintetizou uma abordagem pessoal para as principais contribuições dessa área de conhecimento. Assim, dando crédito a Karl Marx, Bourdieu adotou o conceito de “capital”, porém, generalizado para todas as atividades sociais (e não apenas econômicas), bem como a teoria das classes sociais. Da mesma forma, perfilhou o estilo determinista de Emile Durkheim para o princípio da causalidade, enquanto manteve a importância da dimensão simbólica da legitimidade de toda dominação na vida social de Max Weber – promovendo uma ideia de ordens sociais no conceito dos “campos”.

poderá encontrar uma alta taxa de conversibilidade deste para capital intelectual, porém um valor menor em capital econômico.

É importante ressaltar que cada campo teria regras operacionais específicas, contudo, teriam similaridades estruturais, como a luta entre os agentes dominantes do campo e os novos participantes. Aqueles teriam a intenção de se perpetuar no *status quo*, enquanto esses últimos teriam interesse em subverter as regras particulares do campo sobre a sua distribuição de capitais para aumentar o seu cabedal.

De forma geral, Bourdieu (2013, p. 415) questiona a formação das hierarquias entre grupos sociais e a importância de práticas culturais na luta entre diferentes grupos. Para isso, ele organiza seus conceitos em torno de um mundo social dividido em campos, que constituem locais de competição estruturados em torno de questões específicas. Os agentes inseridos no mundo social teriam como princípio de ação a centralidade do *habitus*, enquanto a violência simbólica<sup>13</sup> inviabilizaria aqueles sob o jugo de dominantes.

### 2.3.1 O Campo Científico

Nesse contexto do mundo social constituído por campos, Bourdieu (1996, p. 87) também definiu o conceito de “campo científico”, no qual, sendo também um espaço de disputas por poder<sup>14</sup>, os agentes concorreriam pelo monopólio da competência e da autoridade científica, mantendo também um certo grau de autonomia com relação aos outros tipos de campos.

O campo científico seria estruturado pela distribuição do capital de competência e autoridade científicas pelas relações entre os agentes em conflito, ou seja, pelas relações de conflito entre os seus diversos agentes. Assim, a cada momento, a égide do campo científico também seria um revérbero da posição dos agentes dominantes, que irão impor a sua

---

<sup>13</sup> O conceito de violência simbólica designa uma agressão internalizada, cuja eficácia está ligada à sua ignorância da soberania de agentes dominantes em um determinado campo e, de maneira mais geral, em sua posição social. Essa violência é subconsciente e não se baseia em uma dominação intersubjetiva (de um indivíduo em outro), mas em uma dominação estrutural (de uma posição segundo outra). Essa estrutura, que é uma função do capital de propriedade dos agentes, faz violência porque não é percebida pelos agentes. É, portanto, fonte de um sentimento de inferioridade ou insignificância que só é sofrido por não ser objetivado.

<sup>14</sup> Esta luta segue as regras do campo. Bourdieu diz que cada campo desenvolve *doxa* e *nomos*. Enquanto o *nomos* pode ser entendido como o conjunto de leis gerais e invariantes de funcionamento do campo, a *doxa* é o senso comum, é aquilo sobre o que todos os agentes estão de acordo, e abrange tudo aquilo que é admitido como “sendo assim mesmo”, como, por exemplo, os sistemas de classificação (BOURDIEU, 1996, p. 119).

perspectiva de ciência de acordo com o capital científico que têm, ao *habitus* que possuem e a violência simbólica que fazem.

As lutas definem a organização do campo, os interesses particulares em cada pesquisa e o histórico de enfrentamentos nas instituições científicas que apoiam os agentes. No entanto, para Bourdieu (1976, p. 88), há um estímulo a essa situação de conflito não só pela impossibilidade de fazer uma distribuição equânime de capital científico, tanto para o prestígio, quanto para a competência ou a autoridade científica, como também faz-se necessário manter uma hierarquia social dos campos científicos – ou seja, das disciplinas/áreas de conhecimento.

É possível notar essa situação hierárquica nos campos científicos através do grau de autonomia que possuem. Quanto mais relativo for, menor o *status* e poder econômico e político de seus agentes quanto a independência e a condução das pesquisas. Essa situação não só acaba orientando as práticas, as escolhas, as estratégias e os investimentos dos agentes e das instituições, como também estimula um conflito entre diferentes tipos de campo no mundo social:

A classe dominante concede às ciências da natureza uma autonomia medida pelo grau de interesse na aplicação das técnicas científicas à economia [*i.e.*, à produção], mas nada tem a esperar das ciências sociais a não ser, no melhor dos casos, uma contribuição particularmente preciosa na legitimação da ordem estabelecida e um reforço do arsenal dos instrumentos simbólicos de dominação. O desenvolvimento tardio e sempre ameaçado das ciências sociais testemunha que o progresso em direção à autonomia real – que condiciona e supõe a instauração de mecanismos constitutivos de um campo científico autorregulado e autárquico – se choca sempre com obstáculos desconhecidos alhures [em outros campos de conhecimento, especialmente nas ciências naturais]. E não pode ser de outro modo, porque o que está em jogo na luta interna pela autoridade científica (o poder de produzir, impor e inculcar a representação legítima do mundo social) é o que está em jogo entre as classes no campo da política. Daí que as posições na luta interna não podem jamais atingir o grau de independência com relação às posições nas lutas externas que se observa no campo das ciências da natureza. A ideia de uma ciência neutra é uma ficção interessada que permite aparentar como científica uma forma neutralizada e eufêmica (simbolicamente muito eficaz porque particularmente irreconhecível) da representação dominante do mundo social (ORTIZ, 2003, p. 136-137).

Dentro desta perspectiva, o campo científico, formado por um universo que compreende alunos, bolsistas, professores, instituições de ensino superior e de pesquisa e universidades até os financiadores, regulações políticas sobre ciências e tecnologia e governos, também se insere dentro de um contexto social que não descarta os seus fundamentos históricos, suas áreas de influência e seus modos de atuação.

Posto isso, seguindo o liame das regras e acordos entre os agentes (*doxa* e *nomos*), toda prática de um campo científico não pode ser considerada neutra ou desinteressada. Toda

atividade exercida, *habitus* praticado e proveito auferido tem o objetivo principal de consubstanciar a aquisição de autoridade científica, sempre definida como “capacidade técnica e poder social” (BOURDIEU, 1983, p. 122). Deste modo, o objetivo final de toda pesquisa se traduz na consideração do cientista sobre o que considera importante para aumentar suas chances de reconhecimentos pelos outros (tanto dentro do campo científico, como pelo mundo social como um todo).

### 2.3.1.1 O Capital Científico

O prestígio, a autoridade, o crédito etc. em um campo científico são o resultado da violência simbólica dos dominantes e são sempre definidos pelas consequências da disputa interna no conflito de interesses de seus agentes. O processo de busca por reconhecimento necessariamente é efetuado por um cientista querendo obter capital científico, reconhecimento e soberania – com relação ao poder dominante.

Portanto, o capital científico pode ser entendido como:

[...] uma espécie particular do capital simbólico (o qual, sabe-se, é sempre fundado sobre atos de conhecimento e reconhecimento) que consiste no reconhecimento (ou no crédito) atribuído pelo conjunto de pares-concorrentes no interior do campo científico (BOURDIEU, 2004, p. 26).

Uma vez que tal conceito depende diretamente da quantidade de trabalho exercido, da qualidade da pesquisa científica e da hierarquia no campo científico, o seu valor é equivalente ao reconhecimento de competência e ao grau de autoridade no campo científico. Por isso mesmo, o capital científico pode ser acumulado, transmitido e convertido.

Há que se falar na importância do *habitus* para o capital científico, pois o conceito não só define as práticas e atividades dos agentes, como influencia diretamente na capacidade de escolher, antecipar ou prospectar pesquisas. De maneira que é através do *habitus* que os indivíduos perpetuam os seus comportamentos, mas também os bloqueios e dificuldades do campo como um todo, definindo assim as “maneiras de ser permanente, duráveis, que podem, em particular, levá-los a resistir, a opor-se às forças do campo” (BOURDIEU, 2004, p. 28)

O sociólogo francês definiu dois tipos de capital científico:

- I. Capital Científico “Institucional”: com características de poder político, institucional e temporal, está vinculado à ocupação de colocação ou cargo importante em instituições científicas, laboratórios de pesquisa, departamentos acadêmicos etc. Identificável através de políticas administrativas, tem a transmissão

efetuada por ditames do capital burocrático – um tipo de capital simbólico objetivado, normatizado, emitido e garantido pelo Estado – podendo ser diretamente, ou por meio de concursos, organizados pelos agentes que detêm o capital científico institucional (BOURDIEU, 2004, p. 35).

- II. Capital Científico “Puro”: relativo ao prestígio pessoal, reconhecimento entre os pares e a competência científica, é ganho por conta de contribuições para o progresso da ciência. Com a transmissão vinculada ao nível íntimo do agente, pode ser transferido por meio da estrutura hierárquica do campo em atividades de parcerias, publicações em colaboração e recomendações acadêmicas. Os detentores do capital científico puro podem ser considerados dominantes em determinado campo, pois possuem o reconhecimento dentre os parceiros de sua autoridade (BOURDIEU, 2004, p. 36).

A luta endógena entre os agentes do campo científico é definida pelos conflitos de poder entre os detentores de capital científico puro e institucionalizado. Segundo Bourdieu (2004), a ciência precisa de autonomia para progredir, evitando intervenções políticas que possam diminuir análise de coerência lógica e verificação experimental. Com o aumento da demanda informacional, a sociedade tem valorizado cada vez mais a informação com embasamento e transparência. Por esse motivo, os campos que produzirem resultados e informações com mais fundamentos acabam por reforçar o seu capital – e, no caso do campo científico, principalmente o capital científico do tipo puro.

#### 2.3.1.2 A definição de Autonomia e de Heteronomia

Bourdieu (2004, p. 55) faz uma distinção entre a autonomia e a heteronomia dos campos científicos. Essas características dependem exclusivamente da capacidade de suportar pressões externas ao campo, sendo esse mais autônomo por ter maior independência ou mais heterônomo se sofrer influências de outros campos da ciência ou de outros níveis (político, cultural, cultural etc.), mas também por conta de conflitos de seus próprios agentes.

As questões internas de tensões e lutas do campo científico também são alteradas por conta desta diferenciação. Em campos mais autônomos, “seus agentes galgam posições de poder por conta de atividades embasadas em trabalhos com produções teóricas ou na produtividade científica. Já naqueles heterônomos, a concorrência é imperfeita e é mais lícito

para os agentes fazer intervir forças não científicas nas lutas científicas” (BOURDIEU, 2004, p. 32).

De forma mais apurada, atualmente, há um aporte do capital científico “puro” na contribuição da autonomia dos campos científicos:

Analisando da perspectiva colocada por Bourdieu, considera-se que, atualmente, com a busca contínua pela produtividade na ciência, o poder científico institucional tem adquirido cada vez mais destaque à medida que este realiza a articulação entre financiamento de pesquisas e sua viabilização pelos pesquisadores. Cada vez mais as buscas por recursos financeiros para compra de material, equipamentos, contratação de bolsistas etc. são definidoras das práticas científicas que resultarão no desenvolvimento do capital científico “puro”.

Neste aspecto, a autonomia do campo está ligada à sua capacidade de dar respostas científicas a problemas encontrados na vida social, relacionados com os interesses de pesquisa do campo. A necessidade da constante articulação política e econômica entre os pesquisadores, instituições de pesquisa e fontes de financiamento e de respostas aos investimentos em pesquisa levam a uma corrida constante por publicações “certificadas” e, em algumas áreas, patentes. (JORGE; ALBAGLI, 2017, p. 214).

Assim, de certa forma, a produtividade científica impacta diretamente na questão da autonomia do campo, demonstrando uma maior independência não só por conta da baixa influência externa ou a justeza no esforço de seus agentes, mas principalmente pelo grau de influência que o campo científico pode ter em campos de outro tipo (político, simbólico etc.).

### 2.3.1.3 Considerações sobre as “Revoluções” Científicas

Sobre as questões de mudanças de *doxa* e *nomos* nos grupos e reivindicações internas por seus agentes, Bourdieu raciocinou que o:

[...] fato de que o campo científico comporte sempre uma parte de arbitrário social na medida em que ele se serve dos interesses daqueles que, no campo e/ou fora dele, são capazes de receber os proveitos, não exclui que, sob certas condições, a própria lógica do campo (em particular, a luta entre dominantes e recém-chegados e a censura mútua que daí resulta) exerça um desvio sistemático dos fins que transforma continuamente a busca dos interesses científicos privados (no duplo sentido da palavra) em algo de proveitoso para o progresso da ciência (BOURDIEU, 1983, p. 141).

Segundo o sociólogo francês, os campos podem sofrer “revoluções” internas, ou mudanças estruturais, provocando desvios sistemáticos permanentes que transformam não só a qualidade dos conflitos entre os agentes, como também a quantidade e as peculiaridades do conhecimento científico produzido pelo campo.

Essas transformações não são instantâneas. Passam por todo um processo de metamorfose da ordenação do campo científico, alterando de tempos em tempos alguns agentes que dominam o ponto de vista de seus concorrentes em questões problemáticas institucionais ou na construção do conhecimento. Assim, a mudança não se restringe apenas às consequências da preferência que os agentes detêm sobre as perspectivas do campo sobre questões científicas, mas, sim, ao reconhecimento de detratores com relação a soluções inovadoras dos indivíduos “revolucionários”.

O convencimento, aliado a críticas racionais, pode modificar a construção coletiva de produção de conhecimento, alterando a conjuntura contemporânea do campo com relação aos agentes dominantes. Dessa forma, “o antagonismo, que é o princípio da estrutura e da transformação de todo campo social, tende a tornar-se cada vez mais radical e fecundo porque o acordo forçado em que se engendra a razão deixa cada vez menos lugar ao impensado e à *doxa*” (BOURDIEU, 1983, p. 144).

Para Jorge e Albagli:

A ideia de autonomia do campo fica mais forte considerando os aspectos concorrenciais de cunhos técnicos e racionais. Na busca pelo capital científico, a ordem coletiva de produção do conhecimento científico pelos agentes do campo torna-o cada vez mais autônomo, pois o embasamento teórico e técnico desenvolvido nas disputas de poder deixam pouco espaço para disputas políticas e temporais. [...] Quando Kuhn propõe que a noção de paradigma seria o equivalente a uma linguagem ou cultura por determinar quais questões podem ser pensadas ou impensáveis, ou ainda, aquelas que podem ser formuladas ou excluídas, o autor coloca os problemas científicos como algo a ser pensado apenas por uma comunidade científica, afastando o restante da sociedade dessas questões.

Essa reinterpretação traz elementos interessantes para auxiliar no debate atual sobre ciência aberta, que indica mudanças (não exatamente uma “revolução”) na maneira como funcionam algumas práticas científicas. Quando Bourdieu [...] argumenta que a “tensão essencial da ciência não faz com que haja uma tensão entre a revolução e a tradição, entre os conservadores e os revolucionários, mas com que a revolução implique a tradição, que as revoluções se enraízem no paradigma” [...], ele acredita que os “revolucionários” são grupos que possuem grande domínio da “tradição” e que estão em processo de crítica a uma ordem que coloca em risco a autonomia do campo e seus próprios processos de produção de conhecimento (JORGE; ALBAGLI, 2017, p. 216).

Por isso, essa modificação da hierarquia também pode acabar destacando vertentes de pesquisa na produção científica do campo. Uma vez que as especificidades de uma disciplina/área de conhecimento acabam por destacar a pressão no desenvolvimento da produtividade do campo científico, as linhas de pesquisa mais rentáveis, por sua vez, encabeçam uma corrida pela prioridade da descoberta e de inovações.

Essa especialização do confronto interno no campo também é fruto das visões antagonistas de autonomia dos pesquisadores puros e dos pesquisadores aplicados. Trata-se de

uma proposta de reflexão coletiva, capaz de mobilizar todas as forças vivas de uma instituição de pesquisa e todos os seus recursos para uma espécie de conversão coletiva, que é a condição de uma verdadeira atualização do fazer na ciência (BOURDIEU, 2004, p. 65).

Dessa forma, antes de uma verdadeira “revolução” no campo da ciência, há um momento em que a produtividade dos agentes pode ramificar em diferentes desdobramentos, aprofundando temas, aperfeiçoando parcerias com disciplinas/áreas de conhecimento de outros campos e aperfeiçoando pesquisas em áreas novas. E é justamente essa noção de desdobramentos e vertentes de um campo científico que será analisada neste trabalho, após a avaliação dos resultados obtidos com a análise de documentos sobre Ciência Aberta, coletados em bases de dados.

### **3 PROCEDIMENTOS METODOLÓGICOS E INSTRUMENTOS DA PESQUISA EMPÍRICA**

A pesquisa desenvolve-se através de uma abordagem exploratória, de natureza qualitativa e quantitativa. Segundo Gil (2008, p. 27), as pesquisas do tipo exploratório têm o escopo voltado para “proporcionar visão geral, de tipo aproximativo, acerca de determinado fato. Este tipo de pesquisa é realizado especialmente quando o tema escolhido é pouco explorado e torna-se difícil sobre ele formular hipóteses precisas e operacionalizáveis”.

A análise de dados de natureza quantitativa permite “referenciar as unidades lexicais nos textos e enumerar automaticamente suas ocorrências” (GIL, 2002, p. 90), de tal forma que se possam descrever com precisão fenômenos analisados. Da mesma forma, a análise de dados de natureza qualitativa seria feita por meio de uma sequência de atividades para reduzir e categorizar a coleta de dados, dependendo de fatores variados, tais como “a extensão da amostra, os instrumentos de pesquisa e os pressupostos teóricos que nortearam a investigação” (GIL, 2002, p. 133).

Feita esta breve acepção da metodologia, passa-se para uma breve elucidação sobre os usos de dados e metadados, perpassando a importância da representação descritiva para a categorização da informação. Em seguida, são explicitados os procedimentos metodológicos para a sistematização na coleta de resultados em pesquisas nas bases de dados, esclarecendo-se a sequência de estágios na estratégia de pesquisa, não só para organizar a coleta de dados, mas, também, para categorizá-los.

Em derradeiro, são esclarecidos os motivos para a escolha de cada base de dados, apontando os seus prêmios e desvantagens, e, em seguida, apresenta-se os programas utilizados para trabalhar os metadados apurados. Vale ressaltar que tal pormenorização tem o objetivo de justificar as inferências que serão trabalhadas no capítulo seguinte deste trabalho, cruzando não só as informações coletadas em cada plataforma, mas também reunindo alguns dos pontos necessários para a Análise de Domínio.

### 3.1 INFORMAÇÃO SOBRE OS DADOS: DA REPRESENTAÇÃO DESCRITIVA AOS METADADOS

O significado mais básico do termo “informação” perpassa diversas áreas do conhecimento, conceituando-se, hierarquicamente, das teorias mais elementares até às mais sofisticadas. No entanto, o conceito de “informação” é bem mais abrangente e tem diferentes significados dependendo do contexto, podendo relacionar-se com educação (FREIRE, 2004, p. 6), representação social (ROCHA, 2011, p. 9), estímulos mentais (RUCKER; STERNTHAL, 2013, p. 225), entropia (DANTAS, 2006, p. 45), dentre outros.

De forma mais genérica, a informação é associada aos dados, pois, uma vez que esses representam valores atribuídos à caracterização de padrões, medidas ou parâmetros, a informação seria depreendida por um conjunto de dados inserido em um contexto, mas também com a incorporação de um ou mais significados. Da mesma forma, a informação se relaciona com o conhecimento, visto que, fornecendo todo repertório para a compreensão de um conceito abstrato ou concreto, a informação, com um propósito ou uma utilidade, pode ser considerada um conhecimento.

Segundo Le Coadic (1996, p. 22), na Ciência da Informação, o valor da informação varia conforme o indivíduo, as necessidades e o contexto em que é produzida e compartilhada. Assim, uma informação pode ser extremamente relevante para uns indivíduos, porém, a mesma pode não ter significado algum para outros sujeitos. Dessa forma, faz-se necessário explicitar como a Representação Descritiva se coloca como uma das principais maneiras de significar, de forma abstrata, a informação.

#### 3.1.1 Representação Descritiva

Ao longo da história dos conhecimentos produzidos pelas várias sociedades humanas, foi necessário criar recursos informacionais para organizar, controlar e facilitar o acesso a informações. Daí, surgiu a Representação Descritiva, como uma forma de interpretação abstrata da informação, de maneira a registrar conhecimentos em suportes específicos, para então armazená-los e, posteriormente, recuperá-los para alguma utilização. Primordialmente, o primeiro indivíduo a utilizar esse recurso foi Calímaco, na Biblioteca de Alexandria.

Segundo Barbier:

[...] Calímaco (c.310-243), filósofo e poeta que veio de Cirene (atual Líbia) para Atenas, depois para Alexandria, onde ministra aulas no Museu e trabalha na biblioteca [...]. Teria, a partir dos anos 270, estabelecido as listas gerais da coleção e inaugurado assim a construção de metadados servindo para facilitar a utilização da biblioteca e para enriquecer seus conteúdos: os 120 *volumina* de *Pinakes* constituem as “tabelas dos autores ilustres em todos os campos do conhecimento e de suas obras”.

No interior de cada seção, os autores eram classificados por ordem alfabética, e cada nome estava acompanhado por uma nota bibliográfica curta completada por uma avaliação crítica dos escritos do autor em questão (BARBIER, 2018 p. 47).

Assim, já na Antiguidade, o catálogo elaborado na biblioteca de Alexandria foi considerado um marco na história do controle bibliográfico. Esse consolidava todo o arcabouço para organização e divulgação de um catálogo especializado na literatura grega e nas fundações basilares da primeira classificação de áreas do conhecimento.

A Representação Descritiva tem a função de interpretar a forma da informação, ou seja, descrever com um conjunto de diferentes tipos de dados as características que formam uma representação física, ou bibliográfica, que descreva certo item. As categorias dos dados podem variar dependendo do objeto que será caracterizado, podendo variar da autoria, do título e da data, até a edição, a editora, a paginação ou o idioma.

Por isso, a representação descritiva segue padronizações de um ou mais códigos de catalogação, que seriam conjuntos de regras que orientam a elaboração da descrição bibliográfica de um documento. Segundo Guinchat e Menou (1994, p. 255), o processo de representação da informação consiste na aplicação de técnicas de maneira que o conteúdo do documento possa ser recuperado pelo usuário. Para que isso aconteça, é preciso criar um Sistema de Recuperação da Informação (SRI), considerando elementos tais como as necessidades do usuário, contexto temático, recursos humanos, materiais disponíveis, produtos e serviços que serão oferecidos e a relação custo-eficácia.

O processo de catalogação se ocupa dos registros bibliográficos de itens informacionais que são representados por uma padronização codificada. Para Mey (1987, p. 23), a catalogação define um processo de decisão multidimensional, que estrutura e padroniza os diferentes aspectos de um item informacional, sendo considerado único e suscetível a recuperação e ao uso.

Segundo Lancaster (2004, p. 19), a Representação Descritiva é um processo no qual se identificam autores, títulos, fontes, entre outros elementos bibliográficos. Dessa forma, o processo de representação é efetivado por meio de código, que também são aproveitados como fontes de consulta para expedientes de representação de várias espécies de objetos de informação.

Por meio de bases heterogêneas e cooperativas, o conhecimento tem se dinamizado, exigindo atualizações e impulsionando a promoção de novos instrumentos com esse fim (MEY, 1987, p. 49). Assim, na conjuntura contemporânea, as inovações tecnológicas de comunicação e informação trouxeram um maior fluxo informacional, não só pelo volume de produção de informação, mas também pelos vários tipos de suporte no atual contexto digital.

A atividade de representar conceitos e significados de uma área do conhecimento demanda esforços para que a sua apresentação seja feita de forma eficiente. Existem metodologias capazes de extrair e representar conceitos e significados criados pela mente humana e registrá-los em suportes por meio de linguagens e símbolos. Os “modelos de representação” se tornam indispensáveis e oportunos, uma vez que objetivam exercer tal função.

Um dos desafios atuais é a construção de modelos de representação do próprio conhecimento. Na realidade, é um desafio muito presente, pelo crescimento exponencial das informações e por consequência dos conhecimentos gerados, embora a existência [...] essas considerações levam a um patamar de necessidade de modelar o próprio conhecimento do conhecimento, a criar novas formas de modelos, tais como modelos epistemológicos, modelos ontológicos e metamodelos (MICHAUD, 2006, p. 234-235).

Assim, o cenário contemporâneo da Representação Descritiva, especialmente em relação à representação e modelagem da informação, conecta-se a temas atuais como metadados, usabilidade de recursos digitais, modelagem de dados etc. Por isso, principalmente em uma pesquisa com base de dados, o cruzamento de informações, a designação de coocorrência, a análise de reincidências e organização dos resultados se fazem cada vez mais necessários. É nesse contexto que se demanda uma conceituação de metadados que possibilite não só trabalhar a multiplicidade de representações descritivas de resultados extraídos de bases dados, mas também inferir com as informações reunidas.

### **3.1.2 Metadados**

Metadados são “dados que fornecem informações sobre outros dados”, mas, literalmente, o prefixo “meta”, preposição grega que significa “depois” ou “além”, assume o sentido epistemológico de “sobre”. Portanto, a semântica mais clássica define os metadados como “dados sobre dados”, ou seja, dados que fornecem informações sobre um ou mais aspectos dos próprios dados (SAYÃO, 2010, p. 4).

Estabelecida essa definição terminológica, os metadados são usados, principalmente, para resumir informações básicas sobre dados que facilitam o rastreamento e o trabalho com dados específicos, tais como: a fonte de criação, a finalidade dos dados, a hora e a data de elaboração, o autor dos dados, a localização na rede de computador que foram criados etc.

O termo “metadados” foi cunhado por Bagley (1968, p. 26), em seu livro “Extensão de Linguagens de Programação”, em uma definição no sentido “tradicional” da ISO 11179, que define “metadados estruturais”, isto é, “dados” sobre os “conteúdos de dados”; em vez do sentido alternativo “conteúdo sobre instâncias individuais de conteúdo de dados” ou *metacontent*, o tipo de dados normalmente encontrados em catálogos de bibliotecas (BAGLEY, 1968, p. 34). Desde então, os campos de gerenciamento de informações, ciência da informação, tecnologia da informação, biblioteconomia e sistemas de geoprocessamento adotaram amplamente o termo.

Reavivado a partir dos estudos de catalogação em bibliotecas e unidades de informação (DAY, 2005, n. p.), o termo “metadados”, em geral, serve como instrumento para, no contexto de sistemas de dados, descrever, organizar e coordenar a gestão e a utilização de dados. Inclusive, na biblioteconomia, os modelos mais tradicionais de catalogação são o MARC 21<sup>15</sup> e o AACR 2<sup>16</sup>.

Relacionado à função de catalogação, o conceito de “metadados” pode servir como suporte para armazenar e gerenciar um banco de dados, geralmente chamado de registro de metadados ou repositório de metadados. No entanto, sem contexto e um ponto de referência, pode ser impossível identificar metadados apenas olhando para eles.

Por exemplo, o *Information Serial Book Number* (ISBN ou “Número Padrão Internacional de Livro”) é um conjunto com 13 dígitos que pode interpretar a classificação do grupo, do editor e do título de um livro. Porém, sem o suporte do banco de dados que interpreta cada código dentro da numeração, essa especificação se torna irrelevante.

A *Dublin Core Metadata Initiative* (DCMI ou Iniciativa Dublin Core Metadados) promove descrição de objetos digitais na web e facilita a recuperação de informações em

---

<sup>15</sup> Criado pela cientista da computação Henriett Avram, enquanto trabalhava na Biblioteca do Congresso, o padrão *Machine Readable Cataloging* (MARC ou Catalogação Legível Por Máquina) é um conjunto de dados digitais que pode descrever itens catalogados por bibliotecas, como livros, periódicos, mapas etc.

<sup>16</sup> Sigla de “*Anglo-American Cataloguing Rules – 2ª Edition*”, ou Código de Catalogação Anglo-Americano – 2ª Edição. Publicada em 1978, pela *American Library Association* (ALA ou Associação de Bibliotecas Americanas), é uma síntese de regras para a criação de descrições bibliográficas e para a escolha, a construção e a atribuição dos pontos de acesso que representem, de maneira uniforme, pessoas, localizações geográficas e entidades coletivas, além de títulos, obras e expressões.

sistemas de buscas mais dinâmicos e adaptáveis. O serviço faz a caracterização de documentos, vídeos, sons, imagens e sites, além de buscar a interoperabilidade entre cada item, para promover padrões e vocabulários especializados que, em um ambiente extenso e fragmentado como a web, acabam por distinguir melhor a correlação de fontes e recursos para recuperar informações.

A *National Information Standard Organization* (NISO ou Organização Nacional de Padronização da Informação) apresenta uma definição de metadados como “a informação estruturada que descreve, explica, localiza, ou possibilita que um recurso informacional seja fácil de recuperar, usar ou gerenciar”.

No entanto, segundo Sayão :

Não se pode afirmar que haja um consenso, mas uma fração significativa dos autores que tratam do assunto concorda que os metadados podem ser divididos em três categorias conceituais: metadados descritivos, metadados estruturais e metadados administrativos. Essa segmentação é útil para uma compreensão mais clara sobre os tipos de informações que eles podem circunscrever, muito embora os seus contornos não possam ser precisamente definidos.

Metadados descritivos: é a face mais conhecida dos metadados, são eles que descrevem um recurso com o propósito de descoberta e identificação; podem incluir elementos tais como título, autor, resumo, palavras-chave e identificador persistente.

Metadados estruturais: são informações que documentam como os recursos complexos, compostos por vários elementos, devem ser recompostos e ordenados. Por exemplo, como as páginas de um livro, digitalizadas separadamente, são vinculadas entre si e ordenadas para formar um capítulo.

Metadados administrativos: fornecem informações que apoiam os processos de gestão do ciclo de vida dos recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação. Nessa categoria, estão metadados técnicos que explicitam as especificidades e dependências técnicas do recurso; inclui também os metadados voltados para apoio à gestão dos direitos relacionados ao recurso (SAYÃO, 2010, p. 5).

Assim, uma pesquisa que envolva a coleta de dados na web precisa classificar a instrumentalização de seus dados para correlacionar as informações de diferentes fontes. Desta forma, para estudar a evolução do escopo temático da Ciência Aberta, utilizaram-se os metadados, segundo a categoria conceitual de “Metadados Descritivos”, com o objetivo de analisar os dados extraídos das bases de dados *Web of Science*, *Scopus* e *Directory of Open Access Journals* (DOAJ). Isso possibilita equiparar possíveis discrepâncias de maneira a manter a integridade dos dados e a equivalência informacional, além de observar a granularidade entre as correlações contidas nos resultados.

### 3.2 ETAPAS METODOLÓGICAS

Esta etapa do trabalho visou realizar uma pesquisa exploratória por meio da análise da produção acadêmica sobre Ciência Aberta. Em uma detalhada coleta de dados com os resultados de pesquisas em bases de dados, objetivou-se encontrar um repertório com documentos que tenham a expressão “*open science*”, bem como suas vertentes, para então posicionar a Ciência Aberta em um diagnóstico sobre a sua posição no estudo de análise de domínio e o conceito de campo.

Com o objetivo de identificar o espectro temático e terminológico associado ao termo Ciência Aberta, este trabalho empregou como fonte as bases de dados digitais que indexam as publicações de trabalhos acadêmicos e catalogam periódicos científicos. Assim, tais plataformas, com dados inter-relacionados e organizados de forma a permitir a recuperação da informação e de suas referências, possibilitou levantar todas as publicações científicas indexadas com o termo “Ciência Aberta” entre o período de 1970 a 2018 como forma de traçar o desenvolvimento do campo e os seus desdobramentos.

Com uma análise aprofundada através do cruzamento de informações dos metadados extraídos da pesquisa, foi efetuado um exame da evolução das correlações em torno do tema e do conceito de Ciência Aberta. Isto tem por finalidade observar a evolução terminológica de palavras-chave, bem como o desenvolvimento dessas ao longo de diferentes períodos, relacionando-as com áreas de conhecimento e o surgimento e evolução de vertentes da Ciência Aberta.

Através da análise dos dados, procurou-se identificar e determinar as etapas da diversificação semântica com relação à Ciência Aberta e seus desdobramentos, assim como também facilitar a observação de associações entre os termos. Isto permite interligar citações e referências de cada definição ao longo dos períodos pesquisados, e ainda distinguir especificidades de publicações acadêmicas segundo países e áreas de conhecimento.

Como as diferentes bases de dados não oferecem uma padronização quanto aos metadados que indexam (ou seja, organizam em forma de índice ou vocabulário controlado) e produzem, é fundamental elencar os resultados das pesquisas levantadas para eleger alguns fatores que reúnam as associações feitas, bem como discriminem as correspondências identificadas.

O mapeamento do universo terminológico da Ciência Aberta foi elaborado por meio do levantamento de resultados em pesquisas de cunho geral, realizada em bases de dados, tendo como principal aporte a *Web of Science*, avançando pela rede *Scopus*, e DOAJ. Desta forma, no dia 12 de outubro de 2019, foram realizadas as buscas pelos verbetes “Ciência Aberta”, “*Ciencia Abierta*” e “*Open Science*” em cada base de dados. No entanto, cabe ressaltar que as três bases de dados escolhidas estão em plataformas com o idioma inglês. Assim, praticamente todos os resultados encontrados pelo verbete em português e espanhol foram ínfimos e absorvidos pelo enorme repertório encontrado com “*Open Science*”<sup>17</sup>.

Quanto ao período de coleta de resultados, se estendeu até 25 de outubro, pois, uma vez que a DOAJ não conta com ferramenta para a extração de dados pesquisados, foi necessário copiar individualmente cada resultado. Os metadados dos resultados das três plataformas foram indexados e padronizados para que, uma vez reunidos, pudessem ser empregados como suporte para definições terminológicas mais básicas, culminando em sua utilização em recortes mais específicos para o prosseguimento de estudos.

Por conseguinte, conforme a metodologia apresentada na Figura 1, optou-se por obter os resultados nas bases de dados, reuni-los e uniformizá-los, para então classificar as informações sobre os documentos coletados. Na sequência, foram feitas planilhas ordenadas por ano e ordem alfabética, facilitando a identificação e eliminação de redundâncias<sup>18</sup>, permitindo a análise descritiva dos dados obtidos. Foram feitas revisões e coletas de novos registros indexados em 05 de novembro, 25 de novembro e 12 de dezembro, sendo a derradeira no dia 20 dezembro de 2019.

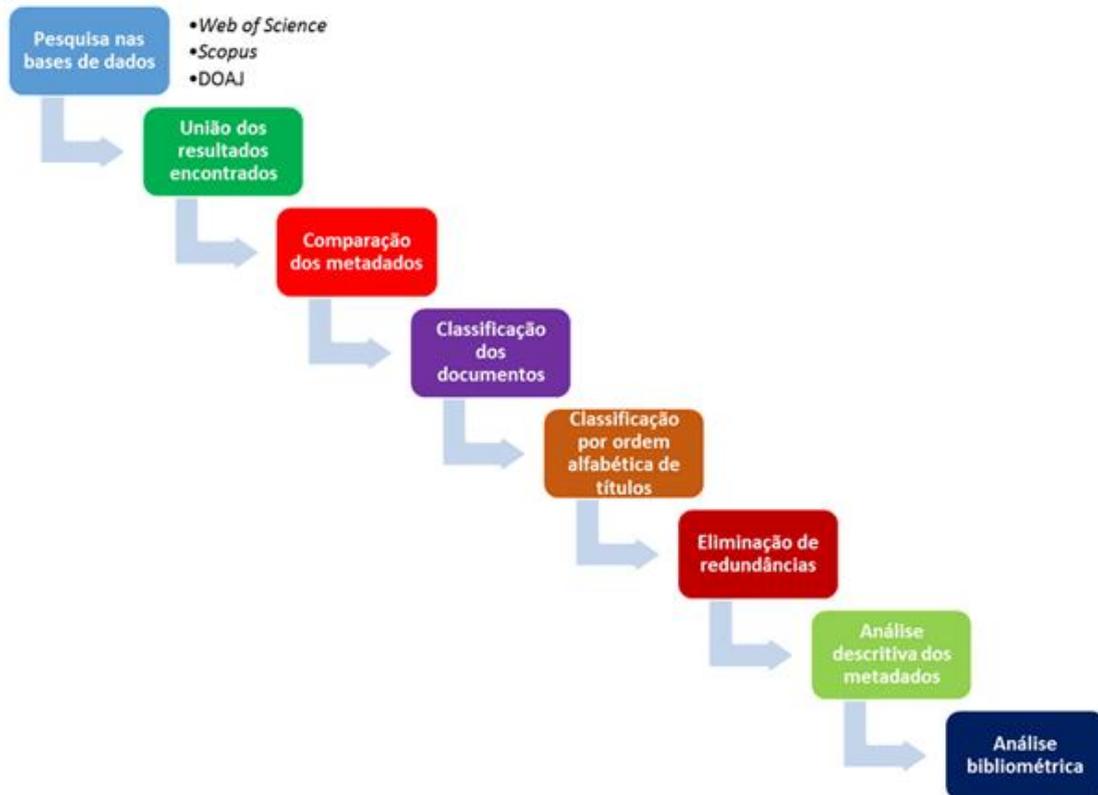
A pesquisa dos documentos com a expressão “*open science*”, no idioma inglês, nas bases de dados *Web of Science*, *Scopus* e DOAJ deu um aporte maior aos resultados que foram obtidos. Portanto, contemplou-se a coincidência de documentos nas três plataformas, bem como a facilidade de extração de recursos, as vantagens de indexações mais pormenorizadas e a especificidade do acervo indexado, ao mesmo tempo em que se contornou problemas tanto na falta de informações em determinados campos de classificação de metadados de documentos, como a completa ausência de informação por deficiências da base de dados.

---

<sup>17</sup> Também cabe apontar que o verbete “*Open Science*” deixa pouca margem para erros de tradução quando traduzido para “Ciência Aberta” ou “*Ciencia Abierta*”.

<sup>18</sup> Na Bibliometria, usa-se o jargão Deduplicação, termo derivado de “*data deduplication*”, que é o nome de uma técnica da Ciência da Computação para eliminar cópias duplicadas de dados repetidos.

Figura 1 – Metodologia para análise bibliométrica do termo de Ciência Aberta.



Fonte: Elaborado pelo autor (2020).

Após a extração do resultado das buscas, optou-se por uma abordagem quantitativa, pois o volume de metadados encontrados se divide em dezenas de classificações, que também se multiplicou pelas centenas de resultados das publicações científicas sobre Ciência Aberta. No entanto, também foi preciso avaliar os dados de forma qualitativa, uma vez que o contexto em que se insere cada dado coletado, bem como a relação destes com termos pesquisados dentre os resultados avaliados, acaba por influir na análise do desenvolvimento nos estudos exploratórios em questão.

De tal forma, reuniu-se os resultados das pesquisas nas bases de dados com o termo “*open science*”, para fazer uma comparação mais apurada dos metadados coletados – definidos por cada uma das plataformas previamente escolhidas. Feito isso, organizou-se os resultados de maneira a estabelecer uma classificação dos documentos, optando-se pelo arranjo dos documentos coletados por título para melhor identificar repetições.

Destacadas as redundâncias nos resultados colhidos, distinguidas primeiro pelo título dos documentos, e confirmadas pelo DOI<sup>19</sup> e local de publicação, eliminou-se as reincidências, destacando as bases em que as repetições foram identificadas. A partir daí, fez-se uma análise descritiva dos metadados reunidos, classificando os documentos de acordo com o local, o idioma e as palavras-chave, para então fazer um exame bibliométrico da pesquisa coletada como marco empírico, de maneira a poder confrontá-lo com o marco teórico.

Assim, pela forma e o conteúdo, efetuou-se uma análise mais profunda e detalhada dos dados reunidos, para então se deter sobre a construção de um mapa conceitual de Ciência Aberta e suas vertentes. Desta forma, o trabalho tem um foco mais voltado para os significados terminológicos, investigando a evolução e as classificações da Ciência Aberta e dos termos originados de seus desdobramentos através do conceito de campo e suas vertentes, à luz de Pierre Bourdieu. Além disso, é efetuada uma análise de domínio do tema principal, nos moldes das ideias de Hjørland, com base na evolução da terminologia, bem como ainda é considerada a abrangência temática das áreas de conhecimento encontradas com o conceito de campo de estudo.

### 3.3 BASES DE DADOS ANALISADAS

Como base na análise de pesquisas em bases de dados, as informações indexadas por essas plataformas puderam fornecer detalhes acerca da produção científica sobre Ciência Aberta nos mais diversos países a partir dos recortes territoriais, temporais, temáticos etc. Isso possibilitou verificar as características sobre o país de publicação dos periódicos, bem como áreas do conhecimento e de autoridade.

Assim, no dia 12 de novembro de 2019, foi refeita a coleta de resultados nas plataformas já citadas, com o fim de reunir os dados necessários para a Análise de Domínio sobre “Ciência Aberta”, aplicada em um estudo sobre o seu Campo Científico. Anteriormente, fez-se uma coleta somente na base de dados *Web of Science* para fins de conceber o marco empírico da qualificação deste trabalho. Isto posto, apresenta-se cada base de dados, porém, devido aos seus objetivos, características e particularidades nos mecanismos de busca, as descrições foram elaboradas separadamente.

---

<sup>19</sup> O *Digital Object Identifier* (DOI) é um identificador único para artigos, anais ou *proceedings* e outras publicações científicas. Esse código também é utilizado pelos órgãos de fomento para identificar a produção científica, o que facilita a contagem de citações entre outras melhorias (GALOÁ, 2020, n. p.).

### 3.3.1 *Web of Science*

A *Web of Science* é uma base de dados que oferece serviço de indexação de citações científicas on-line e que dá acesso a vários bancos de dados institucionais e a periódicos científicos, utilizando instrumentos que identificam metadados de referências interdisciplinares, reconhecendo autores, títulos, assuntos, ano de publicação etc.

Conhecida anteriormente como *Web of Knowledge*, a plataforma era uma interface para manipular os recursos do *Institute of Scientific Information* (ISI). Fundado pelo cientometrista Eugene Garfield (CLARIVATE ANALYTICSWEB OF KNOWLEDGE, 2020), a instituição contava com uma série de índices para metrificar e avaliar publicações científicas, tais como: *Science Citation Index* (SCI, conhecido em português como “Índice de Citação Científica”), *Social Science Citation Index* (SSCI), *Arts and Humanities Citation Index* (AHCI) e o *Science Citation Index Expanded* (SCIE).

Utilizando a lei de distribuição de Bradford<sup>20</sup>, Garfield formulou a sua própria lei de concentração para estabelecer o Índice de Citação Científica (SCI) oferecendo certa credibilidade a uma representação mais justa do que seria considerado uma “Ciência neutra” (GUÉDON, 2010, p.26). Embora este índice estivesse limitado a algumas centenas de títulos, acabou por estabelecer barreiras entre o que seria a “Ciência principal” e o restante das publicações científicas.

Em 1992, a antiga *Web of Knowledge* e todo o seu repertório de recursos passou a fazer parte da *Thompson Scientific & Healthcare*, parte integrante da *Thomson Reuters* (CLARIVATE ANALYTICS, 2020). Esta última empresa, desde 2016, faz parte do setor de Propriedade Intelectual e Ciência da *Clarivate Analytics*. Assim, desde então, os índices eletrônicos de medição cientométrica do *Institute of Scientific Information* passaram a ser realizados e operados pela base de dados e interface *Web of Science*.

Segundo Zattar e Marteletto (2017, p. 105), esta plataforma é uma base de dados multidisciplinar que indexa periódicos mais citados em suas respectivas áreas, além de fornecer um índice de citações, confirma, em cada artigo, os documentos por ele citados e os documentos que o citaram.

---

<sup>20</sup> A lei de Bradford é um padrão estima os retornos exponencialmente decrescentes da busca por referências em revistas científicas. Assim, tal estamento concentra sua descrição no comportamento repetitivo das ocorrências de um determinado campo do saber, de maneira que se observa poucos periódicos publicando muitos artigos e muitos periódicos publicando poucos artigos.

Uma vez que a *Clarivate Analytics* detém o acervo de índices eletrônicos do *Institute of Scientific Information*, a empresa pode operar um acervo de publicações científicas que é baseado por seus serviços de assinaturas, focados principalmente em análises dos dados de pesquisa científica e acadêmica, inteligência de patentes e padrões de conformidade, marca registrada de inteligência em farmacêutica e biotecnologia, domínio e proteção de marca, bem como a propriedade intelectual de todo acervo listado.

Além disso, a *Clarivate Analytics* também utiliza seus recursos para publicar uma lista anual de “Pesquisadores altamente citados”. De acordo com a empresa, tal arrolamento é reconhecido por alguns dos “pesquisadores mais influentes do mundo na última década, demonstrados pela produção de vários artigos altamente citados que figuram no *top*, isto é, 1% por citações de campo e ano na Web of Science” (WEB OF SCIENCE GROUP, 2020b, n. p.). Os pesquisadores são selecionados por seu desempenho excepcional em um ou mais de 21 campos (aqueles usados nos ESI<sup>21</sup>) ou em vários campos.

A plataforma *Web of Science* tem seus dados cadastrados por profissionais que registram e classificam metadados bem variados e detalhados para a aplicabilidade de estudos sobre a literatura formal. No entanto, vale ressaltar que tais registros e indexações, apesar de serem feitos por especialistas, podem apresentar problemas no cadastro e seguir vieses mercadológico de suas proprietárias.

A *Web of Science* foi escolhida não só pela enorme variedade de classificações indexadas nos registros da *Clarivate Analytics*, mas principalmente pelo registro de palavras-chave de autores e por ter duas categorias de rotulagem de área do conhecimento: “Categoria da *Web of Science*” e “Área de pesquisa”. Dessas, enquanto a primeira tem uma classificação mais difusa e que presa pela catalogação multidisciplinar, a última faz uma especificação dentro de uma estrutura hierárquica, nos moldes de um tesouro.

---

<sup>21</sup> *Essential Science Indicators*: é um banco de dados formado por “Indicadores de Ciências Essenciais” que tem a função de revelar tendências científicas emergentes, bem como indivíduos, instituições, artigos, periódicos e países influentes em seu campo de pesquisa. O ESI é um substrato de estatísticas de tendências de pesquisas extraídas de mais de 12 milhões de artigos, de mais de 12.000 periódicos de todo o mundo, oferecendo uma cobertura detalhada para analisar e comparar com eficiência o desempenho da pesquisa, identificar tendências significativas, classificar os melhores desempenhos e avaliar funcionários e colaboradores em potencial. Com base em um arquivo contínuo de 10 anos e atualizado continuamente pelos índices da base de dados, os ESI são atualizados a cada dois meses e usa dados seletivos, estruturados e completos (WEB OF SCIENCE GROUP, 2020a, n. p.).

### 3.3.2 *Scopus*

A *Scopus* é uma base de dados pertencente à *Elsevier*<sup>22</sup> e também tem vieses mercadológicos que oferecem uma fonte referencial de literatura técnica e científica revisada por pares. Porém, assim como a *Clarivate Analytics*, também reúne uma coleção de periódicos eletrônicos, além de banco de dados de citações on-line e mecanismos de pesquisa voltados para profissionais das áreas da saúde. Igualmente, a *Elsevier* oferece ferramentas digitais para gerenciamento, instrução e avaliação de dados, principalmente os indexados em suas bases de dados. Todavia, assim como a *Web of Science*, a plataforma também pode apresentar problemas no registro e na indexação de seus documentos, bem como acompanhar tendências do mercado para a literatura formal.

De tal modo, a *Scopus* é uma base de dados com três tipos de fontes: séries de livros, revistas científicas periódicas e periódicos comerciais. Neste contexto, todas as fontes publicadas em série que estão indexadas em seu banco de dados são revisadas a cada ano para garantir a manutenção dos padrões de qualidade escolhidos pela *Elsevier*. Tal padronização acaba impactando nas palavras-chave e áreas de estudo que a empresa decide publicar em seus relatórios anuais.

A inclusão desta plataforma no repertório da pesquisa atendeu também a grande diversidade de classificações dos registros indexados pela *Elsevier*, dentre essas, destaca-se o gigantesco volume de palavras-chave de autor. Além disso, essa base de dados também traz categorizações mais precisas sobre o suporte material catalogado – se oriundo de periódicos científicos, livros ou congressos –, permitindo qualificar melhor a espécie dos documentos encontrados.

### 3.3.3 DOAJ

O DOAJ é um site que hospeda uma lista com curadoria da comunidade de revistas de acesso aberto, sendo que muitos dos artigos listados foram inseridos por editores de periódicos ou pelos próprios autores dos artigos. O DOAJ define os periódicos de acesso aberto como periódicos científicos e acadêmicos e tem padrões de qualidade definidos pela

---

<sup>22</sup> Editora multinacional com sede na Holanda que faz a publicação e a análise especializada de conteúdos científico, técnico e médico.

revisão por pares ou o controle editorial da qualidade. Assim, apesar de alguns erros, o projeto fornece uma dimensão mais próxima da intenção da classificação do pesquisador ao publicar o artigo.

Mantida pelos *Infrastructure Services for Open Access* (IS4OA ou “Serviços de Infraestrutura para Acesso Aberto”), a plataforma é norteada expressamente pela definição de acesso aberto da Iniciativa de Budapeste em Acesso Aberto:

Por “acesso aberto” a essa literatura, entendemos sua disponibilidade gratuita na internet pública, permitindo que qualquer usuário leia, baixe, copie, distribua, imprima, pesquise ou faça o *link* para os textos completos desses artigos, rastreie-os para indexação, passá-los como dados para o software ou usá-los para qualquer outra finalidade legal, sem barreiras financeiras, legais ou técnicas que não sejam as inseparáveis de obter acesso à própria internet. A única restrição à reprodução e distribuição, e o único papel dos direitos autorais nesse domínio, deve ser o de dar aos autores controle sobre a integridade de seu trabalho e o direito de serem reconhecidos e citados adequadamente (BUDAPEST OPEN ACCESS INITIATIVE, 2002, n. p.).

Tal projeto disponibiliza todo o seu conteúdo gratuitamente, sem demora ou com requisitos de registro do usuário. Da mesma forma, a base de dados oferece garantias para internacionalização de artigos lá cadastrados, com o objetivo de “aumentar a visibilidade, acessibilidade, reputação, uso e impacto de periódicos de pesquisa acadêmica de qualidade, com revisão por pares e acesso aberto” (DIRECTORY OF OPEN ACCESS JOURNALS, 2018, n. p.).

O serviço procura atender aos mais altos padrões de excelência, fazendo tanto a revisão por pares, como um controle editorial da qualidade. Desta forma, os resultados em seus sistemas de pesquisa acabam por afastar de qualquer intencionalidade comercial de grandes empresas que lidam com propriedade intelectual, podendo também repercutir de forma mais apurada na análise do domínio da área pesquisada.

Como o DOAJ é uma base de dados de acesso aberto e tem o processo de publicação totalmente gratuito, os autores ou as instituições de pesquisa precisam preencher e enviar o formulário de publicação, definir a tipologia do periódico cadastrado e esperar a verificação dos dados por parte dos especialistas do DOAJ. Logo, tendo em vista que os artigos e revistas periódicas registradas têm um caráter mais próximo dos objetivos e da classificação temática dos responsáveis pelas pesquisas, os metadados de seu acervo não estão tão voltados para o escopo mercadológico das editoras científicas, tal como a *Web of Science* e a *Scopus*.

Essa particularidade da plataforma com relação ao compromisso de promover o acesso aberto se relaciona diretamente com a sua seleção para figurar no conjunto de pesquisas deste

trabalho. Sem oferecer recursos de filtragem de resultados ou ferramentas para extrair o repertório encontrado, o DOAJ exigiu maior tempo para verificar e copiar a pesquisa, pois foi necessário examinar cada item encontrados e copiá-los um a um. Entretanto, todas as classificações dos registros dessa base de dados estavam preenchidas, o que não ocorreu na *Web of Science* e *Scopus* – apesar da grande quantidade de classificações dessas plataformas, grande parte de seus campos estavam sem informações. Assim, o DOAJ também serviu de extrema valia para confirmar dados em documentos que também foram encontrados nas outras bases de dados.

### 3.4 SOFTWARES UTILIZADOS

A instrumentalização desta pesquisa fez uso de dois programas gratuitos que têm recursos para criar representações gráficas a partir de metadados. A visualização da informação organizada de acordo com indicadores específicos ajuda a construir modelos em que dados abstratos, de diferentes fontes, possam ser cruzados, comparados e destacados.

Os quadros, tabelas e gráficos foram elaborados no Microsoft Excel de modo a permitir retratar as relações abstratas em uma enorme quantidade de dados. Porém, os gráficos criados com o VOSviewer e com o Gephi potencializam a compreensão quantitativa e qualitativa das inúmeras conexões de redes complexas. Assim, utilizou-se os softwares para dar um norte primário com relação aos vínculos entre os metadados coletados nas pesquisas, optando-se por representações mais tradicionais em um segundo momento.

#### 3.4.1 VOSviewer

O VOSviewer é um software gratuito que serve de ferramenta para construir e visualizar mapas bidimensionais de redes bibliométricas. Essas podem abranger, por exemplo, pesquisadores, publicações periódicas ou publicações individuais, com recursos para construí-las “com base em relações de citação, conexões bibliográficas, cocitação ou coautoria” (VOSVIEWER, [2020], n. p.).

Segundo Van Eck *et al.* (2010, p. 2405), a bibliometria e a cientometria utilizam a técnica de *Multidimensional Scaling* (MDS<sup>23</sup>) para construir mapas da ciência que desafiam

---

<sup>23</sup>*Multidimensional Scaling* (MDS) ou escalonamento multidimensional é uma técnica de escalonamento multidimensional que constrói um mapa após as medidas de similaridade serem calculadas. O objetivo do MDS

pesquisadores há décadas. Estes mapas podem exibir relação entre autores, periódicos, documentos ou palavras-chave e são construídos baseados em técnicas que calculam citação, cocitação, acoplamento bibliográfico ou coocorrência.

Segundo seus desenvolvedores, o objetivo da técnica empregada no VOSviewer é o mesmo do MDS, ou seja, localizar itens em um reduzido espaço dimensional de modo que a distância entre quaisquer dos itens reflita a similaridade ou apresente esta similaridade da forma mais acurada possível, diferenciando-se do MDS quanto ao modo como busca alcançar este objetivo (VOSVIEWER, [2020], n.p.).

O programa ainda oferece funcionalidades para mineração de textos, cruzando resultados para destacar a simultaneidade em redes dentro do panorama da literatura científica. Da mesma forma, o VOSviewer pode extrair termos filtrados de suas análises em formatos .txt, permitindo corrigir duplicidades ou erros de ortografia, ou ainda instrumentalizar recursos informacionais – como, por exemplo, tesouros, vocabulários controlados etc.

Como o software tem recursos para utilizar dados extraídos diretamente das bases de dados *Web of Science* e *Scopus* – fato observado após a escolha dessas plataformas –, utilizou-se os arquivos no formato .txt. da primeira e .csv da outra, utilizados como resultados na primeira etapa metodológica: a pesquisa direta nas bases de dados.

No ambiente do programa, após escolher a fonte dos dados, o programa oferece cinco tipos de análise: de coautorias, de coocorrências, de citações, de conexões bibliográficas e de cocitação. Cada opção disponibiliza um tipo de relatório diferente, possibilitando uma adequação melhor ao conteúdo não só pela classificação do campo pesquisado, como também pela abrangência dos termos nos resultados encontrados.

Para concluir o processo de representação gráfica, basta escolher entre a contagem completa ou fracionada dos dados selecionados – sendo esta última um desdobramento do peso das conexões identificadas. As últimas fases permitem escolher e filtrar o número de ligações entre os metadados escolhidos e visualizar o rol de termos que será representado.

---

é “localizar itens em um espaço dimensional reduzido de modo que a distância entre quaisquer dois itens reflita a similaridade ou apresente esta correspondência da forma mais acurada possível. Assim, quanto mais forte for a relação entre dois itens, menor a distância entre eles” (VAN ECK *et al.*, 2010, p. 2405).

Neste trabalho, utilizou-se o VOSviewer para elaborar representações gráficas sobre o repertório de coautores e palavras-chave recolhido na *Web of Science* e na *Scopus*. Dessa forma, as figuras foram formadas com os parâmetros de tipo de análise de coautoria (“*Co-actorship*”), com as unidades de análises de autores (“*Authors*”) e países (“*Countries*”), e tipo de análise de coocorrência (“*Co-occurrence*”), com palavras-chave de autor (“*Author Keywords*”) e das bases de dados (“*KeyWords Plus*” para a *Web of Science* e “*Index keywords*” para a *Scopus*).

As figuras foram formatadas com as ferramentas da aba “*Anaysis*” e com os ajustes visualização (“*Visualization*”), etiquetas (“*Labels*”) e linhas (“*Lines*”) para apresentar a nomenclatura de todos os nós e representar as ligações das arestas. Da mesma forma, optou-se pelas apresentações disponíveis em “*Network Visualization*”, que destacam os agrupamentos estatísticos segundo a modularidade da rede, e “*Overlay Visualization*”, a qual realça a atividade cronológica de cada nó na rede.

Em derradeiro, vale apontar que as representações gráficas foram capturadas com a ferramenta “*Screenshot*” do próprio VOSviewer, sendo importadas para este trabalho com os devidos ajustes de zoom e enquadramento.

### 3.4.2 Gephi

O Gephi também é um software gratuito e de código aberto que tem a finalidade de analisar redes de dados e oferecer suportes para a visualização de agrupamentos de dados (*clusters*), além de isolar singularidades e detectar modularidades. Com instrumentos bem versáteis para importar metadados, personalizar representações gráficas e filtrar informações de gráficos, o programa exige noções sobre o funcionamento de teoria de redes complexas.

Apesar de não ter recursos para ler o resultado extraído diretamente de bases de dados como a *Web of Science* ou a *Scopus*, o software pode interpretar redes se os metadados destes forem adaptados para planilhas com as especificações de “Nós” e “Arestas” da rede que será trabalhada. Assim, mesmo com a desvantagem pela falta de adequação às principais plataformas de pesquisa, o software se torna de enorme proveito para condensar os resultados combinados de diferentes fontes.

Desenvolvido em Java e com base na plataforma NetBeans, o Gephi mapeia dados e oferece recursos para facilitar a análise de perspectivas em representações gráficas. Para estruturar os gráficos de redes de dados, basta carregar arquivos em .csv, podendo alterar

disposições de itens, editar informações importadas e personalizar todas as configurações utilizadas para a construção da representação da relação entre os dados.

O software é bem simples e intuitivo. Contando com instrumentos para destacar, filtrar, nomear e editar os elementos apresentados, o programa permite mapear uma ampla variedade de conjuntos de dados em redes. Para utilizá-lo, basta manejar os recursos e informações em suas três abas: Visão Geral, Laboratório de Dados e Visualização. A primeira faz simulações e configurações na representação dos gráficos que serão criados, a segunda exhibe todo o conteúdo dos metadados da rede de dados, e a última permite personalizar a representação gráfica.

O Gephi foi utilizado nos resultados dessa pesquisa a partir de duas planilhas criadas especificamente para representar uma rede complexa do levantamento acerca dos autores que mais publicaram sobre “Ciência Aberta” dentro do levantamento feito. Enquanto uma planilha representou os nós e suas ligações com outros nós, a segunda planilha determinou a força da ligação das arestas.

Ambas planilhas foram importadas para o Gephi pelo ambiente “Laboratório de Dados”, permitindo gerar a representação gráfica almejada a partir do ambiente “Visão Geral”. A partir desse levantamento, fez-se a figura para demonstrar o volume de publicações de cada um dos autores que mais publicaram em cada plataforma, mas também destacando as interseções entre essas. Ainda em “Visão Geral”, configurou-se a apresentação visual da rede, bem como os destaques necessários para sua análise. Em seguida, construiu-se o modelo para exportação no ambiente “Visualização”.

## 4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Para o delineamento empírico da investigação deste trabalho, buscou-se estruturar uma perspectiva sobre o planejamento e a abrangência dos ambientes em que as informações serão coletadas, analisadas e interpretadas, provendo o trabalho com dados da pesquisa propriamente dita. Tal expediente partiu de uma instrumentalização baseada no confronto de informações coletadas nas bases de dados selecionadas, para então examinar a evolução do escopo temático e terminológico em torno da noção de Ciência Aberta.

As bases de dados, principalmente aquelas de caráter restrito e com viés mercadológico, são ambientes extremamente formais e controlados. Em geral, instituições privadas definem os documentos que entrarão ou não em seus acervos, estabelecendo um regime de informação claramente definido. Os critérios sobre o controle do material que será incorporado fazem parte de um filtro comercial com um viés mercadológico.

Desta forma, a opção pela coleta de dados na *Web of Science* e *Scopus* tem o objetivo principal de confrontar as perspectivas comerciais de duas empresas diferentes. Por outro lado, a coleta dos resultados pesquisados no DOAJ tem o escopo de orbitar os pontos de vista mais próximos dos autores. Uma vez que essa última base de dados tem um controle menos rigoroso e mais próximo de propósitos da divulgação da informação científica, pressupôs-se que a reunião dos resultados traria uma visão mais equilibrada do corolário analisado.

### 4.1 RECORTE HISTÓRICO SOBRE CIÊNCIA ABERTA

Realizada a pesquisa nas bases de dados, organizaram-se os dados coletados de cada plataforma para seguir para a segunda etapa metodológica – a combinação e o cruzamento dos resultados encontrados. No entanto, antes de analisar a repercussão dessa agregação, registrou-se uma caracterização do conjunto de dados reunidos.

Como citado anteriormente, não há uma definição clara da responsabilidade pela cunhagem do termo “Open Science”, e já há registros desta expressão em documentos da *Web of Science* e da *Scopus*, em 1970 e 1979, respectivamente. No entanto, trabalhos sobre “Ciência Aberta” como o conceito que se objetiva estudar nesta investigação só passaram a ser publicadas progressivamente a partir de 2006 – ano em que o DOAJ foi criado.

Antes desta data, o repertório de documentos com o conceito “Ciência Aberta” utiliza esta terminológica de forma variada dependendo da temática do trabalho. Enquanto em alguns “Open Science” para outra expressão, outros tem uma semântica sobre políticas para um acesso mais aberto da ciência que se aproximam da socialização de saberes (Apêndice A).

Essa pormenorização visa não só dar uma dimensão dos padrões observados em cada base de dados, mas também delinear um breve histórico da sistematização efetuada na pesquisa até a organização dos resultados apurados. Uma vez examinada cada plataforma individualmente, as considerações passam a classificar o conjunto de dados coletados de acordo com o material e as especificidades das classificações dos documentos analisados.

#### 4.1.1 *Web of Science*

A base de dados *Web of Science* oferece práticos recursos para extrair todo o seu repertório de metadados sobre cada documento indexado pela *Clarivate Analytics* na base de dados, incluindo endereços de autores e editoras, múltiplos campos de categorias, palavras-chave etc.

Assim, no campo de buscas da *Web of Science*, foi selecionada a opção de resultados por tópico, que pesquisa por termos em títulos, resumos e palavras-chave de todos os documentos indexados em seu acervo. A seguir, efetuou-se a busca pela expressão “*open science*”, resultando em um total de 1.455 documentos encontrados na Principal Coleção da *Web of Science* – conforme demonstrado na Figura 2.

Figura 2 – Coordenadas da pesquisa na *Web of Science*.

TÓPICO: ("open science")
Refinado por: [excluindo] ANOS DE PUBLICAÇÃO: (2019), (2020)
Tempo estipulado: Todos os anos. Índices: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.
Resultados: 1.455 (de Principal Coleção do Web of Science)

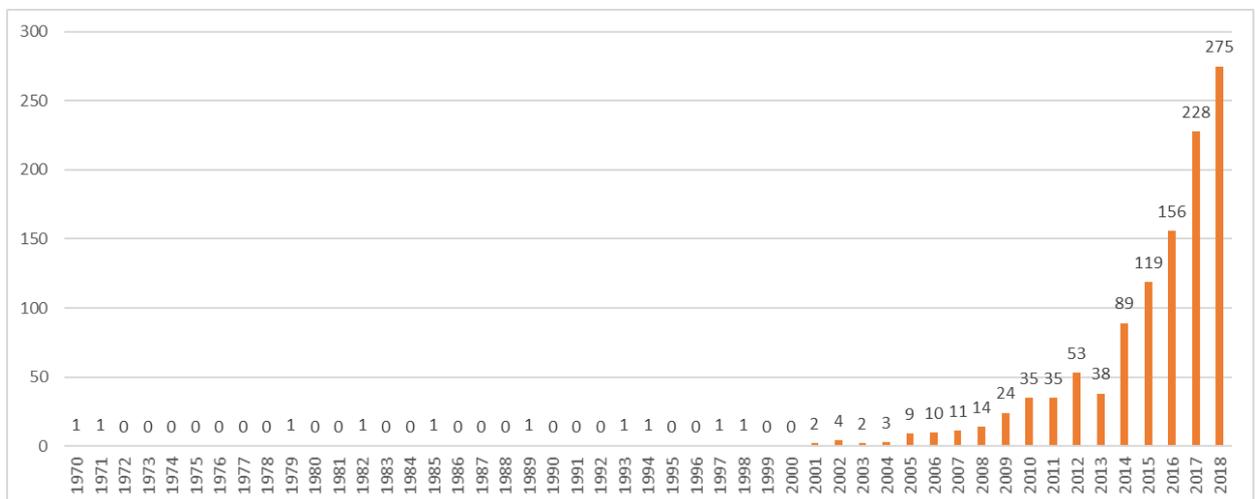
Fonte: Elaborado pelo autor (2020).

Deste montante de 1.455 documentos, com o fim de melhor apurar a investigação sobre o estado da arte da produção científica sobre Ciência Aberta na *Web of Science*, optou-

se por depurar o ano de publicação excluindo “2019” e “2020”, para equalizar o período da investigação em questão para os anos findos. Isto evita o lapso de publicações nos meses de novembro e dezembro de 2019 e incongruências de publicações preambulares programadas para 2020.

Como resultado, obteve-se 1.117 documentos, compreendendo o período de 1970 até 2018, conforme ilustrado na Figura 3. O produto da pesquisa foi extraído por três documentos no formato .txt<sup>24</sup>, pois a *Web of Science* limita o *download* em até 500 registros por vez, com o registro completo e as referências citadas em cada documento. Em seguida, o conteúdo dos arquivos foi copiado e transferido para planilhas do Microsoft Excel, no formato .xlsx<sup>25</sup>. Uma vez nesta padronização, os títulos das colunas, em abreviações, foram transcritos para facilitar a visualização e a organização dos dados para futuras análises.

Figura 3 – Evolução das publicações sobre “open Science” na *Web of Science* no período de 1970-2018.



Fonte: Elaborado pelo autor (2020).

Com base no que foi encontrado, observou-se uma ocorrência mais frequente dos resultados de “open science” identificados com algumas das principais vertentes da Ciência Aberta como: “open access”, “open data”, “open source”, “citizen science” e “open

<sup>24</sup> Abreviatura de “texto simples” ou “texto puro”, é o formato de arquivo mais simples de todos, em que o conteúdo se resume a um arquivo sequencial ordinário e que tem como conteúdo material textual sem muito processamento.

<sup>25</sup> Formato padrão de arquivos do Microsoft Excel 2007 e versões posteriores. Na realidade, é um arquivo compactado Zip com uma estrutura de diretórios de documentos de texto *Extensible Markup Language* (XML ou Linguagem de Marcação Estendida). Funcionando como substituto primário para o formato binário anterior, o *Excel Binary File Format* (XLS ou Formato de Arquivo Binário do Excel), embora não suporte macros do Excel por motivos de segurança.

*education*”, tanto pela frequência destes em títulos, quanto em resumos e palavras-chave. Porém, tal liame será discutido mais à frente.

#### 4.1.2 *Scopus*

Assim como nas buscas feitas na *Web of Science*, foi selecionada a opção de resultados por “*Article title, Abstract, Keywords*”, o que englobou a pesquisa por termos nos títulos, nos resumos e palavras-chave de todos documentos do acervo. Em seguida, foi realizada a pesquisa com a expressão “*open science*” – conforme demonstrado na Figura 4 – e, como resultado, encontrou-se 2.083 itens indexados pela base de dados.

Figura 4 – Coordenadas da pesquisa na *Scopus*.

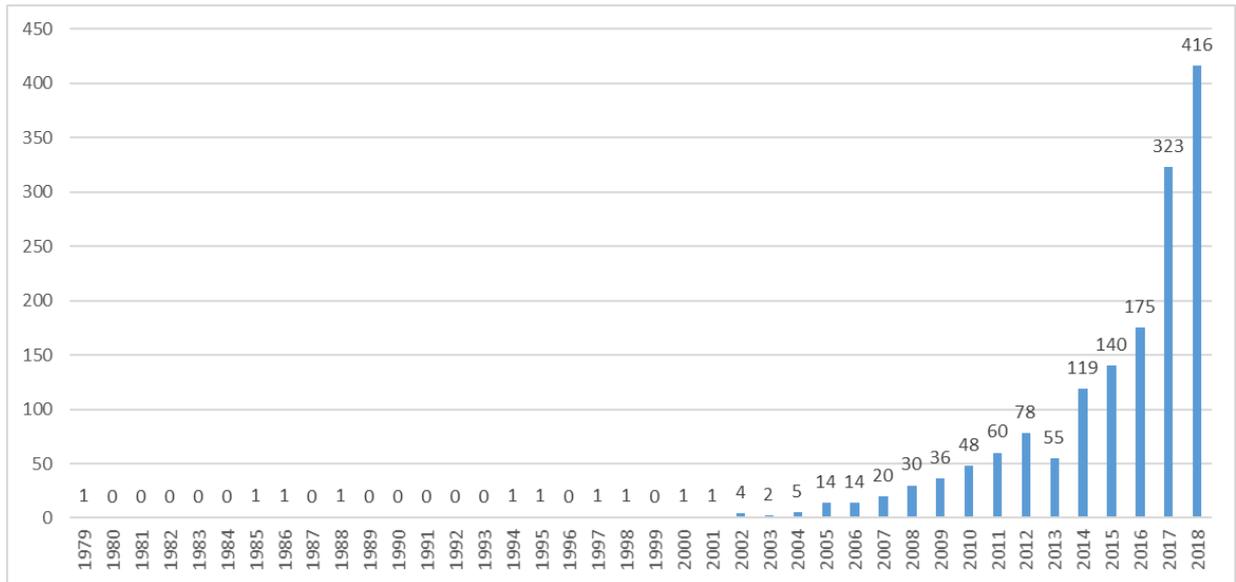
TITLE-ABS-KEY ( "open science" ) AND ( EXCLUDE ( PUBYEAR, 2020 ) OR EXCLUDE ( PUBYEAR, 2019 ) )

Resultados: 1.548

Fonte: Elaborado pelo autor (2020).

Como aplicado anteriormente, optou-se por remover os resultados das publicações dos anos “2019” e “2020”, limitando o período de corte da pesquisa até 2018. Assim, obteve-se o resultado de um total de 1.548 documentos sobre “*open science*” na base de dados *Scopus*, abrangendo um conteúdo de publicações entre 1979 até 2018, conforme apresentado na Figura 5.

Figura 5 – Evolução de publicações sobre “*open science*” na *Scopus* no período de 1979-2018.



Fonte: Elaborado pelo autor (2020).

A base de dados *Scopus* oferece uma enorme quantidade de metadados para coletar, todos selecionáveis e já extraídos no formato .csv<sup>26</sup> para serem trabalhados em planilhas do Microsoft Excel. Cabe destacar que esta base oferece o recurso de extração que limita o *download* em até 2.000 itens de resultado – o que não foi um problema, pois o resultado com o corte do período ficou abaixo deste limite.

#### 4.1.3 DOAJ

A pesquisa no DOAJ foi efetuada nos mesmos moldes daquelas realizadas na *Web of Science* e na *Scopus*. Porém, a base de dados é menos sofisticada que estas de conteúdo restrito e não oferece recurso de refino para as pesquisas. Destarte, foram encontrados 937 artigos com o termo “*open science*” em quaisquer dos campos registrados na base. Do mesmo modo, foram encontrados resultados nos mais variados idiomas, contudo, a indexação das palavras-chave é padronizada pelo idioma inglês, apesar do título e do resumo poderem estar em outras línguas. As coordenadas de pesquisa são apresentadas na Figura 6.

<sup>26</sup> *Comma-separated values* (CSV ou valores separados por vírgulas) é um tipo de arquivo de texto de formato regulamentado pelo RFC 4180 – documentos técnicos desenvolvidos e mantidos pelo *Internet Engineering Task Force* (IETF), instituição que especifica os padrões que serão implementados e utilizados em toda a internet. O formato .csv é, normalmente, utilizado por softwares de organizadores de texto em planilhas, tais como o *Microsoft Excel* e o *LibreOffice Calc.*, pois este tipo de arquivo permite fazer uma ordenação de *bytes* ou um formato de terminador de linha, separando valores com vírgulas.

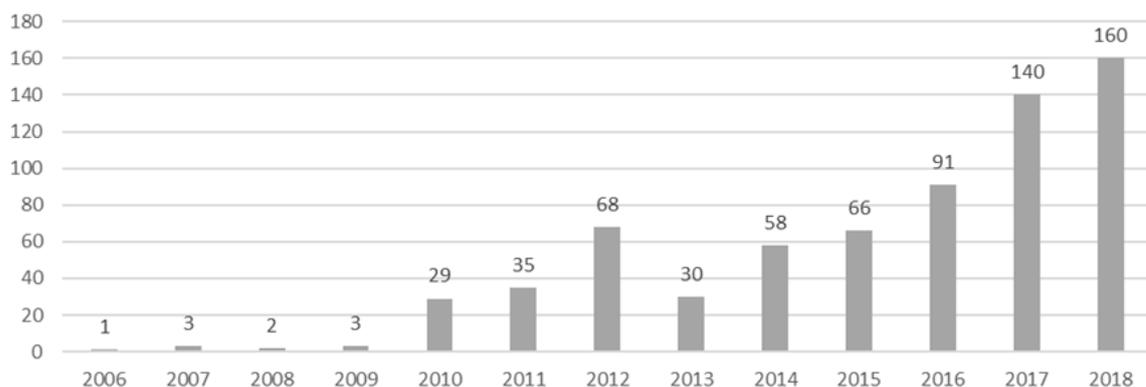
Figura 6 – Coordenadas da pesquisa no DOAJ.

Busca por: “open science”
Resultados: 937

Fonte: Elaborado pelo autor (2020).

Cabe destacar que, como o DOAJ também não conta com recursos para o *download* de resultados, foi necessário delimitar o ano de publicação dos resultados encontrados e, a partir de então, copiar cada resultado individualmente e colá-los em uma planilha do Microsoft Excel, já no formato .xlsx.

Ao organizar a lista transcrita, notou-se que as primeiras publicações registradas na base de dados eram de 2006, enquanto 251 tinham como ano de publicação “2019”. Assim, o resultado se limitou a um total de 686 documentos com os termos “*open science*”, com um período de corte que se inicia em 2006 e vai até 2018, conforme demonstra a Figura 7.

Figura 7 – Evolução das publicações sobre “*open science*” no DOAJ no período de 2006-2018.

Fonte: Elaborado pelo autor (2020).

Tendo em vista o compromisso pelo acesso aberto do DOAJ e o grau de comprometimento dos pesquisadores que registram seus artigos nessa base, fez-se uma breve análise dos resultados pelos verbetes “Ciência Aberta” e “*Ciencia Abierta*”. Encontrou-se 48 resultados para o primeiro e 33 para o segundo, de forma que, como o que se observou previamente com relação às bases de dados *Web of Science* e *Scopus*, os mesmos artigos também já se encontravam no rol da pesquisa feita em inglês. Tal fato demonstra a

preferência pela indexação em inglês, não só pelo grau de internacionalização que o documento alcança, mas também por este ser o idioma oficial da plataforma.

Cabe ainda destacar o diferenciado sistema de classificação de palavras-chave por parte dessa base de dados. Tais termos são incluídos por editores de periódicos científicos, porém, durante o processo de registro de artigos, é preciso escolher os termos disponíveis dentro de um repertório pré-definido pela plataforma. Além desse vocabulário controlado, o DOAJ também faz a importação de classificações da revista científica, que se referem a determinadas áreas de conhecimento, mas não necessariamente do artigo cadastrado.

Assim, a base de dados oferece uma classificação de palavras-chave híbrida e sem termos livre. Há um vocabulário controlado, mas os editores têm liberdade para escolher os vocábulos que melhor se adequam aos artigos cadastrados. Percebe-se certa poluição dos termos incluídos, por conta da importação da tipologia de área de conhecimento das revistas científicas, mas estas indexações são facilmente identificáveis.

#### 4.2 RECORTE MATERIAL DOS DOCUMENTOS SOBRE CIÊNCIA ABERTA

O material coletado da *Web of Science*, *Scopus* e DOAJ foi bem discrepante, menos pelo tipo e qualidade dos documentos reunidos, e mais, principalmente, pelas categorias e dimensões dos metadados obtidos, conforme descrito no Quadro 1. Como a *Web of Science* e a *Scopus* são serviços privados e indexados por especialistas com viés metodológico, a amplitude dos dados foi bastante extensa e variada, enquanto o DOAJ foi limitado a nove campos: “Título”, “Fonte”, “Ano”, “DOI”, “Palavra-chave do autor”, “País”, “Idioma”, “Autores” e “Resumo”.

Quadro 1 – Comparação entre os metadados das bases de dados utilizadas.

	Web of Science	Scopus	DOAJ
<b>Título</b>	Sempre (*)	Sempre	Sempre
<b>Fonte</b>	Sempre	Sempre	Sempre
<b>DOI</b>	Quase sempre (**)	Quase sempre	Sempre
<b>Ano</b>	Sempre	Sempre	Sempre
<b>Palavra-chave do autor</b>	Quase sempre	Sempre	Não tem (***)
<b>Palavra-chave da base de dados</b>	Quase sempre	Quase sempre	Não tem
<b>Palavra-chave híbrida</b>	Não tem	Não tem	Sempre
<b>Idioma</b>	Quase sempre	Quase sempre	Sempre
<b>País de Publicação</b>	Quase sempre	Quase sempre	Sempre
<b>Autor(es)</b>	Sempre	Sempre	Sempre

<b>Resumo</b>	Sempre	Sempre	Sempre
<b>Tipo do documento</b>	Quase sempre	Quase sempre	Não tem
<b>Categoria da publicação</b>	Quase sempre	Sempre	Não tem

(\*) Sempre – Registrado em 100% da amostragem

(\*\*) Quase sempre – Registrado em pelo menos 70% da amostragem

(\*\*\*) Não tem – Classificação inexistente na base de dados

Fonte: Elaborado pelo autor (2020).

A consolidação de uma planilha com todos os dados se pautou nos campos citados, limitando o resultado a uma equiparação aos resultados do DOAJ. No entanto, alguns outros campos da *Web of Science* e da *Scopus* foram considerados, tais como “Tipo de Documento”, “Categoria da Publicação” e “Palavra-chave da base de dados”, que, no caso, seguiram políticas de indexação próprias e diferentes daquelas escolhidas pelos autores.

Vale ressaltar que, enquanto o DOAJ tem um acervo digital de artigos de revistas periódicas com acesso aberto, a classificação com os metadados para os campos “Tipo de Documento” e “Categoria da Publicação” serviram de argumento para qualificar documentos como publicações em revistas periódicas, capítulos de livros ou conferências para o primeiro, e artigos e outros suportes editoriais, tais como cartas, correções, revisões etc. para a outra categoria de campo, conforme apresentado no Quadro 2.

Quadro 2 – Tipos de Documentos e Categorias de Publicação de documentos com a expressão “*open science*” sobre *Ciência Aberta* na *Web of Science* (WOS), *Scopus* e DOAJ, com dupla contagem (1970-2018).

Tipo de documento	Categoria da publicação	WOS	Scopus	DOAJ
	Periódicos	828	1022	686
	Artigos	701	735	686
	Outros	127	287	0
Livros e Conferências		289	526	0

Fonte: Elaborado pelo autor (2020).

Tal levantamento permitiu especificar a porcentagem de cada tipo de documento encontrado na *Web of Science*, *Scopus* e DOAJ, conforme a tipologia documental apresentada no Quadro 3. Essa análise, sem excluir dupla contagem, demonstra a relevância dos artigos científicos para a representação da literatura formal, dentre o material encontrado nas bases de dados selecionadas.

Quadro 3 – Tipos de Documentos com a expressão “*open science*” sobre *Ciência Aberta* na *Web of Science* (WOS), *Scopus* e DOAJ, com dupla contagem (1970-2018).

Tipo de documento	WOS	Scopus	DOAJ	Total
<b>Periódicos</b>	74%	66%	100%	80%
<b>Livros e conferências</b>	26%	34%	0%	20%

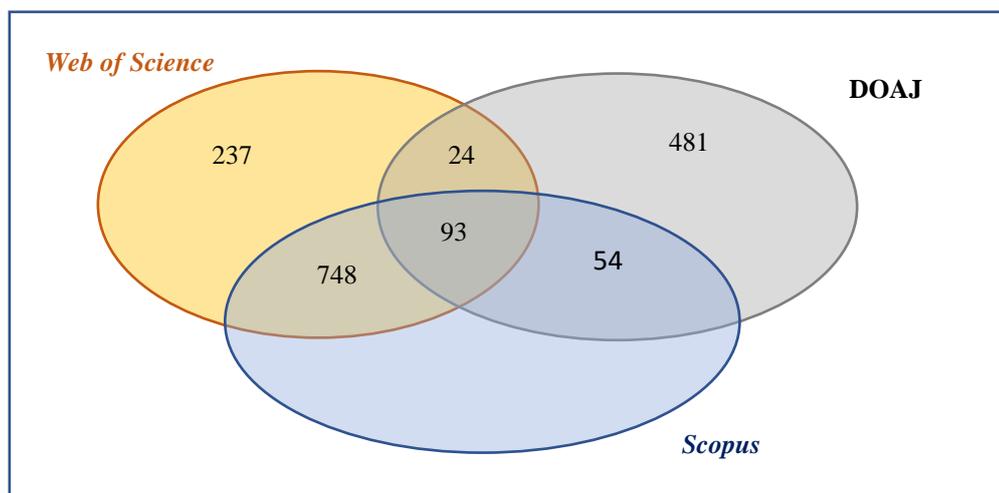
Fonte: Elaborado pelo autor (2020).

Após limitar os campos de classificação das bases de dados para nove, principalmente para aproximar os resultados com os obtidos no DOAJ, fez-se também uma outra planilha em que as pesquisas na *Web of Science* e *Scopus* ficaram com 12 campos, isto é, foram incluídos os campos: “Tipo de Documento”, “Categoria da Publicação” e “Palavra-chave da base de dados”. Assim, as células de cada base de dados receberam uma cor específica e foram organizadas em ordem alfabética pelo título com a finalidade de marcar e facilitar a identificação de duplicidades, para confrontar os dados e, em seguida, classificar e analisar de forma mais detalhada.

Desta forma, de um total de 3.351 documentos registrados, sendo 1.117 da *Web of Science* (de 1970 a 2018), 1.548 da *Scopus* (de 1979 a 2018) e 686 do DOAJ (de 2006 a 2018), identificaram-se as repetições para marcar e comparar a interseção na classificação entre as três bases de dados.

Foram encontrados 93 documentos nas três bases de dados, enquanto 826 apareceram em, pelo menos, duas (748 na *Web of Science* e *Scopus*, 24 na *Web of Science* e DOAJ e 54 na *Scopus* e DOAJ). Da mesma forma, houve 237 documentos encontrados exclusivamente na *Web of Science*, 617 na *Scopus* e 481 no DOAJ. Assim, a reunião do conjunto de documentos encontrados sobre *Ciência Aberta* nestas três bases de dados e sem duplicidades foi de 2.254 documentos, conforme representado na Figura 8.

Figura 8 – Conjunto de documentos sobre *Ciência Aberta* com a expressão “*open science*” encontrados na *Web of Science*, *Scopus* e DOAJ sem duplicidades (1970-2018).



617

 $\Sigma = 2.254$ 

Fonte: Elaborado pelo autor (2020).

Da mesma forma, registraram-se duplicidades nas próprias bases, provavelmente por motivo de recadastramento dos próprios documentos. Em alguns casos, houve um novo cadastro em outro idioma, em outros, foi incluída errata, revisão ou reinserção de novos dados. Porém, em todos estes casos, foi possível identificar essas duplicidades pelo número do DOI.

No entanto, há que se destacar uma recorrente duplicidade no DOAJ, provavelmente para incluir outras instituições de autores envolvidos nos artigos. Uma vez que tais recadastramentos mantiveram o título, a ordem na sequência dos autores e os demais campos de classificação dos documentos, diferindo justamente nas informações sobre as instituições dos pesquisadores, há que se acreditar que tais casos de duplicidade foram uma solução paliativa para contornar algumas insuficiências do processo de cadastramento de artigos na base de dados. Como o DOAJ não permite cadastrar duas ou mais instituições de um mesmo autor, é provável que alguns pesquisadores ou instituições preferiram recadastrar artigos para evidenciar vínculos institucionais. Este resultado pode ser verificado no Quadro 4.

Quadro 4 – Total de documentos sobre Ciência Aberta com a expressão “*open science*” encontrados na *Web of Science*, *Scopus* e DOAJ com duplicidades (1970-2018).

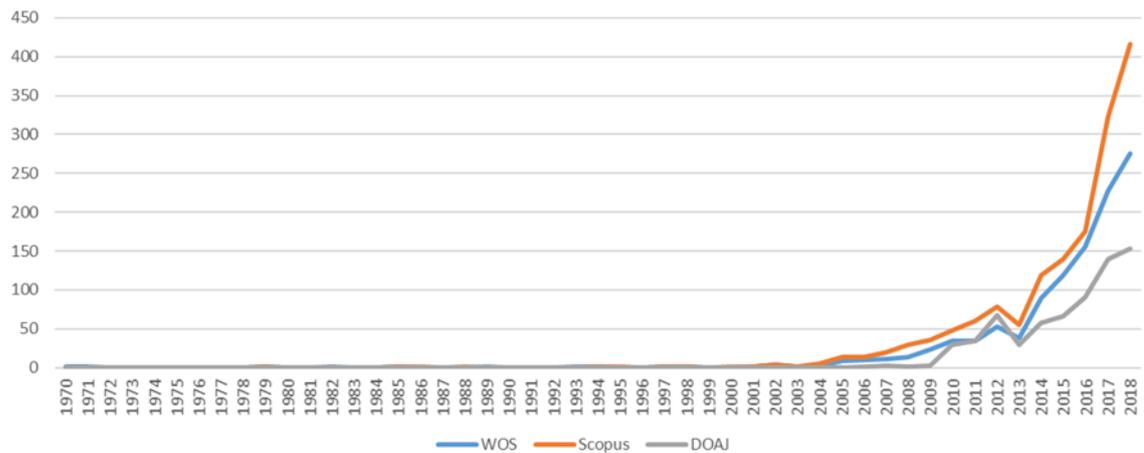
	<b>Web of Science</b>		<b>Scopus</b>		<b>DOAJ</b>		
	Total	%	Total	%	Total	%	
Documentos exclusivos da base	237	21%	617	40%	481	70%	
Duplicidades na própria base	15	1%	36	2%	34	5%	
Com uma outra base	772	69%	802	52%	78	11%	
Com as duas outras bases	93	8%	93	6%	93	14%	
<b>Total</b>	<b>1.117</b>	<b>33%</b>	<b>1.548</b>	<b>46%</b>	<b>686</b>	<b>20%</b>	<b>3.351</b>

Fonte: Elaborado pelo autor (2020).

Encerrando a fase de levantamento, o recorte da pesquisa dos documentos com a expressão “*open science*” nas três bases de dados englobou o período entre 1970 e 2018, conforme apresentado na Figura 9. Verifica-se certa discrepância na recuperação neste recorte

pois o resultado da *Web of Science* inicia em 1970, enquanto a *Scopus* é a partir de 1979 e a DOAJ somente em 2006.

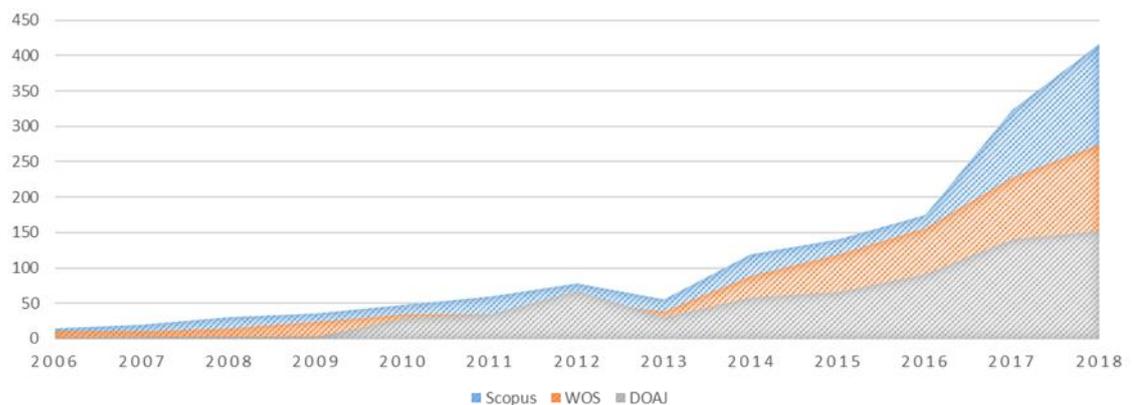
Figura 9 – Evolução das publicações sobre “open science” nas bases de dados *Web of Science*, *Scopus* e *DOAJ* no período de 1970-2018.



Fonte: Elaborado pelo autor (2020).

Assim, tendo em vista o aumento significativo da indexação de documentos com a expressão “open science” justamente a partir de 2006, ano em que aparecem os primeiros artigos indexados no DOAJ, detalhou-se na Figura 10 o crescente aumento de publicações no período de 2006 a 2018:

Figura 10 – Publicações sobre Ciência Aberta nas bases de dados *Web of Science* (WOS), *Scopus* e *DOAJ* no período de 2006-2018.

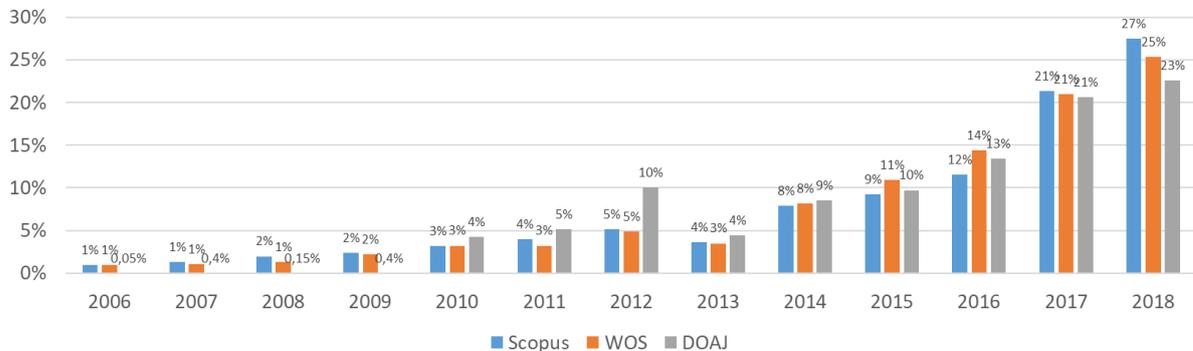


Fonte: Elaborado pelo autor (2020).

Analisando o conjunto de documentos por base de dados, anualmente, conforme apresentado na Figura 11, verifica-se que houve um gradativo aumento de documentos com a

expressão “*open science*” indexados no mesmo período entre 2006 e 2018 que se intensifica progressivamente.

Figura 11 – Publicações sobre Ciência Aberta nas bases de dados *Web of Science* (WOS), *Scopus* e DOAJ no período de 2006-2018, especificando o volume de documentos cronologicamente



Fonte: Elaborado pelo autor (2020).

Analisando o resultado das três bases de dados anteriores, é possível observar ainda um sensível decréscimo no número de publicações em 2013 e uma retomada nos anos seguintes. Porém, há padrões que cabem ser ressaltados, como a equiparação percentual entre a *Scopus* e a *Web of Science* nos anos 2006, 2007, 2009, 2010, 2012, 2014 e 2017. Como ambas bases de dados são restritas e com viés mercadológico, essa regularidade demonstra uma maior tendência para um cotejamento entre elas.

Inclusive, seguindo este exame, há uma propensão estatística para que essas duas bases de dados se sobreponham ao percentual do DOAJ. No entanto, cabe também analisar o destaque do maior percentual de publicação dessa última nos anos de 2010, 2011, 2012, 2013 e 2014. É certo que as bases de dados com viés comercial têm um maior repertório de recursos para incluir documentos em seu acervo, daí a importância de destacar os anos com maiores porcentagens do DOAJ. Uma vez que são os próprios autores ou instituições de pesquisa que registram seus artigos nesse acervo, há um esforço de divulgação por parte da comunidade científica nessa inclusão de documentos.

#### 4.3 RELAÇÕES E ESPECIFICIDADES DA ANÁLISE SOBRE CIÊNCIA ABERTA

Os metadados coletados em cada base de dados não bastam para avaliar as diversas classificações registradas para a caracterização de cada documento. Assim, após os



Notam-se certas particularidades quanto ao levantamento do endereço de publicações em cada uma das bases de dados. A Tabela 1 auxilia na análise dessas particularidades, como, por exemplo, o DOAJ ter aceitado um enorme percentual de edições publicadas na Índia, ou a forte presença de editoras da Alemanha e Holanda na *Scopus* e a grande representatividade da França na *Web of Science*. Além disso, a análise exclusiva de cada base de dados também permite destacar concomitâncias, como as publicações nos Estados Unidos e Reino Unido, bem como Suíça e Itália.

No entanto, cabe ressaltar que, principalmente no caso do maior volume de editoras nos Estados Unidos e no Reino Unido, grande parte dessas são, na realidade, o reflexo da tendência mercadológica do modelo de negócios das editoras científicas. Muitas dessas editoras têm sede naqueles países e em outros da Europa, mas não publicam necessariamente artigos de autores dos mesmos países. O volume de publicações também retrata o interesse de tais editoras em se manter presentes no mercado de publicações de acesso aberto.

Tabela 1 – Países de editoras que publicaram documentos com a expressão “open science” na *Web of Science*, *Scopus* e DOAJ, com dupla contagem (de 1970 a 2018).

(continua)

Países		<i>Web of Science</i>		<i>Scopus</i>		DOAJ		TOTAL	
<b>AMÉRICA</b>									
1º	Estados Unidos	465	47%	476	48%	53	5%	<b>994</b>	33%
9º	Brasil	10	18%	6	11%	41	72%	<b>57</b>	2%
13º	Canadá	18	64%	4	14%	6	21%	<b>28</b>	1%
22º	Colômbia	0	0%	3	43%	4	57%	<b>7</b>	0%
26º	Cuba	4	80%	0	0%	1	20%	<b>5</b>	0%
27º	Argentina	1	25%	0	0%	3	75%	<b>4</b>	0%
41º	México	0	0%	1	100%	0	0%	<b>1</b>	0%
<b>EUROPA</b>									
2º	Reino Unido	266	38%	352	50%	81	12%	<b>699</b>	23%
4º	Alemanha	49	28%	105	59%	24	13%	<b>178</b>	6%
5º	Suíça	75	43%	50	29%	48	28%	<b>173</b>	6%
6º	Holanda	2	1%	142	91%	12	8%	<b>156</b>	5%
7º	França	88	82%	7	7%	12	11%	<b>107</b>	4%
8º	Espanha	41	44%	19	20%	33	35%	<b>93</b>	3%
10º	Itália	22	39%	23	40%	12	21%	<b>57</b>	2%
11º	Áustria	0	0%	15	33%	30	67%	<b>45</b>	1%
12º	Bulgária	4	14%	5	18%	19	68%	<b>28</b>	1%
14º	Finlândia	0	0%	19	70%	8	30%	<b>27</b>	1%
16º	Bélgica	6	55%	4	36%	1	9%	<b>11</b>	0%
17º	Romênia	6	55%	1	9%	4	36%	<b>11</b>	0%
18º	Polónia	4	44%	0	0%	5	56%	<b>9</b>	0%
19º	Rússia	4	44%	2	22%	3	33%	<b>9</b>	0%
20º	Ucrânia	6	67%	0	0%	3	33%	<b>9</b>	0%
30º	Hungria	0	0%	3	100%	0	0%	<b>3</b>	0%
31º	Irlanda	2	67%	1	33%	0	0%	<b>3</b>	0%
33º	Noruega	0	0%	1	33%	2	67%	<b>3</b>	0%

Tabela 2 – Países de editoras que publicaram documentos com a expressão “open science” na *Web of Science*, *Scopus* e *DOAJ*, com dupla contagem (de 1970 a 2018).

(conclusão)

EUROPA									
34°	Portugal	0	0%	2	67%	1	33%	3	0%
36°	Dinamarca	0	0%	1	50%	1	50%	2	0%
37°	Eslováquia	1	50%	1	50%	0	0%	2	0%
38°	República Tcheca	0	0%	1	50%	1	50%	2	0%
40°	Escócia	0	0%	1	100%	0	0%	1	0%
ÁSIA									
3°	Índia	7	3%	5	2%	225	95%	237	8%
15°	Japão	2	13%	13	87%	0	0%	15	0%
21°	China	3	43%	0	0%	4	57%	7	0%
23°	Singapura	6	100%	0	0%	0	0%	6	0%
24°	Turquia	2	33%	1	17%	3	50%	6	0%
25°	Coreia do Sul	3	60%	1	20%	1	20%	5	0%
32°	Israel	2	67%	0	0%	1	33%	3	0%
OCEANIA									
28°	Austrália	0	0%	4	100%	0	0%	4	0%
29°	Indonésia	2	50%	0	0%	2	50%	4	0%
ÁFRICA									
35°	África do Sul	0	0%	2	100%	0	0%	2	0%
39°	Egito	0	0%	0	0%	1	100%	1	0%
<b>TOTAL</b>		<b>1101</b>	<b>36%</b>	<b>1271</b>	<b>42%</b>	<b>645</b>	<b>21%</b>	<b>3017</b>	<b>100%</b>

Fonte: Elaborado pelo autor (2020).

Alguns pontos a serem destacados na análise do local de publicação, com exclusividade de cada base de dados:

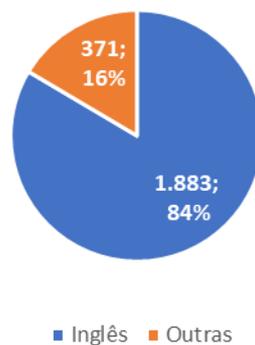
- A Tabela 1 contém as duplicidades excluídas na análise anterior, porém permite visualizar melhor os resultados exclusivos de cada base de dados. Estados Unidos e Reino Unido lideram a lista, com mais de 50% do total de publicações. No entanto, a Índia é o terceiro país com mais publicações de artigos com o termo Ciência Aberta – e praticamente 95% dos documentos analisados foram registrados no DOAJ. Muitas dessas publicações não são necessariamente de autores nativos da Índia, porém o volume de publicações sobre Ciência Aberta se destaca por existirem muitas editoras comerciais sediadas lá, bem como pela proatividade por parte de editores e autores para cadastrar e promover as publicações em plataformas de acesso aberto.
- Cabe destacar também o papel do Brasil, segundo país da América e nono no *ranking* geral. Do exame nas três bases de dados, o país é o único da América Latina que figura entre os dez primeiros, ficando inclusive na frente do Canadá. Além disso, 72% de seus documentos foram encontrados no DOAJ – o que, no caso, corrobora com a importância dessa base de dados para divulgação da pesquisa científica nacional.

- Com um viés mais mercadológico, além do destaque para publicações nos Estados Unidos e Reino Unido, a *Web of Science* e a *Scopus* têm um maior percentual de documentos da Europa Ocidental. No caso, Suíça, França e Espanha, para a primeira, e Alemanha, Holanda e Itália, para a segunda, tem mais de 40% dos registros em uma dessas plataformas.

### 4.3.2 Idioma

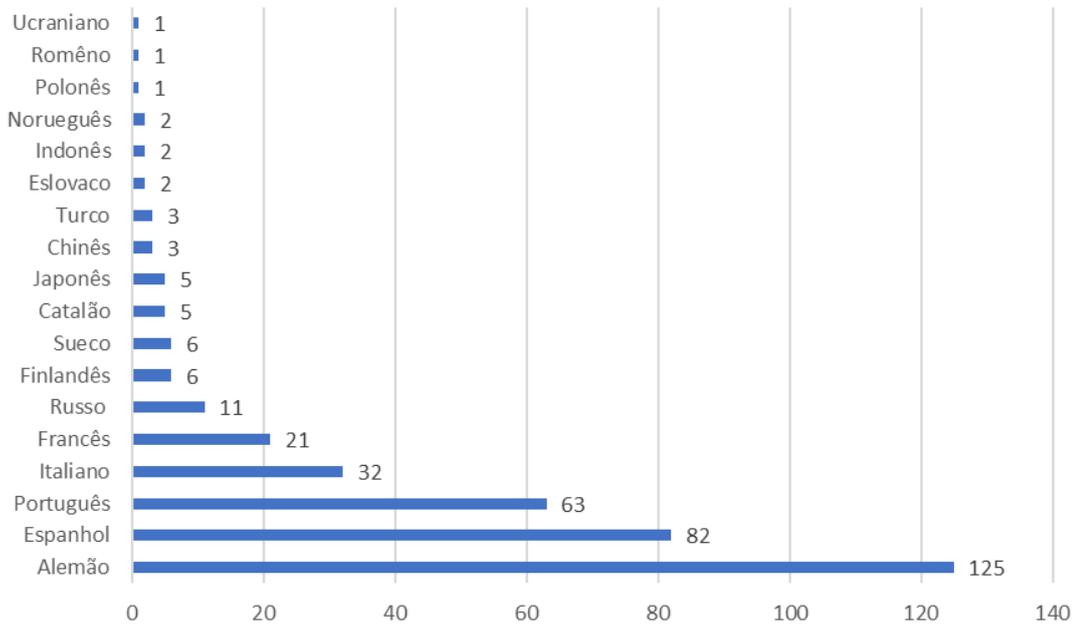
Quanto ao idioma, foi possível encontrar informações sobre os 2.254 registros de documentos sobre “*open science*” recuperados, priorizando sempre o primeiro idioma escolhido pelos autores. A Figura 13 apresenta os resultados encontrados, em que se verifica a preponderância com 84% das publicações do idioma inglês nas três bases, *Web of Science*, *Scopus* e DOAJ. A Figura 14 descreve os principais idiomas que se destacaram em outras línguas nestas bases de dados.

Figura 13 – Publicações sobre “*open science*” na língua inglesa nas bases *Web of Science*, *Scopus* e DOAJ de 1970 a 2018.



Fonte: Elaborado pelo autor (2020).

Figura 14 – Publicações sobre “*open science*” em idiomas diferentes da língua inglesa nas bases *Web of Science*, *Scopus* e *DOAJ* de 1970 a 2018.



Fonte: Elaborado pelo autor (2020).

A Tabela 2 apresenta, por base de dados, o quantitativo de publicações por países, onde se verifica que Estados Unidos, Reino Unido e Índia estão entre os três primeiros locais em que tais documentos foram publicados. Acredita-se que, para atender os ideais de internacionalizações, Suíça, Holanda e França, respectivamente, quinto, sexto e sétimo lugar no *ranking* de locais de publicação, tenham preferido publicar no idioma inglês para atender os padrões de internacionalização de suas pesquisas científicas. No entanto, pode-se cogitar que Alemanha, Espanha e Brasil – respectivamente, quarto, oitavo e nono lugares no *ranking* de locais de publicação – têm resistido aos padrões de internacionalização do idioma inglês, ou publicado em seus próprios idiomas como política de estímulo à produção científica nacional, uma vez que publicações no idioma alemão, espanhol e português estão em segunda, terceira e quarta posições no *ranking* de idiomas de publicação.

Tabela 3 – Idiomas de editoras que publicaram documentos com a expressão “open science” na *Web of Science*, *Scopus* e *DOAJ* (de 1970 a 2018).

Idioma	Web of Science		Scopus		DOAJ		Total	
	Total	%	Total	%	Total	%	Total	%
Inglês	995	90%	1469	95%	512	75%	2976	89%
Alemão	40	4%	33	2%	52	8%	125	4%
Espanhol	35	3%	16	1%	31	5%	82	2%
Português	10	1%	5	0%	48	7%	63	2%
Italiano	14	1%	9	1%	9	1%	32	1%
Francês	2	0%	8	1%	11	2%	21	1%
Russo	4	0%	2	0%	5	1%	11	0%
Catalão	4	0%	0	0%	2	0%	6	0%
Sueco	0	0%	0	0%	6	1%	6	0%
Japonês	2	0%	3	0%	0	0%	5	0%
Turco	2	0%	0	0%	1	0%	3	0%
Chinês	0	0%	1	0%	2	0%	3	0%
Eslovaquo	1	0%	1	0%	0	0%	2	0%
Norueguês	0	0%	1	0%	1	0%	2	0%
Indonês	0	0%	0	0%	2	0%	2	0%
Polonês	0	0%	0	0%	1	0%	1	0%
Romêno	0	0%	0	0%	1	0%	1	0%
<b>Total</b>	<b>1109</b>	<b>100%</b>	<b>1548</b>	<b>100%</b>	<b>684</b>	<b>100%</b>	<b>3341</b>	<b>100%</b>

Fonte: Elaborado pelo autor (2020).

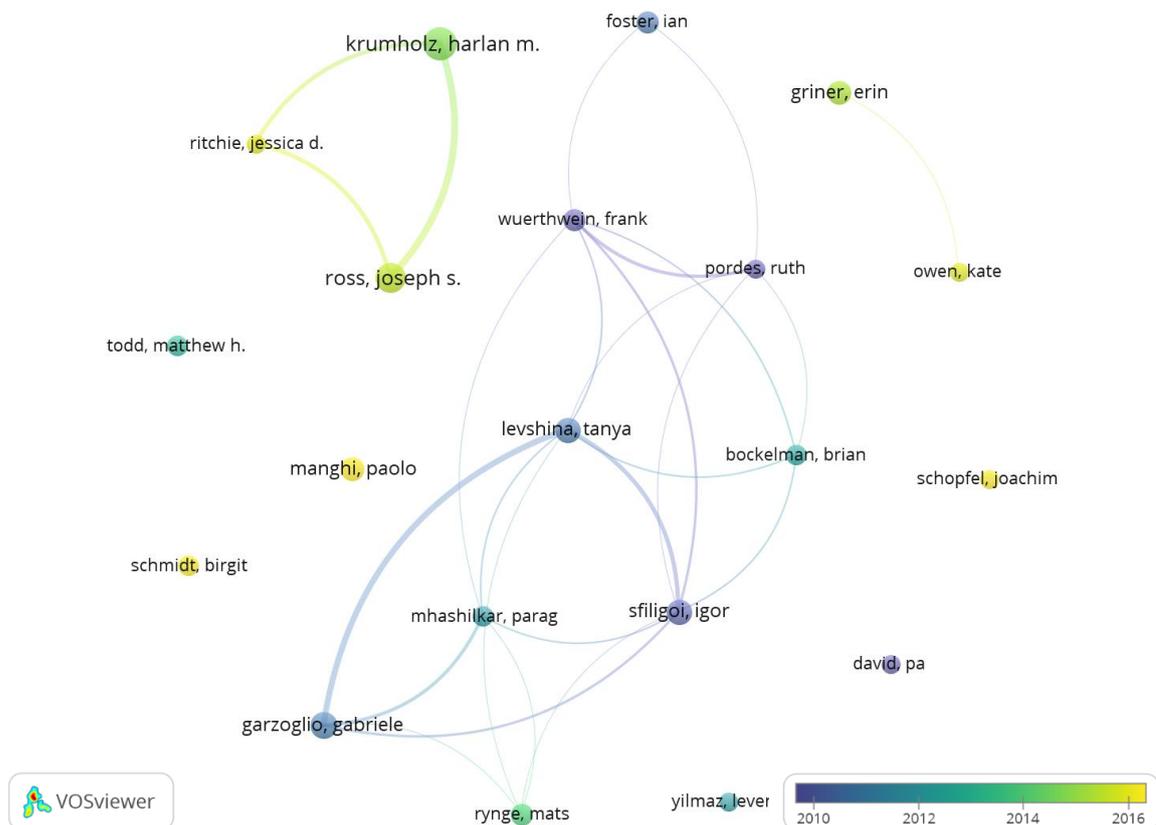
Essa concentração de publicações em língua inglesa transcorre devido aos critérios de indexação das bases de dados utilizadas, preferindo o idioma como padronização para suas fontes de informação. A internacionalização e a uniformização das publicações em idioma inglês têm a vantagem de reforçar a importância do vocabulário utilizado para moldar a produção sobre Ciência Aberta, tanto nos títulos e resumos, quanto nas palavras-chave. Tal reflexão sobre a preponderância do idioma nos metadados coletados neste estudo ajuda a reforçar a relevância destes para definir melhor a autoridade do idioma sobre o Campo Científico da Ciência Aberta, bem como as relações entre os seus agentes.

### 4.3.3 Autores

Com os dados coletados sobre documentos com o termo Ciência Aberta nas plataformas *Web of Science* e *Scopus*, foi possível elaborar representações gráficas com o VOSviewer, refletindo as redes complexas que envolvem os autores. Com o objetivo de demonstrar as ligações entre o universo de autoridades, tal levantamento também foi arquitetado por meio das planilhas de resultados coletados.

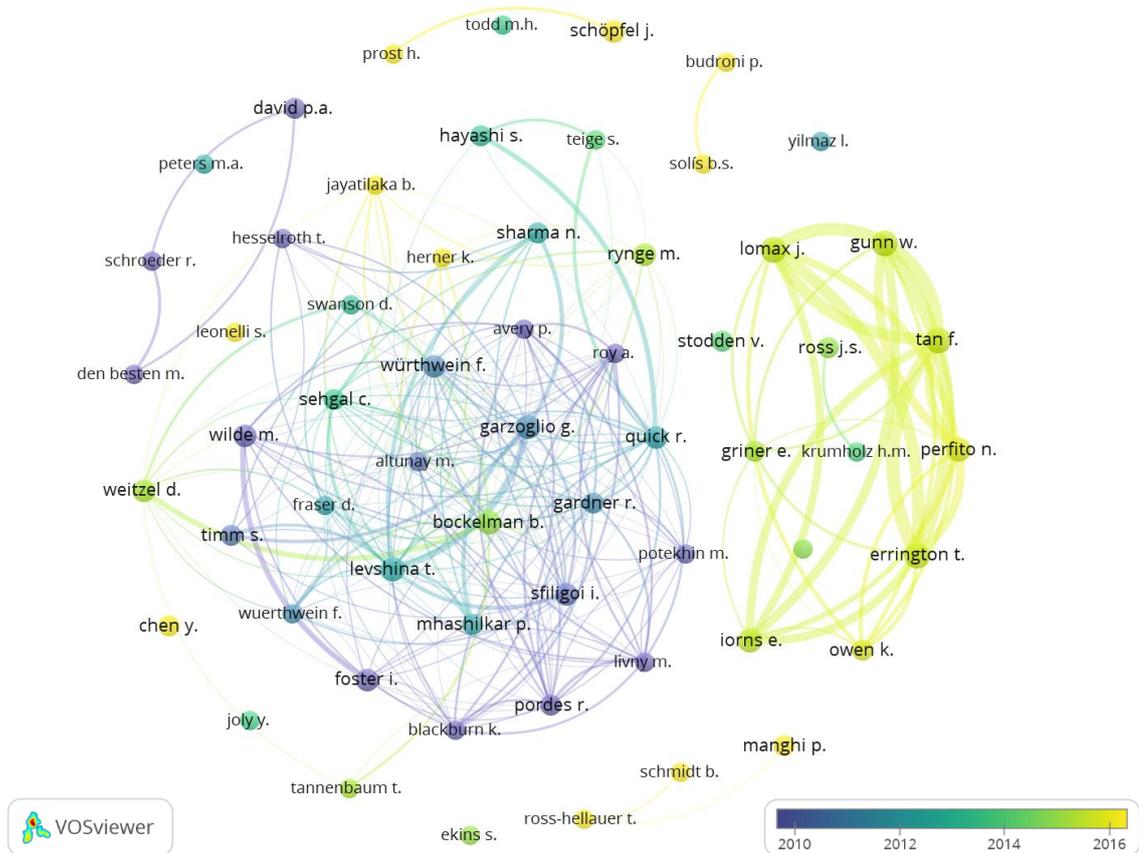
Para indicar as correlações mais distinguíveis, optou-se por iniciar o mapeamento das relações entre autores com o programa. Os mapas criados, conforme demonstrados nas Figuras 15 e 16, refletem os *clusters* e as ligações entre os autores, que são representados pelos nós exibidos em cada modelo. Da mesma forma que foi apresentado nos resultados anteriores sobre os locais de publicação, os dados agregados refletem o quão recente cada artigo indicado foi publicado.

Figura 15 – As relações entre autores de documentos com a expressão “open science” na *Web of Science*, apresentadas pela cronologia da publicação (de 1970 a 2018).



Fonte: Elaborado pelo autor (2020).

Figura 16 – As relações entre autores de documentos com a expressão “*open science*” na *Scopus*, apresentadas pela cronologia da publicação (de 1979 a 2018).



Fonte: Elaborado pelo autor (2020).

A Figura 15 foi formada através dos dados extraídos da *Web of Science*, parâmetros de tipo de análise pela coautoria (“*Co-authorship*”), com a unidade de análise aplicada aos autores (“*Authors*”) e o método de contagem como completa (“*Full counting*”), que acrescenta o mesmo peso para coautoria, coocorrência, reunião bibliográfica ou cocitação. Para tornar a representação gráfica mais detalhada e permitir uma apresentação legível dos nós e das arestas, optou-se por exibir autores com, no mínimo, cinco publicações, o que, em um universo de 3.641, ficaram 20. Da mesma forma, como a amostragem trouxe um extenso período de anos, diminuir o volume de autores foi fundamental para deixar a figura com uma boa visualização razoável dos autores que mais despontaram na contagem do VOSviewer. Destes, nesta representação gráfica, formou-se nove agrupamentos pela modularidade das 265 conexões entre eles.

Cabe ainda ressaltar que essa análise pelo VOSviewer foi realizada com cunho exploratório e que não foi utilizado recurso do Tesouro (“*VOSviewer Thesaurus File*”), de forma que todos os nós representados são exatamente os mesmos que o programa destacou ao analisar os arquivos com a pesquisa extraída das bases de dados *Web of Science* e *Scopus*.

Ambas representações gráficas receberam o filtro “*Overlay Visualization*” do VOSviewer, destacando a atividade dos autores pela média da cronologia de suas publicações ao longo dos anos. Assim, a Figura 15 mostra ainda um *cluster* central com autores que em média tem publicações mais antigas, orbitado por outros pequenos grupamentos desconectados desse. No entanto, a maior parte dos nós fora dessa parte central tem a cor amarela, indicando uma média de publicação ao longo dos anos mais contemporânea na base de dados *Web of Science*.

Quanto à Figura 16, feita com os dados extraídos da *Scopus*, utilizou-se os mesmos parâmetros de tipo de análise pela coautoria e a unidade de análise aplicada aos autores, bem como a configuração de, no mínimo, cinco publicações por autor. Assim, de um total de 4.535, restaram 60, com 16 agrupamentos e 250 ligações entre eles. Nessa representação gráfica, nota-se um *cluster* maior com nós de cores variadas, indicando autores com trabalhos publicados em uma média cronológica mais antiga e outros com uma média de publicações mais recentes ao longo dos anos, e outro grupamento só com nós de cor clara, demonstrando a média de publicações mais contemporâneas desse ajuntamento. Há também outros pequenos *clusters* orbitando ambos os grupos, variando de cor, mas tendendo para os de cor mais clara.

Como tal relação não daria a dimensão do total de publicações de cada autor no repertório de dados coletados, optou-se por calcular o resultado desse montante em cada base de dados escolhida e, em seguida, somá-los. Diante de um total de 8.242 coautores, com 4.217 publicando mais de uma vez, o resultado foi a Tabela 3:

Tabela 3 – Os 20 autores que mais publicaram documentos com a expressão “open science” na *Web of Science* (WOS), *Scopus* e DOAJ (de 1970 a 2018).

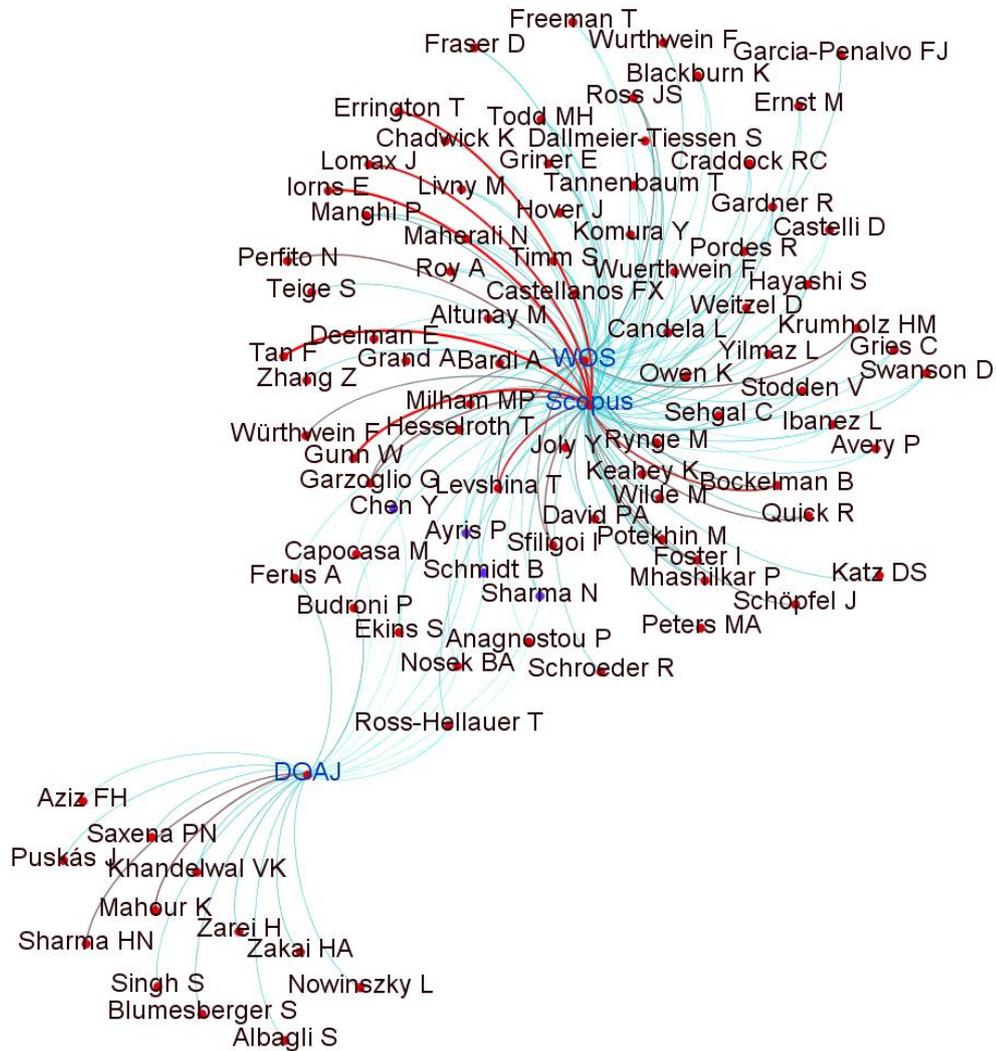
<b>Autores (TOTAL)</b>	<b>Frequência</b>	<b>Autores da WOS</b>	<b>Freq.</b>	<b>Autores da Scopus</b>	<b>Freq.</b>	<b>Autores do DOAJ</b>	<b>Freq.</b>
Levshina T	32	Krumholz HM	14	Tan F	24	Mahour K	15
Bockelman B	31	Levshina T	13	Gunn W	23	Sharma HN	14
Garzoglio G	27	Sfiligoi I	13	Iorns E	23	Budroni P	8
Sfiligoi I	27	Bockelman B	12	Errington T	22	Blumesberger S	7
Tan F	24	Garzoglio G	12	Lomax J	21	Ferus A	7
Foster I	23	Ross JS	12	Bockelman B	19	Zakai HA	7
Gunn W	23	Foster I	9	Levshina T	19	Zarei H	7
Iorns E	23	Griner E	8	Garzoglio G	15	Albagli S	6
Errington T	22	Manghi P	8	Perfito N	15	Khandelwal VK	6
Lomax J	21	Ryng M	8	Foster I	14	Nowinszky L	6
Quick R	21	Owen K	7	Quick R	14	Puskás J	6
Ross JS	21	Quick R	7	Sfiligoi I	14	Saxena PN	6
Krumholz HM	20	Timm S	7	Würthwein F	12	Singh S	6
Ryng M	20	Würthwein F	7	Wilde M	11	Aziz FH	5
Manghi P	17	Garcia-Penalvo FJ	6	Ryng M	10	Bauer B	5
Griner E	16	Mhashilkar P	6	Schöpfel J	10	Clinio A	5
Schmidt B	16	Milham MP	6	Weitzel D	10	Javahery S	5
Wilde M	16	Pordes R	6	David PA	9	Kumar A	5
Mahour K	15	Schmidt B	6	Manghi P	9	Rai AK	5
Mhashilkar P	15	Todd MH	6	Mhashilkar P	9	Singh J	5

Fonte: Elaborado pelo autor (2020).

Com os autores da *Web of Science* e *Scopus* se destacam pelo volume de participações em coautoria, ambas as bases de dados exibiram sete dos 20 autores que mais publicaram documentos com o termo Ciência Aberta. Da mesma forma, esse *ranking* apresentou uma tímida participação de autores do DOAJ no levantamento – com somente dois figurando na contagem total.

Para representar melhor a relação entre o volume de publicação dos autores desse *ranking* em cada base de dados e a relação deles pelo número de coautorias, optou-se por utilizar esse levantamento dos autores que mais publicaram e fazer uma representação gráfica com o programa Gephi (Figura 17):

Figura 17 – As relações entre autores de documentos com a expressão “open science” na *Web of Science* (WOS), *Scopus* e DOAJ (de 1970 a 2018).



Fonte: Elaborado pelo autor (2020).

Essa representação gráfica foi elaborada por meio de duas planilhas, uma indicando os nós e outra as arestas. Na primeira, foram inseridos os autores e o número de vezes em que cada um teve trabalhos publicados na *Web of Science*, *Scopus* e DOAJ, sendo que, desses, somente quatro apareceram nas três bases de dados (nós de cor azul na Figura 17). Quanto à planilha de arestas, inseriu-se o número de publicação por base de dados – representando na cor azul claro os com menos artigos publicados, e na cor vermelha os com mais.

As duas planilhas foram importadas para a ferramenta “Laboratório de dados” no Gephi, para formar a figura que foi trabalhada na aba “Visão Geral”. Lá, modificou-se sua

representação gráfica com as ferramentas de distribuição “*Force Atlas 2*”, “*Contração*”, “*Expansão*”, “*Ajustar Rótulos*” e “*Rotate*”, enquanto sua aparência recebeu as alterações de cor nas ligações das arestas, na fonte de autores e plataformas *Web of Science*, *Scopus* e *DOAJ* e na representação dos nós de autores com publicações nas três bases de dados.

Ao ponderar o que foi apurado pela coleta de dados, é possível estabelecer uma conexão entre as Figuras 15, 16 e 17 com a Tabela 3 para estabelecer dois agrupamentos diferentes de autores na *Scopus* e na *Web of Science*. Enquanto nessa base de dados há um *cluster* que publica a mais tempo com ligações entre “Levshina”, “Garzoglio” e “Foster” (2º, 5º e 7º, respetivamente, lugares no *ranking* de autores com mais publicações nesta base de dados), há outro grupamento menor, com publicações mais recentes, formado por “Krumholz” e “Ross” (1º e 6º do mesmo *ranking*).

Na *Scopus*, há um grande grupo que, em sua maioria, tem publicações mais antigas nessa base dados, contando com “Bockelman”, “Levshina”, “Garzoglio”, “Foster” e “Quick” e que, no *ranking* de autores mais publicados nessa base de dados, consta nos 6º, 7º, 8º, 10º e 11º lugares, na devida ordem. Da mesma forma, o *cluster* com autores que têm trabalhos mais recentes destaca somente dois nomes nesta hierarquia: “Gunn” e “Perfito” – respetivamente 2º e 9º lugares no *ranking*.

Apesar de não haver suporte para os resultados do *DOAJ* no *VOSviewer*, o impacto da publicação dos autores dessa base de dados não alcançou grande repercussão. Somente o primeiro lugar no *ranking* do *DOAJ*, “Mahour”, alcançou um posicionamento entre os 20 autores que mais publicaram das três plataformas. No entanto, o impacto maior do levantamento geral é determinado pelas interseções entre a *Web of Science* e *Scopus*. Ambas compartilham 10 autores entre os mais citados, e todos esses se encontram no *ranking* geral.

Ao analisar os *clusters*, a coincidência de dois grupamentos entre essas bases de dados, um com publicações mais antigas e outro com mais recentes, reflete também a relação da comunidade do Campo Científico da Ciência Aberta. Como os grupamentos com publicações em média pretéritas são maiores e contêm interseções de agentes, pode-se dizer que refletem o ponto de vista dominante na perspectiva do campo sobre as questões científicas. Da mesma forma, como o volume de publicações sobre Ciência Aberta tem aumentado, detecta-se uma tendência com o número de publicações dos autores, mas não uma distribuição equânime entre eles.

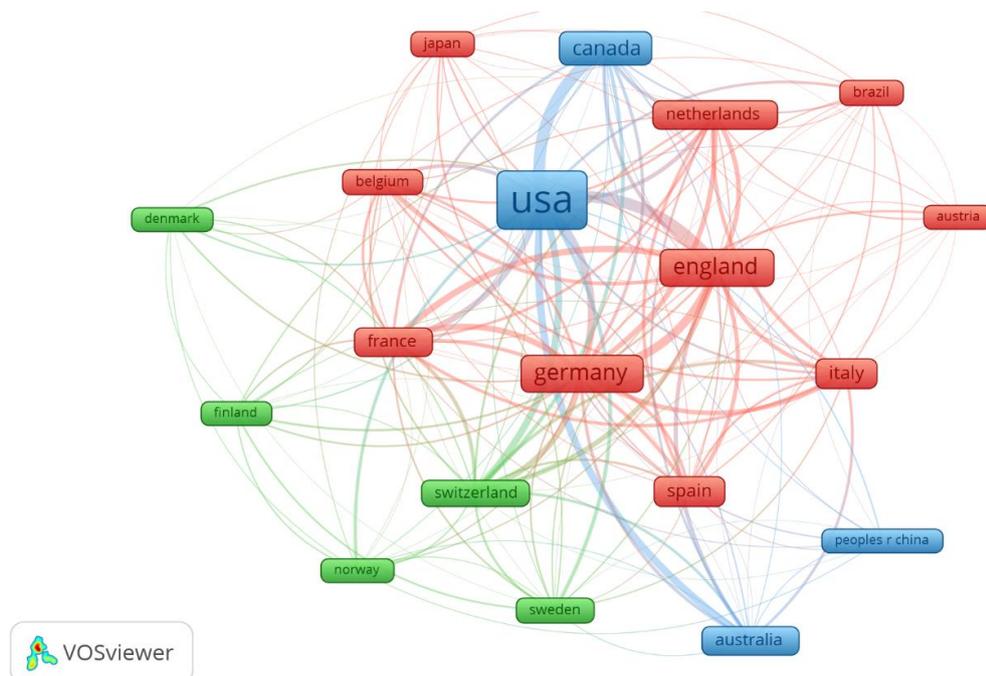
Quanto aos *clusters* da *Scopus* com publicações mais recentes, e sem ligações com autores que publicam a mais tempo, pode ser que represente ramificações, desdobramentos e

aprofundamentos da área de pesquisa em Ciência Aberta. Esse novo grupamento pode indicar uma vertente, como também pode retratar a questão das “revoluções” no campo científico, abordado por Bourdieu e apresentado anteriormente.

#### 4.3.3.1 Nacionalidade de coautores

O VOSviewer também conta com um recurso para visualizar a nacionalidade dos coautores no conjunto de documentos analisados. Com os parâmetros de tipo de análise pela coautoria (“*Co-authorship*”), com a unidade de análise aplicada aos países (“*Countries*”) e o método de contagem como completa (“*Full counting*”), foi possível traçar representações gráficas da nacionalidade do conjunto de autores na *Web of Science* e na *Scopus* – Figuras 18 e 19.

Figura 18 – Nacionalidades de coautores de documentos sobre Ciência Aberta no levantamento feito pela *Web of Science* (1970-2018)

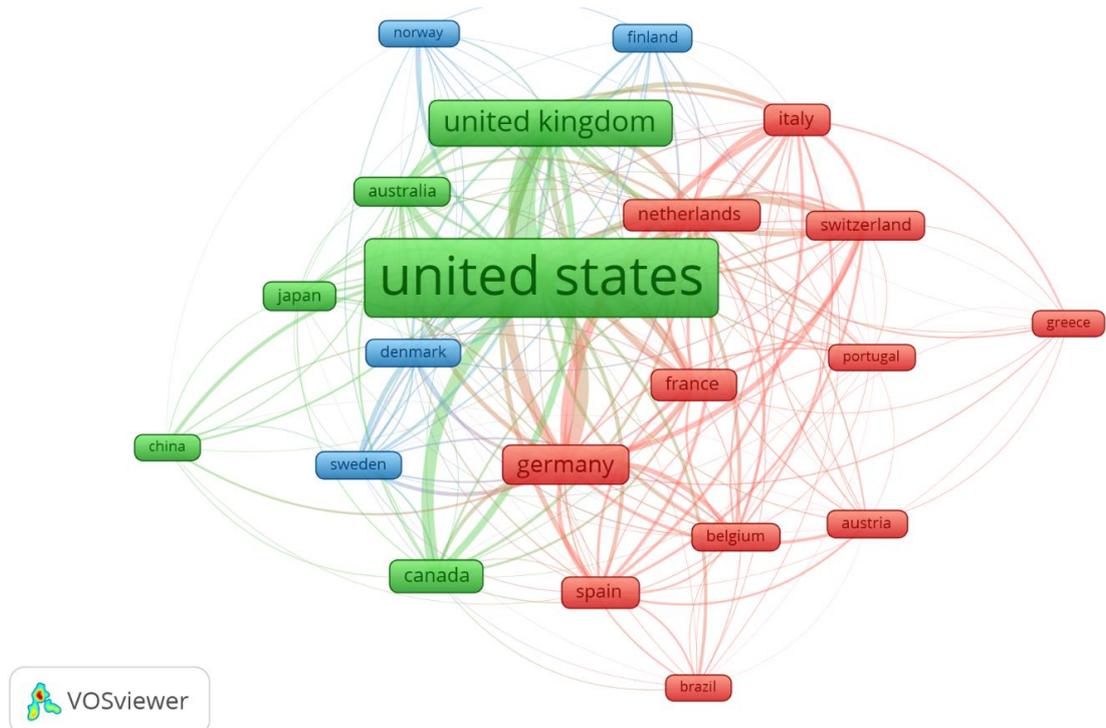


Fonte: Elaborado pelo autor (2020).

Assim como anteriormente, o VOSviewer foi utilizado como um recurso para apresentar um substrato da pesquisa exploratório, sendo necessário configurar a análise do

programa para construir a representação visual em uma estrutura razoável para um exame superficial. Assim, a Figura 18 foi elaborada com a limitação de pelo menos 10 citações de autores de cada país – formando a representação gráfica com 19 nós, em um total de 71 países identificados. Da mesma forma, apresentou-se três *clusters* que indicam 150 ligações entre os países de coautores identificados.

Figura 19 – Nacionalidades de coautores de documentos sobre Ciência Aberta no levantamento feito pela *Scopus* (1979-2018)\*



\* O nó “United Kingdom” nesta figura será comparado com o nó “England” da imagem anterior, pois, a título de classificação, a *Scopus* não utiliza o mesmo padrão que a *Web of Science*, deixando esta discrepância na análise entre as duas bases de dados.

Fonte: Elaborado pelo autor (2020).

Quanto à Figura 19, foi necessário aumentar a limitação de citações de autores por cada país por conta do volume de documentos na *Scopus* e da necessidade de tornar a visualização mais razoável nesta análise exploratória pelo VOSviewer. Assim, optou-se pela configuração de 20 citações, formando a representação gráfica com 21 nós, em um total de 177 países identificados do acervo. O resultado apresentou também três *clusters* com 188 ligações entre os países de coautores identificados.

Em uma análise mais detalhada do agrupamento da modularidade identificada pelo VOSviewer, o Quadro 5 apresenta os *clusters* de cada figura, representando as possíveis correlações que podem ser inferidas coocorrências observadas. Já a Tabela 4 apresenta os cinco países com mais documentos identificados pelo software de representações gráficas, com os dados extraídos da tabela do assistente criação de mapas de redes do VOSviewer.

Quadro 5 – *Clusters* de países de coautores identificado no VOSviewer (Figura 18 e 19), separados por cor.\*

Web of Science	Scopus
Canadá	Canadá
Estados Unidos	Estados Unidos
Austrália	Austrália
China	China
Dinamarca	Reino Unido
Finlândia	Japão
Noruega	Dinamarca
Suíça	Finlândia
Suécia	Noruega
Japão	Suíça
Holanda	Holanda
Brasil	Brasil
Bélgica	Bélgica
Inglaterra	Áustria
Áustria	França
França	Alemanha
Alemanha	Itália
Itália	Espanha
Espanha	Grécia
	Portugal

\*Como explicitado anteriormente, Reino Unido (representado por “*United Kingdom*” na *Scopus*) e Inglaterra (representado por “*England*” na *Web of Science*”) estão sendo comparados, apesar da diferença da classificação de cada base de dados.

Fonte: Elaborado pelo autor (2020).

Tabela 4 – Ranking com os cinco países com mais documentos sobre Ciência Aberta com mais coautores, segundo o VOSviewer.

Web of Science		Scopus	
País	Nº de Documentos	País	Nº de Documentos
Estados Unidos	181	Estados Unidos	659
Inglaterra	74	Reino Unido	254
Alemanha	74	Alemanha	177
Canadá	56	Canadá	104
Espanha	43	Itália	92

\*Como explicitado anteriormente, Reino Unido (representado por “United Kingdom” na Scopus) e Inglaterra (representado por “England” na Web of Science”) estão sendo comparados, apesar da diferença da classificação de cada base de dados.

Fonte: Elaborado pelo autor (2020).

Enquanto o Quadro 5 apresenta uma grande similaridade nos grupamentos identificados pelo VOSviewer, a Tabela 4 destaca os mesmos quatro países no topo do *ranking* daqueles com maior volume de documentos publicados. No caso, as coautorias levantadas representam as parcerias entre os autores, e essa definição do país de origem do autor contribui para o reconhecimento das relações entre os agentes do Campo Científico da Ciência Aberta.

#### 4.3.4 Palavras-chaves

Todos os documentos indexados pelas bases de dados *Web of Science* e *Scopus* contam com dois tipos de palavras-chave: aquelas categorizadas pelos próprios autores e as indexadas pelas equipes de profissionais da *Clarivate Analytics* e da *Elsevier*, respectivamente. As palavras-chave das bases de dados seguem uma padronização mais rígida, porque são condicionadas por um vocabulário controlado único de cada plataforma. Não obstante, as palavras-chave dos autores podem ter uma grande variedade de termos, pois a padronização destas depende da existência de vocabulários controlados nas revistas periódicas em que os artigos científicos estão submetidos.

Como explicitado anteriormente, o DOAJ está a parte deste formato de classificação, oferecendo um terceiro gênero de palavras-chaves, com uma indexação híbrida que é

determinada pelos editores de periódicos científicos ou autores dos artigos lá registrados, mas restrito às palavras-chaves dentro vocabulário controlado oferecido pela plataforma.

Após a coleta dos metadados da pesquisa de documentos sobre Ciência Aberta nas três bases de dados, apurou-se que, nos 1.117 documentos encontrados na *Web of Science*, do período entre 1970 e 2018, encontrou-se 1.466 palavras-chave da *Clarivate Analytics* e 1.978 palavras-chaves dos próprios autores, conforme descrito na Tabela 5. Já na *Scopus*, dos 1.548 registros, de 1979 a 2018, 4.536 palavras-chaves eram da *Elsevier* e 15.883 indexadas pelos autores. Em derradeiro, o DOAJ, dentre os 686 recuperados na pesquisa, abrangendo a data de corte entre 2006 e 2018, encontrou-se, dentro das limitações explicitadas anteriormente, somente 2.598 palavras-chave.

Tabela 5 – Levantamento de palavras-chave nos documentos sobre Ciência Aberta na *Web of Science*, *Scopus* e DOAJ (1970-2018).

	<b>Web of Science</b>	<b>Scopus</b>	<b>DOAJ</b>	<b>TOTAL</b>
Período	1970-2018	1979-2018	2006-2018	
Documentos encontrados	<b>1.117</b>	<b>1.466</b>	<b>686</b>	<b>3.269</b>
Palavras-chave da base de dados	1.466	4.536	X	<b>6.002</b>
Palavras-chave de autores	1.978	15.883	2.598*	<b>20.459</b>
<b>TOTAL POR BASE</b>	3.444	20.419	2.598	<b>26.461</b>
<b>Percentual do total de 26.461</b>	<b>13%</b>	<b>77%</b>	<b>10%</b>	<b>100%</b>

\*Para fins de representação, considerou-se as palavras-chave do DOAJ como de autores.

Fonte: Elaborado pelo autor (2020).

Com a soma do total de palavras-chave fornecidas pelas bases de dados com aquelas registradas pelos autores, tem-se um total de 26.461 palavras-chave de ambas categorias. Deste valor, pode-se extrair de forma mais detalhada o percentual real da contribuição de cada base de dados para a análise em questão. Dessa porcentagem, o levantamento de palavras-chave de autores oferecido pela *Scopus* representa 77% de todo o repertório de metadados levantados, sendo 4.536 palavras-chaves da base de dados (17%) e 15.883 palavras-chave de autores (60%).

Da mesma forma, cabe ressaltar que, dos 1.548 documentos encontrados na *Scopus*, 964 apresentaram o registro de palavras-chave da base de dados e 979 palavras-chave de autor. Desses, 117 tinham mais de 50 palavras-chave listadas, 171 mais de 25 e 675 mais de 10 palavras-chave classificadas por autores. Por causa desse arbítrio na indexação da base de dados, o número apresentado se mostra tão discrepante com relação aos de outras bases, como também ao registro de palavras-chave da própria *Scopus*.

Assim, os metadados coletados da *Scopus* com as palavras-chave de autores representa mais da metade de todo conjunto de dados para esta análise. Como esta seleção está sujeita a muitas repetições, erros ortográficos e alternâncias entre singular e plural, foi necessário compilar todo esse conjunto de dados para uma planilha, permitindo uma análise mais apurada da coletânea reunida.

#### 4.3.4.1 Análise pelo VOSviewer

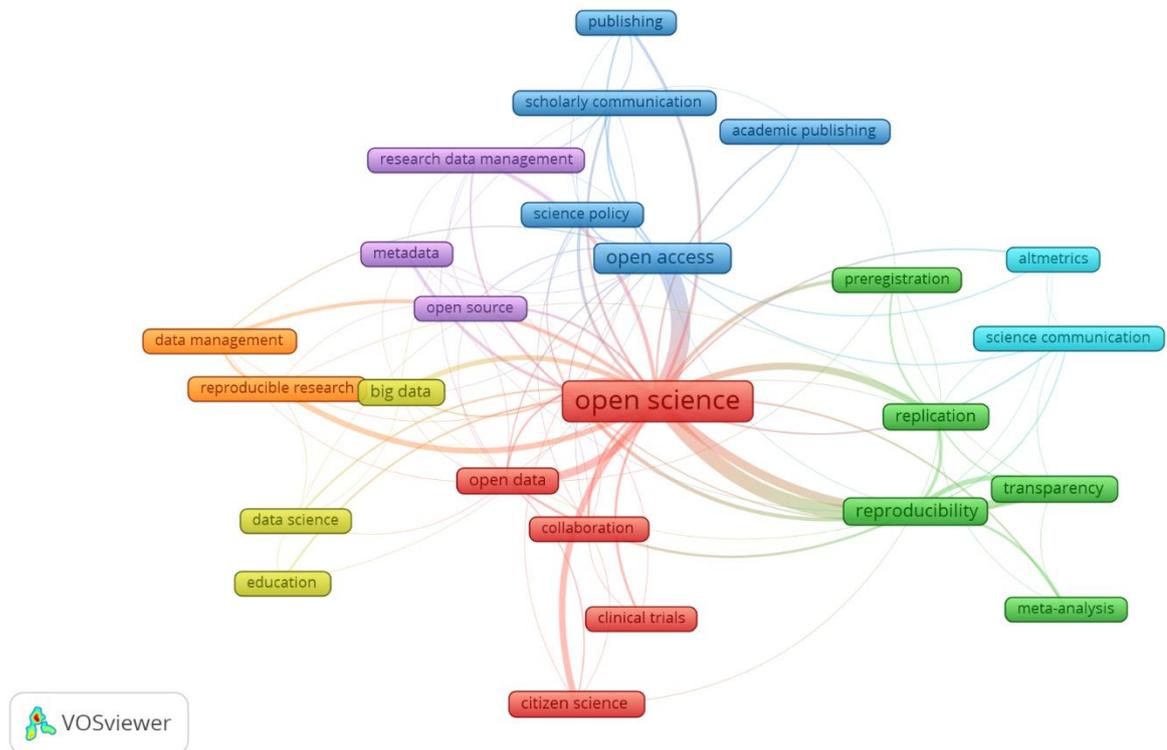
Utilizou-se o programa VOSviewer para criar a representação gráfica da rede complexa dos resultados sobre documentos com o termo Ciência Aberta na *Web of Science* e na *Scopus*. Como realizado anteriormente, limitou-se as ocorrências de palavras-chaves para os resultados de pesquisas de documentos com a expressão “*open science*” em ambas bases de dados, gerando a rede de dados (*cluster*) das correlações entre os termos, conforme ilustrado nas Figuras 20 e 21, respectivamente.

Como ambas as bases de dados contam com metadados que contêm palavras-chave de autores e palavras-chave de seus próprios indexadores, optou-se por montar as representações gráficas com esses dois tipos, tentando equiparar a representação gráfica dos recortes da pesquisa nas duas bases. No entanto, cabe destacar que o programa não reconheceu o percentual de palavras-chave extraído anteriormente. De um total de 3.444 da *Web of Science* e 20.419 da *Scopus* (soma de palavras-chave das bases de dados com palavras-chaves de autores), o VOSviewer só registrou o total de 2.402 na *Web of Science* (1.366 de autores e 1.036 da base) e 9.951 na *Scopus* (3.386 dos autores e 6.565 da base), respectivamente. Cabe destacar a discrepância do número de palavras-chave detectadas pelo programa VOSviewer na *Scopus*, com relação ao levantamento extraído da base. Enquanto no primeiro houve um maior número de termos indexados pela base, nos resultados coletados para a pesquisa foram encontradas mais palavras-chaves de autores.

Feitas essas observações, elaborou-se a representação gráfica dos dados pelo programa. A título de investigação exploratório utilizaram-se os seguintes parâmetros no VOSviewer: de tipo de análise pela coocorrência (“*Co-occurrence*”), com a unidade de análise aplicada primeiro às palavras-chave de autores (“*Authors Keywords*”), e depois às palavras-chave das bases de dados (“*KeyWord Plus*” para *Web of Science* e “*Index Keyword*” para *Scopus*), e o método de contagem como completa (“*Full counting*”).

Optou-se por apresentar a força das ligações entre as palavras-chave registradas em cada base de dados, deixando o termo “Open Science” figurar também para apresentar a correlação deste com os demais, bem como representar o tamanho do nó pela modularidade das ligações, separadas por *clusters* de cores diferentes. Assim, a título de deixar a visualização razoável, configurou-se para apresentação de um número mínimo de pelo menos cinco ocorrências em 1.366 palavras-chave de autores, detectadas pelo VOSviewer no levantamento feito na *Web of Science*, foi possível extrair 22 nós, com 104 ligações entre eles e separados em sete *clusters* (Figura 20).

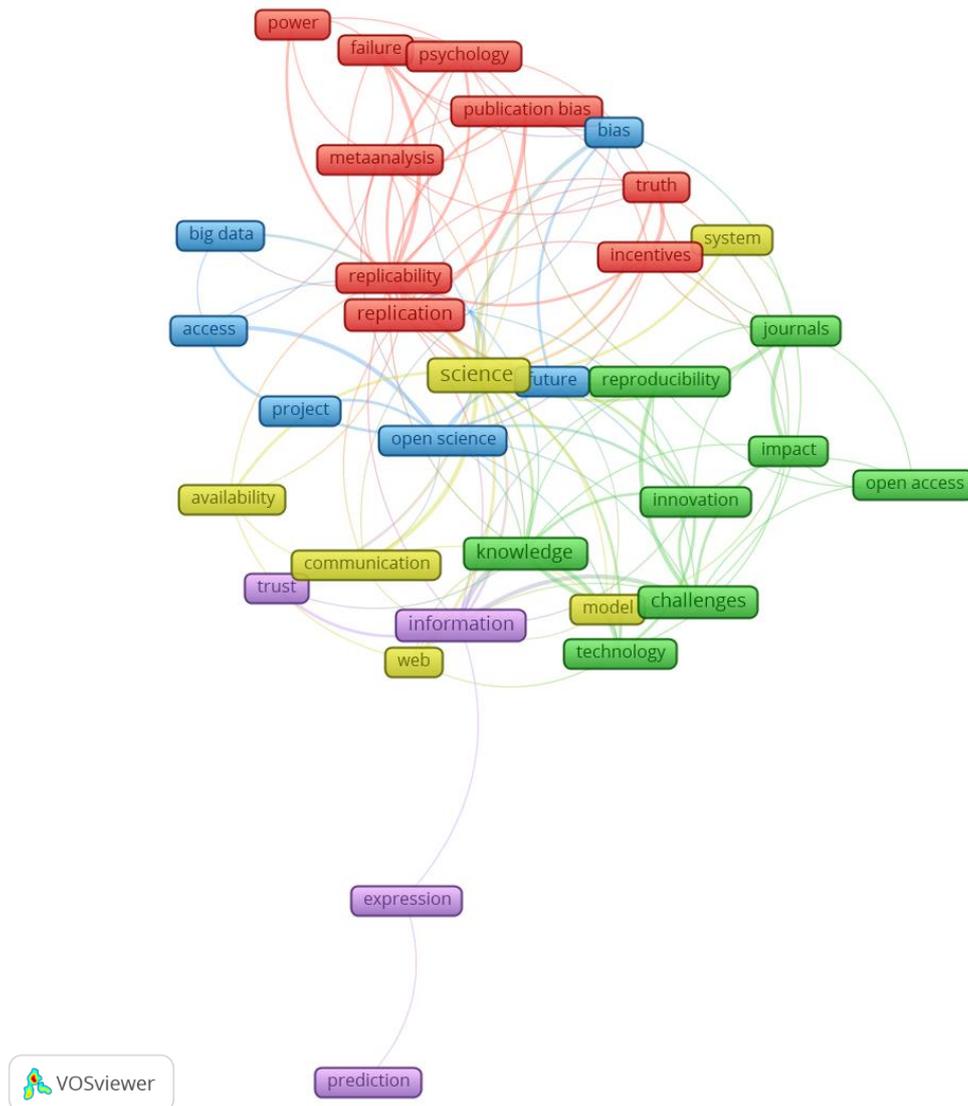
Figura 20 – Ligações entre as ocorrências das palavras-chave de autores em documentos com a expressão “open science” na *Web of Science* (1970-2018).



Fonte: Elaborado pelo autor (2020).

O mesmo levantamento foi feito com as palavras-chave indexadas pela própria base *Web of Science*, limitados também com, pelo menos, cinco ocorrências, mas dentre os 1.036 termos detectados pelo programa (Figura 21). Neste caso, registrou-se 31 nós, com 133 ligações entre eles e separados em cinco *clusters*.

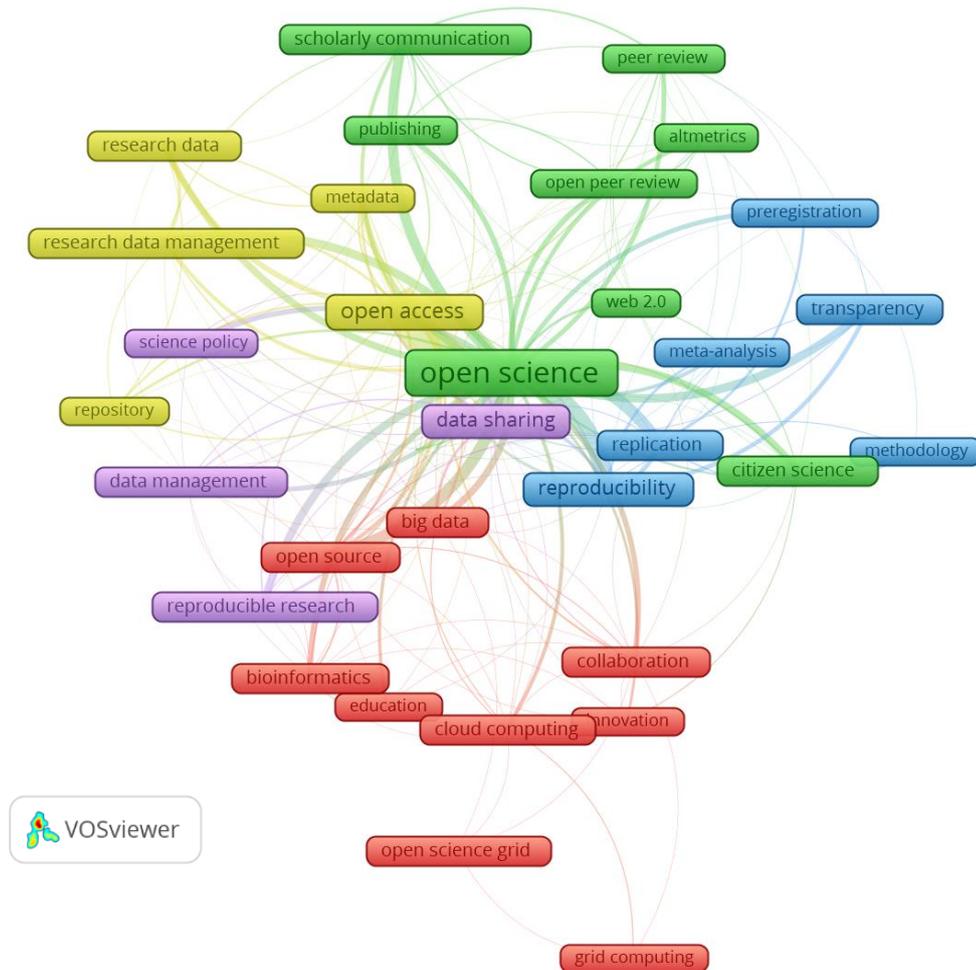
Figura 21 – Ligações entre as ocorrências das palavras-chave da *Web of Science* em documentos com a expressão “*open science*”, indexados pela própria base (1970-2018).



Fonte: Elaborado pelo autor (2020).

Na *Scopus*, ao utilizar o VOSviewer para criar a representação gráfica das palavras-chave de autores, optou-se por limitar, pelo menos, dez ocorrências dentre os 3.386 termos para tornar mais fácil de visualizar a modularidade entre as palavras-chave, facilitando também uma comparação com o mesmo levantamento na *Web of Science*, da Figura 20. Dessa forma, a Figura 22 apresenta 32 nós, com 198 ligações entre eles e separados por cinco *clusters*.

Figura 22 – Ligações entre as ocorrências das palavras-chave de autores em documentos com a expressão “*open science*” na *Scopus* (1979-2018).



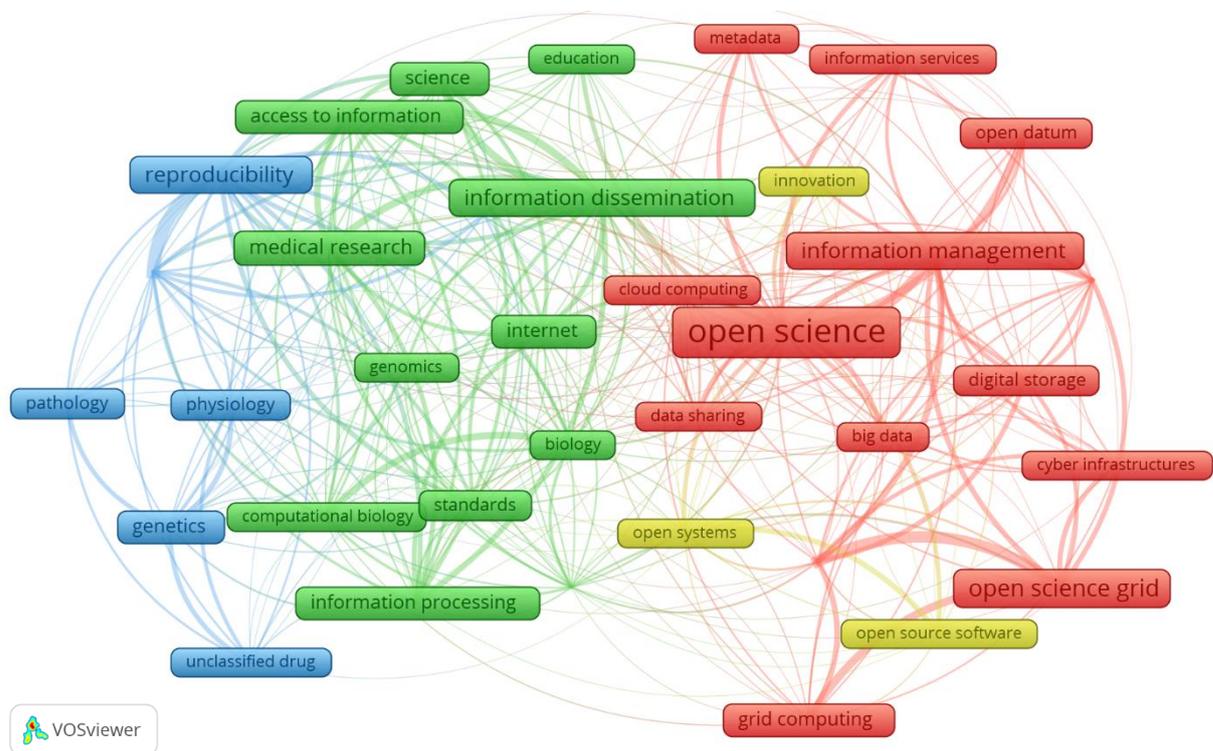
Fonte: Elaborado pelo autor (2020).

Tornou-se mais difícil fazer o mesmo levantamento com as palavras-chave indexadas pela própria *Scopus*, porque essa base de dados utiliza o seu vocabulário temático também para classificar os seus documentos. Termos como “*adult*”, “*animal*”, “*animal cell*”, “*article*”, “*female*”, “*human*”, “*male*” etc. aparecem reiteradas vezes para categorizar a indexação do acervo na base, e acabam poluindo a pesquisa na plataforma.

Desse modo, de um total de 6.565 palavras-chave indexadas pela *Scopus*, limitou-se a termos com pelo menos 20 ocorrências – número maior em virtude da maior proporção das palavras-chave detectadas. Essa medida permitiu registrar 80 nós, porém, conforme explicado anteriormente, depurou-se termos com características classificatórias que se aproximavam de

contaminação no levantamento. Com isso, restaram 35 nós, com 350 ligações entre eles e separados em quatro *clusters* (Figura 23).

Figura 23 – Ligações entre as ocorrências das palavras-chave da *Scopus* em documentos com a expressão “*open science*”, indexados pela própria base (1979-2018).



Fonte: Elaborado pelo autor (2020).

A título de análise exploratório, descartou-se apresentar a aplicação do filtro “*Overlay Visualization*” do VOSviewer nos dados utilizados para elaborar as figuras 20, 21, 22 e 23. Como tal filtro destaca a média cronológica das publicações ao longo dos anos, ao utilizar o recurso, a representação se limitava ao período entre 2017 e 2018, deixando a análise gráfica pouco relevante para a investigação proposta neste trabalho.

Como a avultada pluralidade semântica das palavras-chave da *Web of Science* e *Scopus*, construiu-se o Quadro 6 destacando os *clusters* definidos pelos agrupamentos estatísticos classificados de acordo com a modularidade do VOSviewer. Há certa aproximação entre estes conjuntos de termos, entretanto, fez-se a Tabela 6 para apontar as interseções entre as palavras-chave de autores das duas bases de dados, bem como as coincidências entre os sistemas de indexação de cada plataforma: *Keyword Plus*, na *Web of Science*, e *Keyword Index*, na *Scopus*.

Quadro 6 – *Clusters* identificados em cada representação gráfica de palavras-chave nas bases *Web of Science* e *Scopus* (Figuras 20, 21, 22 e 23).

Palavras-chave da Web of Science		Palavras-chave da Scopus	
De autores	KeyWords Plus	De autores	Index Keywords
publishing	power	scholarly communication	physiology
scholarly communication	failure	publishing	psychology
academic publishing	psychology	peer review	pathology
science policy	publication bias	altmetrics	metabolism
open access	metaanalysis	open peer review	unclassified drug
open science	truth	web 2.0	reproducibility of results
open data	incentives	open science	genetics
collaboration	replicability	citizen science	innovation
clinical trials	replication	research data	open systems
citizen science	bias	metadata	open source software
big data	big data	research data management	education
data science	access	open access	science
education	project	repository	access to information
data management	open science	preregistration	infomation dissemination
reproducible research	future	transparency	medical research
altmetrics	system	meta-analysis	internet
science communication	science	replication	genomics
preregistration	avaliability	reproducibility	biology
replication	communication	methodology	standarts
reproducibility	web	science policy	computation biology
transparency	model	data sharing	bioinformatics
meta-analysis	trust	data management	information processing
	information	reproducible research	metadata
	journals	big data	information services
	reproducibility	open source	open datum
	impact	bioinformatics	information management
	open access	education	cloud computing
	innovation	cloud computing	open science
	knowlegde	collaboration	data sharing
	challenge	innovation	big data
	technology	open science grid	digital storage
		grid computing	distributed computer system
			cyber infraestructure
			open science grid
			grid computing

Fonte: Elaborado pelo autor (2020).

Tabela 6 – Coincidências de palavras-chave na *Web of Science* e na *Scopus* no levantamento do VOSviewer

	Palavras-chave da Web of Science		Palavras-chave da Scopus	
	De autores	KeyWords Plus	De autores	Index Keywords
<b>altmetrics</b>	X		X	
<b>big data</b>	X	X	X	X
<b>bioinformatics</b>			X	X
<b>citizen science</b>	X		X	
<b>cloud computing</b>			X	X
<b>collaboration</b>	X		X	
<b>data management</b>	X		X	
<b>data sharing</b>			X	X
<b>education</b>	X		X	X
<b>grid computing</b>			X	X
<b>innovation</b>		X	X	X
<b>metaanalysis/meta-analysis</b>	X	X	X	
<b>metadata</b>			X	X
<b>open access</b>	X	X	X	
<b>open data/open datum</b>	X			X
<b>open science</b>	X	X	X	X
<b>open science grid</b>			X	X
<b>preregistration</b>	X		X	
<b>psychology</b>		X		X
<b>publishing</b>	X		X	
<b>replication</b>		X	X	
<b>reproducibility/reproducibility of results</b>	X	X	X	X
<b>scholarly communication</b>	X		X	
<b>science</b>		X		X
<b>science policy</b>	X		X	
<b>transparency</b>	X		X	

Fonte: Elaborado pelo autor (2020).

Conforme as três linhas em destaque na Tabela 6, “*big data*”, “*open Science*” e “*reproducibility/reproducibility of results*”, esses quatro termos apareceram no levantamento das palavras-chave de autores e no vocabulário controlado de indexação da *Web of Science* e da *Scopus*. Cabe ressaltar ainda que, dos 26 termos coincidentes encontrados, cinco estavam nas palavras-chave de autores e nas palavras-chave da própria plataforma *Web of Science*, onze estavam nas palavras-chaves de autores e nas palavras-chave da própria plataforma

*Scopus*, seis estavam no vocabulário controlado de cada base de dados (*KeyWords Plus* e *Index Keywords*) e catorze estavam nas palavras-chave de autores de ambas plataformas.

Essa última coincidência não só aponta para a proximidade conceitual da indexação dos autores que publicam documentos sobre “Ciência Aberta”, como também reflete a alta interseção do acervo documental entre as bases de dados *Web of Science* e *Scopus*, conforme já salientado. No entanto, a principal inferência que se alcança é a distância entre o vocabulário controlado, representado pela padronização de palavras-chave indexadas arbitrariamente pelas bases de dados e o que os termos livres, que caracterizam as palavras-chave escolhidas voluntariamente pelos autores.

Tal distinção é fundamental para compreender os assuntos propostos pelos autores, pois os termos livres demoram muito para serem incorporados no vocabulário controlado das bases de dados. Assim, para analisar as tendências das discussões políticas, circunstanciais ou contemporâneas com relação a certas temáticas da Ciência Aberta, deve-se destacar os termos “*Altmetrics*”, “*citizen science*”, “*collaboration*”, “*data management*”, “*education*”, “*metaanalysis*”/“*meta-analysis*”, “*open access*”, “*preregistration*”, “*publishing*”, “*scholarly communication*”, “*science policy*” e “*transparency*”.

No entanto, a análise com o VOSviewer não permitiu englobar todas as palavras-chave das três bases de dados selecionadas para este trabalho, e muito menos reunir diretamente os resultados da *Web of Science* e *Scopus*. Dessa forma, optou-se por fazer um levantamento com comparações mais diretas, a partir das planilhas com os resultados coletados das buscas nas plataformas sobre documentos com a expressão “*open science*”. Cabe lembrar que os dados nesse documento estavam padronizados de forma a permitir uma unificação de classificações de mesmo valor.

#### 4.3.4.2 Análise empírica

Em um total de 3.269 documentos sobre Ciência Aberta analisados no corte da pesquisa, inventariaram-se 6.002 palavras-chave indexadas pelas próprias base de dados, no caso da *Clarivate Analytics* para a *Web of Science* e Elsevier para a *Scopus*, ao mesmo tempo em que se registram 20.459 termos indexadas pelos próprios autores nas três plataformas. Quanto ao DOAJ, não há palavras-chave da base de dados nem de autores, mas, sim, um terceiro gênero em que editores e autores registraram os termos segundo um vocabulário controlado da plataforma.

Antes de apresentar os resultados com frequência de palavras-chave mais encontradas, cabe ressaltar que a *Web of Science* apresentou muitos documentos anteriores a 2017 registrados de forma errática. Dessa forma, do total de 1.117 documentos sobre Ciência Aberta registrados no recorte desta base de dados, somente 349 tinham palavras-chave de autores e 291 possuíam palavras-chaves indexadas pela *Clarivate Analytics*.

Essa ausência de informação da *Web of Science* justifica a discrepância dos seus números quanto às palavras-chave, com um pouco mais de uma palavra-chave da base de dados e dos autores por cada documento encontrado. Em comparação, a *Scopus* apresenta 4.536 palavras-chave da *Elsevier* para 1.466 documentos, e o DOAJ, 2.598 para 686 artigos, em um total de um pouco mais de três e quase quatro palavras-chave, respectivamente.

Com essa ponderação, observa-se, na Tabela 7, logo a seguir, um *ranking* com as cem palavras-chave mais frequentes indexadas pelas próprias bases de dados – no caso, *Web of Science* e *Scopus*. Na apuração dessas palavras, foram separados os termos que estavam nos *clusters* próximos da palavra-chave “*open science*” – Apêndice A.

Tabela 7 – Palavras-chave mais registradas nos documentos sobre Ciência Aberta da *Web of Science* e *Scopus*, com maior proximidade com o tema e indexadas pelas bases de dados (1970-2018).

(continua)

Posição	Frequência	Palavras-chave
1°	448	open science
2°	92	open access
3°	64	open data
4°	59	reproducibility
5°	42	replication
6°	41	data sharing
7°	41	science
8°	26	open source
9°	23	transparency
10°	22	scholarly communication
11°	20	knowledge
12°	20	research data management
13°	18	metaanalysis
14°	18	research data
15°	17	open science grid
16°	17	repository
17°	16	replicability
18°	16	reproducible research
20°	15	information
21°	15	metadata
22°	14	big data
25°	13	collaboration
26°	13	publishing
27°	12	education
32°	11	data management
35°	10	access
37°	10	peer review
38°	10	research infrastructure

40°	9	intellectual property
43°	9	repository
47°	9	workflow

Tabela 7 – Palavras-chave mais registradas nos documentos sobre Ciência Aberta da *Web of Science* e *Scopus*, com maior proximidade com o tema e indexadas pelas bases de dados (1970-2018).

(conclusão)		
Posição	Frequência	Palavras-chave
48°	8	e-infrastructure
50°	8	grid computing
55°	8	open innovation
59°	8	research
64°	7	deposition
66°	7	interdisciplinarity
67°	7	interoperability
69°	7	openness
71°	7	research integrity
72°	7	researchers
73°	7	science communication
75°	7	technology transfer
76°	7	visualization
78°	6	availability
79°	6	citizen science
80°	6	cloud computing
81°	6	communication
89°	6	open educational resources
90°	6	open peer review

Fonte: Elaborado pelo autor (2020).

Dando seguimento à análise das palavras-chave, repetiu-se o processo com a Tabela 7, na qual fez-se o *ranking* com as cem palavras-chave mais frequentes dos próprios autores – agora com as três plataformas: *Web of Science*, *Scopus* e *DOAJ*. Como na Tabela 8, também foram triados os termos mais próximos do agrupamento da palavra-chave “*open science*” – Apêndice B.

Tabela 8 – Palavras-chave mais registradas nos documentos sobre Ciência Aberta da *Web of Science*, *Scopus* e *DOAJ*, com maior proximidade com o tema e indexadas pelos próprios autores (1970-2018).

(continua)		
Posição	Frequência	Palavras-chave
1°	737	science
2°	413	open science
7°	151	information resource
9°	128	library science
10°	127	bibliography
12°	121	information dissemination
14°	109	research
15°	100	reproducibility
19°	85	open science grid
20°	76	open access
21°	75	publication
22°	75	publishing
23°	70	procedures

28°	61	information management
32°	59	controlled study
33°	59	distributed computer systems

Tabela 8 – Palavras-chave mais registradas nos documentos sobre Ciência Aberta da *Web of Science*, *Scopus* e DOAJ, com maior proximidade com o tema e indexadas pelos próprios autores (1970-2018).

(conclusão)		
Posição	Frequência	Palavras-chave
35°	58	medical research
40°	47	peer review
42°	44	grid computing
43°	44	reproducibility of results
45°	42	information processing
46°	41	data sharing
49°	39	review
51°	37	access to information
59°	33	data handling
60°	33	open source
61°	33	standards
63°	31	communication
64°	31	databases
71°	29	digital storage
73°	29	scholarly communication
74°	28	big data
79°	25	computer science
80°	25	open access publishing
81°	25	open data
83°	25	transparency
89°	24	metadata
93°	24	open datum
95°	23	cloud computing
97°	23	information technology
100°	22	citizen science

Fonte: Elaborado pelo autor (2020).

Em uma primeira análise, as palavras-chave podem servir de embasamento para evidenciar a ligação da Ciência Aberta com algumas das suas vertentes, indicadas anteriormente. Tanto que, dentre os 100 termos mais registrados, tanto para os indexados pelas próprias bases de dados, quanto pelos dos autores, registram, respectivamente, em:

- i) Cinco palavras-chaves de bases de dados:
  - a. “*open access*” e seu desdobramento “*openness*”,
  - b. “*open data*”,
  - c. “*open source*”,
  - d. “*citizen science*” e
  - e. “*open educational resources*”; e
- ii) Quatro palavras-chaves de autores:
  - a. “*open access*” e seu desdobramento “*open access publishing*”,

- b. “*open sources*”,
- c. “*open data*” e seu desdobramento “*open datum*” e
- d. “*citizen science*”.

Esta análise pode auxiliar na visualização de novas vertentes, bem como áreas de pesquisa e categorias de estudos utilizadas por algumas bases de dados. Da mesma forma, ao confrontar o universo com todas as palavras-chave, tanto de autores (nas plataformas *Web of Science*, *Scopus* e *DOAJ*) quanto das próprias bases de dados (*Web of Science* e *Scopus*), encontrou-se uma consonância mais abrangente entre os termos listados. A partir desse levantamento, com as palavras-chave coincidentes entre os parâmetros das Tabelas 7 e 8, foi elaborada a Tabela 9.

Tabela 9 – Termos coincidentes na análise das 100 palavras-chave mais registradas nos documentos sobre Ciência Aberta nas bases de dados *Web of Science*, *Scopus* e *DOAJ* (1970-2018).

Palavras-Chave	Soma das frequências
science	778
open science	861
open access	168
reproducibility	159
research	117
open science grid	102
publishing	88
data sharing	82
publication	81
open source	59
peer review	57
grid computing	52
scholarly communication	51
transparency	48
big data	42
metadata	39
standards	39
communication	37
cloud computing	29

Fonte: Elaborado pelo autor (2020).

Os 19 resultados encontrados permitem destacar:

- i) Duas vertentes:
  - a. “*open access*” e
  - b. “*open source*”.
- ii) Oito classificações mais frequentes:
  - a. “*reproducibility*”,

- b. “*open science grid*”,
  - c. “*data sharing*”,
  - d. “*peer review*”,
  - e. “*grid computing*”,
  - f. “*scholarly communication*”,
  - g. “*transparency*”,
  - h. “*metadata*” e
  - i. “*cloud computing*”; e,
- iii) Três tipos de áreas de conhecimento e categorias de estudo:
- a. “*science*”,
  - b. “*research*” e
  - c. “*communication*”.

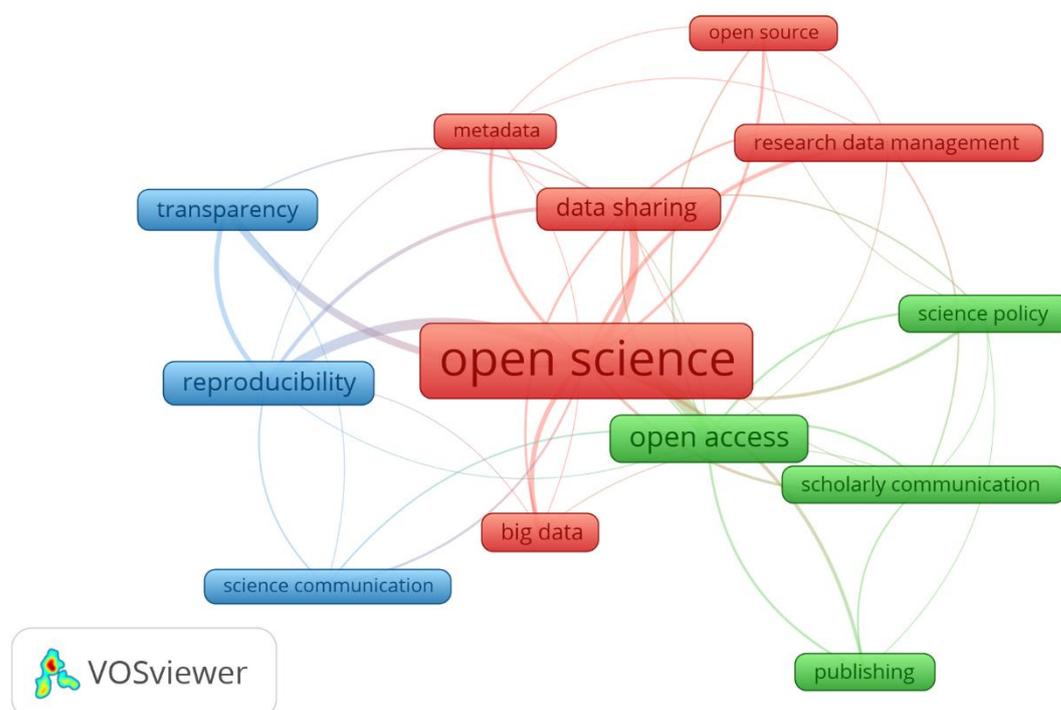
Feita essa exposição, cabe ressaltar a importância das palavras-chave para identificar as vertentes da Ciência Aberta, com destaque para “open access” e “open source”. Da mesma forma, as áreas de conhecimento e categorias de estudo podem indicar uma contaminação das palavras-chave indexadas pelo DOAJ. Uma vez que essa base de dados faz a importação de classificações de revistas científicas para o seu repertório de palavras-chave, o grande volume de termos como “science”, “research” e “communication” no repertório coletado, acabam por evidenciar essa indexação diferenciada do DOAJ, em meio ao resultado das demais plataformas. podem ter tido um grande aumento no repertório coletado.

Quanto às oito palavras-chave de classificações mais frequentes, assim como as demais, vale ressaltar que são indicações das indexações arbitrárias feitas pelas plataformas *Web of Science*, *Scopus* e DOAJ. Assim, esses termos não refletem uma evolução ou desenvolvimento da área, nem a intenção classificatória dos autores ou a exposição das tendências da literatura, mas sim a predisposição das bases de dados para classificar as palavras-chave dos artigos de Ciência Aberta. Tal critério pode ter tendências mercadológicas, mas também pode refletir o avanço das pesquisas na área, mas com menos precisão do que o que reflete uma análise das palavras-chave de autores.

Em derradeiro, fez-se uma última análise no VOSviewer, confrontando os 19 termos da Tabela 7 com o recurso do programa para modelar as redes de palavras-chave de autores e das próprias bases de dados – no caso, a *Web of Science* e *Scopus*. As Figuras 24 e 25 são representações gráficas com os dados coletados na *Web of Science* e *Scopus*, evidenciando a

modularidade das coocorrências de termos pelas ligações dos *clusters* formados, com os termos indexados como palavras-chave de autor. Essa opção foi selecionada em detrimento das palavras-chaves indexadas pelas próprias bases de dados, pois, como observou-se no Quadro 6, há uma maior interseção entre os termos indexados nos documentos pelos autores.

Figura 24 – A relação de palavras-chave de autores em documentos com a expressão Ciência Aberta, mais relevantes sobre o tema, pelo viés da *Web of Science* (1970-2018).



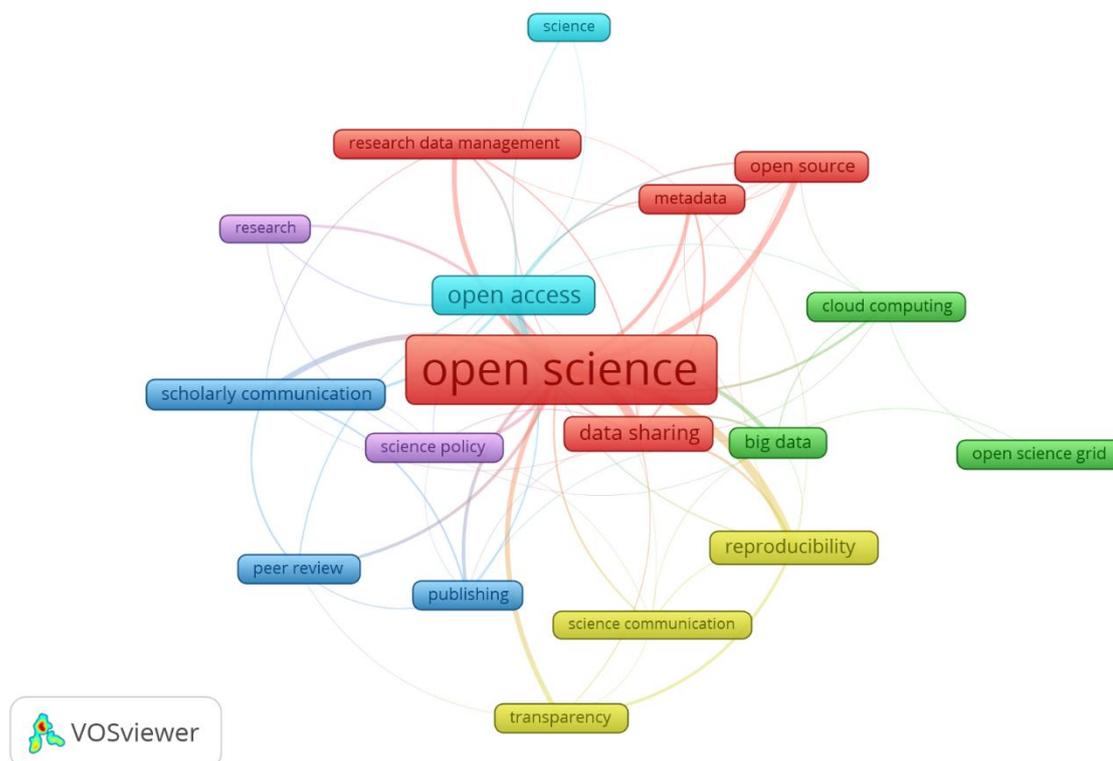
Fonte: Elaborado pelo autor (2020).

Na Figura 24, das 19 palavras-chave da Tabela 7, dos 1.366 termos indexados como palavras-chave de autores na *Web of Science* e detectados pelo VOSviewer, limitou-se o levantamento a, pelo menos, cinco ocorrências na base de dados, registrando-se 13 nós. Tais grupamentos tem 52 ligações entre si e estão separados em três *clusters*, todos conectados pela modularidade.

Não foram encontrados: “*cloud computing*”, “*communication*”, “*grid computing*”, “*open science grid*”, “*peer review*”, “*publication*” e “*standarts*”. Entretanto, optou-se por registrar “*research data management*” no lugar de “*research*” e “*science communication*” e

“*science policy*” no lugar de “*science*”, pois esses termos únicos não foram listados no vocabulário processado pelo programa.

Figura 25 – A relação de palavras-chave de autores em documentos com a expressão Ciência Aberta, mais relevantes sobre o tema, pelo viés da *Scopus* (1979-2018).



Fonte: Elaborado pelo autor (2020).

Quanto à Figura 25, dos 3.386 termos indexados como palavras-chave de autores na *Scopus* e detectados pelo VOSviewer, o levantamento também foi limitado a, pelo menos, cinco ocorrências nessa plataforma, registrando-se 18 nós. Tais grupamentos tinham 65 ligações entre si e estavam separados em seis *clusters*, todos conectados pela modularidade detectada na análise do programa.

Assim como no levantamento anterior, não foram encontrados “*communication*”, “*grid computing*”, “*publication*” e “*standarts*”. E, da mesma forma, também foi escolhido registrar “*research data management*” no lugar de “*research*” e “*science communication*” e “*science policy*” no lugar de “*science*”. Porém, nesse caso, na *Scopus*, os termos únicos “*research*” e “*science*” foram encontrados no vocabulário processado pelo software.

Os modelos gráficos das Figuras 24 e 25 apresentam em comum 13 palavras-chave dos 19 termos da Tabela 7, incluindo nessa contagem “*research data management*”, “*science communication*” e “*science policy*”. Não foi reproduzida a análise cronológica pela

ferramenta de visualização “*Overlay Visualization*” do VOSviewer, porque o registro de variação do uso dos termos era de menos de 1 ano.

No entanto, o Quadro 7 apresenta a aproximação dos *clusters* formados pela modularidade analisada no software:

Quadro 7 – *Clusters* identificados em cada representação gráfica com as palavras-chave de autores nas bases *Web of Science* e *Scopus* (Figuras 24 e 25)

Web of Science	Scopus
transparency reproducibility science communication	transparency reproducibility science communication
open source metadata research data management data sharing open science big data	open source metadata research data management data sharing open science big data
open access science policy scholarly communication publishing	cloud computing open science grid scholarly communication publishing peer review
	open access science
	research science policy

Fonte: Elaborado pelo autor (2020).

Nota-se que há certa aproximação entre os *clusters* registrados em ambas figuras, reproduzindo fielmente, pelo menos, um em ambos levantamentos. É certo que tal registro também se relaciona com a grande interseção entre as duas bases de dados, mas o levantamento pelo VOSviewer, combinado com a análise dos dados extraídos individualmente de cada plataforma, corrobora com a investigação e o destaque que se quer dar para as palavras-chave mais relevantes para os documentos sobre Ciência Aberta, na literatura formal.

Estas últimas análises gráficas contribuem para uma investigação da Análise de Domínio dos três pontos escolhidos, indexação e recuperação de especialidades, estudos bibliométricos e estudos documentais e de gênero. Uma vez realizadas as investigações sobre o material dos resultados encontrados, local de publicação, o idioma dos documentos e os

casos de coautoria, o exame das palavras-chave conclui os subsídios para uma Análise de Domínio com a amostragem.

Assim, o estudo das palavras-chaves no recorte de pesquisas de documentos sobre “Ciência Aberta”, realizados nas bases de dados *Web of Science*, *Scopus* e *DOAJ*, alicerça uma configuração inicial tanto para a Análise de Domínio quanto para o estudo de Campo Científico. Tal suporte é possível justamente pelo repertório de termos específicos que a pesquisa reúne, permitindo as noções iniciais dos estudos que norteiam as configurações da área de conhecimento e a categoria de estudo da Ciência Aberta.

#### 4.3.4.3 Problematização sobre as diferentes definições de Ciência Aberta

Após o levantamento de palavras-chave, cabe ressaltar a questão da dificuldade terminológica do conceito de Ciência Aberta, frente a outras terminologias concorrentes para “*Open Science*”. Atualmente, há uma discussão muito acentuada sobre a definição deste tema e de outros que coexistem, tais como “*Open Research*”, “*Open Scholarship*” e “*Standards*”.

Há mais similaridades do que diferenças entre o emprego desses termos e “*Open Science*”. Entretanto, como alguns destes se destacaram na análise do montante de palavras-chave encontradas na pesquisa sobre documentos de Ciência Aberta nas bases de dados *Web of Science*, *Scopus* e *DOAJ*, é pertinente diferenciar algumas das definições desses conceitos para melhor fundamentar cada terminologia.

No entanto, conforme ocorre com a Ciência Aberta, esses conceitos também têm um limitado repertório na literatura formal para nortear a sua conceitualização. No caso, “*Open Research*”, que literalmente se definiria por “Pesquisa Aberta”, seria, segundo a plataforma *Wellcome Open Research*, o acesso aberto a publicações e compartilhamentos sobre um variado repertório de resultados de pesquisa, tais como protocolos de estudo, notas de dados, cadernos de pesquisa e, inclusive, os tradicionais artigos científicos (KINLEY, MARKIE; 2019, n. p.).

A plataforma *Wellcome Open Research*, que é referência para pesquisadores, principalmente da Europa (LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES, 2018, p. 12), ajuda a sustentar a definição do termo Pesquisa Aberta. Tal conceito se distinguiria da Ciência Aberta por estar mais voltado para o compartilhamento de materiais que são resultados diretos de atividades de pesquisa, mas seria mais uma especificidade conceitual do que uma verdadeira diferença.

“*Open Scholarship*” teria um sentido mais amplo, literalmente de “Aprendizado e Conhecimento Abertos”. Segundo a *League of European Research Universities* (LERU) (2018, p. 4), o conceito seria mais adequado que Ciência Aberta, pois “Ciência” abrangeria somente uma subcategoria de disciplinas acadêmicas. No entanto, entende-se que a utilização desta terminologia representa uma mudança cultural de todos os agentes envolvidos nas “comunidades de pesquisa, educação e troca de conhecimento, criação, armazenamento, compartilhamento e entrega dos resultados de suas atividades” (LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES, 2018, p. 9)

Assim, a definição do termo seria sim mais abrangente do que a de Ciência Aberta, porém mais por uma questão filológica do termo do escopo temático em si. Uma vez que o movimento da Ciência Aberta envolve campos que vão muito além das fronteiras acadêmicas, “Aprendizado e Conhecimento Abertos” tem mais o efeito de uma retificação e adaptação semântica do que uma verdadeira diferença conceitual.

Outro termo que também foi identificado é “*Standards*”, que literalmente seriam “Padrões”. A ementa de um *workshop* organizado pela Diretoria Geral de Pesquisa e Inovação da Comissão Europeia de 2016<sup>27</sup> indica que “toda pesquisa financiada publicamente na União Europeia adere a comum acordo ao *Open Science Standards of Research Integrity* (ou “Padrões da Ciência Aberta da Integridade da Pesquisa”). Em seguida, lista-se quatro protótipos normativos que servem como medidas políticas para promover uma Ciência Aberta mais bem referenciada, transparente, ética e íntegra.

Assim, “Padrões” seria um conceito mais voltado para semântica de políticas institucionais que buscam promover a Ciência Aberta. Da mesma forma, como o campo é formado por múltiplas iniciativas e agentes, o sentido da padronização também caracteriza o constante movimento pela interoperabilidade da Ciência Aberta, não só de seus resultados, mas também de práticas e atividades administrativas e técnicas. Assim, o termo relaciona-se diretamente com a necessidade direta de correspondência entre todos os envolvidos no movimento da Ciência Aberta.

Com essas ressalvas, reconhece-se que seria necessária uma análise mais ampla da problematização dos diferentes conceitos da Ciência Aberta. No entanto, esta pesquisa tem a

---

<sup>27</sup> Workshop de padronização de diretrizes, organizado pela Diretoria Geral de Pesquisa e Inovação da Comissão Europeia e pelos representantes do Estado da Holanda. Disponível em: [http://wavelets.ens.fr/OPEN\\_SCIENCE/MARIE\\_FARGE/CONFERENCES/2016\\_04\\_04\\_OPEN\\_SCIENCE\\_WORKSHOP\\_SCIENCE\\_ACADEMY\\_OF\\_NETHERLANDS/2016\\_04\\_04\\_EC\\_Open\\_Science\\_Workshop\\_Science\\_Academy\\_of\\_Netherlands\\_Report.pdf](http://wavelets.ens.fr/OPEN_SCIENCE/MARIE_FARGE/CONFERENCES/2016_04_04_OPEN_SCIENCE_WORKSHOP_SCIENCE_ACADEMY_OF_NETHERLANDS/2016_04_04_EC_Open_Science_Workshop_Science_Academy_of_Netherlands_Report.pdf). Acesso em: 14 abr. 2020.

intenção de se deter sobre a definição terminológica mais tradicional do tema, verificada diretamente pela investigação do vocábulo “*Open Science*” nas três bases de dados escolhidas. Para investigar a questão das diferenças terminológicas dos outros conceitos seria necessário também ampliar a busca por tais termos, relacionando-os com as palavras-chave encontradas em cada pesquisa.

#### 4.4 ANALOGIAS ENTRE A CIÊNCIA ABERTA E SUAS VERTENTES

Após a análise das palavras-chave, foi feito um levantamento dos documentos na *Web of Science* através da categorização temática de áreas de conhecimento – a base de dados desdobra essa informação em duas classificações. Em “Áreas de pesquisa”, encontram-se áreas de conhecimento categorizadas dentro de cinco grandes assuntos, enquanto “Categorias da *Web of Science*” é uma classificação temática mais abrangente, atendendo a especificações e distinções mais imediatas de pesquisas e áreas de conhecimento. A Tabela 8 apresenta as áreas de pesquisa e Categorias da *Web of Science* de documentos da base de dados com a expressão “*open science*” no período estudado.

O campo “Áreas de pesquisa” conta com 153 categorias diferentes, divididas em cinco grandes classificações: *Arts & Humanities* (Artes e humanidades), com 14 temas; *Life Sciences & Biomedicine* (Ciência e tecnologia: Ciências da vida e biomedicina), com 76 temas; *Physical Sciences* (Ciência e Tecnologia: Ciências físicas), com 17 temas; *Social Sciences* (Ciências sociais), com 25 temas; e *Technology* (Ciência e Tecnologia: Tecnologia), com 21 temas.

As **áreas de pesquisa** formam um esquema de categorização de assunto compartilhado por todas as bases de dados de produtos da *Web of Science*. Como resultado, é possível identificar, recuperar e analisar os documentos a partir de diversas bases de dados relacionadas ao mesmo assunto.

Periódicos e livros cobertos pela Principal Coleção da *Web of Science* são atribuídos a pelo menos uma **categoria da *Web of Science***. **Todas as categorias da *Web of Science* são associadas a uma área de pesquisa.** (CLARIVATE ANALYTICS, 2019a, n. p., grifo nosso)

Já o campo “Categorias da *Web of Science*” se limita a um conjunto de 254 temas, nos quais “todos os periódicos e livros cobertos pela principal coleção da *Web of Science* são atribuídos a, pelo menos, uma das seguintes categorias de assunto” (CLARIVATE ANALYTICS, 2019b, n. p.).

Assim, esses dois campos continham metadados das duas classificações que a base contém: enquanto “Categorias da *Web of Science*” atende às especificações de inovações e

tendências, “Áreas de Pesquisa” organiza a padronização da base de dados de maneira mais apurada, classificando as áreas de conhecimentos de acordo com as cinco famílias hierárquicas de sua organização.

Ao fazer uma análise dentro dos 1.117 documentos com a expressão “*open science*” na base de dados *Web of Science*, de 1970 até 2018, encontrou-se uma preponderância da classificação temática “*Information Science & Library Science*” (Ciência da informação e biblioteconomia), que, na organização do campo “Área de pesquisa”, se encontra categorizada como subárea da área de conhecimento “*Technology*” (Ciência e Tecnologia: Tecnologia). Contudo, cabe ressaltar que essa classificação de área de conhecimento foi uma opção de indexação exclusiva da plataforma.

Assim, a análise da correlação entre a Ciência Aberta e suas vertentes, compreendida nas classificações de “Categorias da *Web of Science*” e “Áreas de Pesquisa” pode indicar uma apreciação das relações entre os termos identificados como seus desdobramentos da Ciência Aberta. No entanto, cabe ressaltar que essa investigação se restringiu a *Web of Science* por conta da distinção dessa base de dados com relação a essas classificações, mas essas foram estabelecidas arbitrariamente pela plataforma.

Da mesma forma, cabe ressaltar uma pequena falha na recuperação da informação na base de dados *Scopus* que, ao buscar por documentos com o termo “*Open Data*”, acaba resgatando todos os documentos que tenham efetivamente a classificação de conteúdo, no acervo da base de dados, como “Dados Abertos” – contaminando de forma irreversível qualquer pesquisa pelo termo.

Assim, por conta desta irregularidade técnica nas buscas da *Scopus* e pela ausência de recursos de refino e coleta de resultados no DOAJ, optou-se pela pesquisa das vertentes da Ciência Aberta na *Web of Science*. De tal forma, foram coletadas as informações sobre “Categoria da *Web of Science*” e “Área de pesquisa”, para poder criar tabelas com os números e a porcentagem de cada classificação de documentos com o termo “*Open Science*” (Ciência Aberta) e, em suas vertentes, a saber: “*Open Access*” (Acesso Aberto), “*Open Source*” (Fontes Abertas), “*Open Data*” (Dados Abertos), “*Citizen Science*” (Ciência Cidadã) e “*Open Education*” (Educação Aberta) – respectivamente as Tabelas 10, 11, 12, 13, 14 e 15, com marcações destacando a classificação “*Information science library science*” e a sua posição dentre as demais áreas de pesquisa e Categorias *Web of Science*.

Tabela 10 – Áreas de pesquisa e Categorias da *Web of Science* de documentos da base de dados com a expressão “open science” (Ciência Aberta) (1970-2018).

<b>Categorias da <i>Web of Science</i></b>	<b>Reg.</b>	<b>% de 1.117</b>	<b>Áreas de pesquisa</b>	<b>Reg.</b>	<b>% de 1.117</b>
<b>Information science library science</b>	<b>120</b>	<b>10,5</b>	<i>Computer science</i>	252	22,2
<i>Computer science information systems</i>	115	10,1	<b>Information science library science</b>	<b>120</b>	<b>10,5</b>
<i>Computer science theory methods</i>	100	8,8	<i>Engineering</i>	98	8,6
<i>Computer science interdisciplinary applications</i>	89	7,8	<i>Psychology</i>	82	7,2
<i>Multidisciplinary sciences</i>	77	6,7	<i>Science technology other topics</i>	81	7,1
<i>Biology</i>	60	5,2	<i>Physics</i>	72	6,3
<i>Neurosciences</i>	55	4,8	<i>Business economics</i>	64	5,6
<i>Engineering electrical electronic</i>	54	4,7	<i>Life sciences biomedicine other topics</i>	60	5,2
<i>Management</i>	47	4,1	<i>Neurosciences neurology</i>	56	4,9
<i>Physics nuclear</i>	46	4,0	<i>Environmental sciences ecology</i>	55	4,8
<i>Education educational research</i>	43	3,7	<i>Education educational research</i>	48	4,2
<i>Physics particles fields</i>	35	3,0	<i>Biochemistry molecular biology</i>	43	3,7
<i>Biochemistry molecular biology</i>	34	3,0	<i>Chemistry</i>	39	3,4
<i>Chemistry multidisciplinary</i>	33	2,9	<i>Mathematical computational biology</i>	30	2,6
<i>Computer science artificial intelligence</i>	33	2,9	<i>Social sciences other topics</i>	28	2,4
<i>Mathematical computational biology</i>	30	2,6	<i>Communication</i>	21	1,8
<i>Psychology multidisciplinary</i>	30	2,6	<i>Genetics heredity</i>	20	1,7
<i>Computer science hardware architecture</i>	28	2,4	<i>Astronomy astrophysics</i>	19	1,6
<i>Ecology</i>	28	2,4	<i>Psychiatry</i>	19	1,6
<i>Environmental sciences</i>	24	2,1	<i>General internal medicine</i>	17	1,5

Fonte: Elaborado pelo autor (2020).

Tabela 11 – Áreas de pesquisa e Categorias da *Web of Science* de documentos da base de dados com a expressão “open access” (Acesso Aberto) (1990-2018).

<b>Categorias da <i>Web of Science</i></b>	<b>Reg.</b>	<b>% de 49.810</b>	<b>Áreas de pesquisa</b>	<b>Reg.</b>	<b>% de 49.810</b>
<i>Optics</i>	4.616	9,2	<i>Engineering</i>	6.304	12,6
<i>Materials science multidisciplinary</i>	2.937	5,8	<i>Optics</i>	4.616	9,2
<i>Energy fuels</i>	2.497	5,0	<i>Materials science</i>	3.952	7,9
<b>Information science library science</b>	<b>1.990</b>	<b>3,9</b>	<i>Physics</i>	3.161	6,3
<i>Engineering multidisciplinary</i>	1.956	3,9	<i>Computer science</i>	2.929	5,8
<i>Multidisciplinary sciences</i>	1.853	3,7	<i>Science technology other topics</i>	2.609	5,2
<i>Education educational research</i>	1.822	3,6	<i>Energy fuels</i>	2.497	5,0
<i>Medicine general internal</i>	1.705	3,4	<i>Environmental sciences ecology</i>	2.177	4,3
<i>Engineering electrical electronic</i>	1.559	3,1	<b>Information science library science</b>	<b>1.990</b>	<b>3,9</b>
<i>Environmental sciences</i>	1.444	2,8	<i>Education educational research</i>	1.928	3,8
<i>Radiology nuclear medicine medical imaging</i>	1.194	2,3	<i>Biochemistry molecular biology</i>	1.915	3,8
<i>Computer science theory methods</i>	1.166	2,3	<i>Business economics</i>	1.842	3,6
<i>Biochemistry molecular biology</i>	1.159	2,3	<i>General internal medicine</i>	1.771	3,5
<i>Economics</i>	1.156	2,3	<i>Neurosciences neurology</i>	1.459	2,9
<i>Computer science information systems</i>	1.134	2,2	<i>Chemistry</i>	1.324	2,6
<i>Physics particles fields</i>	1.084	2,1	<i>Radiology nuclear medicine medical imaging</i>	1.194	2,3
<i>Computer science interdisciplinary applications</i>	1.039	2,0	<i>Pharmacology pharmacy</i>	1.073	2,1
<i>Medicine research experimental</i>	999	2,0	<i>Research experimental medicine</i>	999	2,0
<i>Physics multidisciplinary</i>	983	1,9	<i>Public environmental occupational health</i>	956	1,9
<i>Pharmacology pharmacy</i>	969	1,9	<i>Astronomy astrophysics</i>	930	1,8

Fonte: Elaborado pelo autor (2020).

Tabela 12 – Áreas de pesquisa e Categorias da *Web of Science* de documentos da base de dados com a expressão “open source” (Fontes Abertas) (1973-2018).

<b>Categorias da <i>Web of Science</i></b>	<b>Reg.</b>	<b>% de 41.418</b>	<b>Áreas de pesquisa</b>	<b>Reg.</b>	<b>% de 41.418</b>
<i>Engineering electrical electronic</i>	8.414	20,3	<i>Computer science</i>	20.338	49,1
<i>Computer science theory methods</i>	7.382	17,8	<i>Engineering</i>	12.584	30,3
<i>Computer science information systems</i>	6.491	15,6	<i>Telecommunications</i>	2.769	6,6
<i>Computer science software engineering</i>	6.196	14,9	<i>Biochemistry molecular biology</i>	2.222	5,3
<i>Computer science interdisciplinary applications</i>	4.598	11,1	<i>Education educational research</i>	1.555	3,7
<i>Computer science artificial intelligence</i>	3.409	8,2	<i>Mathematical computational biology</i>	1.543	3,7
<i>Telecommunications</i>	2.769	6,6	<i>Physics</i>	1.455	3,5
<i>Computer science hardware architecture</i>	2.352	5,6	<i>Science technology other topics</i>	1.388	3,3
<i>Biochemical research methods</i>	1.742	4,2	<i>Mathematics</i>	1.327	3,2
<i>Mathematical computational biology</i>	1.543	3,7	<i>Biotechnology applied microbiology</i>	1.185	2,8
<i>Biotechnology applied microbiology</i>	1.185	2,8	<b><i>Information science library science</i></b>	<b>1.110</b>	<b>2,6</b>
<b><i>Information science library science</i></b>	<b>1.110</b>	<b>2,6</b>	<i>Business economics</i>	1.091	2,6
<i>Education educational research</i>	1.080	2,6	<i>Automation control systems</i>	1.079	2,6
<i>Automation control systems</i>	1.079	2,6	<i>Chemistry</i>	1.040	2,5
<i>Multidisciplinary sciences</i>	1.009	2,4	<i>Radiology nuclear medicine medical imaging</i>	881	2,1
<i>Engineering mechanical</i>	976	2,3	<i>Environmental sciences ecology</i>	867	2,0
<i>Engineering multidisciplinary</i>	903	2,1	<i>Remote sensing</i>	843	2,0
<i>Engineering biomedical</i>	881	2,1	<i>Geology</i>	812	1,9
<i>Radiology nuclear medicine medical imaging</i>	881	2,1	<i>Optics</i>	790	1,9
<i>Remote sensing</i>	843	2,0	<i>Robotics</i>	777	1,8

Fonte: Elaborado pelo autor (2020).

Tabela 13 – Áreas de pesquisa e Categorias da *Web of Science* de documentos da base de dados com a expressão “open data” (Dados Abertos) (1978-2018).

<b>Categorias da <i>Web of Science</i></b>	<b>Reg.</b>	<b>% de 4.549</b>	<b>Áreas de pesquisa</b>	<b>Reg.</b>	<b>% de 4.549</b>
<i>Computer science information systems</i>	1.233	27,1	<i>Computer science</i>	2.114	46,4
<i>Computer science theory methods</i>	1.077	23,6	<i>Engineering</i>	708	15,5
<i>Computer science artificial intelligence</i>	668	14,6	<b>Information science library science</b>	<b>430</b>	<b>9,4</b>
<i>Engineering electrical electronic</i>	531	11,6	<i>Psychology</i>	376	8,2
<i>Computer science interdisciplinary applications</i>	471	10,3	<i>Remote sensing</i>	158	3,4
<b>Information science library science</b>	<b>430</b>	<b>9,4</b>	<i>Telecommunications</i>	152	3,3
<i>Psychology multidisciplinary</i>	343	7,5	<i>Business economics</i>	146	3,2
<i>Computer science software engineering</i>	292	6,4	<i>Science technology other topics</i>	140	3,0
<i>Remote sensing</i>	158	3,4	<i>Environmental sciences ecology</i>	136	2,9
<i>Telecommunications</i>	152	3,3	<i>Education educational research</i>	101	2,2
<i>Computer science hardware architecture</i>	134	2,9	<i>Public administration</i>	98	2,1
<i>Multidisciplinary sciences</i>	94	2,0	<i>Geology</i>	93	2,0
<i>Geosciences multidisciplinary</i>	90	1,9	<i>Social sciences other topics</i>	89	1,9
<i>Education educational research</i>	85	1,8	<i>Government law</i>	85	1,8
<i>Environmental sciences</i>	84	1,8	<i>Physical geography</i>	80	1,7
<i>Geography physical</i>	80	1,7	<i>Communication</i>	77	1,6
<i>Communication</i>	77	1,6	<i>Chemistry</i>	76	1,6
<i>Social sciences interdisciplinary</i>	73	1,6	<i>Mathematical computational biology</i>	72	1,5
<i>Mathematical computational biology</i>	72	1,5	<i>Geography</i>	69	1,5
<i>Geography</i>	69	1,5	<i>Robotics</i>	67	1,4

Fonte: Elaborado pelo autor (2020).

Tabela 14 – Áreas de pesquisa e Categorias da *Web of Science* de documentos da base de dados com a expressão “*citizen science*” (Ciência Cidadã) (1997-2018).

<b>Categorias da <i>Web of Science</i></b>	<b>Reg.</b>	<b>% de 2.784</b>	<b>Áreas de pesquisa</b>	<b>Reg.</b>	<b>% de 2.784</b>
<i>Ecology</i>	614	22,0	<i>Environmental sciences ecology</i>	943	33,4
<i>Environmental sciences</i>	489	17,5	<i>Biodiversity conservation</i>	419	15,0
<i>Biodiversity conservation</i>	419	15,0	<i>Computer science</i>	261	9,3
<i>Multidisciplinary sciences</i>	194	6,9	<i>Science technology other topics</i>	212	7,6
<i>Marine freshwater biology</i>	124	4,4	<i>Zoology</i>	199	7,1
<i>Zoology</i>	121	4,3	<i>Engineering</i>	163	5,8
<i>Environmental studies</i>	111	3,9	<i>Education educational research</i>	129	4,6
<i>Astronomy astrophysics</i>	102	3,6	<i>Marine freshwater biology</i>	128	4,5
<i>Computer science theory methods</i>	102	3,6	<i>Astronomy astrophysics</i>	102	3,6
<i>Computer science information systems</i>	98	3,5	<i>Remote sensing</i>	82	2,9
<i>Engineering electrical electronic</i>	89	3,1	<i>Entomology</i>	79	2,8
<i>Remote sensing</i>	82	2,9	<i>Water resources</i>	79	2,8
<i>Ornithology</i>	80	2,8	<i>Geography</i>	78	2,8
<i>Entomology</i>	79	2,8	<i>Physical geography</i>	78	2,8
<i>Water resources</i>	79	2,8	<i>Life sciences biomedicine other topics</i>	77	2,7
<i>Geography</i>	78	2,8	<i>Geology</i>	64	2,2
<i>Geography physical</i>	78	2,8	<i>Public environmental occupational health</i>	63	2,2
<i>Biology</i>	77	2,7	<i>Evolutionary biology</i>	62	2,2
<i>Computer science interdisciplinary applications</i>	75	2,6	<i>Oceanography</i>	60	2,1
<i>Education educational research</i>	74	2,6	<i>Communication</i>	58	2,0
<b><i>Information science library science</i></b>	<b>38</b>	<b>1,3</b>	<b><i>Information science library science</i></b>	<b>38</b>	<b>1,3</b>

Fonte: Elaborado pelo autor (2020).

Tabela 15 – Áreas de pesquisa e Categorias da *Web of Science* de documentos da base de dados com a expressão “*open education*” (Educação Aberta) (1971-2018).

<b>Categorias da <i>Web of Science</i></b>	<b>Reg.</b>	<b>% de 753</b>	<b>Áreas de pesquisa</b>	<b>Reg.</b>	<b>% de 753</b>
<i>Education educational research</i>	495	65,7	<i>Education educational research</i>	529	70,2
<i>Education scientific disciplines</i>	61	8,1	<i>Computer science</i>	133	17,6
<i>Computer science information systems</i>	57	7,5	<i>Engineering</i>	65	8,6
<i>Computer science interdisciplinary applications</i>	57	7,5	<i>Social sciences other topics</i>	42	5,5
<i>Computer science theory methods</i>	53	7,0	<b>Information science library science</b>	<b>26</b>	<b>3,4</b>
<i>Engineering electrical electronic</i>	48	6,3	<i>Psychology</i>	24	3,1
<i>Social sciences interdisciplinary</i>	41	5,4	<i>Business economics</i>	17	2,2
<b>Information science library Science</b>	<b>26</b>	<b>3,4</b>	<i>Arts humanities other topics</i>	14	1,8
<i>Computer science artificial intelligence</i>	21	2,7	<i>Telecommunications</i>	11	1,4
<i>Computer science software engineering</i>	14	1,8	<i>Communication</i>	7	0,9
<i>Humanities multidisciplinary</i>	14	1,8	<i>Social issues</i>	7	0,9
<i>Management</i>	14	1,8	<i>Linguistics</i>	5	0,6
<i>Psychology educational</i>	13	1,7	<i>Science technology other topics</i>	5	0,6
<i>Telecommunications</i>	11	1,4	<i>Automation control systems</i>	4	0,5
<i>Engineering multidisciplinary</i>	9	1,1	<i>Materials science</i>	4	0,5
<i>Communication</i>	7	0,9	<i>Medical informatics</i>	4	0,5
<i>Computer science hardware architecture</i>	7	0,9	<i>Neurosciences neurology</i>	4	0,5
<i>Economics</i>	7	0,9	<i>Operations research management science</i>	4	0,5
<i>Social issues</i>	7	0,9	<i>Philosophy</i>	4	0,5
<i>Psychology multidisciplinary</i>	6	0,7	<i>Sociology</i>	4	0,5

Fonte: Elaborado pelo autor (2020).

Dentre o recorte de áreas de conhecimento e categorias de pesquisa da base de dados na *Web of Science*, destacou-se a relevância de “*information science library science*”, que seria uma abreviação terminológica adaptada para a extração de dados feita na pesquisa para o termo “*Information Science & Library Science*” (Ciência da informação e biblioteconomia).

#### 4.4.1 Discussão sobre as áreas de conhecimento da Ciência Aberta e suas vertentes

Partindo desta análise nos 1.117 resultados de documentos da *Web of Science* sobre “Ciência Aberta”, a “Ciência da Informação” se destaca como 2º lugar nas áreas de conhecimento escolhidas pelos autores (com 120 resultados, equivalente a 10,5% do total) e como 1º lugar nas categorias de pesquisa da base de dados (com os mesmos 120 registros, e também 10,5%).

Tendo como base este primeiro recorte, observou-se que a porcentagem e o número de registros sobre as Áreas de Pesquisa e as Categorias da *Web of Science* se igualavam, sendo mais fácil simplificar a incidência dos campos sobre “Ciência da Informação” em uma representação gráfica mais simplificada – Tabela 16.

Tabela 16 – Incidência do campo “Ciência da Informação” em Áreas de Pesquisa e Categorias da *Web of Science* de documentos da base de dados sobre “Ciência Aberta” e suas vertentes (1970-2018).

	Quantidade	Registros	Porcentagem	Posição na Categoria da <i>Web of Science</i>	Posição na Área de Pesquisa
<b>Ciência Aberta (1970-2018)</b>	1.117	120	10,50%	2º	1º
<b>Acesso Aberto (1990-2018)</b>	49.810	1990	3,90%	4º	9º
<b>Fontes Abertas (1973-2018)</b>	41.418	1110	2,60%	12º	11º
<b>Dados Abertos (1978-2018)</b>	4.549	430	9,40%	6º	3º
<b>Ciência Cidadã (1997-2018)</b>	2.784	38	3,10%	30º	25º
<b>Educação Aberta (1971-2018)</b>	753	26	3,40%	8º	5º

Fonte: Elaborado pelo autor (2020).

Nota-se que há uma preponderância do termo “Ciência da informação e biblioteconomia” nos campos “Área de pesquisa” e “Categoria da *Web of Science*” em Ciência Aberta e nas suas seis vertentes aqui listadas. Porém, tendo em vista a especificação de cada desdobramento da Ciência Aberta, nem todos esses têm o tema “Ciência da Informação e biblioteconomia” dentre as 20 primeiras classificações temáticas de áreas do conhecimento ou categorias da pesquisa dos campos citados.

Em especial, Fontes Abertas e Ciência Cidadã não têm como prioridade a classificação de Ciência da Informação e biblioteconomia justamente pela especialização que estas

vertentes detêm. Enquanto a primeira se aproximou mais de campos de saber da Ciência da Computação e Telecomunicações, a Ciência Cidadã tem um escopo mais voltado para Ecologia e Geografia.

Contudo, há outras coincidências de temas a serem destacadas na análise dos campos “Área de pesquisa e “Categorias da *Web of Science*”, mas no que tange justamente a essas outras áreas do conhecimento. Como há certa correlação nos recortes levantados, vale ressaltar as correlações temáticas dentre as 20 primeiras classificações anteriormente apresentadas no Quadro 8.

Quadro 8 – Coincidências de temas no campo “Categorias da *Web of Science*” entre o recorte de documentos da base de dados sobre “Ciência Aberta” e suas vertentes (1970-2018).

	Ciência Aberta	Acesso Aberto	Fontes Abertas	Dados Abertos	Ciência Cidadã	Educação Aberta
<i>Computer science information systems</i>	X	X	X	X	X	X
<i>Computer science interdisciplinary applications</i>	X	X	X	X	X	X
<i>Computer science theory methods</i>	X	X	X	X	X	X
<i>Education educational research</i>	X	X	X	X	X	X
<i>Engineering</i>	X	X	X	X	X	X
<i>Multidisciplinary sciences</i>	X	X	X	X	X	
<i>Environmental sciences ecology</i>	X	X		X	X	
<i>Computer science artificial intelligence</i>	X		X	X		X
<i>Computer science hardware architecture</i>	X		X	X		X
<i>Computer science software engineering</i>			X	X		X
<i>Mathematical computational biology</i>	X		X	X		
<i>Biology</i>	X				X	
<i>Biochemistry molecular biology</i>	X	X				
<i>Telecommunications</i>			X	X		X

Fonte: Elaborado pelo autor (2020).

No levantamento feito, 14 temas coincidentes foram encontrados. Porém, dentre estes, “*Computer Science, Information Systems*” (Ciência da computação, sistemas da informação) e “*Computer Science, Interdisciplinary Applications*” (Ciência da computação, aplicações interdisciplinares) apareceram concomitantemente nas pesquisas sobre Ciência Aberta e todas

as vertentes selecionadas, enquanto “*Computer Science, Software Engineering*” (Ciência da computação, engenharia de software) e “*Telecommunications*” (Telecomunicações) não foram listados dentre os resultados de Ciência Aberta, mas estavam dentre as três vertentes.

Uma vez que a “Categoria da *Web of Science*” é uma classificação que preza pela especificidade da área do conhecimento pesquisada, deixando de pertencer à hierarquia da indexação pela “Área de pesquisa”, uma análise desse campo aumenta as coincidências por conta de diminuição de temas, mas também pode apresentar maiores aproximações, conforme as indicações apresentadas no Quadro 9.

Quadro 9 – Coincidências de temas no campo “Área de pesquisa” entre o recorte de documentos da base de dados sobre “Ciência Aberta” e suas vertentes (1970-2018)

	Ciência Aberta	Acesso Aberto	Fontes Abertas	Dados Abertos	Ciência Cidadã	Educação Aberta
<b>Computer science</b>	X	X	X	X	X	X
<b>Engineering</b>	X	X	X	X	X	X
Education educational research	X	X	X	X	X	X
Science technology other topics	X	X	X	X	X	X
Environmental sciences	X	X	X	X	X	
Chemistry	X	X	X	X		
Biochemistry molecular biology	X	X	X			
Business economics	X	X	X			
Neurosciences neurology	X	X				X
Mathematical computational biology	X		X	X		
Communication	X			X	X	X
Physics	X	X	X			
Astronomy astrophysics	X	X			X	
Psychology	X			X		X
Social sciences other topics	X			X		X
<b>Telecommunications</b>			X	X		X

Fonte: Elaborado pelo autor (2020).

Como previsto, o campo “Área de pesquisa” aumentou para 16 temas reincidentes dentre as 20 primeiras classificações em quantidade. Porém, cabe ressaltar que 6 dos 12 temas destacados em “Categorias da *Web of Science*” – “*Computer Science, Artificial Intelligence*” (Ciência da computação, inteligência artificial), “*Computer Science, Hardware &*

*Architecture*” (Ciência da computação, hardware e arquitetura), “*Computer Science, Information Systems*” (Ciência da computação, sistemas da informação) e “*Computer Science, Interdisciplinary Applications*” (Ciência da computação, aplicações interdisciplinares), “*Computer Science, Software Engineering*” (Ciência da computação, engenharia de software) e “*Computer Science, Theory & Methods*” (Ciência da computação, teoria e métodos) – encontram-se representados em “*Computer Science*” (Ciência da computação) nesta avaliação, conforme descrito no Quadro 10.

Quadro 10 – Aproximação dos temas de “Categoria da *Web of Science*” com o tema “Ciência da computação” em “Área de pesquisa” da base de dados.

<b>Tema em “Categoria da <i>Web of Science</i>”</b>	<b>Tema em “Área de pesquisa”</b>
Ciência da computação, inteligência artificial	Ciência da computação
Ciência da computação, hardware e arquitetura	
Ciência da computação, sistemas da informação	
Ciência da computação, aplicações interdisciplinares	
Ciência da computação, engenharia de software	
Ciência da computação, teoria e métodos	

Fonte: Elaborado pelo autor (2020).

Vale ainda ressaltar que esse campo, “*Computer Science*” (Ciência da computação) e “*Engineering*” (Engenharia), apareceu tanto para Ciência Aberta quanto para todas as outras vertentes. Da mesma forma, o tema “*Telecommunications*” (Telecomunicações) não apareceu listado dentre os resultados de Ciência Aberta, mas estava dentre três das mesmas vertentes que foram apresentadas na análise do campo “Categoria da *Web of Science*” – Fontes Abertas, Dados Abertos e Educação Aberta.

Na comparação direta dentre os termos que coincidem nessas análises dos campos “Categoria da *Web of Science*” e “Área de pesquisa”, apresentada no Quadro 11, registrou-se, em ambas as classificações, estes sete temas que acabam reiterando a área do conhecimento de “Ciência da informação e biblioteconomia”: “*Telecommunications*” (Telecomunicações), pertencente ao grupo temático “*Technology*” (Ciência e Tecnologia: Tecnologia) de “Área de pesquisa” – juntamente com “*Information Science & Library Science*” (Ciência da informação e biblioteconomia), “*Mathematical & Computational Biology*” (Biologia matemática e computacional) e “*Biochemistry & Molecular Biology*” (Bioquímica e biologia molecular), do grupo temático “*Life Sciences & Biomedicine*” (Ciência e tecnologia: Ciências da vida e

biomedicina), e, finalmente, “*Computer Science*” (Ciência da computação) – representando os seis temas anteriormente comentados, do campo “Categoria do *Web of Science*” –, “*Engineering*” (Engenharia), “*Education & Educational Research*” (Educação e pesquisa educacional), do grupo temático “*Social Sciences*” (Ciências sociais).

Quadro 11 – Correlação dos temas coincidentes em “Categoria da *Web of Science*” e “Área de pesquisa” nos grupos temáticos desse campo, no recorte em documentos da base de dados sobre “Ciência Aberta” e suas vertentes (1970-2018).

<b>Temas coincidentes em "Categoria da <i>Web of Science</i>" e "Área de pesquisa"</b>	<b>Grupos temáticos de "Área de pesquisa"</b>
Telecomunicações	Ciência e Tecnologia: Tecnologia
Ciência da Informação e Biblioteconomia	
Biologia Matemática e Computacional	Ciência e tecnologia: Ciências da vida e Biomedicina
Bioquímica e Biologia Molecular	
Ciência da Computação	Ciências Sociais
Engenharia	
Educação e Pesquisa Educacional	

Fonte: Elaborado pelo autor (2020).

Vale ressaltar que, apesar da quantidade de registros sobre Ciência Aberta na *Web of Science* se restringir a 1.117 documentos, só superou o que foi registrado em Educação Aberta – um total de 753 documentos. Mesmo assim, cabe destacar a anterioridade do registro daquela terminologia na base de dados. Com a indexação de documentos publicados desde 1970, o campo antecede todas aquelas identificadas como suas vertentes, sendo, por isso, escolhido como norte para tal comparação.

#### 4.5 O CONCEITO DE CAMPO CIENTÍFICO COM UMA ANÁLISE DE DOMÍNIO NOS RESULTADOS

Finalizando a apresentação dos resultados coletados nas três bases de dados, bem como as inferências e conclusões após cada exposição, faz-se necessário explanar a evolução do escopo temático e terminológico da noção de Ciência Aberta por meio do repertório teórico já apresentado. A combinação das áreas temáticas das áreas e da exposição do vocabulário utilizado, tanto pelas bases de dados quanto pelos autores, acaba por consolidar

uma representação das tendências do estudo em Ciência Aberta, pelo menos dentro do período do corte deste estudo, de 1970 até 2018.

Assim, parte-se da análise empírica para a abstração conceitual do referencial teórico. Uma vez que cada recorte e apuração da exposição sobre os resultados trouxe breves associações do repertório conceitual, manifesta-se a partir de agora sobre a relação desse conteúdo com as teorias de Análise de Domínio e Campo Científico, para, em seguida, citar algumas considerações sobre os pontos de interseção entre as três bases de dados.

#### **4.5.1 Uma visão geral do Domínio na Ciência Aberta**

Como relatado anteriormente, cada base de dados escolhida possui especificidades únicas que auxiliam a investigação de documentos sobre Ciência Aberta. A pesquisa reuniu e unificou os resultados da *Web of Science*, *Scopus* e DOAJ para embasar a análise de dados e aproveitar a sofisticação de algumas indexações e o registro de classificações de autores. No entanto, é certo que, exclusivamente, cada plataforma também apresenta falhas em registros, padronizações erráticas ou inconsistências nos metadados.

Enquanto a *Web of Science* é uma base de dados de acesso restrito, seu acervo conta com um bom volume de metadados sobre artigos de periódicos e capítulos de livro, apesar de haver algumas falhas na indexação nos campos de palavras-chave de autores e da própria *Clarivate Analytics*. Da mesma forma, a base de dados oferece recursos práticos e confiáveis para a extração de dados e contém um repertório de classificação temática de áreas de conhecimento exclusivo, bem como uma padronização de indexação bem fundamentada e organizada quanto suas categorias das áreas de conhecimentos.

A *Scopus* também tem acesso restrito, mas trouxe um repertório de documentos de diferentes espécies, indo de conferências a capítulos de livros e artigos de periódicos – dos quais, em sua grande maioria, também foram encontrados na *Web of Science*. Igualmente, a base de dados conta com palavras-chave de autores e dos indexadores da *Elsevier* em grande parte dos documentos que foram encontrados, facilitando o estudo sobre este campo. No entanto, os documentos na *Scopus* não contam com nenhuma classificação quanto área de conhecimento.

Quanto ao DOAJ, apesar de ter limitações quanto aos seus recursos, foi a única das bases de dados escolhidas de acesso aberto – além de ser uma das plataformas mais comprometidas em aumentar a visibilidade e acessibilidade de periódicos científicos de

acesso aberto. Todos os seus registros tiveram que ser verificados e copiados manualmente, o que permitiu observar a inconsistência de algumas repetições por conta de políticas de registro. Isto decorre do fato de os autores terem a possibilidade de registrar mais de uma instituição de pesquisa. No entanto, a base de dados confirmou a importância dos documentos coincidentes com a *Web of Science* e a *Scopus*, além de trazer informações sobre idioma e local de publicação de mais de 90% de seus registros.

A unificação dos resultados, com 96 artigos nas três bases – de um total de 2.254 documentos sem duplicidades – reforçou a importância na escolha de cada plataforma para verificar informações quanto ao local e idioma, as palavras-chave e a análise de áreas de conhecimento, quase que individualmente, ou em confronto com as outras duas. No entanto, há uma diferença entre forma e conteúdo dos resultados unificados, pois, uma vez que as informações sobre o formato dos documentos, idioma e local de publicação se relacionam com a forma, as classificações de palavras-chave e as áreas de conhecimento ajudam a fundamentar a investigação sobre a semântica dos documentos com o termo Ciência Aberta.

A base de dados DOAJ foi fundamental para verificação de local de publicação e idioma, pois, apesar da base de dados comprovar a importância do idioma inglês na produção científica sobre Ciência Aberta (fundamental para um estudo quanto às palavras-chave utilizadas), desbancou a preponderância dos Estados Unidos na primazia da publicação (dando destaque para a Índia) e apontou para maiores resultados em outros idiomas – com destaque para o alemão, espanhol e português. Como essa base de dados tem o caráter de promover o acesso aberto, esses resultados apontam para a incidência de estudos sobre a política da Ciência Aberta em países periféricos, bem como a maior valorização de idiomas, diferentes do padrão de internacionalização com o inglês. Assim, aproveitou-se muito do conteúdo dessa base de dados para fazer aplicar os pontos de estudos documentais e estudos bibliográficos da Análise de Domínio.

Tanto a análise das palavras-chave quanto a das áreas de conhecimento procuraram investigar o contexto do conteúdo do material indexado nas bases de dados. O levantamento sobre palavras-chave permitiu levantar as ideias e temas da produção documental das três bases de dados escolhidas para comparação. No entanto, foram fundamentais o repertório e a qualidade das classificações na *Scopus*, tanto dos termos escolhidos pelos autores, quanto dos selecionados pelos indexadores da *Elsevier*. De maneira que, tanto os dados coletados acrescentaram informações sobre o recorte material, como o enorme volume de indexações de

palavras-chave contribuiu para os pontos da Análise de Domínio, bem como nos pontos de estudos documentais, de gênero, indexação e recuperação de especialidades.

Tal análise permitiu se debruçar sobre as áreas de conhecimento na base de dados *Web of Science*, pois essa base contava com dois campos sobre esse assunto – um mais relevante para especificidades dos documentos e outro para reunir e classificar de acordo com uma padronização hierarquizada em cinco naturezas de áreas de conhecimento na produção científica. Assim, foi possível observar quais setores da classificação indexada se inter-relacionavam, permitindo extrair sete categorias de áreas de conhecimento diferentes, mas com ligações semânticas e hierárquicas. Esses dados foram fundamentais para verificar os pontos de estudo bibliográficos, indexação e recuperação de especialidades na Análise de Domínio sobre o recorte da Ciência Aberta.

Desta forma, a investigação dos dados sobre os documentos levantados permitiu relacionar a forma e conteúdo do Domínio da Ciência Aberta. O Quadro 12 apresenta um resumo das considerações levantadas quanto a cada área examinada, correlacionando local de publicação, idioma, autores, palavras-chave e áreas de conhecimento dos resultados analisados.

Quadro 12 – Resultados da Análise de Domínio de Ciência Aberta da amostragem.

(continua)			
Quanto ao material		Registros	% de 2254
	Periódicos	2537	76%
	Livros	763	23%
	Conferências	61	2%
Quanto ao local de publicação		Registros	% de 2254
	Estados Unidos	623	28%
	Reino Unido	417	19%
	Índia	231	10%
	Holanda	206	9%
	Alemanha	157	7%
Quanto ao idioma		Registros	% de 2254
	Inglês	1883	84%
	Alemão	125	6%
	Espanhol	82	4%
	Português	63	3%
	Italiano	32	1%
Quanto aos coautores		Registros	
	Levshina, Tanya	32	
	Bockelman, Brian	31	
	Garzoglio, Gabriele	27	
	Sfiligoi, Igor	27	
	Tan, Fraser Elisabeth	24	



Quadro 12 – Resultados da Análise de Domínio de Ciência Aberta da amostragem.

		(conclusão)	
Quanto às palavras-chave		Registros	26461%
	science	778	2,9%
	open science	861	3,3%
	open access	168	0,01%
	reproducibility	159	0,01%
	research	117	0,004%
Quanto à área de conhecimento		Registros	% de 1117
	Ciência da Computação	252	22,2%
	Ciência da Informação e Biblioteconomia	120	10,5%
	Engenharia	98	8,6%
	Psicologia	82	7,2%
	Ciência da Tecnologia	81	7,1%

Fonte: Elaborado pelo autor (2020).

A relação apresentada demonstra a definição de Domínio da Ciência Aberta em um viés construído por um levantamento da literatura formal. Apesar de muitos conceitos e definições da Ciência Aberta derivarem de manifestos, livros abertos e iniciativas, a formalidade exigida no ambiente extremamente controlado das bases de dados fornece um arcabouço rígido o suficiente para a Análise de Domínio, e, também, acaba subsidiando as considerações necessárias para analisar o Campo Científico da Ciência Aberta.

#### 4.5.2 Os contornos de um Campo Científico na Ciência Aberta

Ciência Aberta é a definição geral de um termo guarda-chuva que deve ser interpretado como um processo que mobiliza vieses e diferentes propensões, tendo, por isso, uma acepção terminológica ampla e sujeita a diferentes perspectivas. No entanto, classificar e qualificar essa definição é uma tarefa complexa, não só pelo volume de produção acadêmica sobre o assunto, mas também pela dificuldade semântica de compreender a definição terminológica da Ciência Aberta, bem como a maneira em que esse movimento se transforma.

Como uma forma de abarcar melhor a evolução do escopo temático e terminológico de Ciência Aberta, haja vista a dimensão e os desdobramentos que o tema tem, optou-se por traçar os contornos do Campo Científico (BOURDIEU, 1976, p. 89) deste movimento. Assim, foi feita um amálgama delineado por meio da Análise de Domínio da literatura formal, coletada por meio de um recorte feito em três bases de dados. Tal arranjo permitiu esboçar a situação hierárquica de seus agentes e mensurar a materialidade dos documentos analisados, o

grau de subordinação do local de publicação das editoras e os idiomas mais publicados dentro do acervo pesquisado, assim como o volume de publicação dos coautores, o *ranking* das palavras-chave mais utilizadas e as áreas de conhecimento classificadas, conforme apresentado no Quadro 12.

Na apresentação sobre o referencial teórico, destacou-se três aspectos sobre o Campo Científico: o capital, a autonomia e as “revoluções”. Sobre esse primeiro ponto, vale ressaltar que atende diretamente a distribuição do capital de competência sobre o campo, bem como a autoridade científica pelas relações entre os agentes. A violência simbólica dos conflitos internos dentro de um campo reflete um campo magnético com linhas de força entre agentes e instituições que se opõem e agregam em determinados lugares e momentos. Uma vez que o recorte analisou a literatura formal sobre a Ciência Aberta, considerou-se a relevância da regionalidade das editoras que publicam o conteúdo acadêmico da área.

Na discussão sobre como a autonomia do campo é impactada pela produtividade científica, vale destacar que um Campo Científico demonstra maior independência quando preserva a igualdade no empenho de seus agentes e mantêm um grau de influência externa em outros campos. A correlação da Ciência Aberta com as suas vertentes demonstra que, apesar de praticamente todas essas terem um maior volume de publicações indexadas, conforme demonstrado na Tabela 16, aquela exerce grande influência nessas, não só pela correspondência nas classificações de áreas de conhecimento, mas também pela anterioridade dos trabalhos publicados, conforme mostrado no Quadro 11. E, mesmo com duas das três bases de dados analisadas terem um viés mercadológico, comprovou-se que o movimento da Ciência Aberta manteve certa justeza no esforço dos coautores na literatura.

Vale ressaltar que não foi possível identificar indícios de “revoluções” internas ou mudanças estruturais no Campo Científico da Ciência Aberta. Apesar do demonstrado com as relações dos grupamentos de autores, de acordo com a cronologia de suas publicações, não houve uma reinterpretação de elementos e práticas científicas no debate atual sobre o movimento da Ciência Aberta. Alguns desdobramentos de pesquisas desse Campo Científico puderam ser identificados pelo volume de palavras-chave e sua publicação ao longo do tempo (de acordo com as Figuras 20, 21, 22, 23, 24 e 25), mas, nesses casos, as publicações tinham um viés mais próximo do aperfeiçoamento de parcerias com diferentes áreas de conhecimento e do aprofundamento do conteúdo de pesquisas.

Assim, relacionando esses três aspectos, o delineamento do Campo Científico da Ciência Aberta demonstra o destaque dos Estados Unidos, Inglaterra e Suíça, não só pelo

volume de publicações, mas também pela anterioridade dessas. Destacam-se alguns países, como Alemanha, Brasil e Holanda, com publicações mais recentes e uma quantidade considerável de publicações, no entanto, o Capital Científico se mantém preponderante nos três países citados previamente.

De igual natureza, o Campo Científico da Ciência Aberta manteve um elevado grau de autonomia, estabelecendo equiparações com vertentes que podem ser identificadas como desdobramentos temáticos e terminológicos, como Acesso Aberto, Fontes Abertas, Dados Abertos, Ciência Cidadã e Educação Aberta. Tal independência também denota preponderância na área de conhecimento da Ciência da Informação e Biblioteconomia, mas também permite estabelecer correlações em matérias de áreas da saúde, além de Ciência da Computação, Engenharia e Educação.

Quanto ao aspecto identificado como “revoluções” do Campo Científico, cabe destacar que há certa equidade no número de coautorias na Ciência Aberta, sem a formação de grupamentos que questionem o ordenamento interno do campo. Destaca-se ainda a relevância de palavras-chaves como “*big data*”, “*standarts*”, “*open access publications*”, “*data sharing*” e “*cloud computing*”, por conta de figurarem na publicação de documentos mais recentes. Tal ocorrência pode ajudar a reafirmar a autonomia do Campo Científico da Ciência Aberta ao caracterizar o aprimoramento de parcerias e a interpenetração em áreas diferentes de Ciência da Informação e Biblioteconomia.

Isto posto, a averiguação dos resultados dessa amostragem – toda fundamentada na literatura formal – estabeleceu alguns dos aspectos que configuram o Campo Científico da Ciência Aberta. A Análise de Domínio auxiliou no exame do grau de autonomia e na análise do Capital Científico, enquanto o exame do produto das pesquisas sustentou os contornos para apresentar o campo, as relações entre seus agentes e o ordenamento interno do campo.

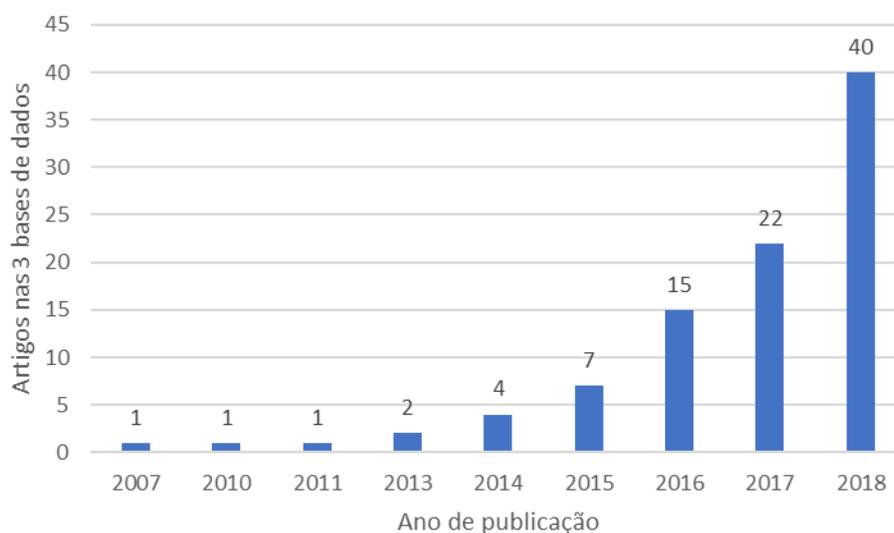
#### **4.5.3 Uma breve análise sobre a intercessão entre as bases de dados**

Com apontado anteriormente, excluindo as duplicidades, foram coletados um total de 2.254 documentos sobre Ciência Aberta nas bases de dados *Web of Science*, *Scopus* e *DOAJ*. Desses, apenas 93 artigos estavam nas três plataformas – o que representa um percentual de 0,04%.

Apesar do pequeno valor, há que se tecer alguns comentários sobre a interseção entre as bases de dados da amostra. A Figura 26 apresenta a evolução da indexação desses artigos

ao longo dos anos, enquanto a Tabela 17 exhibe os países em que tais documentos foram publicados e a Tabela 18 retrata os periódicos em que mais se destacaram.

Figura 26 – Artigos sobre Ciência Aberta registrados simultaneamente nas bases de dados *Web of Science*, *Scopus* e DOAJ no período de 2007-2018.



Fonte: Elaborado pelo autor (2020).

Tabela 17 – Países em que foram publicados os artigos sobre Ciência Aberta registrados simultaneamente nas bases de dados *Web of Science*, *Scopus* e DOAJ no período de 2007-2018.

Países	Registros comuns
Estados Unidos	32
Suíça	25
Inglaterra	23
Alemanha	3
Brasil	2
França	2
Holanda	2
Itália	2
Coreia do Sul	1
Noruega	1

Fonte: Elaborado pelo autor (2020).

Tabela 18 – Periódicos que mais se destacaram por publicar os artigos sobre Ciência Aberta registrados simultaneamente nas bases de dados *Web of Science*, *Scopus* e DOAJ no período de 2007-2018

<b>Periódicos</b>	<b>Artigos Comuns</b>
PeerJ	7
Frontiers in Psychology	6
IUCrJ	4
Publications	4
Collabra: Psychology	3
Data in Brief	3
Frontiers in Neuroinformatics	3
Frontiers in Neuroscience	3
PLoS ONE	3
AIB Studi	2
Data	2
EPJ Web of Conferences	2
Frontiers in Human Neuroscience	2
Human Genomics	2
Insights: the UKSG Journal	2
ISPRS International Journal of Geo-Information	2
PeerJ Computer Science	2
PLoS Biology	2

Fonte: Elaborado pelo autor (2020).

Ao mesmo tempo em que a Figura 26 reflete o acréscimo da interseção do processo de indexação entre a *Web of Science*, *Scopus* e DOAJ, a Tabela 17 destaca a preponderância dos mesmos países no *ranking* com todos os documentos coletados anteriormente.

Quanto a Tabela 18, cabe destacar que houve outras 39 revistas científicas que só foram contabilizadas uma vez. Porém, do montante apresentado no levantamento, grande parte das revistas tem a temática com áreas de conhecimento relacionadas à saúde, entretanto algumas fazem parte da Ciência da Informação e da Ciência da Computação.

Cabe ressaltar também que, em um levantamento sobre coautoria, encontrou-se 942 autores, porém, desses, 14 tiveram duas publicações indexadas pelas três plataformas. Do mesmo recorte, três publicaram juntos em dois artigos (primeiro em 2015, depois em 2016), dois foram coautores em duas publicações (ambas em 2017) e outros dois em somente um (em 2014).

Esse levantamento reproduz como a interseção entre as três bases de dados tem uma relevância que aumenta progressivamente, e que, com o tempo, terá estrutura com embasamento o suficiente para refletir o Campo de Domínio sobre Ciência Aberta pela literatura formal.

## 5 CONSIDERAÇÕES FINAIS

A presente pesquisa objetivou analisar o escopo temático e terminológico do conceito de Ciência Aberta, evidenciando as vertentes dentro de seu campo semântico. Para tanto, o trabalho buscou investigar, na literatura formal, os documentos com o termo “*open science*” para, a partir dos resultados, elaborar uma Análise de Domínio que oferecesse subsídios sobre o Campo Científico da Ciência Aberta. Tal constructo teórico, associado aos procedimentos da metodologia, permitiram traçar um delineamento histórico sobre o conceito de Ciência Aberta nas três bases de dados escolhidas para a amostragem experimental. O mapeamento de coocorrências, correlações e iterações entre os documentos analisados permitiu organizar os recortes temporais e temáticos extraídos.

A análise empírica permitiu verificar, com relação aos suportes materiais, a preponderância dos artigos na literatura formal examinada, de maneira que há um predomínio de documentos escritos na língua inglesa. Da mesma forma, confirmou-se a hegemonia de publicações sobre Ciência Aberta em editoras científicas situadas nos Estados Unidos, Reino Unido e Índia. No entanto, esse último país se destacou devido ao esforço de seus autores no registro e divulgação de seus trabalhos na base de dados DOAJ.

Ao apurar a evolução cronológica e temática das palavras-chave, os resultados combinados das bases de dados, com destaque para o sofisticado repertório de metadados da plataforma *Scopus*, permitiu cartografar os vínculos entre os termos “*open science*”. Enquanto palavras-chave como “*transparency*”, “*big data*” e “*open access*” têm uma convergência mais recente, termos como “*open source*”, “*metadata*”, “*data sharing*” e “*scholarly communication*” se correlacionam a mais tempo. Uma vez que os resultados da base de dados *Web of Science* ofereceram classificações sobre a área de conhecimento dos documentos encontrados, foi possível traçar as correspondências entre a Ciência Aberta e algumas de suas vertentes. Com destaque para “Ciência da Informação e Biblioteconomia”, verificou-se também uma forte correlação com “Ciência da Computação”, “Engenharia” e “Educação e Pesquisa Educacional”.

De igual modo, o levantamento contabilizou um total de 15.094 autores, permitindo constatar a rede de ligações entre eles, destacando os grupos que refletiriam as relações da comunidade do Campo Científico da Ciência Aberta. Essa inferência sobre as redes complexas dos coautores, assim como as deduções sobre o gênero dos materiais analisados e

as ilações sobre as palavras-chave e áreas de conhecimento foram fundamentais para elaborar o constructo da Análise de Domínio.

Ao escolher três dos 11 modos sugeridos por Hjørland, teve-se a intensão de descomplexificar a tarefa hercúlea de fazer uma Análise de Domínio na literatura formal da Ciência Aberta. Uma vez que analisar os registros da literatura geral, com livros abertos, manifestos e iniciativas, seria como a mitológica limpeza dos Estábulos de Augias, padronizar e comparar os resultados de bases de dados do recorte selecionado seriam uma operação mais pragmática e condizente com escopo desse trabalho.

Assim, realizada a Análise de Domínio nos documentos com o termo Ciência Aberta, comprovou-se a dimensão multicultural e regional da produção acadêmica dos resultados analisados. Visto que o movimento da Ciência Aberta busca promover a pesquisa transparente, sem filtros, barreiras ou preconceitos, as comunidades discursivas são impactadas diretamente por tal incentivo, pois sua designação conceitual na literatura se torna cada vez mais complexa, justamente pelo grau de abrangência que a granularidade que a liberdade de participação que a Ciência Aberta defende. Daí a dificuldade de, além de definir a terminológica semântica sobre “ciência aberta”, compreender a sua transformação.

A fundamentação do Campo Científico da Ciência Aberta é essencial para a análise do desenvolvimento de sua literatura formal, bem como da evolução conceitual do termo “Ciência Aberta. É importante destacar que os resultados apontam para um Campo com autonomia e muito bem estruturado por seus Capitais Científicos, e que ainda oferece muitos subsídios para suas vertentes.

Como a literatura formal é regida por um processo de divulgação, com uma série de processos de autonomização e um rígido sistema regido por leis próprias, qualquer evolução no escopo analisado é bem sensível e demonstra mais tendências do que abruptos corolários. Algumas definições encontradas suportam certa autossuficiência quanto a semântica terminológica investigada, porém, a amostragem não demonstrou uma evolução declivosa em direção a conceitos que fossem estranhos à tradicionalmente ampla definição de Ciência Aberta.

Não surpreendentemente, o conteúdo de documentos analisados nas bases de dados é formado quase que exclusivamente por artigos, quase todos em língua inglesa. Apesar destes contribuírem para montar um retrato da literatura formal, tal qual este trabalho comprometeu-se a apresentar, é necessário também ponderar sobre o processo de indexação das plataformas. Uma vez que as bases de dados, principalmente no caso da Scopus e da Web of Science, tem

processos de registro, classificação e indexação arbitrários, o material em seus acervos não representa exatamente o que foi pesquisado, mas sim a maneira em que a plataforma o indexou.

Assim, a pesquisa em bases de dados representa perspectivas e orientações arbitrárias das instituições por trás dessas. No entanto, para um estudo cientométrico mais aprofundado, é necessário extrair o conteúdo buscado em mais de uma plataforma, destacando período em que foi feita a pesquisa, para se tirarem as conclusões com relação às classificações e indexações efetuadas individualmente por cada base de dados, bem como em conjunto. De certa maneira, os títulos, os resumos e as palavras-chave de autores são os critérios mais independentes da autoridade classificatória das plataformas utilizadas. À vista disto, cabe salientar que este trabalho se comprometeu a examinar a literatura formal da Ciência Aberta, mas não deixa de que pode ter havido critério de indexação discricionária por parte das bases de dados.

Levando-se em consideração esses aspectos, a presente pesquisa é relevante pelo fato de analisar as publicações científicas, sob o viés da Análise de Domínio e do Campo Científico, justamente para compreender a atual situação da Ciência Aberta na literatura formal, segundo o viés das três bases de dados analisadas. Assim, a investigação da amostragem coletada tem o propósito de pormenorizar o conceito e a representação do movimento da Ciência Aberta nas publicações científicas.

Da mesma forma, observa-se a necessidade de pautar uma agenda com possibilidades futuras de pesquisa, principalmente quanto aos critérios para determinar os principais atores do Campo Científico da Ciência Aberta. Enxerga-se uma senda plausível na análise sobre o predomínio do contexto anglófono dos artigos encontrados, bem como a hegemonia da classificação das áreas de conhecimento pela Web of Science, a saber: Telecomunicações, Ciência da Informação e Biblioteconomia, Biologia Matemática e Computacional, Bioquímica e Biologia Molecular, Ciência da Computação, Engenharia e Educação e Pesquisa Educacional.

Outra oportunidade de investigação sobre as autoridades no campo poderia ser feita com uma análise mais metódica com relação aos pesquisadores e autores que publicaram os documentos nas bases de dados. Examinar estes agentes permitiria analisar o potencial de influência de cada um, de acordo com os seus grupos de estudo, campos de pesquisa e escolha de periódicos para submissão. De igual natureza, uma verificação acerca dos pesquisadores de temas específicos, que não são estudiosos específicos sobre a Ciência Aberta, pode revelar de maneira mais apurada qual o verdadeiro papel deste tema para suas pesquisas.

Igualmente, há que se falar nos profissionais que trabalham indexando os documentos nas bases de dados segundo diretrizes de editores, técnicos e diretores dessas plataformas. Uma vez que esses agentes, tanto as bases de dados quanto os seus indexadores, provocam uma forte interferência na configuração da organização do conhecimento ao indexar documentos que figuram na literatura formal, o exame de suas atividades é fundamental para compreender os seus papéis como agentes no Campo Científico.

Dado o exposto, apesar desse trabalho despertar a sensação da necessidade de uma constante observação das publicações em base de dados, que tem o acervo constantemente ampliado e revisado, teve-se a finalidade de contribuir para a representação da literatura formal da Ciência Aberta, na perspectiva acadêmica. De tal forma, a investigação visa contribuir para o desenvolvimento de aspectos teóricos, metodológicos e práticos nos estudos, produção e representação da instrumentalização das bases de dados para estudar a própria produtividade acadêmica. Portanto, em trabalhos futuros, sugere-se uma ampliação da pesquisa qualitativa e quantitativamente, acrescentando mais bases de dados na análise e aprofundando os recortes temporais, materiais e temáticos, bem como as relações e especificidades na análise da amostragem.

## REFERÊNCIAS

ALBAGLI, S. Ciência aberta em questão. *In*: ALBAGLI, S.; MACIEL, M. L (Orgs.) **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. Disponível em: [http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta\\_questoes%20abertas\\_PORTUGUES\\_DIGITAL%20%285%29.pdf](http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20%285%29.pdf). Acesso em: 20 fev. 2020.

ALBAGLI, S.; APPEL, A. L.; MACIEL, M. L. E-science e ciência aberta: questões em debate. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais eletrônicos...** Florianópolis: UFSC, 2013. Disponível em: <https://ridi.ibict.br/handle/123456789/465>. Acesso em 20 fev. 2020.

ALBAGLI, S.; APPEL, A. L.; MACIEL, M. L. E-Science, ciência aberta e o regime de informação em ciência e tecnologia. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 7, p. 1-20, 2014. Disponível em: <https://ridi.ibict.br/bitstream/123456789/854/1/124-540-1-PB.pdf>. Acesso em 20 fev. 2020.

BAGLEY, P. **Extension of programming language concepts**. Philadelphia: University City Science Center, 1968. Disponível em: <https://apps.dtic.mil/dtic/tr/fulltext/u2/680815.pdf>. Acesso em 20 fev. 2020.

BARBIER, F. **História das bibliotecas**. São Paulo: Editora da Universidade de São Paulo, 2018.

BUDAPEST OPEN ACCESS INITIATIVE. **Budapest Open Access Initiative**, 2020. Disponível em <https://www.budapestopenaccessinitiative.org/>. Acesso em: 20 fev. 2020

BOURDIEU, P. **A distinção: crítica social do julgamento**. 2. ed. Porto Alegre: Zouk, 2013.

BOURDIEU, P. Le champ scientifique. **Actes de la Recherche en Sciences Sociales**, n. 2/3, p. 88-104, 1976. Disponível em: [https://www.persee.fr/doc/arss\\_0335-5322\\_1976\\_num\\_2\\_2\\_3454](https://www.persee.fr/doc/arss_0335-5322_1976_num_2_2_3454). Acesso em 10 dez. 2019.

BOURDIEU, P. **Os Usos sociais da ciência: por uma sociologia clínica do campo científico**. São Paulo: Unesp, 2004.

BOURDIEU, P. **Questões de Sociologia**. Rio de Janeiro: Marco Zero, 1983.

BOURDIEU, P. **Razões práticas: sobre a teoria da ação**. São Paulo: Papius, 1996.

CARVALHO, L. A abordagem sociocultural da produção de conhecimento científico. *In*: MARTELETO, R. M.; PIMENTA, R. M. (Orgs.) **Pierre Bourdieu e a produção social da cultura, do conhecimento e da informação**. - Rio de Janeiro: Garamond, 2017, p.190-206. Disponível em: [https://edisciplinas.usp.br/pluginfile.php/4952291/mod\\_resource/content/1/pierre\\_bourdieu\\_e\\_book.pdf](https://edisciplinas.usp.br/pluginfile.php/4952291/mod_resource/content/1/pierre_bourdieu_e_book.pdf). Acesso em 20 fev. 2020.

CLARIVATE ANALYTICS. **About us**, 2020. Disponível em: <https://https://clarivate.com/webofsciencegroup/about-us/>. Acesso em 25 maio 2020.

CLARIVATE ANALYTICS. Principal Coleção do Web of Science ajuda: áreas de pesquisa. **Clarivate Analytics – Images Web of Knowledge**, 2019a. Disponível em: [https://images.webofknowledge.com/WOKRS511B5/help/pt\\_BR/WOS/hp\\_research\\_areas\\_ea\\_sca.html](https://images.webofknowledge.com/WOKRS511B5/help/pt_BR/WOS/hp_research_areas_ea_sca.html). Acesso em 10 abr. 2020.

CLARIVATE ANALYTICS. Principal Coleção do Web of Science Ajuda: Categorias Web of Science. **Clarivate Analytics – Images Web of Knowledge**, 2019b. Disponível em: [https://images.webofknowledge.com/WOKRS511B5/help/pt\\_BR/WOS/hp\\_research\\_areas\\_ea\\_sca.html](https://images.webofknowledge.com/WOKRS511B5/help/pt_BR/WOS/hp_research_areas_ea_sca.html). Acesso em 10 abr. 2020.

DANTAS, M. Informação como trabalho e como valor. **Revista da Sociedade Brasileira de Economia Política**, Rio de Janeiro, n. 19, pp. 44-72, 2006. Disponível em: <http://marcosdantas.com.br/conteudos/wp-content/uploads/2013/03/informacao-trabalho-valor.pdf>. Acesso em 24 fev. 2020

DAVID, P. A. Can "Open Science" be Protected from the Evolving Regime of IPR Protections?. **Journal of Institutional and Theoretical**, Tubinga. v. 160, n. 1, p. 09-34, 2004. Disponível em: <https://www.jstor.org/stable/40752435?seq=1>. Acesso em 20 fev. 2020.

DAVID, P. A. The historical origins of ‘open science’: An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution. **Capitalism and Society**. v. 3, n. 2, p. 1-106, 2008. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.7499&rep=rep1&type=pdf>. Acesso em: 20 fev. 2020.

DAY, M. **DCC Digital curation manual**: installment on “metadata”. Bath: University of Bath, 2005. Disponível em: <https://pdfs.semanticscholar.org/1f03/ec11c0d7010548e474e5d1f2de1c37c65ec1.pdf>. Acesso em: 20 fev. 2020

DELFANTI, A. **Biohackers**: the politics of Open Science. London: Pluto Press, 2013. Disponível em: [http://delfanti.org/wp-content/uploads/2013/05/biohackers\\_book.pdf](http://delfanti.org/wp-content/uploads/2013/05/biohackers_book.pdf). Acesso em: 20 fev. 2020.

DIAS, C. C. A análise de domínio, as comunidades discursivas, a garantia da literatura e outras garantias. **Informação e Sociedade**, João Pessoa, v. 25, n. 2, p. 7-17, maio/ago. 2015. Disponível em: <http://www.ies.ufpb.br/ojs/index.php/ies/article/view/7/13741>. Acesso em: 20 fev. 2020.

DOAJ - DIRECTORY OF OPEN ACCESS JOURNALS. **About DOAJ (Directory of Open Access Journals)**, 2020. Disponível em: <https://doaj.org/about>. Acesso em 20 fev. 2020

FREIRE, I. M. A responsabilidade social da ciência da informação na perspectiva da consciência possível. **DataGramZero**, v.5, n.1, 2004. p.51-68. Disponível em: [https://ridi.ibict.br/handle/123456789/229?locale=pt\\_BR](https://ridi.ibict.br/handle/123456789/229?locale=pt_BR). Acesso em 22 fev. 2020.

GALOÁ. **Como incluir um DOI no Lattes?**, 2020. Disponível em: <https://galoa.com.br/deposito-doi-crossref/como-incluir-um-doi-no-lattes>. Acesso em 20 fev. 2020.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GIL, A. **Métodos de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.,

GUÉDON, Jean-Claude. Acesso aberto e divisão entre ciência predominante e ciência periférica. In: FERREIRA, Sueli Mara Soares Pinto; TARGINO, Maria das Graças (orgs.). **Acessibilidade e visibilidade de revistas científicas eletrônicas**. São Paulo, SP: Cengage Learning; Senac.

GUINCHAT, C.; MENOU, M. **Introdução geral às ciências e técnicas da informação e documentação**. Brasília: IBICT, 1994. Disponível em: <http://livroaberto.ibict.br/bitstream/1/1007/1/Introdu%C3%A7%C3%A3o%20geral%20%C3%A0s%20ci%C3%A7%C3%A2ncias%20e%20t%C3%A9cnicas%20da%20informa%C3%A7%C3%A3o%20e%20documenta%C3%A7%C3%A3o.pdf>. Acesso em 20 fev. 2020.

HJØRLAND, B. Concept theory. **Journal of the American Society for Information Science and Technology**, v. 60, n. 8, p. 1519–1536, 2009. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21082>. Acesso em: 20 fev. 2020.

HJØRLAND, B. Domain analysis in information science: eleven approaches – traditional as well as innovative. **Journal of Documentation**, v. 58, p. 422-462, 2002. Disponível em: [https://www.researchgate.net/publication/249366184\\_Domain\\_analysis\\_in\\_information\\_science\\_Eleven\\_approaches\\_-\\_Traditional\\_as\\_well\\_as\\_innovative](https://www.researchgate.net/publication/249366184_Domain_analysis_in_information_science_Eleven_approaches_-_Traditional_as_well_as_innovative). Acesso em: 20 fev. 2020.

HJØRLAND, B. Fundamentals of knowledge organization. **Knowledge Organization**, v. 30, n. 2, p. 87-111, 2003. Disponível em: <http://ppggoc.eci.ufmg.br/downloads/bibliografia/Hjorland2003.pdf>. Acesso em: 20 fev. 2020.

HJØRLAND, B. Library and information science and the philosophy of science. **Journal of Documentation**, v. 61 n. 1, p. 05-10. 2005. Disponível em: <https://doi.org/10.1108/00220410510577970>. Acesso em 20 fev. 2020.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in Information Science: domain analysis. **Journal of the American Society for Information Science**, v. 46, n. 6, p. 400425, 1995. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/%28SICI%291097-4571%28199507%2946%3A6%3C400%3A%3AAID-ASI%293E3.0.CO%3B2-Y>. Acesso em: 20 fev. 2020.

JACOBS, N. **Open Access: key strategic, technical and economic aspects**. Oxford: Elsevier, 2006.

JORGE V.; ALBAGLI, S. Pierre Bourdieu e a produção: do conhecimento científico. *In*: MARTELETO, R. M.; PIMENTA, R. M. (Orgs.) **Pierre Bourdieu e a produção social da cultura, do conhecimento e da informação**. Rio de Janeiro: Garamond, 2017, p.205-228.

Disponível em:

[https://edisciplinas.usp.br/pluginfile.php/4952291/mod\\_resource/content/1/pierre\\_bourdieu\\_book.pdf](https://edisciplinas.usp.br/pluginfile.php/4952291/mod_resource/content/1/pierre_bourdieu_book.pdf). Acesso em 20 fev. 2020.

KINLEY, R.; MARKIE, M. Wellcome Open Research, the future of scholarly communication? **LSE Impact of Social Sciences**, 2019. Disponível em:

<https://blogs.lse.ac.uk/impactofsocialsciences/2019/02/22/wellcome-open-research-the-future-of-scholarly-communication/>. Acesso em: 4 abr. 2020.

KURAMOTO, H. Acesso livre à informação científica: novos desafios. **Liinc em Revista**, v.4, n. 2, 2008. p.154-157. Disponível em:

<https://ridi.ibict.br/bitstream/123456789/154/1/Kuramoto%2c%20H.pdf>. Acesso em 20 fev. 2020.

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

LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES. Open Science and its role in universities: a roadmap for cultural change. **Advice Paper**, n. 24, 2018. Disponível em: <https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf>. Acesso em 6 abr. 2020.

LE COADIC, Y. F. **A Ciência da Informação**. Brasília: Briquet de Lemos, 1996.

MANN, S. *et al.* "Squeakeys": a friction idiophone, for physical interaction with mobile devices. *In*: IEEE CONSUMER ELECTRONICS SOCIETY GAMES, ENTERTAINMENT, MEDIA CONFERENCE, 7., 2015, Toronto. **Electronic proceedings** [...]. Toronto: University of Toronto, 2015. Disponível em:

[http://www.eyetap.org/docs/SqueaKEYS\\_MannJanzen2015.pdf](http://www.eyetap.org/docs/SqueaKEYS_MannJanzen2015.pdf). Acesso em 30 mar. 2020.

MEY, E. S. A. **Catálogo e descrição bibliográfica**: contribuições a uma teoria. Brasília: ABDF, 1987.

MICHAUD, C. Modelos e conhecimento. *In*: TARAPANOFF, K. (Org.). **Inteligência, informação e conhecimento**. Brasília: IBICT, 2006. p. 211-239. Disponível em:

<https://livroaberto.ibict.br/bitstream/1/465/1/Inteligencia,%20informa%20c3%a7%20a3o%20e%20conhecimento.pdf>. Acesso em: 20 fev. 2020

NIELSEN, M. **Reinventing discovery**. Nova Jersey: Princeton University Press, 2011.

ORTIZ, R. **A Sociologia de Pierre Bourdieu**. São Paulo: Olho D'Água, 2003.

PRIETO-DÍAZ, R. Domain analysis: An introduction. **ACM SIGSoft Software Engineering Notes**, v. 15, n. p: 47-54, 1990. Disponível em:

<https://dl.acm.org/doi/10.1145/382296.382703>. Acesso em 20 fev. 2020.

ROCHA, S. R. M. **Construção e análise do inventário do patrimônio religioso paraibano: informação como representação social**. 2011. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal da Paraíba, João Pessoa, 2011. Disponível em: <https://repositorio.ufpb.br/jspui/handle/tede/3903>. Acesso em: 25 fev. 2020.

RUCKER, D. D.; STERNTHAL, B. Estratégia de propaganda. *In*: CALDER, B. J.; TYBOUT, A. M. (Orgs). **Marketing**. São Paulo: Saraiva, 2013, p.223-230.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli**, Florianópolis, v. 15, n. 30, p. 1-31, out. 2010. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1>. Acesso em: 20 fev. 2020.

SMIRAGLIA, R. P. Universes, dimensions, domains, intensions and extensions: knowledge organization for the 21<sup>o</sup> century. *In*: NEELAMEGHAN, A.; RAGHAVAN, K. S. (Eds.) **Categories, contexts and relations in knowledge organization: Twelfth International ISKO Conference**. Mysore, Índia: ISKO, 2012. Disponível em: [https://www.ergon-verlag.de/isko\\_ko/downloads/aiko\\_vol\\_13\\_2012.pdf](https://www.ergon-verlag.de/isko_ko/downloads/aiko_vol_13_2012.pdf). Acesso em: 20 fev. 2020.

SOARES, M. D.; SANTOS, R. D. C. Ciência cidadã: o envolvimento popular em atividades científicas. **Ciência Hoje**, v. 47, p.38-43, 2011. Disponível em: [https://www.researchgate.net/publication/236952978\\_Ciencia\\_Cidada\\_o\\_envolvimento\\_popular\\_em\\_atividades\\_cientificas](https://www.researchgate.net/publication/236952978_Ciencia_Cidada_o_envolvimento_popular_em_atividades_cientificas). Acesso em: 10 dez. 2019.

THIRY-CHERQUES, H. R. Pierre Bourdieu: a teoria na prática. **Revista de Administração Pública**, Rio de Janeiro, v. 40, n. 1, p. 27-53, fev. 2006. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-76122006000100003&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-76122006000100003&lng=en&nrm=iso). Acesso em: 20 fev. 2020.

VAN ECK, N. J., *et al.* J. A Comparison of two techniques for bibliometric mapping: multidimensional scaling and VOS. **Journal of the American Society for Information Science and Technology**, v. 61, p. 2.405-2.416, 2010. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21421>. Acesso em: 20 fev. 2020.

VICKERY, B. C. **Faceted classification: a guide to the construction and use of special schemes**. London: ASLIB, 1960.

VOSVIEWER. **Welcome to VOSviewer**, 2020. Disponível em: <https://www.vosviewer.com/>. Acesso em 20 fev. 2020.

WEB OF SCIENCE GROUP. **Essential Science Indicators**, 2020a. Disponível em: <https://clarivate.com/webofsciencegroup/solutions/essential-science-indicators/>. Acesso em: 20 fev. 2020.

WEB OF SCIENCE GROUP. **Web of Science: confident research begins here**, 2020b. Disponível em: <https://clarivate.com/webofsciencegroup/solutions/web-of-science/>. Acesso em: 20 fev. 2020.

WEB OF KNOWLEDGE. **About us**, 2020. Disponível em:  
<https://www.webofknowledge.com/aboutus>. Acesso em: 25 maio 2020.

ZATTAR, M.; MARTELETO, R. M. Pierre Bourdieu no campo de estudos da informação: uma revisão da literatura. *In*: MARTELETO, R. M.; PIMENTA, R. M. (Orgs.) **Pierre Pierre Bourdieu e a produção social da cultura, do conhecimento e da informação**. Rio de Janeiro: Garamond, 2017. P. 102-133.

## APÊNDICE A – Documentos sobre Ciência Aberta publicados antes de 1998

(continua)

Web of Science	Scopus	Título do documento	Ano	Resumo
x		Psychiatry as open science -	1970	[No abstract available]
x	x	Education is the key to open science	1979	[No abstract available]
x		Open Science V National-Security	1982	[No abstract available]
x	x	Open science and closed science - tradeoffs in a democracy	1985	[No abstract available]
	x	The future of sports: challenge for the science of sport	1986	The conditions for a further positive development of the science of sport are very favorable as both the interest in sports and the demand for scientific knowledge will increase. However, if the science of sport is to achieve a prominent standing, new problems must be faced. Sport is developing into an open system in that its boundaries to other fields (economy, media, etc.) and to other forms of spare-time activity (e.g. tourism, entertainment) have become blurred. The consequence must be the development of an open science of sport. New groups of people are becoming interested in sports, and new innovative models of sport activity must be developed for them. With new strategies of planning, new forms of sport activity must be organized and realized. Emerging from these elements is the concept of the science of sport as an open, innovative, planning science. © 1986, Sage Publications. All rights reserved.
	x	Keller's gender/science system: is the philosophy of science to science as science is to nature?	1988	I argue that although in "The Gender/Science System," Keller intends to formulate a middle ground position in order to open science to feminist criticisms without forcing it into relativism, she steps back into objectivism. While she endorses the dynamic-object model for science, she endorses the static-object model for philosophy of science. I suggest that by modeling her methodology for philosophy on her methodology for science her philosophy would better serve her feminist goals. 1988 by Hypatia, Inc.
x		Open science, open skies	1989	
x		World laboratory - animal liberation week - protests fail to weaken scientists resolve - biomedical researchers vow to go on using animal subjects in their labs, despite acts of personal intimidation	1993	<b>CORPORATE RESEARCH HUBS:</b> Historically, the biotech industry has had close ties with the academic community, a fact that is still strongly reflected in cultural similarities between biotechs and university laboratories. Differences do emerge, however, over attitudes toward publishing: When are research findings open science and when are they proprietary secrets?

### Documentos sobre Ciência Aberta publicados antes de 1998

(continua)

Web of Science	Scopus	Título do documento	Ano	Resumo
x	x	Toward a new economics of science	1994	<p>Science policy issues have recently joined technology issues in being acknowledged to have strategic importance for national 'competitiveness' and 'economic security'. The economics literature addressed specifically to science and its interdependences with technological progress has been quite narrowly focused, and has lacked an overarching conceptual framework to guide empirical studies and public policy discussions in this area. The emerging 'new economics of science', described by this paper, offers a way to remedy these deficiencies. It makes use of insights from the theory of games of incomplete information to synthesize the classic approach of Arrow and Nelson in examining the implications of the characteristics of information for allocative efficiency in research activities, on the one hand, with the functionalist analysis of institutional structures, reward systems and behavioral norms of 'open science' communities-associated with the sociology of science in the tradition of Merton-on the other. An analysis is presented of the gross features of the institutions and norms distinguishing open science from other modes of organizing scientific research, which shows that the collegiate reputation-based reward system functions rather well in satisfying the requirement of social efficiency in increasing the stock of reliable knowledge. At a more fine-grain level of examination, however, the detailed workings of the system based on the pursuit of priority are found to cause numerous inefficiencies in the allocation of basic and applied science resources, both within given fields and programs and across time. Another major conclusion, arrived at in the context of examining policy measures and institutional reforms proposed to promote knowledge transfers between university-based open science and commercial R&amp;D, is that there are no economic forces that operate automatically to maintain dynamic efficiency in the interactions of these two (organizational) spheres. Ill-considered institutional experiments, which destroy their distinctive features if undertaken on a sufficient scale, may turn out to be very costly in terms of long-term economic</p>

				performance.
--	--	--	--	--------------

### Documentos sobre Ciência Aberta publicados antes de 1998

(continua)

Web of Science	Scopus	Título do documento	Ano	Resumo
	x	role of the slovak academy of sciences during the period of transition in the slovak republic	1995	The transformation of a socialist-type Academy of Sciences into an institution capable of functioning in a democratic market economy is described. Prior to 1989, the Slovak Academy of Sciences was the state mandated coordinator of science and technology in Slovakia and was funded directly by the state budget. Since 1990, the Academy has had to share many of its prerogatives with other authorities and institutions, as in the case of the universities in regard to doctoral programmes, or to cede them outright, as in the case of the Ministry of Education and Science in regard to the coordination of basic research. Its budget has been drastically cut. It has also had to contend with the introduction of a western type of grant programme and system of evaluation for its subordinate institutes, some of which have been closed. In short, the Slovak Academy of Sciences must compete in an increasingly open science market in which it must give proof both of the quality of its work and of the relevance of the latter to the needs of society. © 1995 by Taylor & Francis Group. All rights reserved.
x	x	Open science meeting of the human dimensions of global environmental change research community	1997	[No abstract available]
	x	Common agency contracting and the emergence of "open science" institutions	1998	[No abstract available]
x	x	Environmental processes of the ice age: land, oceans, glaciers (EPILOG)	2001	Knowledge of the state of the earth at the Last Glacial Maximum (LGM, an interval around 21,000 years ago) is an important benchmark for understanding the sensitivity of global environmental systems to change. Much progress in understanding climates of the LGM has occurred in the similar to 20 years since the end of the CLIMAP project of the 1970s (Climate Long-range Investigation, Mapping and Prediction). Here we review this progress, based on presentations and discussion at a first open science meeting of the EPILOG project (Environmental Processes of the Ice age: Land, Oceans, Glaciers) held in Delmenhorst, Germany, May 1999. We outline key controversies and

				document protocols for EPILOG contributions, so that a new synthesis of the LGM Earth can emerge as an open project of the world's community of scientists. (C) 2001 Elsevier Science Ltd. All rights reserved.
--	--	--	--	---

**Documentos sobre Ciência Aberta publicados antes de 1998**

(continua)

<b>Web of Science</b>	<b>Scopus</b>	<b>Título do documento</b>	<b>Ano</b>	<b>Resumo</b>
x		Special issue - Land-Ocean Interactions in the Coastal Zone - Proceedings of the Fourth LOICZ Open Science Meeting held in Bahia Blanca, Argentina, November 1999 - Preface	2001	[No abstract available]
x		Challenges of a changing Earth - Proceedings of the Global Change Open Science Conference, Amsterdam, the Netherlands, 10-13 July 2001 - Closing address	2002	[No abstract available]
x	x	Legal and policy questions for international collaboration in neuroscience	2002	The requirements for neuroinformatics to make a significant impact on the field of neuroscience as a whole are not simply technical - the hardware, software, and protocols for collaborative research - they also include the legal and policy frameworks within which research is conducted. This is not least because the creation of large collaborative scientific databases amplifies the complicated interactions between proprietary, for-profit R&D and public "open science". In this paper, we draw on experiences from the field of bioinformatics to examine some of the likely consequences of these interactions in neuroscience. The widespread sharing of data and tools for neuroscientific research is a key step in the development of neuroinformatics. We consider some of the areas in which policy frameworks can facilitate these forms of collaboration, by easing legal and cultural barriers which have slowed developments to date.
x	x	Regimes of regional and global coastal change A selection of papers from the 4th LOICZ Open Science Meeting, Bahia Blanca, Argentina, Nov 1999	2002	[No abstract available]
x	x	The access of researchers from developing countries to international science and technology	2002	Different problems of access to scientific knowledge by researchers from developing countries are surveyed. Obstacles are classified according to their source: those inherent to research processes in general and having special inhibitory effects on researchers from developing countries, and those related to the institutional framework and economic environment of the developing

				country itself. The trend towards privatisation of scientific results has led many open-science institutions in industrialised countries to restrict access to the knowledge they use and produce. Nonetheless, the inability of scientific and technological communities of
--	--	--	--	--

**Documentos sobre Ciência Aberta publicados antes de 1998**

(continua)

<b>Web of Science</b>	<b>Scopus</b>	<b>Título do documento</b>	<b>Ano</b>	<b>Resumo</b>
x	x	The access of researchers from developing countries to international science and technology	2002	(continuação) developing countries to endogenise economic and political support for their activities is presented as the main obstacle for their access to knowledge. Insufficient resources for research, even in countries that could easily allocate important funds for these activities, are shown to be an important obstacle in a world where scientific and technological information is increasingly marketed and expensive. The dynamics leading to a low efficiency equilibrium of scientific and technological activities are characterised.
	x	The public understanding of science - A rhetorical invention	2002	This article contributes to the development of a rhetorical approach to the public understanding of science or science literacy. It is argued that rhetoric promises an alternative approach to deficit models that treat people as faulty scientists. Some tensions in the relevant rhetorical literature need resolution. These center on the application to science of an Aristotelian conception of rhetorical reasoning as enthymematic, without breaking from the Platonic/Aristotelian division between technical and public spheres. The former opens science to the potential of public critique
x		Neuroscience data and tool sharing - A legal and policy framework for neuroinformatics	2003	The requirements for neuroinformatics to make a significant impact on neuroscience are not simply technical-the hardware, software, and protocols for collaborative research-they also include the legal and policy frameworks within which projects operate. This is not least because the creation of large collaborative scientific databases amplifies the complicated interactions between proprietary, for-profit R&D and public "open science." In this paper, we draw on experiences from the field of genomics to examine some of the likely consequences of these interactions in neuroscience. Facilitating the widespread sharing of data and tools for neuroscientific research will accelerate the development of neuroinformatics. We propose approaches to

				overcome the cultural and legal barriers that have slowed these developments to date. We also draw on legal strategies employed by the Free Software community, in suggesting frameworks neuroinformatics might adopt to reinforce the role of public-science databases, and propose a mechanism for identifying and allowing "open science" uses for data whilst
--	--	--	--	---

### Documentos sobre Ciência Aberta publicados antes de 1998

(continua)

Web of Science	Scopus	Título do documento	Ano	Resumo
x	x	Neuroscience data and tool sharing - A legal and policy framework for neuroinformatics	2003	(continuação) still permitting flexible licensing for secondary commercial research.
x	x	The advance of technology and the scientific commons	2003	The advance of technology proceeds through an evolutionary process, with many different new departures in competition with each other and with prevailing practice, and with ex-post selection determining the winners and losers. In modern times what gives power to the process is the strong base of scientific and technological understanding and technique that guides the efforts of those seeking to advance the technology. Most of that base is part of a commons open to all who have expertise in a field. The proprietary aspects of technology traditionally have comprised a small topping on the commons. But recently parts of the commons have become privatized. While the justification for the policies and actions that have spurred privatization of the commons is that this will spur technological progress, the argument here is that the result can be just the opposite.
x	x	Can "open science" be protected from the evolving regime of IPR protections?	2004	Increasing access charges and transactions costs arising from monopoly rights in data and information adversely affect the conduct of science, especially exploratory research programs. The latter are critical for the sustained growth of knowledge-driven economies, and are most efficiently pursued in the "open science" mode. In some fields, informal cooperative norms for timely sharing of access to raw data-streams and documented database resources are being undermined by legal institutional innovations that accommodate the further privatizing of the public domain in information. A variety of corrective measures are needed to restore proper balance to the IPR regime.
	x	Good usage of piracy. Free culture, open science [Du bon usage de la piraterie. Culture libre, science ouverte]	2004	What is a pirate ? In this excerpt from a forthcoming book, Florent Latrive uses a series of historical examples to show that the pirate copy, as a legal and artificial construction, is

				a notion that varies in time and space. Who still remembers that as recently as 1976, the Hollywood majors dragged a diabolical technological invention, the tape deck, into the courts, and proved that if everyone could freely pirate the films broadcast over television, the entire film industry would soon bite the dust?. © Association Multitudes.
--	--	--	--	---

### Documentos sobre Ciência Aberta publicados antes de 1998

(continua)

Web of Science	Scopus	Título do documento	Ano	Resumo
	x	JGOFS makes significant advances, meets new challenges	2004	The Joint Global Ocean Flux Study (JGOFS) Open Science Conference was held at the National Science Foundation in Arlington, Virginia, USA, in May 2003. The main topic of discussion at the conference was the reduction of uncertainties in the oceans' absorption of atmospheric carbon. JGOFS aims at determining and understanding the processes that control the time-varying fluxes of carbon and associated biogenic elements in the ocean. It also aims to develop the capability to predict the response of oceanic biogeochemical (BGC) processes on a global scale. The World Ocean Circulation Experiment (WOCE) and JGOFS worked towards understanding ocean carbon uptake. JGOFS started its work in 1988 by establishing the Hawaiian Ocean Time-Series (HOT) and the Bermuda Atlantic Time Series (BATS). Combined global observing systems and the new generation of general circulation models allowed daily time-dependent estimates of ocean circulation. The global observing system put in place by WOCE, if sustained, will give improved circulation estimates. It has now been made possible to focus mainly on BGC processes. Partnerships among industry, academia, and government scientists will be essential to study these processes on the required time, space, and taxonomic levels.
x	x	Networks of inventors and the role of academia: an exploration of Italian patent data	2004	This paper proposes a quantitative analysis of social distance between Open Science and Proprietary Technology. A few general properties of social networks within both realms are discussed, as they emerge from the new economics of science and recent applied work on "small worlds". A new data-set on patent inventors is explored, in order to show that social networks within Proprietary Technology are much more fragmented than Open Science ones, except for science-based technologies. Two propositions are then put forward on the "open" behaviour expected

				from academic inventors, namely university scientists getting involved in Proprietary Technology networks by signing patents. Both propositions are confirmed by data, which show academic inventors to be more central and better connected than non-academic ones. The database and methodology produced for this paper are suggested to be relevant for the more general debate on the role of geographical and cognitive distance in
--	--	--	--	--

### Documentos sobre Ciência Aberta publicados antes de 1998

(continua)

Web of Science	Scopus	Título do documento	Ano	Resumo
x	x	Networks of inventors and the role of academia: an exploration of Italian patent data	2004	(Continuação) university-industry technology transfer. (C) 2003 Elsevier B.V. All rights reserved.
x	x	Understanding the emergence of 'open science' institutions: functionalist economics in historical context	2004	This essay exposes the limitations of the 'logical origins' approach that has found favour among economists who seek to understand the workings of institutions in the past and present. A different approach is pursued here, applying functionalism in historical context to explain the emergence of the characteristic ethos and institutions of 'open science'. The emergence during the late sixteenth and early seventeenth centuries of the idea and practice of 'open science' represented a break from the previously dominant ethos of secrecy in the pursuit of 'Nature's secrets'. It was a distinctive and vital organizational aspect of the scientific revolution, from which crystallized a new set of norms, incentives and organizational structures that reinforced scientific researchers' commitments to rapid disclosure of new knowledge. To understand how this came about, it is necessary to examine the economics of patronage and the roles of asymmetric information and reputation in the early modern reorganization of scientific activities. The rise of 'cooperative rivalries' in the revelation of new knowledge is seen as a functional response to heightened asymmetric information problems posed for the Renaissance system of court patronage of the arts and sciences; pre-existing informational asymmetries had been exacerbated by increased importance of mathematics and the greater reliance upon sophisticated mathematical techniques in a variety of

				practical contexts of application. Analysis of the court patronage system of late Renaissance Europe, within which the new natural philosophers found their support, points to the significance of the feudal legacy of fragmented political authority in creating conditions of 'common agency contracting in substitutes'. These conditions are shown to have been conducive to more favorable contract terms (especially with regard to autonomy and financial support) for the agent-client members of western Europe's nascent
--	--	--	--	---

**Documentos sobre Ciência Aberta publicados antes de 1998**

(continua)

<b>Web of Science</b>	<b>Scopus</b>	<b>Título do documento</b>	<b>Ano</b>	<b>Resumo</b>
x	x	Understanding the emergence of 'open science' institutions: functionalist economics in historical context	2004	(Continuação) scientific communities. Some lessons may be drawn for contemporary science and technology policy debates, in which the open science mode of pursuing knowledge often seems to be viewed a robust concomitant of the power of scientific research techniques- whereas it is a fragile cultural legacy of western Europe's history upon which rests the ascendancy of modern science as a driver of long-term economic growth.
x	x	Authorization and account management in the open science grid	2005	An attribute-based authorization infrastructure developed for the Open Science Grid is presented. The infrastructure integrates existing identity-mapping and group-membership service using concepts prototyped in the PRIMA system. Authorization scenarios for requests to compute and data resources are detailed. A new SAML obligated authorization decision statement is introduced that attaches an XACML obligation to the authorization decision. The use of obligations enables site-centralized, service-independent policy management. Authorization decisions are enforced via a Workspace Service that creates constrained execution environments configured in accordance with the obligations and other attribute-based information. Finally, an experimental PRIMA authorization service that extends and simplifies the infrastructure is described.
	x	DI-GRUBER: A distributed approach to grid resource brokering	2005	Managing usage service level agreements (USLAs) within environments that integrate

				<p>participants and resources spanning multiple physical institutions is a challenging problem. Maintaining a single unified USLA management decision point over hundreds to thousands of jobs and sites can become a bottleneck in terms of reliability as well as performance. DIGRUBER, an extension to our GRUBER brokering framework, was developed as a distributed grid USLA-based resource broker that allows multiple decision points to coexist and cooperate in real-time. DIGRUBER addresses issues regarding how USLAs can be stored, retrieved, and</p>
--	--	--	--	---

**Documentos sobre Ciência Aberta publicados antes de 1998**

(continua)

Web of Science	Scopus	Título do documento	Ano	Resumo
	x	DI-GRUBER: A distributed approach to grid resource brokering	2005	<p>(Continuação)</p> <p>disseminated efficiently in a large distributed environment. The key question this paper addresses is the scalability and performance of DI-GRUBER in large Grid environments. We conclude that as little as three to five decision points can be sufficient in an environment with 300 sites and 60 VOs, an environment ten times larger than today's Open Science Grid. © 2005 IEEE.</p>
x		From keeping "nature's secrets" to the institutionalization of "open science"	2005	[No abstract available]
	x	Koyaanisqatsi in cyberspace: The economics of an "out-of-balance" regime of private property rights in data and information	2005	<p>Koyaanisqatsi is a Hopi Indian word that translates into English as "life out of balance," "crazy life," "life in turmoil," all meanings consistent with indicating "a way of life which calls for another way of living." While not wishing to suggest either that the international regime of intellectual property rights protection of scientific and technical data and information is "crazy" or that it is "in turmoil," this chapter argues that the persisting drift of institutional change towards a stronger, more extensive and globally harmonized system of protection has dangerously altered the balance between private rights and the public domain. In this regard, we have embarked upon "a way of life which calls for another way of living." High access charges imposed by holders of monopoly rights in intellectual property have overall consequences for the conduct of science that are particularly damaging to programs of exploratory research, which are</p>

				<p>recognized to be critical for the sustained growth of knowledge-driven economies. The urgency of working towards a restoration of proper balance between private property rights and the public domain in data and information arises from considerations beyond the need to protect the public knowledge commons, upon which the vitality of open science depends. Policymakers who seek to configure the institutional infrastructure to better accommodate emerging commercial opportunities of the information-intensive “new economy” – in the developed and developing countries alike.</p>
--	--	--	--	--

**Documentos sobre Ciência Aberta publicados antes de 1998**

(continua)

<b>Web of Science</b>	<b>Scopus</b>	<b>Título do documento</b>	<b>Ano</b>	<b>Resumo</b>
x	x	Material transfer agreements: open science vs. proprietary claims	2005	For academia and industry, MTAs are the means to widely differing ends.
x	x	Minorities striving and pursuing higher degrees of success in earth system science (MS PHD'S) initiatives professional development program	2005	<p>The NASA and NSF-funded Minorities Striving and Pursuing Higher Degrees of Success in Earth System Science (MS PHDS) initiative was developed by and for underrepresented minorities with the purpose of facilitating our increased participation in Earth system science. The initiative's goal is to provide professional development and mentoring experiences that facilitate the advancement of minorities committed to achieving outstanding Earth system science careers. The 2003 MS PHDS in Ocean Sciences Program facilitated the meaningful engagement of 25 student participants at the final Joint Global Ocean Flux Study (JGOFS) Open Science Meeting. During the 2004 American Geophysical Union (AGU) fall meeting, the MS PHDS initiative completed Phase I of its newest endeavor, entitled the MS PHDS Professional Development Program (MS PHDS PDP). The 3 MS PHDS PDP phases are connected by engagement in virtual community building activities that occur through the website, listserv and asynchronous/synchronous web-based dialogues. Leveraging computer mediated communication to enhance peer to peer and peer to protege mentoring is one of the best practices explored in detailed in this paper. Students who successfully complete all three phases and virtual community activities of the MS PHD'S PDP will be better prepared to</p>

				achieve their academic and professional goals and will likely remain actively engaged in their fields of specialization and respective professional societies. Preliminary findings from this ongoing investigation are presented.
x	x	Open science - Combining open data and open source software: Medical image analysis with the Insight Toolkit	2005	[No abstract available]
x	x	Proceedings of the 6th IEEE/ACM International Workshop on Grid Computing	2005	The proceedings contain 51 papers. The topics discussed include: wide area data replication for scientific collaborations

### Documentos sobre Ciência Aberta publicados antes de 1998

(continua)

Web of Science	Scopus	Título do documento	Ano	Resumo
	x	Pros and Cons of Faculty Participation in Licensing	2005	In this chapter we provide a general overview of the university licensing process and its dramatic growth over the past decade. We then discuss the role faculty play in commercialization through the licensing process. Concerns have been voiced in recent years over the possibility that the recent growth in university licensing suggests that the traditional role of faculty in the generation of basic research results - as well, possibly, as their role in open science - has been compromised. We discuss the available evidence for this downside to faculty licensing. Finally, we consider several impediments to the licensing process. © 2005 Elsevier Ltd. All rights reserved.
x		PROS AND CONS OF FACULTY PARTICIPATION IN LICENSING	2005	In this chapter we provide a general overview of the university licensing process and its dramatic growth over the past decade. We then discuss the role faculty play in commercialization through the licensing process. Concerns have been voiced in recent years over the possibility that the recent growth in university licensing suggests that the traditional role of faculty in the generation of "basic" research results - as well, possibly, as their role in "open science" - has been compromised. We discuss the available evidence for this downside to faculty

				licensing. Finally, we consider several impediments to the licensing process.
x	x	R&D cooperation between firms and universities. Some empirical evidence from Belgian manufacturing	2005	Using Community Innovation Survey data for Belgium we analyze which firm and industry characteristics are conducive to cooperation with universities. We take into consideration that the decision to cooperate with universities cannot be analyzed in isolation from the overall innovation strategy of the firm. Cooperating with universities is complementary to other innovation activities such as performing own R&D, sourcing public information and cooperating with other partners. In the econometric analysis we fully

**Documentos sobre Ciência Aberta publicados antes de 1998**

(continua)

<b>Web of Science</b>	<b>Scopus</b>	<b>Título do documento</b>	<b>Ano</b>	<b>Resumo</b>
x	x	R&D cooperation between firms and universities. Some empirical evidence from Belgian manufacturing	2005	(Continuação) account for the impact of these simultaneously determined other R&D strategies. In addition, we examine the theoretically conflicting effect of the appropriation conditions on the likelihood of cooperating with universities. Our analysis confirms that large firms and firms in the chemical and pharmaceutical industry are more likely to be actively involved in industry science links. Furthermore, we find that these cooperative agreements are formed whenever risk is not an important obstacle to innovation and typically serve to share costs. Consistent with the open science paradigm, we find no evidence for the importance of the capacity to appropriate the returns from innovation for explaining cooperative agreements with universities. Finally, the results are consistent with complementarity of these agreements with other innovation activities. (c) 2005 Elsevier B.V. All rights reserved.
	x	Science on a large scale	2005	The TeraGrid, the U.S. National Science Foundation's multi-year project to build a distributed national cyberinfrastructure, entered full production mode in the fall of 2004, providing a coordinated set of services for the science and engineering community. TeraGrid operates a unified user support infrastructure and software environment across its eight resource partner sites, which together provide more than 40 teraflops of

				<p>computing capability and mass storage capability in the petabytes, linked by networks operating at tens of Gigabit/sec. This unified environment allows TeraGrid users to access storage and information resources as well as over a dozen major computing systems via a single allocation, either as stand-alone resources or as components of a distributed application using Grid software capabilities. Many lessons can be drawn from the dual pursuit of high performance and close integration. The next phase will be even more exciting, with the roll out of a wide range of science gateways and additional advanced applications. Science gateway projects are aimed at supporting access to TeraGrid via</p>
--	--	--	--	--

**Documentos sobre Ciência Aberta publicados antes de 1998**

(continua)

<b>Web of Science</b>	<b>Scopus</b>	<b>Título do documento</b>	<b>Ano</b>	<b>Resumo</b>
	x	Science on a large scale	2005	<p>(Continuação)</p> <p>web portals, desktop applications or via other grids. An initial set of 10 gateways will address new scientific opportunities in fields from bioinformatics to nanotechnology as well as interoperation between TeraGrid and other Grid infrastructures. TeraGrid is also enabling an impressive array of large scale science applications, where researchers can perform complex simulations and manipulate enormous data sets in novel ways to gain new insights into research questions and societal problems, for example, finding the most efficient and least expensive ways to clean up groundwater pollution. Effort in these and other related areas will allow more researchers and educators access to TeraGrid capabilities and advance compatibility between TeraGrid and other major Grid deployments such as Open Science Grid, Network for Earthquake Engineering Simulation (NEES), and major European and Asian Grid deployments. © Springer-Verlag Berlin Heidelberg 2005.</p>
	x	TeraGrid: A foundation for US cyberinfrastructure	2005	<p>TeraGrid is a collaboration of partners providing a high-performance, nationally distributed capability infrastructure for computational science. The TeraGrid team has utilized multiple surveys of user requirements to develop five-year roadmaps describing new capabilities and services,</p>

			organized into several new initiatives: Deep, Wide, and Open. TeraGrid is managed by the University of Chicago and includes resources at eight partner sites (Argonne National Laboratory, Indiana University, National Center for Supercomputing Applications, Oak Ridge National Laboratory, Pittsburgh Supercomputing Center, Purdue University, San Diego Supercomputer Center, and Texas Advanced Computing Center). TeraGrid Deep aims to assist scientists with applications that require the combination of multiple leadership class systems- including TeraGrid storage, computing, instruments, visualization, etc. - working in concert. A team of roughly 15 staff is providing hands-on assistance to application teams pursuing TeraGrid Deep projects. TeraGrid Wide is a set of partnerships with
--	--	--	--

**Documentos sobre Ciência Aberta publicados antes de 1998**

(continua)

<b>Web of Science</b>	<b>Scopus</b>	<b>Título do documento</b>	<b>Ano</b>	<b>Resumo</b>
	x	TeraGrid: A foundation for US cyberinfrastructure	2005	(Continuação) peer Grid projects and prototype science gateways that are aimed at making TeraGrid resources available to, and tailored to, entire communities of users. Science gateways are driving policy, process, and technology standards to enable web portals, desktop applications, campus clusters, and other grid infrastructure projects to seamlessly use TeraGrid resources, Initial TeraGrid science gateway projects include community portals and desktop tools supporting life sciences and biomedicine, high-energy physics, neutron science, astronomy, nanotechnology, atmospheric and climate sciences, and environmental and emergency decision-support. TeraGrid Open involves the rapid evolution of the TeraGrid software and services toward interoperability with peer Grids and campus resources. Currently TeraGrid is partnering with the Open Science Grid as well as partners in Europe (e.g. UK eScience, DEISA) and Asia-Pacific (e.g. Naregi, K*Grid). © IFIP International Federation for Information Processing 2005.
	x	The unacknowledged convergence of open source, open access, and open science	2005	A number of open initiatives are actively resisting the extension of intellectual property rights. Among these developments, three prominent instances - open source software,

				open access to research and scholarship, and open science - share not only a commitment to the unrestricted exchange of information and ideas, but economic principles based on (1) the efficacy of free software and research
x	x	The University of Virginia campus grid: Integrating grid technologies with the campus information infrastructure	2005	Grid software often unfortunately requires significant changes in existing infrastructure, both in terms of policy and mechanism, instead of accommodating and leveraging existing information servers such as enterprise LDAP servers and enterprise authentication infrastructures. The University of Virginia Campus Grid (UVaCG) has been designed explicitly to re-use as much existing infrastructure in the campus environment as possible in creating a Grid based on the Web Services Resource Framework (WSRF), specifically the Globus Toolkit v4 and

### Documentos sobre Ciência Aberta publicados antes de 1998

(conclusão)

Web of Science	Scopus	Título do documento	Ano	Resumo
x	x	The University of Virginia campus grid: Integrating grid technologies with the campus information infrastructure	2005	(Continuação) WSRF.NET. We report on the design and the current status of the UVaCG, with particular emphasis on the challenge of creating explicit policy expression, negotiation, and enforcement. When fully operational, campus researchers will be able to seamlessly utilize resources within the campus enterprise and expand on-demand to larger Grids such as the TeraGrid and the Open Science Grid.
x	x	What do pupils think of open science investigations? A study of Singaporean primary 6 pupils	2005	The purpose of this study was to find out: (a) pupils' attitudes towards open-ended science investigations, and (b) the problems that pupils encounter when carrying out such investigations in groups. The study was conducted in a class of 39 primary 6 pupils of mixed ability who carried out four investigations. Data were based on pupils' responses on an attitude survey questionnaire, interviews, video tapes of pupils performing the investigations and field notes. The majority of the pupils liked conducting such investigations. Reasons for positive responses were that the pupils had the freedom to explore and devise their own procedures to find out things that they wanted to know about, they could work together and discuss with friends, the investigations were interesting and it was a valuable learning

			<p>experience. Reasons for negative responses included the difficulty of coming up with good ideas to design the investigations, and group conflicts. The findings also indicate that group dynamics played an important role in influencing pupils' attitudes towards the investigations.</p>
--	--	--	--

**APÊNDICE B – 100 palavras-chave mais registradas nos documentos sobre Ciência Aberta da *Web of Science* e *Scopus*, indexadas pelas bases de dados (1970-2018)**

(continua)

Posição	Frequência	Palavras-chave
1°	448	<b>open science</b>
2°	92	<b>open access</b>
3°	64	<b>open data</b>
4°	59	<b>reproducibility</b>
5°	42	<b>replication</b>
6°	41	<b>data sharing</b>
7°	41	<b>science</b>
8°	26	<b>open source</b>
9°	23	<b>transparency</b>
10°	22	<b>scholarly communication</b>
11°	20	<b>knowledge</b>
12°	20	<b>research data management</b>
13°	18	<b>metaanalysis</b>
14°	18	<b>research data</b>
15°	17	<b>open science grid</b>
16°	17	<b>repository</b>
17°	16	<b>replicability</b>
18°	16	<b>reproducible research</b>
19°	15	impact
20°	15	<b>information</b>
21°	15	<b>metadata</b>
22°	14	<b>big data</b>
23°	14	challenges
24°	14	preregistration
25°	13	<b>collaboration</b>
26°	13	<b>publishing</b>
27°	12	<b>education</b>
28°	12	innovation
29°	12	registered report
30°	12	software
31°	12	web 2.0
32°	11	<b>data management</b>
33°	11	science policy
34°	11	semantic web
35°	10	<b>access</b>
36°	10	drug discovery
37°	10	<b>peer review</b>
38°	10	<b>research infrastructure</b>
39°	9	genomics
40°	9	<b>intellectual property</b>
41°	9	osg
42°	9	psychologist
43°	9	<b>repository</b>
44°	9	simulation
45°	9	systematic review
46°	9	technology
47°	9	<b>workflow</b>
48°	8	<b>e-infrastructure</b>
49°	8	framework

**100 palavras-chave mais registradas nos documentos sobre Ciência Aberta da *Web of Science* e *Scopus*,  
indexadas pelas bases de dados (1970-2018)**

(conclusão)

Posição	Frequência	Palavras-chave
50°	8	<b>grid computing</b>
51°	8	incentives
52°	8	journals
53°	8	methodology
54°	8	model
55°	8	<b>open innovation</b>
56°	8	patents
57°	8	power
58°	8	publication bias
59°	8	<b>research</b>
60°	8	science 2.0
61°	8	social media
62°	7	crowdfunding
63°	7	cyberinfrastructure
64°	7	<b>deposition</b>
65°	7	future
66°	7	<b>interdisciplinarity</b>
67°	7	<b>interoperability</b>
68°	7	linked data
69°	7	<b>openness</b>
70°	7	prediction
71°	7	<b>research integrity</b>
72°	7	<b>researchers</b>
73°	7	<b>science communication</b>
74°	7	sustainability
75°	7	<b>technology transfer</b>
76°	7	<b>visualization</b>
77°	7	web
78°	6	<b>availability</b>
79°	6	<b>citizen science</b>
80°	6	<b>cloud computing</b>
81°	6	<b>communication</b>
82°	6	design
83°	6	distributed computing
84°	6	ecology
85°	6	escience
86°	6	ethics
87°	6	expression
88°	6	modeling
89°	6	<b>open educational resources</b>
90°	6	<b>open peer review</b>
91°	6	parkinson's disease
92°	6	policy
93°	6	project
94°	6	publication
95°	6	recommendations
96°	6	science gateway
97°	6	standards
98°	6	trust
99°	6	web service
100°	5	attention

Fonte: Elaborado pelo autor (2020).

**APÊNDICE C – 100 palavras-chave mais registradas nos documentos sobre Ciência Aberta da Web of Science, Scopus e DOAJ, indexadas pelos próprios autores (1970-2018)**

(continua)

Posição	Frequência	Palavras-chave
1°	737	<b>science</b>
2°	413	<b>open science</b>
3°	331	biology
4°	303	human
5°	261	microbiology
6°	195	humans
7°	151	<b>information resource</b>
8°	145	medicine
9°	128	<b>library science</b>
10°	127	<b>bibliography</b>
11°	124	article
12°	121	<b>information dissemination</b>
13°	113	priority journal
14°	109	<b>research</b>
15°	100	<b>reproducibility</b>
16°	94	female
17°	91	psychology
18°	87	male
19°	85	<b>open science grid</b>
20°	76	<b>open access</b>
21°	75	<b>publication</b>
22°	75	<b>publishing</b>
23°	70	<b>procedures</b>
24°	68	software
25°	67	internet
26°	63	animals
27°	62	technology
28°	61	<b>information management</b>
29°	60	animal
30°	60	education
31°	60	nonhuman
32°	59	<b>controlled study</b>
33°	59	<b>distributed computer systems</b>
34°	58	high energy physics
35°	58	<b>medical research</b>
36°	54	mouse
37°	52	metabolism
38°	50	scientist
39°	48	genetics
40°	47	<b>peer review</b>
41°	45	genotype
42°	44	<b>grid computing</b>
43°	44	<b>reproducibility of results</b>
44°	43	quality control
45°	42	<b>information processing</b>
46°	41	<b>data sharing</b>
47°	40	genomics
48°	40	physiology
49°	39	<b>Review</b>

**100 palavras-chave mais registradas nos documentos sobre Ciência Aberta da *Web of Science*, *Scopus* e DOAJ, indexadas pelos próprios autores (1970-2018)**

(conclusão)

Posição	Frequência	Palavras-chave
50°	38	neuroscience
51°	37	<b>access to information</b>
52°	37	methodology
53°	36	mice
54°	36	united states
55°	35	adult
56°	34	brain
57°	34	language and literature
58°	34	social sciences
59°	33	<b>data handling</b>
60°	33	<b>open source</b>
61°	33	<b>standards</b>
62°	31	biomedical research
63°	31	<b>communication</b>
64°	31	<b>databases</b>
65°	31	neuroimaging
66°	30	linguistics
67°	30	mass media
68°	30	pathology
69°	30	philology
70°	30	protein expression
71°	29	<b>digital storage</b>
72°	29	ecology
73°	29	<b>scholarly communication</b>
74°	28	<b>big data</b>
75°	28	philosophy
76°	26	human experiment
77°	26	research design
78°	26	social media
79°	25	<b>computer science</b>
80°	25	<b>open access publishing</b>
81°	25	<b>open data</b>
82°	25	rat
83°	25	<b>transparency</b>
84°	25	unclassified drug
85°	24	animal cell
86°	24	bioinformatics
87°	24	computer program
88°	24	drug effect
89°	24	<b>metadata</b>
90°	24	middleware
91°	24	models
92°	24	neoplasm
93°	24	<b>open datum</b>
94°	24	religion
95°	23	<b>cloud computing</b>
96°	23	gene expression regulation
97°	23	<b>information technology</b>
98°	23	middle aged
99°	22	aged

100°	22	<b>citizen science</b>
------	----	------------------------

Fonte: Elaborado pelo autor (2020).