

# prisma.com

Revista de Ciências e Tecnologias de Informação e Comunicação



**CIC.DIGITAL PORTO**

CENTER FOR RESEARCH IN COMMUNICATION,  
INFORMATION AND DIGITAL CULTURE

**U. PORTO**  
**FLUP** FACULDADE DE LETRAS  
UNIVERSIDADE DO PORTO

## Direção

**Elisa Cerveira**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal

## Conselho Editorial

1. **António Machuco Rosa**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
2. **Armando Malheiro da Silva**, Universidade do Porto - Faculdade de Letras / CIC.Digital (Porto), Portugal
3. **Fernando Zamith**, Universidade do Porto - Faculdade de Letras / CIC.Digital (Porto), Portugal
4. **Helena Sousa**, Universidade do Minho, Instituto de Ciências Sociais, Portugal
5. **Maria Manuela Cardoso**, Instituto Politécnico do Porto, ISCAP, / Instituto Superior de Contabilidade e Administração do Porto, Portugal
6. **Óscar Mealha**, Universidade de Aveiro, Departamento de Comunicação e Arte, Portugal
7. **Paulo Faustino**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
8. **Sara Jesus Gomes Pereira**, Universidade do Minho, Instituto de Ciências Sociais, Portugal

## Gestão da Informação

1. **Mariana Paula Martins Selas**, Universidade do Porto - Faculdade de Letras, Portugal
2. **Raquel Graça**, CIC.Digital – Porto, Portugal

## Comissão Científica

1. **Alfredo Pena-Vega**, IIAC - Institut Interdisciplinaire d'Anthropologie du Contemporain - Centre Edgar Morin-EHESS/CNRS, França
2. **Ana Isabel Reis**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
3. **Ana Lúcia Terra**, Instituto Politécnico do Porto, ISCAP / CIC.Digital (Porto), Portugal
4. **Ana Margarida Pisco Almeida**, Universidade de Aveiro, Departamento de Comunicação e Arte, Portugal
5. **António Machuco Rosa**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
6. **Armando Malheiro da Silva**, Universidade do Porto - Faculdade de Letras / CIC.Digital (Porto), Portugal
7. **Brasilina Passarelli**, Universidade de São Paulo, Escola de Comunicações e Artes, Brasil
8. **Carla Conti de Freitas**, Universidade Estadual de Goiás (Campus Inhumas), Brasil
9. **Carlos Ávila de Araújo**, Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Brasil
10. **Carlos Felimer del Valle Rojas**, Facultad de Educación y Humanidades, Universidad de la Frontera, Chile
11. **Cláudio Roberto Magalhães Pessoa**, Universidade FUMEC, Belo Horizonte, Brasil
12. **Cristina Ponte**, Universidade Nova de Lisboa, Departamento de Ciências da Comunicação, Portugal
13. **Edileuza Regina Pena**, Universidade Federal de Mato Grosso, Instituto de Ciências Humanas e Sociais (Campus de Rondonópolis), Brasil
14. **Edson Luiz Riccio**, Universidade de São Paulo, Faculdade de Economia, Administração e Contabilidade, Brasil
15. **Enói Dagô Liedke**, Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Brasil
16. **Fernanda da Silva Martins**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
17. **Fernanda Ribeiro**, Universidade do Porto - Faculdade de Letras / CIC.Digital (Porto), Portugal
18. **Fernando Ramos**, Universidade de Aveiro, Departamento de Comunicação e Arte, Portugal
19. **Fernando Zamith**, Universidade do Porto - Faculdade de Letras / CIC.Digital (Porto), Portugal
20. **Francisco Alberto Severo de Almeida**, Universidade Estadual de Goiás (Campus Inhumas), Brasil
21. **Francisco Carlos Palleta**, Universidade de S. Paulo, Escola de Comunicações e Artes, Brasil
22. **George Leal Jamil**, Universidade FUMEC, Belo Horizonte, Brasil

23. **Helder Bastos**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
24. **Helena Lima**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
25. **Helena Santos**, Universidade do Porto, Faculdade de Economia / CIC.Digital (Porto), Portugal
26. **Inês Amaral**, Universidade Autónoma de Lisboa, Instituto Superior Miguel Torga, Portugal
27. **Inês Peixoto Braga**, Instituto Politécnico do Porto, ISCAP / CIC.Digital (Porto), Portugal
28. **Jorge Ferraz de Abreu**, Universidade de Aveiro, Departamento de Comunicação e Arte, Portugal
29. **José António Moreiro González**, Universidad Carlos III, Facultad de Humanidades, Comunicación y Documentación, Espanha
30. **José Azevedo**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
31. **José Simões de Almeida Júnior**, Universidade Federal de Minas Gerais, Faculdade de Educação, Brasil
32. **Laura Rosseti Ricapito**, Universidad Autónoma Metropolitana, Xochimilco, México
33. **Lidia Barboza Norbis**, Universidad de Montevideo, Facultad de Humanidades y Educación, Uruguay
34. **Lídia Oliveira**, Universidade de Aveiro, Departamento de Comunicação e Arte, Portugal
35. **Luc Quoniam**, Université Sud – Toulon Var, França
36. **Lucivaldo Barros**, Universidade Federal do Pará, Faculdade de Biblioteconomia do Instituto de Ciências Sociais Aplicadas, Brasil
37. **Luís Borges Gouveia**, Universidade Fernando Pessoa, Porto / CIC.Digital (Porto), Portugal
38. **Lynn Gama Alves**, Universidade do Estado da Bahia, SENAI - CIMATEC, Brasil
39. **Marcos Galindo**, Universidade Federal de Pernambuco, Departamento de Ciência da Informação, Brasil
40. **Maria Beatriz Marques**, Universidade de Coimbra, Faculdade de Letras / CIC.Digital (Porto), Portugal
41. **Maria Irene Fonseca e Sá**, Universidade Federal do Rio de Janeiro, Faculdade de Administração e Ciências Contábeis, Brasil
42. **Maria Manuel Borges**, Universidade de Coimbra, Faculdade de Letras, Portugal
43. **Maria Manuela Pinto**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
44. **María Victoria Carrillo Durán**, Universidad de Extremadura, Facultad de Biblioteconomía y Documentación, Espanha
45. **Moisés Rockembach**, Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Brasil
46. **Olívia Pestana**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
47. **Paulo Frias**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
48. **Pedro Almeida**, Universidade de Aveiro, Departamento de Comunicação e Arte, Portugal
49. **Renata Baracho**, Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Brasil
50. **Silvana Vidotti**, Universidade Estadual Paulista, Faculdade de Filosofia e Ciências (Campus de Marília), Brasil
51. **Tom Linden**, University of North Carolina at Chapel Hill, School of Media and Journalism, Estados Unidos da América do Norte
52. **Vasco Ribeiro**, Universidade do Porto, Faculdade de Letras / CIC.Digital (Porto), Portugal
53. **Zeny Duarte**, Universidade Federal da Bahia, Instituto de Ciência da Informação, Brasil

**Periodicidade:** semestral (edição de 1 ou 2 números temáticos por ano)

**ISSN:** 1646-3153

**Contacto:** [prisma.com@letras.up.pt](mailto:prisma.com@letras.up.pt)

## Editorial

**Sônia Caregnato**

**Universidade Federal do Rio Grande do Sul**  
[sonia.caregnato@ufrgs.br](mailto:sonia.caregnato@ufrgs.br)

Este número especial da Revista Prisma.com reúne relatos de investigações realizadas por membros do Grupo de Pesquisa em Comunicação Científica da Universidade Federal do Rio Grande do Sul (UFRGS), em Porto Alegre, Brasil.

A trajetória do Grupo teve início em 1996, sob a coordenação da professora Dra. Ida Regina Chittó Stumpf, pouco depois do surgimento do Programa de Pós-graduação em Comunicação e Informação da UFRGS, ao qual está vinculado desde então. Assim como o Programa, o grupo reúne pessoas oriundas das áreas da Informação e da Comunicação e, eventualmente, também egressos de outras disciplinas. Estudar os fenômenos emergentes e as práticas consolidadas na comunicação da ciência e, também, desenvolver competências para a pesquisa são seus propósitos maiores. Para isso, prioriza quatro temáticas de estudo: Comportamento informacional na ciência, Organização do conhecimento para a comunicação científica, Estudos métricos e Estudos em ética e em sociologia para a compreensão da comunicação na ciência.

O Grupo começou conciso, mas ao longo desses 22 anos cresceu para além até dos padrões de um país com o tamanho do Brasil. Em atividade hoje, encontram-se 11 professores, uma pós-doutoranda, 18 alunos de doutorado e mestrado, além de vários alunos da graduação que realizam iniciação científica.

Parte representativa dos estudantes são autores dos artigos deste número especial, o que faz com que se consiga abranger as temáticas de trabalho, assim como apresentar aos leitores um panorama dos estudos que estão sendo realizados no Grupo.

O artigo de Rodrigo Silva Caxias de Sousa, que abre esta edição, discute as funções dos links em blogs de pesquisadores brasileiros, identificando a sua utilização pelos autores com propósitos tanto argumentativos como referenciais, embora com pouca exploração das potencialidades do hipertexto enquanto manifestação comunicativa.

Também enfocando as práticas dos pesquisadores, Ana Gabriela Clipes Ferreira e Clóvis Milton Duval Wannmacher abordam a escolha de publicação em livros e capítulos de livros por professores universitários da área de Educação e afins e concluem que, na comunidade estudada, os livros são importantes veículos de circulação da informação científica.

A adaptação dos estilos Vancouver e APA na elaboração de referência e citação de atos normativos brasileiros é o tema debatido por Natascha Helena Franz Hoppen, Ana Paula Medeiros Magnus e Samile Andréa de Souza Vanz no texto que encerra as discussões sobre comportamento informacional neste número especial.

Dois textos abordam formas de organizar o conhecimento para uso pelas comunidades científicas. O trabalho de autoria de Luciana Monteiro Krebs, Rita do Carmo Ferreira Laipelt e Samuel dos Santos Rosa emprega a *folksonomia* para identificar variações terminológicas que podem ser utilizadas para atualização de vocabulários controlados na área da Pediatria. Lisiane Braga Ferreira, Marina Rodrigues Martins e Moisés Rockembach discutem o arquivamento da informação científica na web, ou seja, a coleta e preservação de objetos digitais para recuperação e uso posterior, a fim de garantir a efetivação de seu potencial informacional e probatório na ciência.

Os estudos métricos, que representam uma parte significativa dos esforços dos membros do grupo de pesquisa, estão contemplados em três artigos deste número especial. Rodrigo Aquino de Carvalho, Sônia Elisa Caregnato, Maria de Fátima Santos Maia e Raquel Oroski tratam de estabelecer a relação entre as referências presentes na lista final dos artigos científicos e as menções aos documentos feitas no corpo do texto, a partir de um estudo exploratório em artigos indexados na base BRAPCI (Base de Dados de Periódicos Brasileiros em Ciência da Informação). Empregando a mesma fonte de dados, Fernanda Bochi dos Santos, Gonzalo Rubén Alvarez, Rene Faustino Gabriel Junior e Ana Maria Mielniczuk de Moura estudam as redes de citação constituídas a partir dos artigos sobre altmetria, no artigo intitulado "Altmetria no Brasil: estudo de citação e cocitação na base de dados BRAPCI".

Finalmente, Raquel Oroski, Maria de Fátima Santos Maia e Rodrigo Aquino de Carvalho abordam as características da produção científica sobre o tema preservação digital indexada na base de dados LISA - Library and Information Science Abstracts.

Para encerrar o número especial, Karen Ribeiro de Freitas Irizaga, Luciana Monteiro Krebs, Maiara Bettio e Moisés Rockembach discutem questões éticas nas práticas científicas, enfocando particularmente a ciência aberta, o acesso aberto aos dados de pesquisa, a privacidade dos dados, a múltipla filiação institucional e os impactos na avaliação da produtividade dos autores e instituições.

A atividade colaborativa na academia é construída com o empenho de cada um dos participantes, mesmo daqueles que no momento não se inscrevem como autores. Neste sentido, o entusiasmo dos membros do grupo com a tarefa de produzir este número especial foi fundamental para atingir o resultado, assim como foram importantes o apoio institucional da UFRGS e o valioso trabalho da revisora de línguas, Carolina Georg Dressler.

Ao final, cabe-nos agradecer a quem esteve na gênese deste número especial. Ao Professor Armando Malheiro da Silva, que desafiou o grupo na empreitada, e à diretora da Revista Prisma.com, Professora Elisa Cerveira, nosso muito obrigado pela oportunidade de mostrar a produção corrente do Grupo de Pesquisa em Comunicação Científica da UFRGS aos colegas portugueses.

# O uso de links em blogs de pesquisadores brasileiros

*The use of links in blogs of Brazilian researchers*

**Rodrigo Silva Caxias de Sousa**

**Faculdade de Biblioteconomia e Comunicação da Universidade Federal  
do Rio Grande do Sul**  
[rodrigo.caxias@ufrgs.br](mailto:rodrigo.caxias@ufrgs.br)

## Resumo

Analisa o uso de *links* nas postagens dos *blogs* de pesquisadores profissionais brasileiros como manifestações que viabilizam possíveis interpretações acerca de um novo paradigma de comunicação científica. Baseia a metodologia na análise de conteúdo, observando as funções que cumprem os *links* em meio às postagens, identificando categorias emergidas do fenômeno. Os resultados indicam que as categorias suscitadas do fenômeno e as categorias obtidas *a posteriori* são pertinentes no que se refere à observação de fenômenos de comunicação científica.

**Palavras-chave:** Comunicação científica. Webometria. *Blogs*. Pesquisadores Brasileiros.

## Abstract

*It analyzes the use of links in blog posts of Brazilian professional researchers as demonstrations that enable possible interpretations about a new paradigm for scientific communication. The methodology is based on content analysis, noting the functions that the links comply through the posts, identifying subcategories emerged from the phenomenon. The results indicate that the categories raised from the phenomenon and the ones subsequently obtained are relevant with regard to observation of the scientific communication phenomena.*

**Keywords:** *Scientific communication. Webometrics. Blogs. Brazilian researchers.*

## 1. Introdução

No que diz respeito ao comportamento de pesquisadores em relação à produção e divulgação do conhecimento científico, é possível identificar estudos que elencam os motivos que os levam à composição e socialização dos resultados de suas pesquisas em periódicos científicos. Outras contribuições investigam tanto as razões quanto as motivações do uso de citações na composição dos escritos científicos. (Weinstock, 1971; Bavelas, 1978; Case & Higgins, 2000; Oliveira, 2010; Silveira, Caregnato & Bufrem, 2014). Ainda que essa díade de interesses seja parte diminuta de um amplo universo de questões atinentes aos estudos de Comunicação Científica, as mesmas tornam-se referência no sentido de que outros fenômenos, agora

materializados na web 2.0, possam ter como base distintos processos de comunicação científica, considerando os elos hipertextuais neles encontrados.

Cabe destacar que o presente estudo não se presta a distinguir conceitualmente tais termos (divulgação científica, difusão científica, comunicação científica) considerando seus usos por pesquisadores da área da Ciência da Informação, tampouco estabelecer proposições apontando qual termo é o mais adequado a ser adotado (razão, motivo, motivação) em relação ao uso de tais conectivos hipertextuais. Este estudo analisa o uso de links nas postagens dos blogs de pesquisadores brasileiros, entendendo que tais documentos web se concretizam como alternativa inusitada de publicação de informações relativas à ciência. Tal perspectiva implica, também, em considerar que pensemos na ampliação de distintos processos de comunicação científica, envolvendo os pesquisadores, a diversidade de públicos aos quais incidem as informações e a relativização das formas de composição textual habitualmente efetivada por esses cientistas.

## 2. Elos de informações sobre ciência

Considerando o fato de que os links são também encontrados em documentos científicos disponibilizados eletronicamente e em meio digital, a similaridade entre link e citação, foi tratada por diversos autores (Mckiernan, 1996; Rosseau, 1997; Vaughan & Shaw, 2003; Smith, 2004; Thelwall, 2004), servindo de base interpretativa para as primeiras relações sobre o fenômeno aqui investigado. Entretanto, comparações entre *links* e citações não devem desconsiderar o propósito e os meios nos quais esses conectivos são inseridos, nem mesmo as intencionalidades subjacentes à produção desses hipertextos e os elementos que condicionam esses fenômenos comunicativos.

O primeiro desses elementos está relacionado ao fato de que os fluxos de informação, a partir dos *links* em *blogs*, não se materializam, e tampouco são tributados, enquanto indicadores que tragam visibilidade ao meio de comunicação e ao pesquisador, o que pode ser observado nas práticas sedimentadas nos canais formais. Ao considerarmos os argumentos mencionados, é prudente afirmar que os links de blogs estão alheios ao circuito tradicional de arbitragem característico das produções científicas. Ademais, no qual seja parte diminuta de um universo de preocupações concernentes à área da Ciência da Informação, no tocante à produção hipertextual nos blogs, é preciso pontuar que a perspectiva cumulativa, baseada na



consulta a textos pregressos, não pode ser entendida como uma convenção a ser respeitada pelos seus autores. Tanto citações em meio às postagens, quanto inserções de *links*, não se constituem em obrigatoriedades, no sentido de que o texto deva ser composto a partir das mesmas racionalidades que embasam os escritos científicos. A inserção de links baseados em diferentes recursos hipermídia compõe um quadro no qual a textualidade dos escritos científicos tem não apenas sua linearidade dirimida, mas também determina outra lógica de composição. Isso requer que o pesquisador rearticule sua dinâmica de composição de textos relativos à ciência e passe a relativizar seus *habitus* até então consolidados.

Especificamente, em relação aos blogs, proposições teóricas pontuaram acerca de suas potencialidades como meio de comunicação entre pesquisadores. (Lawley, 2004; Alcará & Curty, 2008; Luzón, 2009; Wilkins, 2008; Kjellberg, 2009, 2010; Kouper, 2010). No âmbito da Ciência da Informação, existe a necessidade de interpretações que busquem problematizar o uso dos links em blogs a partir de abordagem metodológica de pesquisa quanti-qualitativa. (Kjellberg, 2009, 2010; Thelwall 2010; Wilkinson, Harries, Thelwall & Price, 2003; Wang, Jiang & Ma, 2010). Tais inquietações fundamentam a metodologia deste estudo. Em razão das questões citadas, discutiremos, a seguir, o uso dos *links* nos *blogs*, como conectivos que delineiam inusitados processos de comunicação científica. Decorre dessa questão o fato de que, nos *blogs*, os *links* se configuram como alternativa de composição hipertextual, sugerindo que inusitados comportamentos decorrentes de práticas de comunicação e divulgação científica sejam identificados na web.

Tais possibilidades estão relacionadas às formas de construção textual, nas quais pode ocorrer a dissolução da propriedade privada do texto. Nos *blogs*, as inserções que pervertem as fronteiras da autoria são objetivadas a partir dos comentários e dos *links* inseridos pelo leitor, ampliando a interlocução quanto à publicação de informações sobre ciência. Além disso, os links das postagens indicam intencionalidades por parte do autor, ao permitir o acesso a outras informações que estejam tanto em seu blog, quanto em outros espaços na web. Os links caracterizam a essência do hipertexto. Cabe esperar que no futuro todos os sistemas de hipertexto permitam estabelecer links com lexias sobre as que não se tem direito de fazer trocas, verbais ou de outro tipo, não tendo analogia no mundo impresso. Um dos efeitos dos links é que criam um domínio intermediário entre escritor e o leitor, indefinindo ainda mais a distinção entre ambas as funções. (Landow, 2009, p. 436).

Essa questão já era considerada por autores que anunciavam potencialidades em virtude dessas formas de composição hipertextual. Relativo à escrita e leitura hipertextual, ambas atividades sofrem apropriações e influências de um universo de estudiosos que buscam problematizar o domínio da textualidade, estando relacionadas às alterações nas formas de compor e conceber a construção de narrativas. Os *links* permitem que outros meandros da informação científica possam ser compostos pelo autor e por seu leitor, dando continuidade ao processo de escrita do texto encontrado no *blog* ou em outros espaços. Desse modo, alteram-se as racionalidades comunicativas adotadas na produção de textos científicos que apresentem resultados de pesquisa validados pelos pares, passando, os pesquisadores, a construir seu hipertexto em razão de referenciais que permitam ao leitor essa interferência, contrapondo-se às normas de composição textual herdadas de sua formação científica.

Ultrapassando a lógica da citação, a composição de fluxos de informação, evidenciada a partir dos *links* em *blogs*, não se constitui em comunicações relacionadas apenas à literatura científica, aos canais e aos atores integrantes do ciclo de comunicação científica formal. Conectivo característico de composição do hipertexto produzido em documentos *web*, o *link* cumpre funções que variam conforme o espaço a partir do qual esse conectivo é elencado, relacionado a motivações no tocante aos direcionamentos à fontes que o autor pretenda disponibilizar ao seu leitor. Para Fragoso, Recuero e Amaral (2011), a herança das Ciências da Informação, em especial a Biblioteconomia, confere centralidade à comparação entre *hiperlinks* e citações, e com isso, enfatiza um subconjunto das possíveis motivações e usos desses conectivos. A ideia de que o trabalho referido é suficientemente importante para ser mencionado, por sua vez, ultrapassa todas essas motivações (p. 142-143).

Neste estudo, consideramos tal similaridade, porém não a estabelecemos em nossas análises, pois como será vista, a complexidade na qual estão envolvidos os *links* usados nos *blogs* transcende a noção de processos argumentativos que se valem de elos entre partes de contribuições científicas como mecanismo de produção textual.

Em razão dessas observações, intencionamos interpretar, tanto a partir das ocorrências quanto das explícitas ausências, possíveis elementos que motivem os pesquisadores a se valer desses dispositivos hipertextuais, inferindo acerca dos fins para os quais os links são utilizados

nas postagens. Tais interpretações se devem ao fato de que o *link*, em meio às postagens, constitui-se em estratégia de socialização de informações relativas ao campo científico, que reverberam para além do próprio campo científico, atribuindo outros capitais científicos ao *blog* e ao pesquisador.

Neste caso, as ligações esboçam não apenas a incorporação de hábitos oriundos da dinâmica cibercultural da *web 2.0*, em que a participação e a troca de *links* determinam a reputação de um determinado ator social na rede, mas a possibilidade de que atores e instituições científicas passem a considerar a implicação que essas ligações têm no impacto dos processos de comunicação científica compostos por pesquisadores. O estudo dos *blogs* engloba o estudo dos fluxos de informação a partir desses sistemas na *web 2.0*. Esses estudos também focam nos usos dos *blogs* como ferramentas de publicação e os tipos de informação que são trabalhados entre os blogueiros.

Sendo assim, os *links* indicam alternâncias quanto aos objetos de estudo, as composições discursivas, aos canais adotados, aos públicos atingidos. Podem passar a imprimir diferentes graus de notoriedade aos pesquisadores e outras audiências, aos canais e as temáticas para as quais os leitores são direcionados.

Diante dessas evidências, é possível pontuar que os *links*, nos *blogs*, cumprem propósitos que variam conforme o espaço a partir dos quais são elencados (Luzón, 2009), referindo-se às intenções que influenciam o direcionamento que o autor pretende atribuir ao bloco de texto no qual esse elo está inserido e às possibilidades fornecidas ao leitor quanto às alternativas de navegação e recomposição do hipertexto.

### **3. Metodologia**

Estudo de natureza quanti-qualitativo baseado na análise de conteúdo de páginas da *web* (THELWALL, 2004). Fundamenta-se na análise de *links* identificados em meio ao conteúdo das postagens dos *blogs* de pesquisadores brasileiros. A análise de conteúdo está baseada nas categorias emergidas (Análise de Conteúdo Indutiva) dos textos das postagens, a partir da leitura flutuante.

O processo viabilizado, atinente à Análise de Conteúdo desses hipertextos, foi composto a partir dos 155 *blogs* que compunham o universo do *Anel de Blogs Científicos*.

Metodologicamente optamos por analisar os *blogs* de pesquisadores profissionais, utilizando os seguintes critérios: pesquisador(es) brasileiro(s) que tenha(m) atualizado seu *blog* nos últimos 12 meses; vínculo profissional do pesquisador a uma instituição de ensino/pesquisa/profissional no Brasil; vínculo, como pesquisador, a pelo menos um projeto de pesquisa cadastrado no Diretório de Grupos de Pesquisa do CNPq; explícita identificação encontrada no *blog* do(s) pesquisador(es) como autor(es).

Analisando a incidência dos critérios mencionados, foi possível chegar a um total de 22 *blogs* de pesquisadores. Consideradas essas análises, optamos pela coleta das dez últimas postagens produzidas em cada um dos 22 *blogs* de pesquisadores brasileiros, utilizando de um parâmetro numérico adotado do estudo de Luzón (2008). Tais opções permitiram que se chegasse a 220 postagens e total de 640 *links* nelas inclusos.

Inicialmente foi realizada, na íntegra, a “leitura flutuante” (BARDIN, 2004) das postagens, observando as diferentes alternativas que cada pesquisador se valeu para compor os hipertextos. Posteriormente, as postagens foram “salvas”, organizadas e nomeadas, separadamente, em arquivos do software Word (.doc), com as respectivas identificações dos *blogs*. Foram efetuadas outras duas releituras das postagens, buscando inferir sobre as funções dos *links* nesses escritos e a existência de possíveis padrões utilizados pelos pesquisadores em suas produções hipertextuais. Nesses momentos, as postagens eram lidas na íntegra e clicávamos nos links como forma de depreender as intencionalidades dos pesquisadores. Decorrentes dessas releituras, foram arroladas as categorias relativas às intencionalidades pelas quais os links foram inseridos nas postagens. O tópico subsequente apresenta as análises efetivadas.

#### **4. Análise dos dados**

Quanto ao número total de *links* entre as postagens, foram observados picos de maior ocorrência nos *blogs* Cultura Científica, Notas em CFD e ULE, respectivamente com 113, 74 e 64 *links*, o que mostra que a média de *links* teve um número superior a dois. Quanto àqueles que estão no extremo inferior, os *blogs* Os Humanos, *Blog* do Mércio, Geógrafos e Modos de Fazer Mundos totalizam 13 *links* entre a soma de suas 30 postagens, contabilizando 0,4 *links* em média por *blog*. Diante desses dados, evidenciamos que distintos comportamentos

podem, através dessas incidências, ser identificados, não sendo possível evidenciar um padrão em relação ao comportamento dos pesquisadores.

A amodal encontrada foi de 04 links, identificada entre as postagens dos *blogs* Modos de Fazer Mundos, *Blog* do Mércio, Os Humanos e Geófagos, respectivamente das áreas de Ciências Humanas (Filosofia e Antropologia) e Ciências Naturais (Biologia). A aproximação da amodal, com o número mínimo de *links* encontrados, mostra que a ocorrência de *links* entre as postagens é bastante reduzida, o que implica em uma composição hipertextual pouco atrelada às estratégias de troca de informações característica da web 2.0.

Ausência de *links* que remetesse o leitor a outros blocos de textos ou hipermídias, a partir das postagens, foi encontrada em 48 das 220 postagens, correspondendo a um total de 21,8%.

Dos 22 *blogs* de pesquisadores brasileiros que fazem parte da amostra intencional, 07 *blogs* continham *links* em todas as suas postagens, quais sejam: Ciência Brasil, Cultura Científica, Sem Ciência, Via Gene, Por Dentro da Ciência, Laboratório Itinerante de Ensino de Sociologia e *No Wires*. Por sua vez, esses dados mostram que apenas 31,8% dos *blogs* contêm *links* em todas as suas postagens, o que é reforçado pela grande incidência de postagens sem *links* encontradas respectivamente nos *blogs* Os Humanos, Geófagos, Modos de Fazer o Mundo e *Blog* do Mércio.

Essas ausências permitem observar que, as potencialidades comunicativas anunciadas em função dos *links* se mostram subutilizadas nos *blogs* de pesquisadores brasileiros estudados, e as mesmas podem estar indicando que não exista, por parte dos pesquisadores, identificação de que os *links* sejam elementos determinantes em relação às potencialidades que esses recursos hipertextuais podem proporcionar aos processos de comunicação científica. Além disso, os comportamentos em relação ao uso de *links* denotam que os pesquisadores não estão, necessariamente, motivados no que diz respeito à participação em uma economia de *links*, se consideradas as ocorrências nas quais os *links* não são encontrados.

A seguir, estabelecemos uma interpretação de cunho qualitativo que resultou na emergência de categorias que compõem a complexidade do fenômeno, estas baseadas em inferências relacionadas a aspectos alusivos às funções a que os pesquisadores se propõem, em razão do uso dos *links* como conectivo característico da comunicação científica.

#### 4.1. Análise qualitativa: os *links* inseridos nas postagens

Nesse espaço de escrita, optamos em analisar os *links* em meio às postagens, devido ao fato de os mesmos serem manifestações similares às citações, em relação à produção da literatura científica estabelecida pelos pesquisadores como parte de seu ofício.

Diferentemente dos processos de comunicação dos resultados das pesquisas, o uso de *links* nas postagens dos *blogs* reveste-se de singularidades relacionadas a formas de uso e compartilhamento de informações não convencionalmente identificadas entre as práticas de produção e comunicação de narrativas científicas, que é nos *blogs*, ao contrário, composta por pesquisadores e correlacionada a outras contribuições intelectuais através dos *links*.

Nossa análise buscou identificar com quais finalidades os pesquisadores se valem dos *links* para composição do texto em meio digital. Ressaltamos que nossas análises foram refinadas devido às releituras, articuladas às impressões sobre o fenômeno, realizadas durante o tratamento dos dados; de tal forma, que não se basearam apenas nos conteúdos explícitos e na interpretação que fizemos das maneiras pelas quais pesquisadores se valem de *links* na composição dos textos em seus *blogs*, mas também na aproximação com questões teóricas em nível interpretativo. Isso se deve ao fato de que hipertexto, é depois de tudo, texto, escritura, e desse fato resulta a difícil distinção de muitas das qualidades da boa escritura que se apresenta com a escritura com *links*. Em outras palavras, qualidade do hipertexto não depende somente do *link*. Há um ponto considerável, o texto que rodeia o *link* também é importante, já que o caráter do *link* e das imagens dentro de uma lexia individual está relacionado com a qualidade (Landow, 2009, p. 262, tradução nossa).

Sendo assim, buscamos entender os conteúdos e os usos dessas conexões a partir da leitura da postagem na íntegra, da ativação dos *links* e da consequente remessa do leitor a outros contextos, visto que nossas interpretações indicam que o *link* cumpre funções de caráter argumentativo e referencial nos textos em que está inserido e que outras informações podem ser acessadas nos espaços para os quais o leitor é direcionado.

Manifestações em que os *links* são utilizados com a função exclusiva de **adicionar informação** foram encontradas em 18 ocorrências dos 640 links. Exemplos da não existência de combinação dessa categoria a outras podem ser encontrados no *blog* Estudos Hum(e)anos, nas postagens que trazem as informações sobre as conferências de Luiz Eduardo Soares,

Rubem Barboza Filho, Werneck Vianna e Ronaldo Lessa. Em todas há um padrão de composição da mensagem em que varia apenas o nome dos conferencistas, como pode ser observado na passagem a seguir: conferência de Rubem Barboza Filho em evento do Laboratório de Estudos Hum(e)anos e do Centro de Estudos em Direito e Sociedade comemorativo dos 120 anos da República Federativa do Brasil. Para ver toda essa conferência, clique aqui.

Os *links* compostos a partir da expressão “clique aqui” remetem para parte ou totalidade do conteúdo das conferências, nas quais, em todas as ocorrências, o leitor tem acesso ao vídeo da mesma conferência com as falas dos respectivos conferencistas; materiais estes que estão disponibilizados no *YouTube*. Essa forma de composição não trouxe subsídios para que pudéssemos inferir da mensagem outros elementos, que não apenas a função mencionada, em razão de seu caráter instrumental na composição das postagens.

A função **autocitar** está mais ligada ao fato de os pesquisadores permitirem que seus leitores tenham acesso a outras postagens por eles compostas, e menos à citações em relação a produções que divulguem resultados de pesquisa, o que pode ser compreendido como uma forma de compor as postagens a partir de contribuições precedentes que são recorrentes como temática do *blog*.

Tais interpretações se deram na medida em que a leitura da postagem permitia depreender que, explicitamente, o pesquisador estava a se valer do *link* como mecanismo de autoatribuição de valor a uma determinada fonte ou documento. Trazemos como manifestação dessa categoria a postagem “Como são e como deveriam ser os vencimentos dos servidores da UNB”, datado de 10 de maio de 2010, em que o autor referenda na medida em que coloca assim na sua postagem: oi pessoal. O texto abaixo me foi enviado hoje, mas ANTES do final da greve dos professores, que ocorreu mais ou menos ao meio dia. Vale muito a pena ser lido. O autor do texto, o Prof. Sidio, ressalta que este *blog* tem discutido e muito os gastos excessivos da UnB com a rubrica Pessoa Física. (*Blog Ciência Brasil*, 2010).

Embora o pesquisador exponha, na sequência da postagem, o texto de outro blogueiro, antes disso ele não se furta em inserir um *link* que remete o leitor para uma postagem dele. A forma com que o texto é composto permite entender que a postagem que vale a pena ser lida é a

do professor mencionado, enquanto, contraditoriamente, o autor do *blog* faz menção a uma postagem anterior.

Cumprindo a função de **comprovar informações**, os *links* aparecem entre as postagens de diferentes maneiras. Na postagem “Seminário teoria ou teorias da evolução”, do *blog* Coletivo Ácido Cético, o *link* cumpre a função de comprovar os dados do pesquisador que foi o responsável pela fala no seminário, visto que a postagem tem a proposta de divulgar um evento que está sendo retomado na Universidade Federal do Rio Grande do Sul. O hipertexto está assim composto: no próximo dia 7/4 teremos, para reativar as atividades do Café Científico, o seminário (bastante informal) do Prof. Aldo Mellender de Araújo, do Instituto de Genética da UFRGS: Título: Teoria ou teorias da evolução biológica? Local: Anfiteatro Antônio Cabral Instituto de Física – UFRGS Campus do Vale. Horário: 12:30 Do [currículo Lattes](#) do Aldo: Possui graduação em História Natural pela Universidade Federal do Rio Grande do Sul (1967) e doutorado em Genética e Biologia Molecular pela Universidade Federal do Rio Grande do Sul (1973). [...] Atualmente é professor titular da Universidade Federal do Rio Grande do Sul. [...] (*Blog Ciência Brasil*, 2010).

Também são observadas situações recorrentes entre as postagens nas quais o *link* tenha a função **definir conteúdo ou conceito**; fundamentalmente, quando nas postagens o texto remete o leitor à um verbete da *Wikipédia*. Cabe-se destacar que não foram encontradas manifestações, entre os 640 *links*, que remetesse a conceitos em fontes formais de informação, corroborando uma atmosfera de informalidade característica da web 2.0. A categoria acima mencionada foi identificada sempre atrelada à outra categoria (adicionar informação), visto que, ao migrar para uma definição ou conceito, o leitor sempre se deparava com uma informação adicional. Manifestação assim composta pode ser identificada através dos *links* da postagem “*Cartoons* de Fernando Gonsales sobre as ideias de Darwin”, do *blog* ULE. A mensagem foi assim composta: Seguem alguns *cartoons* de Fernando Gonsales, em comemoração aos 150 anos desde a publicação de *Origem das Espécies* de [Charles Darwin](#) em 1859. As tiras foram publicadas em novembro do ano passado no caderno *Mais!* do jornal [Folha de São Paulo](#). (*Blog ULE*, 2010).



Tanto o primeiro *link* quanto o segundo, inclusos, respectivamente, através das expressões destacadas, remetem a verbetes da *Wikipédia*, esclarecendo quanto ao conteúdo ou ao conceito ao qual o pesquisador se refere.

Outra subcategoria de função emergida a partir da inserção de *links* em postagens foi **exemplificar**, observada quando o *link* remetia explicitamente ao exemplo que estava sendo mencionado em uma parte da postagem. A ocorrência dessa função pôde ser identificada nos *links* encontrados na postagem intitulada “Boletim da SBL, o mais novo endereço para divulgação científica” ([http://bafanaciencia.blog.br/bafana-divulga/boletim-da-sbl-o-mais-novo-endereco-para-divulgacao-cientifica.](http://bafanaciencia.blog.br/bafana-divulga/boletim-da-sbl-o-mais-novo-endereco-para-divulgacao-cientifica)), do *blog* Bafana Ciência (<http://bafanaciencia.blog.br/>). Nela, qual o pesquisador divulga alternativas de editoração e apresenta exemplos das diferentes categorias de artigos encontrados, de acordo com categorias arroladas no próprio boletim da Sociedade Brasileira de Limnologia.

Ao remeterem a publicações com formatos distintos tais como: artigo de periódico encontrado na postagem “A estrutura das Teias Tróficas”), categorizado como artigo de formação; artigo de informação, que relata “algum acontecimento” e que o *blog* traz como *link*), com informações sobre o XII Congresso Brasileiro de Limnologia, baseado no “formato de apresentação” de comunicação científica semelhante ao do artigo de periódico, embora o próprio conceito encontrado na postagem esclareça que a distinção é a ausência do principal elemento característico dos processos de comunicação científica (*peer review*). O texto da postagem está assim elencado: artigos de Divulgação: são necessariamente curtos, como bem sabe o leitor do Bafana, com poucas citações ou nenhuma e que têm o objetivo de mostrar ao público leigo (e à imprensa) algumas aplicações e particularidades da limnologia”, fazendo posterior linkagem ao texto “Nenhum peixe é uma ilha” (). O *link* permite acesso ao conteúdo do artigo, exemplificando a tipologia proposta pelos autores, ao arrolar categorias de distintos processos de comunicação científica e seus respectivos formatos.

Tendo como função elencar ocorrências em relação à subcategoria **identificar autoria**, arrolamos como indícios os seis *links*, de um total de oito contabilizados segundo a combinação com outras categorias, encontrados na postagem “A principal diferença entre gripe suína e gripe comum”, do *blog* Sem Ciência. Nesses casos, os *links* permitiam que o leitor identificasse a autoria de outros textos mencionados na postagem. Todos os *links inclusos*

*eram baseados na* mesma lógica de composição, pois remetiam aos nomes dos autores dos respectivos estudos ou a instituição responsável pela autoria do estudo mencionado.

À função de **identificar estudos** foram contabilizadas 39 ocorrências, sempre combinadas a outras categorias. Um exemplo dessas combinações pode ser encontrado através do único *link* elencado na postagem “[Essa mistura terrena grosseira: filosofia e vida comum em David Hume](#)”, por Marcos Balieiro, com data de 30 de maio de 2010, postada no *blog* Estudos Hum(e)anos. A forma como a postagem foi escrita, e como o *link* está destacado em clique aqui, encontra-se a seguir. O Laboratório de Estudos Hum(e)anos dá notícia da defesa de tese, pelo departamento de filosofia da USP, do amigo Marcos Balieiro, já próximo ao IUPERJ, quando nele esteve para o I Encontro Hume. Em função da felicidade da boa notícia, apresentamos o resumo da tese de Marcos e o *link* para a leitura. Resumo: Ainda que muitos trabalhos tenham sido escritos sobre a filosofia de David Hume, é bastante raro vermos comentários sobre o que seria, para ele, a própria filosofia. Na maior parte das vezes, os intérpretes da obra desse filósofo limitam a caracterizá-lo como cético, naturalista, realista, sentimentalista, entre outras categorias. [...] Além disso, veremos que, nos textos posteriores ao Tratado da natureza humana, Hume considerou a filosofia não como algo que deveria ficar restrito às universidades, mas como uma ferramenta poderosa de formação moral para o homem comum. Para continuar a ler, [clique aqui](#). (*upload* para uma tese sem sumário Ao acionar o *link*, o leitor é remetido para uma tese em extensão .pdf, em que é possível ter acesso ao conteúdo completo do estudo.

Outra subcategoria encontrada a partir dos *links*, no contexto das postagens, é **ilustrar**. Na postagem “Apenas humanos”, do *blog* Um Longo Argumento, é possível visualizar imagens, em meio ao conteúdo, que apenas identificam quem é o autor ao qual vem sendo feita uma determinada referência no texto. Nesse caso específico, aparecem as imagens de Richard Dawkins e de Carl Sagan, que se constituem em elementos ilustrativos, permitindo ao leitor, através de uma informação adicional, identificar quem são esses pesquisadores. Por outro lado devemos considerar que não se tratam de quaisquer cientistas, mas de nomes proeminentes da ciência, indicando o quanto o blogueiro (pesquisador) se filia ou credita valor a esses indivíduos.

O autor repete essa estratégia de linkagem na postagem “O hipopótamo de Tahl” (<http://charlesmorphology.blogspot.com/2010/01/o-hipopotamo-de-tahl.html>) no qual está inserida uma imagem do enxadrista Mikhail Tahl, imprimindo uma relação não apenas semântica, mas também visual, tendo a intenção de esclarecer o leitor acerca de quem é o indivíduo citado.

**Permitir contato** é uma das funções dos links encontradas entre as postagens, o que demonstra que a postagem se constitui em alternativa de comunicação entre o pesquisador e o leitor do *blog*. Formam-se, fundamentalmente, a partir dos *links* das postagens que remetem para formulários de contato, que pode ser realizado através de *e-mails* ou mesmo *feeds* RSS. Uma ocorrência dessa categoria pode ser identificada no *blog* “Bafana Ciência” que tem a intenção, com o uso do *link*, de recrutar pessoal para colaborar com o boletim. A sentença na qual o *link* se encontra é a seguinte: “Se quiser colaborar com o Boletim, entre em contato com os editores (eu entre eles) clicando aqui”, sendo o *link* inserido na palavra “aqui”. Ao ser acionado pelo leitor, é possível chegar ao formulário que permite enviar um *e-mail* para a página da Sociedade Brasileira de Limnologia.

**Referenciar fontes** é a subcategoria que, ao emergir das análises dos dados, está relacionada ao fato de o autor fazer referência explícita e de diferentes formas às fontes de informação ou a outros textos, através do uso de um *link*. Por exemplo, na postagem “Foi por medo de avião”, do *blog* “Um Longo Argumento”, em que o *link* é assim inserido através do número do PMID, identificador de padrão do registro PUBMED, base de dados na área de saúde. O contexto de inserção do link está assim apresentado: Soliman F, Glatt CE, Bath KG, Levita L, Jones RM, Pattwell SS, Jing D, Tottenham N, Amso D, Somerville L, Voss HU, Glover G, Ballon DJ, Liston C, Teslovich T, Van Kempen T, Lee FS, & Casey BJ (2010). A Genetic Variant BDNF Polymorphism Alters Extinction Learning in Both Mouse and Human. *Science (New York, N.Y.)* PMID: 20075215. Ao ativar o *link*, o leitor é remetido para o *abstract* da publicação, viabilizando a leitura no *site* da base de dados e permitindo acesso ao outro documento.

A subcategoria **referenciar fontes** surgiu devido ao fato de identificarmos que, em meio ao conteúdo da postagem, o autor atribui respaldo à fonte de informação ou documento a que o leitor é remetido, como no *link* apresentado na postagem “Religulous” do *blog* “Um Longo Argumento”. A postagem está assim escrita: acabei de assistir ao documentário “Religulous”,

de 2008, dirigido por Larry Charles, escrito e apresentado pelo comediante norte-americano Bill Maher. O filme foi indicação do grande [Dedalus](#) (do *blog Atlas*) durante uma conversa nos corredores da universidade em que lecionamos. Abusando do sarcasmo, Maher faz um trabalho semelhante ao de Richard Dawkins no documentário “The Root” (uma síntese das ideias presentes em “Deus, um Delírio”), porém com maior ênfase nos aspectos cômico-trágicos das crenças religiosas. Em tom satírico, nem por isso pouco sério ou raso, o sujeito mostra que, se interpretadas literalmente, muitas das religiões não passam de arremedos de péssimas histórias de ficção. No conteúdo da mensagem, o autor é explícito ao mencionar que se trata do “grande” *Dedalus*, autor de outro *blog* e seu colega de trabalho, sendo que o *link* incluso na postagem que o segue pode ser ativado a partir da palavra *Atlas*, que dá nome ao *blog* mencionado.

Por último, identificamos a subcategoria **suscitar relações**. As ocorrências relativas a essa função foram computadas quando nos deparamos com a inserção de imagens nas quais poderiam ser depreendidas intencionalidades subjacentes por parte do autor, em razão de seus vínculos com o contexto de escrita, como no caso da postagem “Corpo e Mente: uma linguagem unificada” em que o autor do *blog* Os Humanos disponibiliza o que ele atribui como sendo um artigo de sua autoria, inclusive com as referências bibliográficas utilizadas, embora na análise das publicações que constam em seu currículo lattes não seja encontrada ocorrência dessa produção intelectual.

Figura 1 – **Imagem retirada da postagem Corpo e Mente: uma linguagem unificada**



Fonte: *Blog Os Humanos*

Nesta postagem, a imagem pouco contribui para a compreensão do texto, embora a função do *link* seja **suscitar relações**, tendo em vista que o link permite a inferência acerca da noção de disjunção entre corpo e mente, em razão da imagem mostrar os sentidos e partes do corpo a eles relacionadas em quadrados distintos. No entanto, entendemos que, pelo fato de a imagem não permitir que o leitor possa continuar avançando na composição de percursos hipertextuais na web 2.0, a imagem cumpre uma função meramente instrumental em relação ao contexto.

#### **4.2. Observações acerca do fenômeno investigado**

O estudo aqui apresentado permite afirmar que, em atenção aos *blogs* de pesquisadores brasileiros, os links, enquanto manifestações hipertextuais, têm variações expressivas no que se refere às ocorrências. A dinâmica de participação anunciada através dos serviços web 2.0, que tem nos links uma de suas manifestações mais importantes, é estruturada segundo uma lógica reducionista e instrumental. As postagens, efetivadas nesses meios de comunicação, mostram, através da inserção de *links*, que a composição do hipertexto é ainda feita de forma

bastante primária. A baixa incidência ou até mesmo ausência dos mesmos entre as postagens inviabiliza a composição bidirecional de fluxos de informação e a possibilidade de que o leitor possa avançar para outras informações disponibilizadas em rede. Dessa forma, identificamos tanto propósitos argumentativos quanto propósitos referenciais (instrumentais).

**Quadro 1 – *Categorias obtidas das análises dos links das postagens***

<b><i>Categorias dos links argumentativos-referenciais</i></b>
Adicionar informação
Autocitar
Comprovar informação
Definir conteúdo/conceito
Exemplificar
Identificar autoria
Identificar estudo
Ilustrar o hipertexto
Permitir contato
Referenciar fonte
Referendar fonte
Suscitar relações

Fonte: Dados da Pesquisa

As categorias de funções extraídas dos *links* das postagens, baseadas em verbos e expressões verbais, permitem-nos inferir que os *links* são utilizados preponderantemente como estratégias argumentativo-referencial de composição hipertextual.

Quanto às ocorrências, as mesmas apontam para um comportamento em relação ao uso de conectivos (*links*), nas postagens, que pouco explora as potencialidades do hipertexto enquanto manifestação comunicativa, permitindo que seja questionado se preocupações dessa natureza fazem parte do imaginário dos pesquisadores brasileiros.

O uso dos *links* como conectivos e dos *blogs* como meio de comunicação, por pesquisadores, é manifestação de inovação em relação ao seu comportamento, indicando que os mesmos estão coadunados à economia de compartilhamento de informações, dado essas formas de inovação.

Ao viabilizar o reuso de uma fonte de informação que tenha o aval do pesquisador, o *link* é o manifesto da pertinência da navegação no próprio contexto e nos espaços dessas migrações. Mais do que um mecanismo de reconhecimento expresso pelo pesquisador, os *links* cumprem a função de direcionar o leitor e enfatizar fontes de informação na reconstrução do hipertexto, sendo marcas cognitivas validadas por diferentes “blogueiros” quanto à pertinência da informação e de seus elos relativos a determinada área do conhecimento.

As interpretações dos links identificados nos blogs de ciência, em especial, nos blogs de pesquisadores brasileiros, permitem que novas questões sejam colocadas aos estudiosos das áreas da Comunicação e Ciência da Informação, deixando por suspenso certezas quanto ao uso dos blogs como meios de comunicação científica.

## Referências Bibliográficas

---

- ALCARÁ, A. R. & CURTY, R. G. (2008). Fontes de informação na internet. Londrina: EDUEL.
- AMARAL, A., & RECUERO, R., & PORTELLA, S. BLOGS.COM: estudos sobre blogs e comunicação. São Paulo: Momento Editorial, 2009. p. 27-53.
- BARDIN, L. Análise de conteúdo (2004, 4ª edição). São Paulo: Martins Fontes.
- BAUWENS, M. A economia política da produção entre pares. Disponível em: <[http://www.p2pfoundation.net/A\\_Economia\\_Pol%C3%ADtica\\_da\\_Produ%C3%A7%C3%A3o\\_entre\\_Pares](http://www.p2pfoundation.net/A_Economia_Pol%C3%ADtica_da_Produ%C3%A7%C3%A3o_entre_Pares)>. Acesso em 10 de jan. de 2010.
- BAVELAS, J. B. (1978). The Social Psychology of Citations. *Canadian Psychological Review*, Calgary, 19, 158-163.
- CASE, D. O., & HIGGINS, G. M. (2000). How can we investigate citation behavior?: a study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, 51, 635-645.
- CHU, H. (2005). Taxonomy of inlinked web entities: what does it imply for webometric research? *Library & Information Science Research*, 27, 8-27.
- FRAGOSO, S., & RECUERO, R., & AMARAL, S. Métodos de pesquisa para internet. (2011). Porto Alegre: Sulina.

- KIM, H. J. (2000). Motivations for hyperlinking in scholarly electronic articles: a qualitative study. *Journal of the American Society for Information Science*, Washington, 51, 887-899.
- KOUPER, I. (2010). Science blogs and public engagement with science: practices, challenges and opportunities. *Journal of Science Communication*, 9. Recuperado em 27 julho, 2009, de, [http://jcom.sissa.it/archive/09/01/Jcom0901\(2010\)A02](http://jcom.sissa.it/archive/09/01/Jcom0901(2010)A02)
- LANDOW, G. P. (1995). Hipertexto: la convergencia de la teoría crítica contemporánea y la tecnología. Barcelona: Paidós, 1995.
- LUZÓN, M. J. (2008). Scholarly hyperwriting: the function of links in academic weblogs. *Journal of the American Society for Information Science and Technology*, New York, 60, 75-89.
- ROSSEAU, R. (1997). Citations: an exploratory study. *Cybermetrics*, n. 1, v. 1, 1997. Recuperado em 27 julho, 2009, de, <http://www.cindoc.csic.es/cybermetrics/>
- THELWALL, M. Link analysis: an information science approach. Amsterdam: Elsevier Academic, 2004.
- WANG; X.; JIANG, T; MA, F. (2010). Blog-supported scientific communication: an exploratory analysis based on social hyperlinks in a Chinese blog community blog-supported scientific communication: an exploratory analysis. *Journal of Information Science*. Recuperado em 27 julho, 2009, de, <http://jis.sagepub.com/content/early/2010/09/16/0165551510383189>
- WEINSTOCK, M. (1971). Citation Index. In: KENT, Allen; LANCOUR, Harold (Ed.). *Encyclopedia of Library and Information Science*. New York: M.Dekker, 1971. V. 5, p. 19.
- WILKINSON, D., & HARRIES, G., & THELWALL, M., & PRICE, E. (2003) Motivations for academic web site interlinking: evidence for the web as a novel source of information on informal scholarly communication. *Journal of Information Science*, v. 29, 59-66.

## Hiperlinks

- KJELLBERG, S. (2009). Scholarly blogging practice as situated genre: an analytical framework based on genre theory. *Information Research*, 14, paper 410. Recuperado em 27 julho, 2009, de, <http://informationr.net/ir/14-3/paper410.html>
- KJELLBERG, S (2010). I am a blogging researcher: motivations for blogging in scholarly context. *First Monday*, Bridgman, v. 15, n. 8, 2 Aug. 2010. Recuperado em 27 julho, 2009, de, <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2962/2580>.



- WILKINS, J. (2008). The Roles, reasons and restrictions of science blogs. *Trends in Ecology & Evolution*, Amsterdam, 23, 411-413 Recuperado em 27 julho, 2009, de, [http://www.cell.com/trends/ecology-evolution/abstract/S0169-5347\(08\)00200-0](http://www.cell.com/trends/ecology-evolution/abstract/S0169-5347(08)00200-0)
- SILVEIRA, M. A. A. da., & CAREGNATO, S. E., & BUFREM, L. S. (2014) Estudo das razões de citação na Ciência da Informação: proposta de classificação. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 7. Recuperado em 27 julho, 2017, de, [http:// http://basessibi.c3sl.ufpr.br/brapci/v/a/17423](http://basessibi.c3sl.ufpr.br/brapci/v/a/17423)
- SMITH, A. G (2004). Web links as analogues of citations. *Information Research*, v. 9, n. 4, 2004. Recuperado em 23 junho, 2009, de, <http://www.informationr.net/ir/9-4/paper188.html>

# Comunicação científica em livros e capítulos na área de educação e afins

*Scientific communication through books and book chapters in education and related fields*

**Ana Gabriela Clipes Ferreira**

Faculdade de Educação da Universidade Federal do Rio Grande do Sul  
[anaclipes@ufrgs.br](mailto:anaclipes@ufrgs.br)

**Clóvis Milton Duval Wannmacher**

Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul  
[clovisdw@ufrgs.br](mailto:clovisdw@ufrgs.br)

## Resumo

Este artigo tem como objetivo analisar as publicações de uma Faculdade de Educação brasileira em livros e capítulos de livros. Contextualiza a comunidade estudada e o veículo de comunicação científica formal livro, bem como faz breve discussão sobre o comportamento de publicação da área da educação e áreas afins. Pesquisa cientométrica, de caráter exploratório e descritivo. Com recorte temporal de 2000 a 2015, os livros e capítulos representaram um total de 2740 registros, ou 17,97% do total da produção. Quando verificados os itens com maior número de circulação na biblioteca, verifica-se que existe ligação entre estes e a produção docente da unidade, sendo que dos 50 títulos com maior número de empréstimos, 15 pertencem à produção intelectual da faculdade. Conclui-se que na comunidade pesquisada, os livros mantêm importância como veículo de comunicação científica, bem como pôde ser verificado a circulação destas informações.

**Palavras-chave:** Comunicação científica. Livro. Produção científica. Faculdade de Educação.

## Abstract

*This paper aims to analyze the publications of a Brazilian School of Education through books and book chapters. It contextualizes the studied community and the book as a vehicle for formal scientific communication, as well as makes a brief discussion about the publication behavior of the area of education and related areas. Scientometric research, exploratory and descriptive. With a time cut from 2000 to 2015, the books and chapters represented a total of 2740 records or 17.97% of the total production. When the items with the highest number of circulation in the library are checked, there is a link between these items and the teaching production of the unit, of which 50 belong to the intellectual production of the university. It is concluded that in the community surveyed, books maintain importance as a vehicle for scientific communication, as well as the circulation of this information, could be verified.*

**Keywords:** Scientific communication. Book. Scientific production. Faculty of education.

## 1. Introdução

As áreas e subáreas do conhecimento pesquisam e publicam de maneiras diferentes. Cada canal de comunicação científica pode ser o mais adequado dentro da comunidade na qual está inserida. O livro é tido como o meio mais utilizado pelos autores das áreas das ciências humanas e sociais para realizar e divulgar suas pesquisas, afirmações corroboradas por Meadows (1999), Carvalho e Manoel (2006) e Giménez-Toledo, Tejada-Artigas e Mañana-Rodríguez (2013), para quem os livros são inerentes à natureza destas áreas e são mais adequados para as características de comunicação dos resultados de suas pesquisas.

Meadows em sua obra discute e apresenta o comportamento de publicação dos pesquisadores e os principais veículos em que cada uma das áreas tem a tendência e tradição de divulgar os resultados. Nas ciências médicas, são os periódicos científicos, já algumas áreas consideram e utilizam meios menos formais, como os *papers* de eventos, como importantes meios de comunicação e divulgação (Odlyzko, 2002). As pesquisas apresentadas em congressos podem ser ou não posteriormente publicadas em um canal formal de comunicação científica (Meadows, 1999). Porém, os resultados da pesquisa de algumas áreas, bem como o espaço necessário para informar o acúmulo de conhecimento e descrever todas as variáveis nem sempre são adequados para a apresentação em artigos ou conferências (Giménez-Toledo, Tejada-Artigas, & Mañana-Rodríguez, 2013), ainda sendo o livro o meio ideal.

Porém, já é perceptível uma mudança na forma de publicar nas diversas áreas do conhecimento, existindo uma possível preferência na publicação em periódicos, ou mesmo uma migração para este veículo, provocado por diversas causas. Entre os motivos, a rapidez da edição que as revistas proporcionam em relação ao livro, questionável quando é observado o tempo da avaliação por pares em algumas áreas e em periódicos nacionais brasileiros, a adequação dos modelos das agências de fomento ou ainda o acesso livre à informação, mais presente se comparado à edição dos livros. Ainda assim, a publicação em livros e capítulos de livros ocorre, é presente e significativa dentro da comunicação científica nos diversos campos do conhecimento, e não é diferente na educação e áreas afins. Questiona-se, assim, como se dá a produção em livros e a utilização dos mesmos na comunidade acadêmica de uma Faculdade de Educação brasileira.

Dessa forma, neste artigo, o objetivo é analisar e apresentar as principais características da produção da Faculdade de Educação da Universidade Federal do Rio Grande do Sul (FACED/UFRGS) neste tipo específico de documento, qual seja, livros e capítulos de livros, observando: tipo de autoria e número de autores, temáticas mais recorrentes, departamento com maior número de produções, ano com maior frequência, origem (nacionais e estrangeiros) e idioma, através das editoras destas publicações. Por fim, verificar os títulos de maior circulação na biblioteca da unidade e conferir se há alguma relação com a produção intelectual.

## **2. Contextualizando o objeto: FACED/UFRGS e a comunicação científica em livros**

A FACED, selecionada para este trabalho, pertence à Universidade Federal do Rio Grande do Sul. Foi fundada em 1970, e tinha no seu quadro de recursos humanos, em julho de 2017, 143 docentes, de acordo o Sistema de Recursos Humanos (SRH) da Universidade. Os professores possuem formações acadêmicas diversificadas, podendo ser na área das ciências humanas, das ciências exatas e das ciências sociais aplicadas. Estão divididos em 3 departamentos da seguinte maneira: 39 no Departamento de Estudos Básicos (DEBAS, código EDU01), 51 no Departamento de Ensino e Currículo (DEC, código EDU02) e 53 no Departamento de Estudos Especializados (DEE, código EDU03). O código EDU0 é referente aos técnico-administrativos em educação e demais colaboradores, como docentes aposentados, vínculos temporários, tutores, entre outros. O Programa de Pós-Graduação em Educação (PPGEdu) foi fundado em 1974, e possui 101 professores orientadores de Mestrado acadêmico e Doutorado.

A comunicação científica é tida como processo de comunicação clássico, descrito, em 1949, por Shannon e Weaver, onde há um emissor da mensagem, um canal e um receptor (Caribé, 2015). Entre as funções da comunicação científica está a transferência de informações científicas, tecnológicas ou associadas a inovações destinadas aos especialistas em determinadas áreas do conhecimento (Bueno, 2010). Sendo o livro o meio de comunicação científica formal ou estruturada, ou ainda planejada, o formato se dá principalmente através de documentos do formato impresso (Targino, 2000). Autores como Le Coadic (2004) utilizam o termo comunicação impressa. A coleção impressa de livros é a principal da biblioteca setorial da Faculdade de Educação da UFRGS em números e circulação.

Os livros e os periódicos são os mais conhecidos e provavelmente os mais utilizados de comunicação científica formais. Entre os outros canais formais pode-se citar: obras de referência, relatórios técnicos, revisões de literatura, bibliografias de bibliografias. As principais vantagens desse meio de comunicação, apresentadas por Meadows (1999) e Targino (2000), estão descritas a seguir:

*O alcance de um público amplo.* Livros podem ser encontrados nas bibliotecas e em livrarias por preços razoavelmente acessíveis, porém os periódicos científicos são mais utilizados nas unidades de informação, uma vez que, normalmente, a modalidade de venda é através de assinatura, muito dispendiosa para um usuário comum, em especial estudantes. Muitas vezes também há interesse em um único artigo apenas, não sendo interessante ao usuário possuir o fascículo inteiro.

*Armazenagem e recuperação mais seguras.* Novamente volta-se aos centros de informação e documentação, unidades de informação, bibliotecas e arquivos. O papel apresenta uma segurança, pois representa para muitas pessoas uma segurança formalizada, impressa, existente fisicamente. Com certeza é uma vantagem, porém os meios onde essa informação é armazenada necessitam de cuidados específicos, pois, como vários estudos apontam, o papel não é eterno e ao sofrer com a ação de agentes externos sua durabilidade diminui drasticamente.

*Volume moderado de informações redundantes, maior rigidez e controle via avaliação prévia.* As edições que passam por uma editoração ou avaliação por pares possuem credibilidade. Para uma editora conceituada aceitar publicar um livro que contenha conhecimento científico, o editor comercial avalia não somente o retorno comercial, mas também a qualidade do conteúdo intelectual.

A principal desvantagem do livro é um nível de desatualização da informação. O processo editorial envolve muitas etapas e quando impresso ainda deve-se incluir o tempo das provas. Mesmo os periódicos, considerados de publicação mais rápida, demandam tempo para a publicação. Outra desvantagem é o pouco retorno para o autor, porém se houver reconhecimento pela sua comunidade, as recompensas serão recebidas de outras maneiras.

### 3. Métodos

Estudo cientométrico, de caráter descritivo e exploratório, com presença de dados quantitativos. Utiliza métricas para o tratamento dos dados quantitativos, caracterizando como um estudo sobre a produtividade de um determinado grupo de pesquisadores.

Os dados são referentes à produção intelectual da FACED, fornecidos pelo Centro de Processamento de Dados (CPD) da UFRGS através da coleta realizada com o software Aleph versão 22 em maio de 2017. Contém os dados da produção intelectual, campos específicos da catalogação e diferenciais da UFRGS, a saber, campos Marc 090, que identifica as áreas do conhecimento de acordo com o Conselho Nacional de Desenvolvimento Científico e Tecnológico, e 909, com mais subcampos identificadores: 'a' (departamento/unidade/órgão), 'b' (tipo de produção), 'c' (programa de pós-graduação), 'd' (curso de especialização), 'e' (curso de graduação), 'f' (órgão financiador), e 'g' (curso de ensino profissional) (Oliveira *et al.*, 2004). Neste trabalho, os tipos de produção serão **pa** (livro) e **pb** (capítulo de livro). Não é diferenciado o suporte, ou seja, são incluídos livros ou capítulos impressos e e-books, pois ambos são considerados tipo de produção **pa** ou **pb**.

O recorte temporal foi entre os anos de 2000 e 2015.

Organizados em planilha eletrônica, os dados foram tratados utilizando o recurso de tabela dinâmica, foram organizadas análises simplificadas dos tipos de documentos somente para fins ilustrativos deste trabalho. Os dados foram organizados em tabela e estão em ordem de documento com maior número até o menor número de ocorrências. A nuvem com as principais temáticas dos títulos foi elaborada com o recurso Wordl<sup>1</sup> e foi excluído o termo educação, a fim de destacar os demais temas.

Para verificar se há ligação entre os itens do tipo livro com maior circulação na biblioteca setorial e os livros e/ou da produção intelectual, utilizou-se o relatório do módulo Menu de Serviços do Sistema de Automação de Bibliotecas (SABi). Este módulo é destinado para controle e gerenciamento de diversos tipos de serviços da biblioteca como circulação, aquisição e produção intelectual. O intervalo de tempo será o mesmo, qual seja 2000 a 2015. A lista contou com os 50 títulos com maior número de empréstimos no período. As informações que constam nos itens com maior circulação são: o número de registro no

sistema, o título da obra, a categoria de usuário que retirou e o número de empréstimos por categoria. Serão apresentados somente aqueles títulos da produção intelectual.

A escolha metodológica pela produção intelectual da FACED ocorreu por mais de um motivo. Na UFRGS, a produção dos servidores docentes registrada no catálogo de bibliotecas é utilizada para fins de progressão funcional e distribuição de vagas nos departamentos; o depósito é realizado nas bibliotecas setoriais da respectiva unidade, sendo umas das atividades que, além de atender as demandas profissionais, ainda possibilita a preservação da memória institucional e dá acesso à informação (Universidade, 2006; Universidade, 2001; Oliveira *et al.*, 2004).

Outro fator decisivo foi o controle da autoria. Pelo controle de produção intelectual, somente foi contabilizado o pesquisador enquanto docente da Universidade, devido aos campos de catalogação diferenciais do MARC. Ainda há uma vantagem: padronização das autoridades e ausência da necessidade de limpeza de dados de autoria, uma vez que grande parte dos esforços realizados no tratamento de dados nas análises métricas por são despendidas nesta etapa (Mugnaini, 2006).

Além disso, o catálogo de bibliotecas da UFRGS não possui nenhum viés de idioma, assunto ou limitação de pesquisador, como, por exemplo, limitar os resultados somente a pesquisadores de elite, bolsistas, ligados aos programas de pós-graduação, entre outros resultados. Também não há inserção de artigos no formato *pré-print*, os documentos são categorizados por profissionais da área da ciência da informação, ao invés da inserção de dados pelo próprio pesquisador, como ocorre no Currículo Lattes, por exemplo, no qual as informações podem estar incompletas ou equivocadas. Certamente existem limitações na coleta realizada no Aleph, como falta de documentos pelo motivo de não realização de depósito pelo docente e equívoco na catalogação pelos bibliotecários, uma vez que nenhuma base de dados é a mais completa, todavia esta foi a metodologia mais adequada para atender aos objetivos deste trabalho e identificar a comunidade da FACED na sua totalidade quanto à publicação de livros e capítulos de livros.

## 4. Resultados

Do total de 15.254 registros, 2.740 são livros ou capítulos de livros, ou seja, 17,97% do total. Os capítulos são a maioria dos registros, sendo 1.996 (13,09%) e os livros 744 registros (4,88%). A Tabela 1 apresenta estes resultados e os demais tipos de documentos produzidos durante o período de 2000 a 2015. Em destaque, os tipos de documentos nos quais há maior produção.

**Tabela 1 – Documentos depositados por tipo, de 2000 a 2015**

Tipo de documento	Código na catalogação	Número	Porcentagem (%)
Trabalho publicado em anais de evento no país	ph	2627	17,22
Capítulo de livro	pb	1996	13,09
Trabalho de conclusão de curso de especialização	pn	1496	9,81
Trabalho de conclusão de curso de graduação	po	1461	9,58
Dissertação de mestrado acadêmico	pe	1055	6,92
Artigo de periódico nacional indexado	pgi	950	6,23
Tese	pd	866	5,68
Livro	PA	744	4,88
Texto de exame de qualificação	pex	728	4,77
Texto de apresentação/encerramento	pta	563	3,69
Artigo de periódico nacional não indexado	pgn	480	3,15
Trabalho de conclusão de disciplina de graduação	pod	429	2,81
Trabalho publicado em anais de evento fora do país	pj	386	2,53
Verbete	pve	238	1,56
Entrevista	pab	150	0,98
Artigo de divulgação	pp	148	0,97
Relatório técnico e de pesquisa	pl	138	0,90
Tradução	paf	106	0,69
Artigo de periódico estrangeiro indexado	pfi	103	0,68
Outros Documentos	pz	102	0,67
Apostilas	pc	94	0,62
Material gráfico ou visual	px	89	0,58
Artigo de periódico estrangeiro não indexado	pfn	86	0,56
Trabalho técnico	pad	81	0,53
Resenha	pre	58	0,38
Palestra	pac	32	0,21
Documento administrativo/de planejamento	pm	14	0,09
Recurso Educacional	poa	8	0,05
Dissertação de mestrado profissional	pep	7	0,05
Periódico	pt	6	0,04
Arquivo de computador	pi	5	0,03
Catálogo de evento	ps	4	0,03
Disciplina de mestrado acadêmico	pem	2	0,01
Relatório administrativo	pq	1	0,01
Trabalho de disciplina de doutorado	ped	1	0,01
<b>Total</b>		<b>15254</b>	<b>100</b>

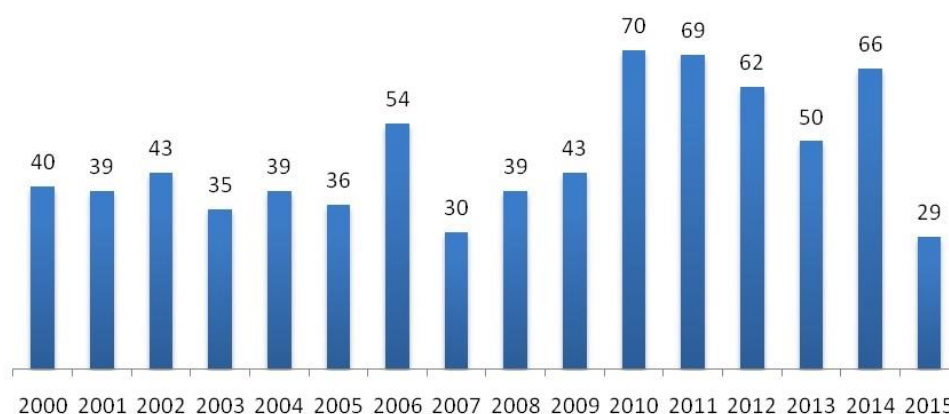
Fonte: dados da pesquisa



É possível observar que há destaque dos livros e capítulos de livros, além de textos de apresentação e encerramentos (prefácios, introdução, posfácios, entre outros), trabalhos publicados em eventos (nacionais) e trabalhos produzidos no âmbito acadêmico para fins de conclusão de cursos, tais como as teses e dissertações, nos quais a produção é obtida como orientador ou coorientador.

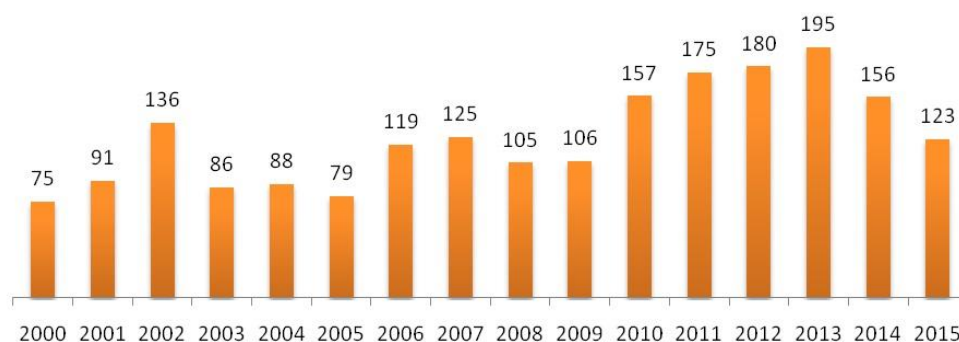
As Figuras 1 e 2 ilustram os resultados das publicações em livros e capítulos em livros, respectivamente, no intervalo da amostra. O ano com maior produção de livros foi 2010 (70 registros) e com maior número em capítulos de livros foi 2013 (195 registros).

**Figura 1 – Número de publicações em livros por ano (n = 744, 2000 a 2015)**



Fonte: dados da pesquisa

**Figura 2 – Número de publicações em capítulos por ano (n=1996, 2000 a 2015)**



Fonte: dados da pesquisa

Ao analisar o tipo de autoria dos livros, constata-se que 204 dos itens são livros autorais, dos quais 129 são autorias únicas, ou seja, 63,23%. Situação descrita como frequente na área das humanidades, a autoria única pode então ser comprovada em diversos desses registros. O número máximo de autores é de 12 pessoas em 16 registros. A autoria na área de educação e afins, aqui sendo representada pela FACED, está de acordo com a literatura e não possui número de autores elevados, ao contrário do que é observado na área das ciências duras, por exemplo, como indicado por diversos autores, como Rubio (1992), Meadows (1999), Kyvik (2003) e Nederhof (2006). Este resultado também corrobora os resultados da pesquisa de Santos (2010), embora a sua análise tenha ocorrido em periódicos, também é da área de humanas e ciências sociais, e mostra um expressivo número de autoria única: 47,6%.

Os demais livros são obras organizadas por autores ou da FACED ou de diferentes instituições, mas foram considerados como produção intelectual por conter capítulo de docente ou técnico com vínculo a um dos departamentos ou à Faculdade. Embora a produção dos técnicos não fosse um objetivo deste trabalho e seja um resultado pouco significativo diante da produção dos docentes, explicado não só pela diferença da natureza do trabalho dos dois tipos de profissionais, como também pela quantidade de técnicos que atuam na instituição e que estão ligados a alguma pesquisa, os resultados recuperaram as contribuições destes. Assim, há presença de técnicos administrativos da FACED em coautoria com os docentes, seja na escrita dos livros ou capítulos, ou em outras contribuições, como desenhista e ilustrador, e também em trabalhos técnicos (tradução ou consultorias, por exemplo).

Quando a análise é dos capítulos em livros, a autoria única também é significativa: dos 1.996 registros, 1.039 são escritos por autores únicos, 52,05%. O número máximo de autores é de 10 pessoas, em 7 registros. O resultado não foi muito diferente em relação aos livros.

**Quadro 1 – Número de livros publicados por departamento ano a ano**

DEP.	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total geral
EDU0	16	9	10	7	8	6	9	2	5	12	7	11	10	4	3	2	121
EDU01	5	6	10	9	3	7	8	7	13	9	29	12	11	12	33	8	182
EDU02	8	13	13	10	14	9	21	11	11	10	22	26	25	23	16	10	242
EDU03	11	11	10	9	14	14	16	10	10	12	12	20	16	11	14	9	199
<b>Total</b>	<b>40</b>	<b>39</b>	<b>43</b>	<b>35</b>	<b>39</b>	<b>36</b>	<b>54</b>	<b>30</b>	<b>39</b>	<b>43</b>	<b>70</b>	<b>69</b>	<b>62</b>	<b>50</b>	<b>66</b>	<b>29</b>	<b>744</b>

EDU0 - Vínculos diversos, EDU01 - Departamento de Estudos Básicos, EDU02 - Departamento de Estudo e Currículo, EDU03 - Departamento de Estudos Especiais

Antes de analisar a produtividade por departamento (Quadro 1), assim como nas áreas do conhecimento, destaca-se que os departamentos possuem diferenças entre si. Desde o número de docentes, primeiro fator a ser enfatizado, como as áreas de formação. Portanto, por se tratar de um estudo descritivo e exploratório, não é feito aqui qualquer julgamento além de apontar os dados absolutos da pesquisa. A produção em livros por departamento sinalizou como mais produtivo o Departamento de Ensino e Currículo (DEC). Entre os anos de 2000 a 2009, o número de publicações foi praticamente regular, com um aumento significativo entre os anos de 2010 e 2013. Nos dois últimos anos da amostra, 2014 e 2015, os números são similares aos de 2000 a 2009, todavia, não é possível mensurar se o depósito está completo ou mesmo se os livros foram publicados, uma vez que a editoração de livros pode demorar um pouco mais, se comparada a outros veículos de comunicação científica, em especial os veículos informais.

Os demais departamentos mantêm um número de publicações regular durante os anos, mas observando novamente o intervalo de 2010 a 2013, entre todos os departamentos, houve uma produção relevante. Em 2014 houve um aumento acentuado no Departamento de Estudos Básicos, assim como em 2010. O resultado é similar na análise das publicações em capítulos de livros, sendo novamente o DEC o mais produtivo (Quadro 2).

**Quadro 2 – Número de capítulos de livros publicados por departamento ano a ano**

DEP.	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total geral
EDU0	30	25	33	30	13	15	21	14	5	15	15	9	25	13	6	4	273
EDU01	8	18	20	16	12	15	26	41	27	24	40	39	39	48	41	24	438
EDU02	21	26	53	27	42	26	33	42	49	39	60	74	59	74	43	57	725
EDU03	16	22	30	13	21	23	39	28	24	28	42	53	57	60	66	38	560
<b>Total</b>	<b>75</b>	<b>91</b>	<b>136</b>	<b>86</b>	<b>88</b>	<b>79</b>	<b>119</b>	<b>125</b>	<b>105</b>	<b>106</b>	<b>157</b>	<b>175</b>	<b>180</b>	<b>195</b>	<b>156</b>	<b>123</b>	<b>1996</b>

EDU0 - Vínculos diversos, EDU01 - Departamento de Estudos Básicos, EDU02 - Departamento de Estudo e Currículo, EDU03 - Departamento de Estudos Especiais

Ao analisar as principais temáticas dos livros, ficam evidenciados os principais assuntos tratados nas obras (Figura 3). Há uma diversidade de temas, como formação de professores, ensino, gênero, sexualidade, corpo, escola, educação de jovens e adultos, educação infantil, prática pedagógica, currículo, entre muitos outros. Observa-se também a interdisciplinaridade



15	Modernidade líquida	121
16	Documentos de identidade : uma introdução às teorias do currículo	117
17	Psicologias : uma introdução ao estudo de psicologia	113
18	Pedagogia da autonomia : saberes necessários à prática educativa	111
19	Pesquisa social : teoria, método e criatividade	108
20	História social da criança e da família	107
21	Corpo, gênero e sexualidade : um debate contemporâneo na educação	106
22	Desenvolvimento psicológico e educação	104
23	A formação social da mente : o desenvolvimento dos processos psicológicos superiores	103
24	Sabores, cores, sons, aromas : a organização dos espaços na educação infantil	101
25	A adolescência	100
26	Convite à filosofia	100
27	Emílio ou Da educação	100
28	Mil platôs : capitalismo e esquizofrenia	100
29	Caminhos investigativos II : outros modos de pensar e fazer pesquisa em educação	99
30	Brincar : crescer e aprender : o resgate do jogo infantil	96
31	As cem linguagens da criança : a abordagem de Reggio Emilia na educação da primeira infância	95
32	Identidade e diferença : a perspectiva dos estudos culturais	95
33	Alfabetizar : fundamentos e práticas	94
34	Inclusão e escolarização : múltiplas perspectivas	94
35	Escola e sala de aula : mitos e ritos : um olhar pelo avesso do avesso	93
36	Sociologia da educação	93
37	Ciclos na escola, tempos na vida : criando possibilidades	91
38	Documentos de identidade : uma introdução às teorias do currículo	90
39	A paixão de conhecer o mundo : relatos de uma professora	89
40	Epistemologia genética	88
41	Pensamento e linguagem	88
42	Educação infantil : pra que te quero?	87
43	Histórias e memórias da educação no Brasil	86
44	500 anos de educação no Brasil	83
45	Fadas no divã : psicanálise nas histórias infantis	83
46	O corpo educado : pedagogias da sexualidade	83
47	Gestão escolar democrática : concepções e vivências	82
48	Seis estudos de psicologia	82
49	Caminhos investigativos II : outros modos de pensar e fazer pesquisa em educação	80
50	Psicogênese da língua escrita	80
<b>TOTAL GERAL</b>		<b>5724</b>

Fonte: Menu de Serviços SABI

O total dos 50 itens do tipo livro com maior número de circulação no intervalo selecionado na pesquisa, de 2000 a 2015, foi 5.724 empréstimos de um universo total de 236.544, sendo que os 50 títulos de publicações periódicas somaram 4.204 empréstimos no mesmo período. Dos 50 livros com maior número no intervalo, o primeiro é o título "Documentos de Identidade: uma introdução às teorias do currículo", 2.ª edição, 2005, de autoria de Tomaz Tadeu da Silva, foi produção intelectual da FAGED até 2010 (primeiro do Departamento de Ensino e Currículo e após, do Programa de Pós-Graduação), uma vez que o autor é docente da FAGED. O livro

não é, atualmente, bibliografia básica essencial do curso, porém tem grande circulação entre os discentes de graduação, assim como os de pós-graduação. O mesmo item é registrado novamente em mais duas edições, sendo o décimo sexto e o trigésimo oitavo (3.ª edição, 2009, e a primeira, 1999). O 4º item com maior número é "Alienígenas na sala de aula: uma introdução aos estudos culturais em educação", também de autoria Tomaz Tadeu da Silva.

"Caminhos investigativos III: riscos e possibilidades de pesquisar nas fronteiras" é o 8º item da listagem e é organizado por Marisa Vorraber Costa e Maria Isabel Edelweiss Bujes, ano da obra 2005. Faz parte da coleção Caminhos Investigativos e o outro item, II, aparece neste quadro. O 9º item é "Projetos pedagógicos na educação infantil", de Maria Carmen Silveira Barbosa e Maria da Graça Souza Horn, de 2008. "Educação infantil: pra que te quero?" organizado por Carmem Maria Craidy e Gládis Elise P. da Silva Kaercher, de 2001, é o décimo item e também o 42º, numa edição do ano de 1998.

O 12º item é um tipo de produção em livro como consultoria "Transgressão e mudança na educação: os projetos de trabalho", 1998, autoria de Fernando Hernández, sem vínculo com a FAGED, e consultoria, supervisão e revisão técnica de Maria da Graça Souza Horn, com vínculo. O 21º item é uma produção intelectual da FAGED e da Escola de Educação Física, "Corpo, gênero e sexualidade: um debate contemporâneo na educação" organizado por Guacira Lopes Louro, Jane Felipe Neckel, e Silvana Vilodre Goellner, de 2003, e com diversos capítulos também com produções docentes. O item a seguir, 22º, "Desenvolvimento psicológico e educação" organizado por Cesar Coll, Álvaro Marchesi, Jesús Palacios, é uma consultoria, supervisão e revisão técnica de Cláudio Roberto Baptista, docente da FAGED, do ano de 2004.

O 29º item é "Caminhos investigativos II: outros modos de pensar e fazer pesquisa em educação", 2007, e Marisa Vorraber Costa é a organizadora, também sendo o 49º numa edição de 2002. O 31º item é, novamente, uma consultoria e revisão técnica de Maria Carmen S. Barbosa, "As cem linguagens da criança: a abordagem de Reggio Emilia na educação da primeira infância", de Carolyn Edwards, Lella Gandini, George Forman, de 1999. A seguir, o 32º item, é organizado e possui traduções de Tomaz Tadeu da Silva na obra "Identidade e diferença: a perspectiva dos estudos", de 2009. O 33º item é "Alfabetizar: fundamentos e práticas", organizado por Maria Isabel H. Dalla Zen e Maria Luisa M. Xavier, todas docentes da

FACED. A 34ª obra é "Inclusão e escolarização: múltiplas perspectivas", 2006, organizado por Cláudio Roberto Baptista. "Escola e sala de aula: mitos e ritos: um olhar pelo avesso do avesso", 2008, Ivany Souza Àvila é o 35º item com mais retiradas no período.

"Ciclos na escola, tempos na vida: criando possibilidades", 2004, organizado por Jaqueline Moll é o 37º item com maior número de empréstimos no intervalo de 2000 a 2015. O 43º livro é "Histórias e memórias da educação no Brasil", publicado entre 2004 e 2005, em 3 volumes organizados por Maria Stephanou e Maria Helena Camara Bastos. "O corpo educado: pedagogias da sexualidade" organização de Guacira Lopes Louro e tradução de Tomaz Tadeu da Silva de 1999 é o 46º item com maior número de empréstimos.

Dos 50 itens com maior número de empréstimo, 15 títulos pertencem de alguma forma à produção intelectual, seja por autoria, organização ou alguma espécie de trabalho técnico, e alguns deles listados mais de uma vez em diferentes edições. O item mais antigo data de 1998 e o mais recente de 2009, lembrando que o ano dos itens não tem ligação com o ano da amostra. Não foram observados capítulos de todas as obras devido a especificidade de cada registro, mas somente esse resultado, mesmo que superficial em um primeiro momento, proporciona a dimensão da importância dos livros no formato impresso para a comunidade acadêmica da FACED e, entre essas obras, as produzidas pelos docentes que são objetos de pesquisa deste estudo. Esses 15 itens combinam com a nuvem temática da Figura 3, com assuntos similares aos em destaque.

Quanto à origem dos livros, ao observar as editoras com maior presença nas obras, conclui-se que foram as mais diversas. Todavia, há destaque para editoras especializadas em livros da área de ciências humanas, como Mediação, e editoras regionais, como a Artmed ou Grupo A (mudança de nome do grupo editorial), Evangraf e a Editora da Universidade (UFRGS). A maioria maciça é de editoras nacionais brasileiras e de obras em idioma português, similar à coleção da biblioteca.

## 5. Considerações Finais

Publicar, seja em qual for o canal, é um meio de receber reconhecimento dos resultados de pesquisas pelos pares e pela sociedade. Também possibilita que, com o passar dos anos, novos trabalhos melhorem o anterior, reconheçam o autor como pioneiro ou critiquem e apresentem uma nova contribuição. Estas funções assemelham-se aos 15 motivos para citar de Weinstock, citados por Davis (2009): homenagear pioneiros; dar crédito para os trabalhos relacionados; identificar metodologias, equipamentos, entre outros; oferecer leitura básica; retificar ou melhorar o seu próprio trabalho; retificar ou melhorar os trabalhos dos outros; criticar ou analisar trabalhos anteriores; sustentar declarações; informar os pesquisadores de trabalhos futuros; dar destaque a trabalhos pouco disseminados, inadequadamente indexados ou desconhecidos (não citados); validar dados e categorias de fatos, constantes físicas, entre outros; identificar publicações originais nas quais uma ideia ou conceito foram discutidos; identificar publicações originais que descrevem ideias ou conceitos epônimos, por exemplo, Doença de Hodgkin; contestar trabalhos ou ideias de outros; e debater a primazia das declarações de outros.

Os livros, na área das humanidades e afins, são relevantes, possuem importância e circulação, como pôde ser comprovado com os dados desta pesquisa, e na literatura por Giménez-Toledo, Tejada-Artigas e Mañana-Rodríguez (2013), Leydesdorff e Felt (2012), e Meadows (1999). Dentro da própria comunidade estudada, há uma produção significativa de livros e capítulos e há uma circulação dessa produção entre esta comunidade. Os docentes da FACED podem influenciar os discentes com suas pesquisas, pois os livros produzidos estão entre os mais retirados para empréstimos no intervalo selecionado.

Ainda que os dados apontem para números expressivos em publicações em outros veículos, tais como trabalhos em eventos e periódicos, os livros e capítulos possuem parcela significativa do total das produções no período, mostrando que os pesquisadores optam por publicar neste canal, mesmo sendo a produção editorial de um livro mais demorada em relação aos demais itens, e por vezes mais custosa financeiramente. Para Leydesdorff e Felt (2012), os livros circulam mais devagar se comparados aos periódicos, a leitura também é mais demorada. Logo, recebem menos citações, segundo os autores. Assim, pode ser negativo em campos com frentes de pesquisa e com pressão de publicação, como ciências biomédicas, por



exemplo. Porém, os livros atendem às funções básicas da comunicação científica, sistematizadas por Menzel, em 1958, e citadas por Targino (2010), entre elas: fornecimento de respostas a perguntas específicas; atualização profissional do cientista no campo específico de sua área de atuação; estimulação da descoberta e da compreensão de novos campos de interesse; fornecimento aos cientistas de ideia da relevância de seu trabalho através da divulgação das tendências de áreas emergentes; teste da confiabilidade de novos conhecimentos, diante da possibilidade de testemunhos e verificações; redirecionamento ou ampliação do rol de interesse dos cientistas; fornecimento de retroalimentação para aperfeiçoamento da produção do pesquisador.

Considera-se também que foi um trabalho relevante para as atividades da unidade de informação da faculdade. Os dados da produção intelectual e serviços do SABi proporcionam dados e informações de inigualável relevância para o sistema de bibliotecas da UFRGS, além dos próprios pesquisadores, com potencial para elaboração de diferentes pesquisas além dos habituais relatórios de gestão informacional da unidade, sejam da biblioteca ou da FACED. Os resultados puderam comprovar, mesmo que superficialmente, que as obras estão cumprindo as funções da comunicação científica, estão circulando entre a comunidade acadêmica e atendendo a outras partes do ciclo da comunicação, além de produzir e publicar, que é circular.

A pesquisa atendeu aos objetivos propostos, com as limitações encontradas no método, amostra e recorte temporal. É sugerido trabalhos cientométricos com análises em diferentes documentos, como livros, e estudos de citações destes documentos, procurando demonstrar a importância deste veículo, independente do formato impresso ou eletrônico, dentro da comunidade científica, através de outros indicadores e impactos, como citações, por exemplo, em que documentos esses livros são citados e qual o uso dos livros nas referências e citações dos trabalhos acadêmicos da área da educação e afins.

## Referências Bibliográficas

---

- BUENO, W. C. (2010). Comunicação científica e divulgação científica: aproximações e rupturas conceituais. *Informação & Informação*, 15(nesp), 1-12. doi:10.5433/1981-8920.2010v15nesp.
- CARVALHO, Y., & MANOEL, E. (2006). Para além dos indicadores de avaliação da produção intelectual na grande área da saúde. *Movimento*, 12(3), 193-225. Recuperado de <http://www.seer.ufrgs.br/index.php/Movimento/article/view/2915>
- CARIBÉ, R. D. (2015). Comunicação científica: reflexões sobre o conceito. *Informação & Sociedade: Estudos*, 25(3), 89-104. Recuperado de <http://www.ies.ufpb.br/ojs/index.php/ies/article/view/23109>
- DAVIS, P. M. (2009). Reward or persuasion? The battle to define the meaning of a citation. *Learned Publishing*, 22(1), 5-11. doi:10.1087/095315108X378712
- GIMÉNEZ-TOLEDO, E., TEJADA-ARTIGAS, C., & MAÑANA-RODRÍGUEZ, J. (2013). Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey. *Research Evaluation*, 22(1), 64-77. doi:10.1093/reseval/rvs036
- KYVIK, S. (2003). Changing trends in publishing behaviour among university faculty, 1980-2000. *Scientometrics*, 58(1), 35-48. doi:<https://doi.org/10.1023/A:1025475423482>
- LE COADIC, Y.-F. (2004). *A Ciência da Informação*. Brasília, DF: Briquet de Lemos.
- LEYDESDORFF, L., & FELT, U. (2012). "Books" and "book chapters" in the book citation index (BKCI) and science citation index (SCI, SoSCI, A&HCI). *Proceedings of the Association for Information Science and Technology*, 49(1), 1-7.
- MEADOWS, A. (1999). *A Comunicação Científica*. Brasília: Briquet de Lemos.
- MUGNAINI, R. (2006). Caminhos para adequação da avaliação da produção científica brasileira: impacto nacional versus internacional. Tese (Doutorado em Cultura e Informação), 254 p. São Paulo: Universidade de São Paulo. Recuperado de <http://www.teses.usp.br/teses/disponiveis/27/27151/tde-11052007-091052/>.
- NEDERHOF, A. J. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: a review. *Scientometrics*, v. 66, n. 1, p. 81-100. doi: <https://doi.org/10.1007/s11192-006-0007-2>
- ODLYZKO, A. (2002). The rapid evolution of scholarly communication. *Learned Publishing*, 15(1), 7-19. doi:10.1087/095315102753303634
- OLIVEIRA, Z. P. et al. (2004). O uso do campo MARC 9XX para controle bibliográfico institucional. *Ciência da Informação*, 33(2), p. 179-186. Recuperado de <http://revista.ibict.br/ciinf/article/view/1060/1145>

RUBIO, A. V. (1992). Scientific production of Spanish universities in the fields of Social Sciences and Language. *Scientometrics*, 24(1), 3-19. doi:<https://doi.org/10.1007/BF02026470>

SANTOS, S. M. (2010). Perfil dos periódicos científicos de Ciências Sociais e de Humanidades: mapeamento das características extrínsecas. *Dissertação (Mestrado) - Escola de Comunicação e Artes, Pós-graduação em Ciência da Informação*, 176 f. Universidade de São Paulo.

TARGINO, M. (2000). Comunicação científica: uma revisão de seus elementos básicos. *Informação e Sociedade: Estudos*, 10(2), 37-85. Recuperado de <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/326/248>

UFRGS. UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL. (30 jan. 2017). Conselho de Ensino, Pesquisa e Extensão. Recuperado de Resolução 38/2006, de 6 de setembro de 2006: <http://www.ufrgs.br/cepe/legislacao/resolucao-no-38-2006-de-06-09-2006>

UFRGS. UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL. (17 jan. 2017.). Conselho Universitário. Recuperado de Decisão 118/2001, de 17 de agosto de 2001: <http://www.ufrgs.br/consun/legislacao/documentos/decisao-no-118-2001>

## Notas

---

<sup>1</sup> [www.wordle.net](http://www.wordle.net)

# Referências de atos normativos brasileiros na comunicação científica internacional: estilos Vancouver e APA

*Bibliographic references of Brazilian legal acts in international scholarly communication: Vancouver and APA styles*

**Natascha Helena Franz Hoppen**

Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil  
[na.hoppen@hotmail.com](mailto:na.hoppen@hotmail.com)

**Ana Paula Medeiros Magnus**

Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil  
[magnus.ana@gmail.com](mailto:magnus.ana@gmail.com)

**Samile Andréa de Souza Vanz**

Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil  
[samile.vanz@terra.com.br](mailto:samile.vanz@terra.com.br)

## Resumo

Este trabalho apresenta proposta de normalização de referências de atos administrativos e normativos brasileiros nos estilos Vancouver e APA, utilizados internacionalmente na comunicação científica. Propõe-se uma adaptação das normas Vancouver e APA segundo a norma brasileira NBR 6023, a partir de duas linhas lógicas: primeiro, normalizar conforme as regras do estilo do suporte documental do ato administrativo. Segundo, verificar o tipo de entidade autora do ato (se de administração direta ou indireta) e então proceder conforme o estilo de normalização que se quer utilizar. Para o estilo Vancouver, utilizar a unidade federativa como autoria (e entrada) dos atos emanados por órgãos de administração direta, e diretamente pela instituição para atos emanados da administração indireta. Já para o estilo APA, permanecer o que foi consolidado com o uso: sem jurisdição para os órgãos de administração direta e

## Abstract

*This work presents a proposal for the standardization of bibliographic references of Brazilian administrative and normative acts in Vancouver and APA, internationally used, bibliographic styles. It proposes an adaptation of Vancouver and APA styles into the Brazilian bibliographical references standard NBR 6023 using two logical strands: first, to standardize according to the style's rules for the type of document. Secondly, verify the type of entity that authored the document (direct or indirect administrative bodies) and then proceed differently according to the style of normalization that one wants to use. For the Vancouver style, to use the federative unit as authorship, and entry for the acts originated from direct administrative bodies, and directly from the institution for acts originated from indirect administrative bodies. For the APA style, to continue what was consolidated by the use: no jurisdiction for the direct*

entrada pela entidade para os atos da administração indireta. Conclui que é competência do bibliotecário encontrar soluções eficazes para o uso correto da informação por seus usuários, enfatizando práticas de referência acurada de documentos, contribuindo com os processos de comunicação na ciência.

*administration bodies and entry by the institution for the indirect administration ones. It concludes that it is a competency from the librarian to find effective solutions for the correct use of information by its users, emphasizing documents' accurate referencing practices in order to contribute with the communication process in the science.*

**Palavras-chave:** Comunicação científica, Normalização – Referências, Documentos jurídicos.

**Keywords:** *Scholarly communication, Standardization – References, Legal documents.*

## 1. Introdução

Uma das características fundamentais da ciência é a sua necessidade de comunicação (Ziman, 1979; Meadows, 1999; Targino, 2000). Toda e qualquer investigação precisa ser comunicada, divulgada e publicada, sob a pena de perder seu objetivo ou padecer no desconhecimento. Publica-se a fim de se manter o ciclo da ciência: novas pesquisas, relatos, publicações, que, por sua vez, são lidas, criticadas e englobadas à massa de conhecimento, desenvolvendo seu avanço, ou rejeitadas e esquecidas. Tal necessidade produz uma grande quantidade de documentos e publicações que crescem exponencialmente (Solla Price, 1976) e neste contexto, torna-se fundamental que os pesquisadores utilizem padrões para facilitar a comunicação científica, a disseminação e a recuperação da informação. A correta descrição é fundamental na organização e recuperação da informação, pois cumpre a função de possibilitar o acesso a itens específicos (Café & Bräscher, 2008). A normalização das publicações ganhou força neste ambiente e hoje é um critério essencial para a publicação de qualquer documento científico. Conforme afirma Meadows (1999, p. 120): “O movimento no sentido de promover a normalização da apresentação [...] reflete as pressões crescentes exercidas sobre a comunicação científica [...] resultado de sua rápida expansão, que dificultou ainda mais para os leitores a localização de informações relevantes.”

No contexto da comunicação científica, Vanz e Santos (2011, p. 124) ressaltam a necessidade e:

[...] a importância não somente de indicar o autor das ideias e trechos transcritos num determinado documento, mas igualmente a referência acurada dos documentos citados. As referências com todos os dados completos, da autoria até a indicação dos títulos dos periódicos (se for o caso), locais de publicação [quando a

norma assim requerer] e datas, são elementos que elevarão o grau de confiança no documento por parte daqueles que o usarão.

As autoras ainda advertem que é competência do bibliotecário tanto aplicar as normas quanto disseminar sua existência e ensinar seu uso, visto que não são “propriedade” do profissional, mas foram criadas para toda a comunidade (Vanz & Santos, 2011). Compete também ao bibliotecário, portanto, buscar soluções para os problemas recorrentes no uso dessas normas a fim de auxiliar na sua correta utilização.

A normalização contribui para harmonizar as especificidades de cada área do conhecimento em diferentes veículos e ainda, possibilitar a interoperabilidade de dados entre diferentes sistemas computacionais. Por isso é reconhecida como atividade fundamental para a comunicação científica. Para Rodrigues, Lima e Garcia (1998), a normalização surge como possibilidade metodológica para uniformizar a expressão escrita de diversas ciências, que possuem conteúdos, objetos e metodologias diversificadas.

Dentre os documentos que podem ser referenciados estão os documentos jurídicos, que são considerados documentos especiais na referência, visto a dificuldade comum em elencar seus elementos de referência, como autoria, título, data, entre outras informações. Têm-se como documento jurídico:

[...] o documento textual produzido com observância da lei em esfera pública ou privada, que se constitui no próprio ato jurídico, ou, ainda, em registro ou estudo de fato jurídico, cuja produção e forma textual variam em função de sua finalidade, conteúdo e contexto de produção. Desdobra-se em seis categorias documentais: os atos legais, os atos administrativos, os atos judiciais, os atos negociais, os atos notariais e de registro e a doutrina jurídica. (Souza, 2013, p. 114).

Em outras palavras, trata-se de todo documento que sirva de suporte para informação jurídica, legislativa ou doutrinária emanado de órgãos ou entidades da administração pública, direta ou indireta. Na administração direta brasileira têm-se a União, seus Estados, o Distrito Federal e órgãos diretamente subordinados (como ministérios e secretarias). Na administração indireta têm-se as entidades com personalidade jurídica própria, mas que exercem funções do Estado de forma descentralizada, como as autarquias – dentre elas as autarquias especiais, agências reguladoras ou executivas, por exemplo, a Agência Nacional de Vigilância Sanitária (ANVISA) e a Agência Nacional de Aviação Civil (ANAC) –, as fundações, sociedades de economia mista, agências públicas e empresas públicas.

O formato Vancouver, de origem canadense, é um estilo de normalização de referências documentais comumente usado nas publicações científicas da área de ciências médicas e da saúde, o que inclui periódicos internacionais, assim como livros, teses e dissertações, anais, dentre outras. O estilo foi adaptado pela National Library of Medicine (NLM), dos Estados Unidos da América (EUA) com base nos trabalhos do International Committee of Medical Journal Editors (ICMJE, ou Comitê Internacional de Editores de Periódicos Médicos, tradução nossa), cujo primeiro encontro ocorreu na cidade de Vancouver, em 1978 (British Medical Association [BMA], [20--]). Conforme a ICMJE (2010), seu estilo de referência é ainda baseado nas normas da American National Standards Institute (ANSI) e adaptado pela NLM.

O estilo American Psychological Association (APA) foi criado em 1929, em conferência realizada na cidade estadunidense de Washington, por editores e gestores de periódicos das áreas de Antropologia e Psicologia. Nessa conferência, publicou-se na Psychological Bulletin, revista da American Psychological Association, o primeiro artigo com regras que visaram à economia de tempo e dinheiro por parte dos editores de revistas, visto que, na época não existia um padrão para os manuscritos submetidos às publicações científicas (Breitenbach, 2016). Portanto, o estilo APA surgiu para melhorar a qualidade dos documentos não somente no que diz respeito à redação, mas também em seu formato de referências citadas. Hoje, a norma em estilo APA é publicada pela American Psychological Association e está em sua sexta edição.

O pesquisador e o bibliotecário envolvidos com a comunicação científica e que trabalham em disciplinas ligadas às áreas da saúde devem conhecer e saber utilizar a padronização Vancouver. Da mesma forma, o estilo APA, muito difundido em publicações da área de psicologia, também está sendo utilizado em muitas publicações lusófonas das áreas das humanidades. Todavia, o Brasil possui um sistema próprio para referenciação, ditado pela Associação Brasileira de Normas Técnicas (ABNT). Por um lado, a NBR 6023, norma brasileira de elaboração de referências, abrange aspectos pertinentes à realidade brasileira, como a forma de redação de atos oficiais. Por outro, pesquisadores e editores da área biomédica, ciências sociais aplicadas e humanas precisam utilizar as normas já consagradas internacionalmente neste meio – Vancouver e APA. Dada a dificuldade encontrada pelos pesquisadores e bibliotecários para citar e referenciar atos administrativos e normativos brasileiros de acordo com as normas Vancouver e APA, utilizadas para a comunicação

científica internacional, este artigo apresenta uma proposta de normalização deste tipo de referência, atos administrativos e normativos brasileiros, nos estilos Vancouver e APA.

## 2. Os estilos internacionais de referências e os atos normativos brasileiros

O formato Vancouver não possui indicações sobre qualquer elemento dos atos administrativos e normativos brasileiros, atos estes que possuem regras específicas de redação (Brasil, 2002; Passos & Barros, 2009). O *Citing Medicine*, importante guia do formato Vancouver para a área biomédica, já prevê a necessidade de adaptação das referências dos dispositivos legais ao indicar uma obra específica à parte para referência de obras legislativas, um tipo de obra que não possui correspondente nacional:

Os profissionais da área jurídica [nos EUA] utilizam um sistema particular de citação diferentemente da medicina e ciências. Esse estilo de citação é descrito em detalhes em “The bluebook: a uniform system of citation”, 18ª edição, da Harvard Law Review Association. Como essa norma legal é bem estabelecida e seu formato de citação identifica com precisão os documentos legais para a recuperação de informações legais e em bibliotecas gerais, nenhuma tentativa foi feita para referenciar materiais legais, como leis e audiências públicas, em um formato tradicional. (Patrias, 2007, doc. não paginado, tradução nossa).

A seguir, o *Citing Medicine* dá apenas alguns exemplos de referências de dispositivos legais, lembrando ainda que:

Os exemplos [...] referem-se ao estatuto legal federal dos Estados Unidos da América e documentos legislativos. Eles representam as normas seguidas nos Estados Unidos e não podem ser aplicáveis a documentos legais em outros países. Documentos produzidos por estados dos EUA [também] variam em formato [...]. (Patrias, 2007, doc. não paginado, tradução nossa).

Ciente de que o próprio guia para utilização do formato Vancouver orienta pela utilização de outra norma preestabelecida para referência de dispositivos legais e jurídicos, a Faculdade de Saúde Pública da Universidade de São Paulo (FSP-USP), em seu *Guia de apresentação de teses*, indica uma adaptação da norma, que conforme se pode notar, tem como base as regras da ABNT:



**Figura 1 – Leis, adaptação da Biblioteca Virtual em Saúde (BVS)**

Vancouver [adaptado]
<ul style="list-style-type: none"><li>■ Brasil. Lei n.º 8.080, de 19 de setembro de 1990. Dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes e dá outras providências. Diário Oficial da União. 20 set 1990; Seção 1:018055.</li><li>■ São Paulo (Estado). Lei n.º 10.241, de 17 de março de 1999. Dispõe sobre os direitos dos usuários dos serviços e das ações de saúde no Estado e dá outras providências. Diário Oficial do Estado de São Paulo. 18 mar 1999; Seção 1:1.</li></ul>
ABNT
<ul style="list-style-type: none"><li>■ BRASIL. Lei n.º 8.080, de 19 de setembro de 1990. Dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes e dá outras providências. <b>Diário Oficial da União</b>, Brasília, DF, 20 set. 1990. p. 018055.</li><li>■ SÃO PAULO (Estado). Lei n.º 10.241, de 17 de março de 1999. Dispõe sobre os direitos dos usuários dos serviços e das ações de saúde no Estado e dá outras providências. <b>Diário Oficial do Estado de São Paulo</b>, São Paulo, 18 mar. 1999. Seção 1, p. 1.</li></ul>

Fonte: Universidade de São Paulo (2006)

Assim como no formato Vancouver, o estilo APA também não foi criado visando sua aplicação às especificidades legislativas brasileiras. O livro *Regras essenciais de estilo da APA*, publicado pela American Psychological Association ([APA], 2012), pontua que há uma diferença entre a posição e o formato das referências nos periódicos jurídicos e periódicos que aplicam as normas da APA nos EUA. Quando se fala em posição, quer dizer que em periódicos jurídicos usualmente se utiliza as referências em notas de rodapé, já nos periódicos que utilizam o estilo APA, as referências são colocadas em uma lista única no final do documento. Em relação ao formato, nos EUA, na área jurídica utiliza-se o já citado *The Blue Book: a uniform system of citation* (Harvard Law Review Association, 2010) para a formatação de referências. Portanto, na norma da APA não há indicação própria para o uso da norma em documentos jurídicos. Da

mesma forma que o guia oficial do estilo Vancouver, a APA sugere, em seus exemplos, o uso do *Blue Book* para a padronização de referências jurídicas. No Brasil, a Fundação Escola de Comércio Álvares Penteado (FECAP) elaborou, em 2016, o *Manual APA: regras gerais de estilo e formatação de trabalhos acadêmicos*, em que sugere o seguinte formato para documentos jurídicos:

**Figura 2 – Documentos Jurídicos em formato APA**

**7.3.6.1 Constituição Federal**

*Constituição da República Federativa do Brasil: Promulgada em 5 de outubro de 1988. (2003). (32a ed.). São Paulo: Saraiva.*

*Constituição da República Federativa do Brasil de 1988. Recuperado de [http://www.planalto.gov.br/ccivil\\_03/Constituicao/Constituicao.htm](http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm)*

**7.3.6.2 Leis**

*Lei n. 6.404, de 15 de dezembro de 1976. Dispõe sobre as sociedades por ações. Recuperado de [http://www.planalto.gov.br/ccivil\\_03/leis/16404compilada.htm](http://www.planalto.gov.br/ccivil_03/leis/16404compilada.htm)*

*Lei n. 11.638, de 28 de setembro de 2007. Altera e revoga dispositivos da Lei n. 6.404, de 15 de dezembro de 1976, e da Lei n. 6.385, de 7 de dezembro de 1976, e estende às sociedades de grande porte disposições relativas à elaboração e divulgação de demonstrações financeiras. Recuperado de [http://www.planalto.gov.br/ccivil\\_03/\\_ato2007-2010/2007/lei/11638.htm](http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/lei/11638.htm)*

**Fonte:** Almeida, Lopes, Camilo, & Choi (2016, p. 75)

Observa-se que no manual da FECAP parte-se do princípio de que as referências sejam retiradas da internet, sem fornecer exemplos de documentos consultados em outro formato.

No trabalho de editoração são feitas muitas adaptações a fim de atender às demandas de cada obra, seus autores e seu público. Para isso, entretanto, as normas e instruções de normalização devem ser consistentes, caso contrário, as adaptações acabam tornando-se provisórias ao invés de se efetivarem como soluções pertinentes e fundamentadas. Tal fato pode ser percebido nas tentativas de normalização de documentos jurídicos com o formato Vancouver. Como o formato foi criado fora do contexto lusófono, diversas particularidades dos atos administrativos brasileiros não são contempladas, o que gera uma dispersão de entendimentos do que e como se deve referenciar esses atos em estilos estrangeiros e internacionais, resultando em contradições na normalização entre autores, às vezes de uma mesma obra e até mesmo dentro do mesmo texto. Desta forma surge a necessidade de definir

um procedimento padrão a ser seguido na normalização de atos normativos brasileiros através do estilo de referências Vancouver e APA.

Há algumas diferenças também entre o que os estilos estabelecem e o que a comunidade acadêmica consagra com o uso. Por exemplo, o formato Vancouver estabelece que se deve referenciar até o sexto autor. Já a adaptação dessas normas utilizadas nas bases de dados da NLM dita que todos os autores devem ser referenciados.

**Figura 3 – Exemplo de referência Vancouver que lista até seis autores seguidos de et al. com nota avisando que a NLM cita todos os autores**

1. *Standard journal article*  
List the first six authors followed by et al. (Note: NLM now lists all authors.)

[...]

Rose ME, Huerbin MB, Melick J, Marion DW, Palmer AM, Schiding JK, et al. Regulation of interstitial excitatory amino acid concentrations after cortical contusion injury. *Brain Res.* 2002;935(1-2):40-6.

Fonte: Adaptado de U. S. National Library of Medicine ([NLM], 2011)

**Figura 4 – Indicação na Citing Medicine que indica "Liste todos os autores/editores, independentemente do número"**

**Author/Editor for Entire Books (required)**

**General Rules for Author/Editor**

- List names in the order they appear in the text
- Enter surname (family or last name) first for each author/editor
- Capitalize surnames and enter spaces within surnames as they appear in the document cited on the assumption that the author approved the form used. For example: Van Der Horn *or* van der Horn; De Wolf *or* de Wolf *or* DeWolf.
- Convert given (first) names and middle names to initials for a maximum of two initials following each surname
- ➔ Give all authors/editors, regardless of the number
- Separate author/editor names from each other by a comma and a space
- If there are no authors, only editors, follow the last named editor with a comma and the word editor or editors; see [Editor and Other Secondary Authors](#) below if there are authors and editors
- End author/editor information with a period

Fonte: Citing Medicine (Patrias, 2007)

Já no estilo APA, as regras de citações das referências utilizadas ao longo do texto são bem consolidadas nos periódicos e demais documentos que utilizam o estilo, no entanto exigem atenção por parte do escritor (ver Figura 5). Por exemplo, citam-se todos os nomes quando houver um ou dois autores no documento. Quando houver de três a cinco autores, indica-se o sobrenome de todos eles na primeira citação e nas citações subsequentes, indica-se

somente o primeiro autor seguido da expressão latina abreviada “et al.”, entre outros. No caso de citações com seis ou mais autores, coloca-se apenas o sobrenome do primeiro autor seguido da expressão “et al.” (Almeida et al., 2016).

Salienta-se que na lista de referências, quando o documento indicar até sete autores, todos os nomes são indicados na referência, separando o sobrenome do último autor do penúltimo pelo sinal tironiano, “&”. E quando o documento indicar oito ou mais autores, indicam-se os seis primeiros, seguidos de três pontos (reticências) e do nome do último autor, não utilizando a expressão “et al.”.

**Figura 5 – Estilos básicos de citação utilizando o estilo APA**

Tipo de citação	Primeira citação no texto	Citações subsequentes no texto	Formato entre parênteses, primeira citação no texto	Formato entre parênteses, citações subsequentes no texto
Trabalho de um autor	Walker (2007)	Walker (2007)	(Walker, 2007)	(Walker, 2007)
Trabalho de dois autores	Walker and Alien (2004)	Walker and Alien (2004)	(Walker & Alien, 2004)	(Walker & Allen, 2004)
Trabalho de três autores	Bradley, Ramirez, and Soo (1999)	Bradley et al. (1999)	(Bradley, Ramirez, & Soo, 1999)	(Bradley et al., 1999)
Trabalho de quatro autores	Bradley, Ramirez, Soo, and Walsh (2006)	Bradley et al. (2006)	(Bradley, Ramirez, Soo, & Walsh, 2006)	(Bradley et al., 2006)
Trabalho de cinco autores	Walker, Alien, Bradley, Ramirez, and Soo (2008)	Walker et al. (2008)	(Walker, Alien, Bradley, Ramirez, & Soo, 2008)	(Walker et al., 2008)
Trabalho de seis ou mais autores	Wasserstein et al. (2005)	Wasserstein et al. (2005)	(Wasserstein et al., 2005)	(Wasserstein et al., 2005)
Entidades (identificadas por abreviaturas) como autores	National Institute of Mental Health (NIMH, 2003)	NIMH (2003)	(National Institute of Mental Health [NIMH], 2003)	(NIMH, 2003)
Entidades (sem abreviatura) como autores	University of Pittsburgh (2005)	University of Pittsburgh (2005)	(University of Pittsburgh, 2005)	(University of Pittsburgh, 2005)

Fonte: APA (2012, p. 203)

### 3. Procedimentos metodológicos

A questão de pesquisa que originou o presente trabalho materializou-se nas tarefas cotidianas de bibliotecárias que trabalham com editoração de obras acadêmicas e científicas. Dentre as atividades de editoração de manuscritos pertinentes à comunicação científica estão: a revisão do texto para checagem de citações e/ou referências inconsistentes; a normalização de referências e citações em diferentes estilos de normalização bibliográfica; a validação dos elementos destas referências (títulos de periódicos, grafia dos nomes dos autores, etc.); a correspondência com autores e pesquisa documental para dirimir elementos faltantes ou dúvidas quanto aos documentos citados; a diagramação do texto; a marcação em linguagem

eXtensible Markup Language (XML); entre outras. As principais inconsistências observadas nos documentos que referenciam documentos jurídicos são:

- a) inconsistência na indicação da autoria do documento;
- b) formas variantes e incongruentes para se referenciar o mesmo tipo de ato e/ou documento jurídico;
- c) inexistência/falta de elementos essenciais da referência.

Fez-se um levantamento dos tipos de documentos jurídicos brasileiros, com ênfase nos atos normativos (mais citados e utilizados em textos acadêmicos). A partir da pesquisa de como se constituem e como são redigidos os documentos jurídicos brasileiros, foi feita uma pesquisa dentro dos manuais dos formatos Vancouver e APA a fim de diagnosticar a diretriz mais adequada e pertinente aos documentos brasileiros. Para elaboração da proposta utilizaram-se a NBR 6023 (Associação Brasileira de Normas Técnicas [ABNT], 2002) e o estilo Vancouver e APA, com base também no que é proposto pela FSP-USP (USP, 2006), FECAP (ALMEIDA et al., 2016) e pelas Regras essenciais de estilo APA (APA, 2012).

#### **4. Proposta de normalização**

O guia da FSP USP (USP, 2006) indica que se utilize algumas características gráficas próprias da norma Vancouver (como por exemplo, a ausência de destaques e uso de maiúsculas), mas que se mantenha a estrutura das normas da ABNT, com qualquer que seja o tipo de documento jurídico. Tal orientação pode ser bastante eficiente a priori (seguir o formato ABNT, com características gráficas de Vancouver), mas, por ser muito geral, permite o levantamento de diversas particularidades que devem ser pensadas e definidas a fim de não se voltar às disparidades de normalização advindas da interpretação pessoal de cada autor ou editor. É verdadeiro que algumas diferenças sempre serão encontradas, visto que nem tudo pode ser previsto na teoria, e os documentos e as formas de publicação evoluem, demandando a atualização das normas. Contudo, se propõe aqui a solução para dois dos principais pontos de conflito encontrados na formatação das referências e citações de documentos jurídicos: estabelecimento de autoria, de quais elementos devem compor a referência e estabelecimento de qual dentre todas as regras utilizar para cada tipo de documento.

A ABNT caracteriza como documento jurídico a legislação, a jurisprudência e a doutrina (ABNT, 2002) e, conforme é apresentado em seus exemplos, estabelece que cada documento seja referenciado de acordo com o suporte de sua informação: monografia no todo ou parte, artigo de jornal ou periódico, entre outros. A entrada (ponto de acesso no caso da catalogação e autoria no caso das referências), segundo a ABNT, é feita pela entidade autora (a unidade federativa no caso dos atos da administração direta) utilizando-se qualificador para distinguir homógrafos (como no exemplo a seguir com São Paulo). A *Declaração de princípios internacionais de catalogação* da International Federation of Library Associations and Institutions ([IFLA], 2009) também prevê que a entrada seja feita pela jurisdição quando se trata de uma coletividade (entidade) que seja parte dessa jurisdição ou autoridade territorial.

**Figura 6 – Legislação ABNT**

<b>7.9 Documento jurídico</b>	
Inclui legislação, jurisprudência (decisões judiciais) e doutrina (interpretação dos textos legais).	
<b>7.9.1 Legislação</b>	
Compreende a Constituição, as emendas constitucionais e os textos legais infraconstitucionais (lei complementar e ordinária, medida provisória, decreto em todas as suas formas, resolução do Senado Federal) e normas emanadas das entidades públicas e privadas (ato normativo, portaria, resolução, ordem de serviço, instrução normativa, comunicado, aviso, circular, decisão administrativa, entre outros).	
<b>7.9.1.1</b> Os elementos essenciais são: jurisdição (ou cabeçalho da entidade, no caso de se tratar de normas), título, numeração, data e dados da publicação. No caso de Constituições e suas emendas, entre o nome da jurisdição e o título, acrescenta-se a palavra Constituição, seguida do ano de promulgação, entre parênteses.	
Exemplos:	<p><b>A</b> SÃO PAULO (Estado). Decreto nº 42.822, de 20 de janeiro de 1998. <b>Lex:</b> coletânea de legislação e jurisprudência, São Paulo, v. 62, n. 3, p. 217-220, 1998.</p> <p><b>B</b> BRASIL. Medida provisória nº 1.569-9, de 11 de dezembro de 1997. <b>Diário Oficial [da] República Federativa do Brasil</b>, Poder Executivo, Brasília, DF, 14 dez. 1997. Seção 1, p. 29514.</p> <p><b>C</b> BRASIL. Decreto-lei nº 5.452, de 1 de maio de 1943. <b>Lex:</b> coletânea de legislação: edição federal, São Paulo, v. 7, 1943. Suplemento.</p> <p><b>D</b> BRASIL. <b>Código civil</b>. 46. ed. São Paulo: Saraiva, 1995.</p>

**Fonte:** adaptado de ABNT (2002)

Nos exemplos da figura 6, tem-se:

- legislação estadual publicada e referenciada como artigo de periódico, com entrada pela jurisdição e com qualificador (“Estado”);
- legislação federal publicada e referenciada como artigo de jornal com entrada pela jurisdição;
- legislação federal como suplemento de periódico;
- legislação federal como monografia no todo.

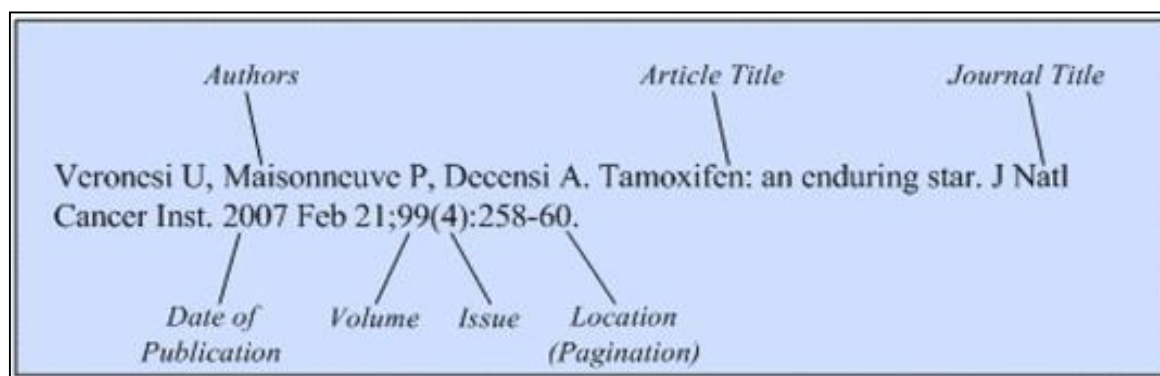
Para normalização dos mesmos documentos em formato Vancouver ou APA, propõe-se o mesmo tipo de sistemática: considerar o suporte do documento e assim, proceder conforme os estilos indicam para cada um dos tipos de publicação. A seguir se apresenta a proposta de modelo de normalização nos formatos Vancouver e APA para os exemplos A, B, C e D presentes na Figura 6.

- a) **VANCOUVER** – São Paulo (Estado). Decreto nº 42.822, de 20 de janeiro de 1998. Lex: coletânea de legislação e jurisprudência. 1998;62(3):217-20.

**APA** – Decreto nº 42.822, de 20 de janeiro de 1998. (1998). *Lex: coletânea de legislação e jurisprudência*, 62(3), 217-220.

Nos exemplos, 20 de janeiro de 1998 é a data do ato (e faz parte de seu título), não é a data de publicação (considerada apenas o ano, 1998, único dado explícito na referência original);

Figura 7 – **Exemplo de artigo de periódico em Vancouver**



Fonte: Citing Medicine (Patrias, 2007)

- b) **VANCOUVER** – Brasil. Medida provisória nº 1.569-9, de 11 de dezembro de 1997. Diário Oficial [da] República Federativa do Brasil. 1997 dez 14;(Seção 1):29514.

Neste exemplo, a data de publicação do ato é mencionada (14 de dezembro de 1997, sendo a data do ato dia 11 de dezembro);

**APA** – Medida provisória nº 1.569-9, de 11 de dezembro de 1997. (1997). *Diário Oficial [da] República Federativa do Brasil, Seção 1*, 29514.

**Figura 8 – Exemplo da Citing Medicine para artigo com órgão governamental como autoria (com qualificador no primeiro e hierarquia de jurisdição de órgão de administração direta no segundo)**

8. Journal article with governmental body as author

National Institutes of Health (US). End-of-life care. National Institutes of Health statement on the state of the science. AWHONN Lifelines. 2005 Feb-Mar;9(1):15-22.

United States District Court, S.D. Florida, Miami Division. Greenberg v. Miami Children's Hospital Research Institute. Wests Fed Suppl. 2003;264:1064-78.

Fonte: Citing Medicine (Patrias, 2007)

- c) **VANCOUVER** – Brasil. Decreto-lei nº 5.452, de 1 de maio de 1943. Lex: coletânea de legislação: edição federal. 1943;7 Sup.

**APA** – Decreto-lei nº 5.452, de 1 de maio de 1943. (1943). *Lex: coletâneas de legislação: edição federal, 7 Sup.*

Observe que, para APA, a entrada não inclui a jurisdição. Além disto, nestes exemplos o suplemento não tem paginação, como no exemplo da NLM em seguida;

**Figura 9 – Volume com suplemento, exemplo da NLM**

6. Volume with supplement

Geraud G, Spierings EL, Keywood C. Tolerability and safety of frovatriptan with short- and long-term use for treatment of migraine and in comparison with sumatriptan. Headache. 2002;42 Suppl 2:S93-9.

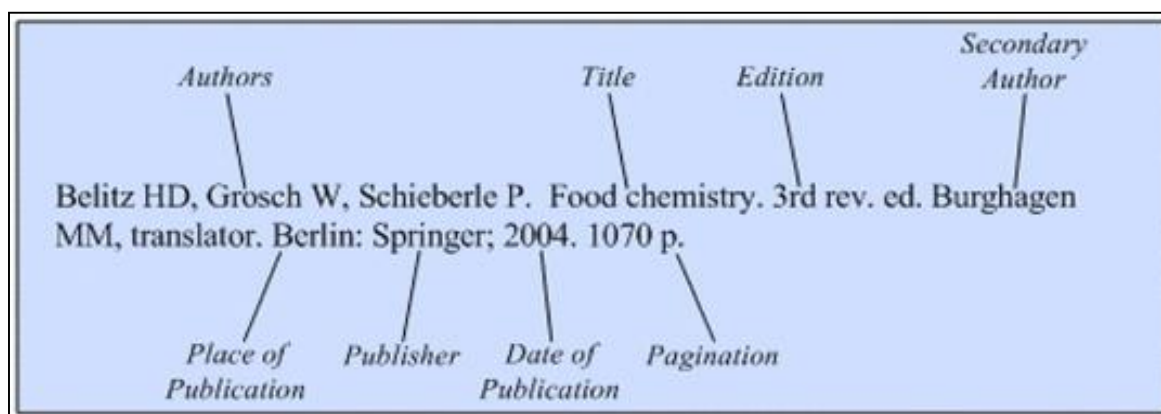
Fonte: NLM (2011)

- d) **VANCOUVER** – Brasil. Código civil. 46. ed. São Paulo: Saraiva; 1995.

**APA** – Brasil. (1995). *Código civil* (46. ed.). São Paulo: Saraiva.

Neste caso, a autoria aparece.

**Figura 10 – Monografia no todo, exemplo Citing Medicine**



Fonte: Citing Medicine (Patrias, 2007)



Todos os casos listados incluem legislação de órgãos de administração direta, matéria legal sempre de autoria de pessoas públicas políticas (União, Estados ou municípios). A entrada (autoria) proposta para o estilo Vancouver, nestes casos: sempre pela unidade federativa (estados ou distrito federal), municípios ou país. Para APA, manter o consolidado com o uso: não incluir a jurisdição (nos documentos de entidades de administração direta, como nos exemplos da FSP USP e FECAP, que adaptaram as normas de acordo com a sexta edição do Manual da APA).

A ABNT, dentro do tópico de legislação, afirma que compreende a Constituição, suas emendas e textos legais infraconstitucionais, assim como “[...] normas emanadas das *entidades públicas e privadas* (ato normativo, portaria, resolução, ordem de serviço, instrução normativa, comunicado, aviso, circular, decisão administrativa, entre outros).” (ABNT, 2002, p. 8, grifo nosso). Mas, na norma de referências da ABNT – NBR 6023 – não há qualquer exemplo com documentos de entidades privadas ou, como se pode aferir que era a intenção da norma, de órgãos de administração indireta. Todos os atos normativos que constam nos exemplos da ABNT são de órgãos de administração direta, que tem como entrada a jurisdição dos mesmos.

Aqui se propõe que, tanto para o estilo Vancouver quanto para APA, para as referências de entidades de administração *indireta*: a entrada das referências (ou seja, a autoria) seja feita diretamente pela entidade, e não pela unidade federativa (conforme o próprio regime jurídico dessas entidades supõe, visto que são regimes jurídicos próprios e descentralizados do Poder Executivo). A normalização segue, então, conforme o suporte da informação do documento. Exemplos:

- a) resolução em jornal (Diário Oficial da União) disponível online, de órgão de administração indireta;

**VANCOUVER** – Agência Nacional de Vigilância Sanitária. Resolução RDC nº 17, de 16 de abril de 2010. Dispõe sobre as boas práticas de fabricação de medicamentos. Diário Oficial da União. 2010 abr 19;(Seção 1, 73):94-110. Disponível em: [http://portal.anvisa.gov.br/documents/33880/2568070/res0017\\_16\\_04\\_2010.pdf/b9a8a293-f04c-45d1-ad4c-19e3e8bee9fa](http://portal.anvisa.gov.br/documents/33880/2568070/res0017_16_04_2010.pdf/b9a8a293-f04c-45d1-ad4c-19e3e8bee9fa).

**APA** – Agência Nacional de Vigilância Sanitária [ANVISA]. (2010). Resolução RDC nº 17, de 16 de abril de 2010: Dispõe sobre as boas práticas de fabricação de medicamentos. *Diário Oficial da União, Seção 1(73)*, 94-110. Recuperado de [http://portal.anvisa.gov.br/documents/33880/2568070/res0017\\_16\\_04\\_2010.pdf/b9a8a293-f04c-45d1-ad4c-19e3e8bee9fa](http://portal.anvisa.gov.br/documents/33880/2568070/res0017_16_04_2010.pdf/b9a8a293-f04c-45d1-ad4c-19e3e8bee9fa)

- b) nesta portaria de autarquia federal disponível no formato eletrônico, a ementa da norma não consta no ato, mas é elaborada pela própria instituição e consta no portal oficial da instituição (considerada fonte confiável neste caso);

**VANCOUVER** – Universidade Federal do Rio Grande do Sul. Portaria nº 1.429, de 28 de abril de 2008. Fixa os valores dos benefícios e taxas relativos aos programas vinculados às pró-reitorias acadêmicas. Disponível em: <http://www.ufrgs.br/prograd/Portaria%201429%20de%20280408.pdf>.

**APA** – Universidade Federal do Rio Grande do Sul [UFRGS]. (2008). *Portaria nº 1.429, de 28 de abril de 2008: Fixa os valores dos benefícios relativos aos programas vinculados às pró-reitorias acadêmicas*. Recuperado de <http://www.ufrgs.br/prograd/Portaria%201429%20de%20280408.pdf>

- c) instrução normativa de autarquia federal;

**VANCOUVER** – Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis. Instrução normativa nº 3, de 31 de janeiro de 2013. Diário Oficial da União. 2013 fev 1;(Seção 1, 73):88-9.

**APA** – Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis [IBAMA]. (2013). Instrução normativa nº 3, de 31 de janeiro de 2013. *Diário Oficial da União, Seção 1(73)*, 88-89.

- d) resolução de fundação pública disponível online no portal oficial da instituição.

**VANCOUVER** – Fundação Instituto Nacional de Geografia e Estatística. Resolução PR nº 1/2005, de 25 de fevereiro de 2005. 2005. Altera a caracterização do Sistema Geodésico Brasileiro. Disponível em: [http://www.inde.gov.br/images/inde/rpr\\_01\\_25fev2005.pdf](http://www.inde.gov.br/images/inde/rpr_01_25fev2005.pdf)

**APA** – Fundação Instituto Nacional de Geografia e Estatística [IBGE]. (2005). *Resolução PR nº 1/2005, de 25 de fevereiro de 2005: Altera a caracterização do Sistema Geodésico brasileiro*. Recuperado de [http://www.inde.gov.br/images/inde/rpr\\_01\\_25fev2005.pdf](http://www.inde.gov.br/images/inde/rpr_01_25fev2005.pdf)

## 5. Considerações finais

O uso de autoria com a jurisdição (unidades federativas), para órgãos de administração direta, e diretamente pela entidade, para instituições de administração indireta, deve conduzir para uma referenciação com padrões mais lógicos quando da normalização de documentos jurídicos brasileiros em estilos internacionais como Vancouver. Para a APA, ao invés de adotar este padrão, sugere-se manter o consolidado no uso: sem a jurisdição nas referências de órgãos de administração direta. Em seguida, considerar o tipo de suporte da informação onde o ato foi consultado (o documento) e prosseguir a normalização da referência conforme o tipo

desse documento seguindo as orientações de originais do estilo, também é uma proposta que leva em conta parâmetros lógicos e de fácil assimilação.

Entretanto, é necessário ressaltar que as propostas feitas no presente trabalho não extinguem todas as lacunas de normalização de documentos jurídicos brasileiros que precisam ser sanadas. Algumas das incongruências advêm dos próprios documentos jurídicos (por exemplo, o uso de “n.” ou “nº” para indicar “número”) e lacunas dos estilos de normalização (como a NBR 6023, que indica que as regras de normalização de legislação abrangem também as normas de entidades públicas ou privadas, mas não dá qualquer exemplo de referência deste tipo). As dificuldades aumentam quando se intenta normalizar referências e citações de uma nacionalidade aplicando-se um estilo de normalização de outra nacionalidade (caso da legislação brasileira referenciada nos formatos APA e Vancouver).

Sugere-se continuidade a este trabalho no quesito de outros detalhes de normalização de referências, com a fundamental observação de se proporem regras com embasamento lógico, a fim de não problematizar ainda mais o entendimento das normas por parte de autores e leitores e de se estabelecer uma maior consistência no ato de normalizar. A simplificação das normas e regras de normalização através de diretrizes lógicas que abrangem diversas situações deverão auxiliar na correta e consistente citação e referência de fontes de pesquisa, facilitando a identificação de fontes, a interoperabilidade de sistemas e melhorando a qualidade da comunicação científica. A ampliação das publicações científicas brasileiras no cenário mundial através da indexação em grandes bases de dados impõe a tarefa de cuidar dos aspectos de conteúdo (mérito) juntamente com a forma de apresentação destas publicações, baseada na adequação às normas internacionais.

## Referências Bibliográficas

---

- ALMEIDA, A., LOPES, E. S. S., CAMILO, J. T. S., & CHOI, V. P. (Orgs.). (2016). *Manual APA: Regras gerais de estilo e formatação de trabalhos acadêmicos*. São Paulo: FECAP.
- AMERICAN PSYCHOLOGICAL ASSOCIATION [APA]. (2012). *Regras essenciais de estilo da APA* (6. ed.). Porto Alegre: Penso.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS [ABNT]. (2002). *NBR 6023: Informação e documentação – referências – elaboração*. Rio de Janeiro: ABNT.
- BRASIL. (2002). *Manual de redação da presidência da república* (2. ed.). Brasília: Presidência da República.
- BREITENBACH, A. (2016). *The origins of APA style*. Apa Style. Recuperado de <http://blog.apastyle.org/apastyle/2016/07/the-origins-of-apa-style.html>
- BRITISH MEDICAL ASSOCIATION [BMA]. [20--]. Reference styles (BMA library factsheet). In British Medical Association, *BMA*. London: BMA.
- CAFÉ, L., & BRÄSCHER, M. (2008). Organização da informação e bibliometria. *Encontros Bibli*, (n. esp.), 54-75.
- HARVARD LAW REVIEW ASSOCIATION (Comp.). (2010). *The bluebook: A uniform system of citation* (19th ed.). Cambridge: The Harvard Law Review Association.
- INTERNATIONAL COMMITTEE OF MEDICAL JOURNAL EDITORS [ICMJE]. (2010). *Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication*. ICMJE.
- INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS [IFLA]. (2009). *Declaração de princípios internacionais de catalogação*. München: K. G. Saur. Disponível em: <[http://www.ifla.org/files/assets/cataloguing/icp/icp\\_2009-pt.pdf](http://www.ifla.org/files/assets/cataloguing/icp/icp_2009-pt.pdf)>. Acesso em: 20 ago. 2017.
- MEADOWS, A. J. (1999). *A comunicação científica*. Brasília: Briquet de Lemos.
- PASSOS, E., & BARROS, L. V. (2009). *Fontes de informação para pesquisa em direito*. Brasília: Briquet de Lemos.
- PATRIAS, K. (2007). *Citing medicine: The NLM style guide for authors, editors, and publishers* (2nd ed.). Bethesda: NLM. Recuperado de <http://www.ncbi.nlm.nih.gov/books/NBK7256/>
- RODRIGUES, M. E.; LIMA, M. H. T. F.; & GARCIA, M. J. O. (1998). A normalização no contexto da comunicação científica. *Perspectivas em Ciência da Informação*, 3(2), 147-156.
- SOLLA PRICE, D. J. (1976). *O desenvolvimento da ciência: Análise histórica, filosófica, sociológica e econômica*. Rio de Janeiro: Livros Técnicos e Científicos.

- SOUZA, S. T. (2013). *A caracterização do documento jurídico para a organização da informação*. (Dissertação de mestrado, Universidade Federal Minas Gerais, Belo Horizonte).
- TARGINO, M. G. (2000). Comunicação científica: Uma revisão de seus elementos básicos. *Informação e Sociedade*, 10(2), 37-85.
- UNIVERSIDADE DE SÃO PAULO [USP]. (2006). *Guia de apresentação de teses* (2. ed.). São Paulo: USP.
- U. S. NATIONAL LIBRARY OF MEDICINE [NLM]. (2011). *International Committee of Medical Journal Editors (ICMJE) uniform requirements for manuscripts submitted to biomedical journals: Sample references*. Bethesda: NLM, 2011. Recuperado de [http://www.nlm.nih.gov/bsd/uniform\\_requirements.html](http://www.nlm.nih.gov/bsd/uniform_requirements.html)
- VANZ, S. A. S., & SANTOS, J. P. (2011). Ética na normatização de documentos. In M. O. E. Oliveira, G. I. S. Ferreira, & R. S. A. Lunardelli (Orgs.), *Ética profissional na prática do bibliotecário* (pp. 114-131). Brasília: Usina de Letras.
- ZIMAN, J. (1979). *Conhecimento público*. São Paulo: Ed. da Universidade de São Paulo.

# O uso da folksonomia na atualização de vocabulários controlados da área da Pediatria<sup>1</sup>

*The use of folksonomy in the updating of controlled vocabulary in the area of Pediatrics*

**Luciana Monteiro Krebs**

Universidade Federal do Rio Grande do Sul, Brasil  
[luciana.monteiro@ufrgs.br](mailto:luciana.monteiro@ufrgs.br)

**Rita do Carmo Ferreira Laipelt**

Universidade Federal do Rio Grande do Sul, Brasil  
[rita.laipelt@ufrgs.br](mailto:rita.laipelt@ufrgs.br)

**Samuel Santos da Rosa**

Universidade Federal do Rio Grande do Sul, Brasil  
[samuel.sdrosa@ufrgs.br](mailto:samuel.sdrosa@ufrgs.br)

## Resumo

Apresenta análise transversal de folksonomias para estudo da linguagem dos usuários de informação na área de Pediatria. Tem como pressuposto de pesquisa que as tags podem ser fonte de coleta de candidatos a termos para atualização de vocabulários controlados. O objeto de estudo são as tags armazenadas e compartilhadas através do ambiente online CiteULike. Usou-se dois conjuntos de dados, cada um formado por um corpus de pesquisa (tags) e um corpus textual (artigos científicos compartilhados por pelo menos dois usuários da ferramenta). As tags foram validadas tanto no corpus textual quanto em um vocabulário controlado da área médica. Foram analisadas 195 tags de 2011, das quais 93% foram validadas, e 282 tags de 2017, com validação de 79%. Conclui-se que as tags são boas fontes para a identificação de candidatos a termos, assim como de variantes para enriquecer vocabulários controlados e alimentar sistemas de remissivas de catálogos de autoridades de assunto.

## Abstract

*The paper presents a transversal analysis of folksonomies to investigate the language of information users in the area of Pediatrics. It has as research hypothesis the idea that the tags can be a source for collecting term candidates to update controlled vocabularies. The object of study is the tags stored and shared through the CiteULike online platform. Two sets of data were used, each consisting of a research corpus (tags) and a textual corpus (scientific papers shared by at least two tool's users). The tags were validated both in the textual corpus and in a controlled vocabulary of the medical area. We analyzed 195 tags from 2011, of which 93% were validated, and 282 tags from 2017, with validation of 79%. It is concluded that tags are good sources for identifying term candidates as well as variants for enriching controlled vocabularies and feeding referral systems from subject authority catalogs.*

**Palavras-chave:** Representação da informação. **Keywords:** *Information representation. Knowledge Organization do conhecimento. Vocabulários organization. Controlled vocabularies. Folksonomy. controlados. Folksonomi.*

## 1. Introdução

Em 2007, 94% das informações produzidas no mundo foram armazenadas em formato digital (Hilbert & López, 2011). Uma previsão da Cisco (2013) indica que em 2017 serão produzidos 7,7 zetabytes de dados. Enquanto fazemos uso de recursos tecnológicos que possibilitam atravessar fronteiras, buscar conhecimento, relacionamentos e experiências que talvez pessoalmente não possamos realizar, tudo isso fica registrado, gerando *logs* de dados em grandes volumes, e muitas vezes acessível a públicos que o próprio autor desconhece. Em várias plataformas é possível compartilhar deliberadamente o registro destas atividades *online* e, na maioria dos casos, é o que se deseja, o que revela uma mudança social relevante em termos de privacidade. Além disto, a informação é registrada em escalas tão grandes, que encontrar uma informação que se busca pode se tornar uma tarefa demasiadamente complicada.

O aumento exponencial de informação disponibilizada através da internet sinaliza para uma expansão significativa de competências, cujos profissionais da informação devem se apropriar se quiserem continuar exercendo um papel importante para os usuários. *Blogs*, fóruns de discussão e até redes sociais podem ser fontes de informação para novos usuários, e estes recursos não só cresceram em quantidade, mas revelam-se complexos em termos tecnológicos, que também avançam todos os dias. Os processos de representação e recuperação da informação tornam-se cada vez mais sofisticados, com o incremento de inteligência artificial nos motores de busca, por exemplo. Com o advento de tecnologias para compartilhamento de informações e meta-informações através da web, podemos estar diante de uma significativa oportunidade de melhoria no processo de indexação, contando com a contribuição de metadados fornecidos pelos próprios usuários em linguagem natural.

A presente pesquisa visa refletir sobre a possível contribuição da folksonomia nos processos formais de indexação de documentos (especificamente na representação temática) enquanto fontes de coleta de candidatos a termos, objetivando aproximar a linguagem utilizada nestes

processos daquela conhecida pelo usuário. Almeja-se com isso, alcançar maior eficácia e assertividade no momento da busca, poupando o tempo do leitor.

A utilização do CiteULike como plataforma para coleta de dados justifica-se em função de ser um ambiente exclusivo para compartilhamento de literatura científica. Ele é utilizado, portanto, por um público altamente especializado. Diferente da plataforma social ResearchGate, o CiteULike permite a atribuição de *tags* aos conteúdos compartilhados (a rede social acadêmica previamente mencionada permite atribuição de *tags* apenas como “temas de interesse” do pesquisador). Já o Academia.edu permite a associação de “*bookmarks*” a um artigo, porém apenas quem fez o *upload* – provavelmente seu autor – pode atribuir as *tags/bookmarks* ao documento, não sendo o mesmo possível para os demais usuários.

## 2. Folksonomias e a indexação: possíveis aproximações

O desafio de quem procura conhecimento atualmente, não é mais o acesso e permissão para ler determinado documento como ocorria na era medieval. Hoje, o maior desafio de leitores e pesquisadores está em sua capacidade de encontrar, entre tantas publicações, aquela que realmente tem relevância para suprir uma necessidade informacional. Tornou-se mister então, encontrar, de fato, a informação precisa no tempo adequado. Agilidade e precisão na recuperação da informação têm ainda mais valor no cenário atual com os recursos tecnológicos disponíveis e a descentralização de entidades produtoras de informação propiciadas pela internet.

Nesse contexto, a realidade dos usuários que buscam por informação também se modificou significativamente, com a possibilidade de acessarem o conteúdo de catálogos e bases de dados de qualquer unidade de informação à distância. Conseqüentemente, o bibliotecário passa a ter de lidar com outros recursos para tentar dar conta de “conhecer” este usuário que raramente vê pessoalmente, considerando alternativas de aproximação com esses usuários que se encontram cada vez mais distantes fisicamente.

Essa realidade exige um esforço redobrado por parte dos profissionais, para manter a unidade de informação, e o seu próprio trabalho, relevantes. É vital aprender como a internet funciona, como as pessoas se comunicam através dela, e o que se pode aprender dali. O conceito em si não é novo. A importância de aproximar o máximo possível a linguagem de indexação do



repertório da comunidade usuária já foi sinalizada há muito tempo. A literatura da Ciência da Informação, frequentemente, preconiza a aproximação da linguagem utilizada na indexação com a linguagem natural do usuário, para que este obtenha, cada vez mais, sucesso nos seus objetivos de busca.

Em 1876, Cutter, em sua obra *“Rules for a dictionary catalog”* determina regras para a formação de cabeçalhos alfabéticos de assuntos<sup>2</sup>. Fujita, Rubi & Boccato (2009, p. 26) resumem as regras em três princípios básicos, entre os quais se chama atenção para o primeiro: “Princípio do uso: as descrições devem ser feitas da forma usada pelo usuário”. Gomes, Motta & Campos (2006, *online*) encontram na quarta lei de Ranganathan (Poupe o tempo do leitor) uma premissa para o tratamento dos termos no processo de indexação:

[...] dois dispositivos contribuem para o atendimento a esta Quarta Lei:

- 1) adoção do termo mais plausível de ser buscado pelo usuário, quanto ao uso corrente e quanto à especificidade;
- 2) adoção de um dispositivo no sistema de recuperação de informação de sorte que, mesmo buscando por um termo não-preferido, o sistema automaticamente aceite o termo e recupere a informação via termo preferido. Assim, com um único passo, o leitor acessaria a base de dados. (Gomes, Motta & Campos, 2006, *online*).

O termo não-preferido pode ser entendido aqui como as variações terminológicas, variações do termo escolhido, ou seja, variantes podem tornar-se pontos de acesso no catálogo, sem destituir o termo escolhido como descritor. Assim, o bibliotecário deve aproximar-se do usuário tanto quanto possível, e dentro do ambiente específico da internet, isto significa compreender o uso das ferramentas utilizadas para se comunicar e também para representar a informação, como as *tags*.

Essa aproximação é importante, sobretudo, para impedir que a informação fique “oculta” dentro dos acervos – devido a processos de indexação pouco eficazes. Por isso, é recomendável que sejam realizados estudos de comunidades e usuários para compreender sua linguagem e incluí-la nos sistemas de recuperação da informação através de remissivas. Sendo os sistemas eletrônicos de informação binários, a simples ausência de uma remissiva dentro do conjunto de termos presentes no sistema pode apresentar um resultado vazio para a busca, mesmo que o documento desejado esteja no acervo. Espera-se do bibliotecário que realize esse filtro sofisticado, para agregar cada vez mais formas de tornar a busca do usuário eficaz.

A folksonomia pode ser entendida como uma representação simples de conteúdos na internet (páginas, *links*, fotografias, textos, músicas, vídeos, documentos em geral), que reforça laços sociais e expressa linguagem natural dos usuários desta informação. O termo foi cunhado por Thomas Vander Wal, arquiteto da informação, que define a etiquetagem atribuída a conteúdos na *web* pelos usuários, com o objetivo de recuperar este conteúdo em um dado momento. Segundo Vander Wal (2007), folksonomia é o resultado da marcação livre e pessoal de informações e objetos (qualquer coisa com uma URL) para uma recuperação própria [tradução nossa]. Neste caso, a marcação a que o autor se refere são as etiquetas (ou *tags*), atribuídas pela pessoa que consome a informação. A marcação é feita em um ambiente social (compartilhado e aberto aos outros).

Normalmente, nos sistemas *online* de compartilhamento de documentos (artigos, textos, fotos, vídeos, etc.) é possível, além de atribuir as *tags* para representar a informação, recuperar documentos através da *tag* (quando esta é transformada em *link*). Este *link*, ao ser clicado, retorna uma lista de documentos aos quais a *tag* foi atribuída, permitindo que o usuário visualize e continue sua navegação através desta forma de busca. Para Amaral e Aquino (2008), no momento em que a prática das *tags* permite a qualquer usuário representar e recuperar informações através de etiquetas criadas livremente e com base nos significados dos dados etiquetados, ela surge como uma alternativa de gerenciamento de informação.

Dentro do processamento técnico de uma unidade de informação, o estabelecimento dos descritores dos itens da coleção é uma das atividades mais importantes, pois é o que garante a recuperação - ou não - dos documentos pelos usuários. Esta etapa deve ser feita de tal forma que garanta precisão evitando equívocos e inconsistências no resultado apresentado ao leitor, e que fica disponível no catálogo. O quadro 1 tem a função de delimitar as diferenças entre folksonomias e vocabulários controlados, pois ambos são utilizados nesta pesquisa.

**Quadro 1 – Tabela comparativa de folksonomias e vocabulários controlados**

	Folksonomias	Vocabulários controlados
<b>Objetivo</b>	Representar e recuperar a informação.	Representar e recuperar a informação.
<b>Plataforma de registo</b>	Exclusivamente na web.	Sistemas de organização do conhecimento (catálogos, bibliotecas), sejam online ou off-line.
<b>Autoria da classificação</b>	Usuários da informação e autores.	Bibliotecários e profissionais especializados.
<b>Linguagem</b>	Natural.	Artificial.
<b>Momento em que o material é classificado</b>	Durante ou após o uso da informação.	Antes de disponibilizar a informação para o uso.

Fonte: os autores

Como se pode observar, há diferenças entre as folksonomias e os vocabulários controlados, no que tange à plataforma de registo, autoria da classificação, e outros, porém, ambas são utilizadas com o objetivo de representar e recuperar a informação.

Levanta-se, portanto, a questão sobre se a informação disponível em forma de *tags* pode ser aproveitada em estudos e eventualmente utilizada. Especialmente em relação às expressões empregadas pelos usuários na representação do conhecimento - e que fazem parte do repertório pessoal individual destes - como contributo para enriquecer sistemas de catalogação de documentos enquanto recursos para aproximar o usuário da informação que ele necessita ou deseja.

### 3. Metodologia

Essa seção apresenta as etapas metodológicas realizadas na pesquisa. Os resultados apresentados são fruto de um estudo comparativo entre conjuntos de dados, sendo um de 2011 e outro de 2017. Cada um dos conjuntos foi formado por um *corpus* de pesquisa (*tags* atribuídas pelos usuários) e um *corpus* textual (texto completo dos artigos para validação). Foi utilizado também, para validação, um vocabulário controlado da área médica. Optou-se por utilizar uma ferramenta de armazenamento e compartilhamento *online* de documentos com o recurso de *social tagging*, o CiteULike<sup>3</sup>, para avaliar os resultados de atribuição de etiquetas por usuários nos conteúdos, de forma que se pudesse avaliar a relação entre estas etiquetas e termos de vocabulários controlados, além de identificar sua presença ou não dentro do *corpus* textual.

### 3.1. Planejamento e critérios de busca

A composição dos conjuntos de dados ocorreu de acordo com as seguintes diretrizes: a) foram admitidos artigos científicos a que se obteve acesso ao texto integral; b) artigos categorizados pela *tag* “*pediatrics*” por qualquer usuário da ferramenta CiteULike e em qualquer data; c) artigos compartilhados por pelo menos dois usuários do ambiente.

Nesta pesquisa optou-se por utilizar a busca simples, que inclui todos os artigos públicos e autenticados no CiteULike. Sabe-se que a busca simples no CiteULike não leva em consideração apenas as *tags* atribuídas pelos usuários. Se fosse assim, teríamos na lista de *tags* tantas ocorrências da *tag* “*pediatrics*” quantos documentos recuperados no *corpus*, o que não ocorre.

Também foram analisadas as *tags* atribuídas a cada documento, e ocorre que vários documentos não receberam a *tag* “*pediatrics*” de nenhum usuário. Assim, podemos inferir que a ferramenta de busca simples do CiteULike leva em consideração também outras informações que não somente *tags*, provavelmente título do documento, nome do periódico em que o artigo foi publicado, informações do texto ou resumo etc. Atualmente é possível fazer uma busca avançada na ferramenta apenas pelas *tags* atribuídas aos documentos (usando a expressão *tag:pediatrics*), no entanto, como o objetivo desse trabalho é realizar uma análise transversal, comparando os dados de 2011 e 2017, optou-se por manter o uso da busca simples.

A primeira coleta ocorreu em 27 de agosto de 2011. A segunda em 8 de agosto de 2017. Na página inicial do CiteULike, digitou-se a palavra-chave “*pediatrics*” no campo “*Search citeulike*” no canto direito superior da tela. Os resultados são listados aparentemente sem ordenamento - nem cronológico, nem alfabético, nem por relevância. A seguir estão detalhadas as etapas de coleta, limpeza e organização de cada um dos *corpora*.

### 3.2. Coleta, limpeza e organização do *corpus* textual

Em relação ao *corpus* textual, foram recuperados 903 documentos em 2011 e 850 em 2017, dos quais foram selecionados apenas os compartilhados com dois ou mais membros, o que ocorreu em 177 artigos em 2011, e 190 em 2017. Copiaram-se as seguintes informações de

cada artigo (metadados fornecidos pelo CiteULike): título, referência bibliográfica, autor, *tags* atribuídas ao documento, número de pessoas que compartilham o documento.

Foi realizado o *download* de todos os artigos encontrados no resultado de busca que atendiam aos critérios estabelecidos, com exceção dos que, por alguma restrição de acesso, não se pôde realizar o *download*. Ao pesquisar nas bases de dados disponíveis no Portal de Periódicos da Capes<sup>4</sup> e no Google Acadêmico<sup>5</sup>, tivemos acesso a **156** dos 177 artigos inicialmente selecionados em 2011 e **185**<sup>6</sup> dos 190 inicialmente selecionados em 2017. Todos os documentos se encontram em língua inglesa. Cada documento recebeu um número.

Os arquivos foram então convertidos do formato .PDF para .TXT para que pudessem ser mais facilmente analisados e lidos pelas ferramentas de processamento de *corpus*. Foram retiradas dos arquivos de texto todas as partes dos artigos não necessárias para a posterior análise terminológica, como as tabelas, referências bibliográficas, ilustrações, datas, notas de rodapé, currículo dos autores e paginações, para que permanecesse no arquivo apenas o artigo em si.

Na próxima etapa realizou-se o *upload* do *corpus* textual na ferramenta e-Termos<sup>7</sup>. O e-Termos é um ambiente colaborativo *online* de acesso gratuito cujo objetivo é auxiliar na gestão terminológica. Foi cadastrado um projeto com o nome Pediatria, feito *upload* dos textos (Gênero “Científico” e Tipo Textual “Artigo”) e em seguida foi feita a compilação dos *corpora*.

### **3.3. Coleta, limpeza e organização do *corpus* de pesquisa**

Para coletar o *corpus* de pesquisa (*tags*), clicou-se no registro de cada documento, copiou-se o trecho “*posting history*” e colaram-se todas estas informações em um arquivo de texto. Em *posting history* são exibidas as informações do histórico de compartilhamento, ou seja, data, *nickname* do usuário e *tags* atribuídas por cada um deles no momento do compartilhamento. Os conjuntos de *tags* por artigo receberam a numeração definida para o artigo correspondente.

A atribuição de *tags* no CiteULike ocorre quando o usuário compartilha o documento. Um campo de texto livre é usado para este fim, e as *tags* são separadas por espaço simples, ou seja, a única forma de usar expressões como *tag* é introduzindo um sinal gráfico como underline ( \_ ) ou hífen ( - ) ou ainda escrevendo os dois termos sem separação de espaço,

porque juntos representam um sintagma. São exemplos: *intervention\_services*, *alcohol-abuse* e *familyphysician* (para o conceito “médico de família”). Enquanto alguns usuários utilizam o *underline*, outros preferem escrever as expressões sem o espaço, e outros ainda ignoram esta limitação da ferramenta e mantêm os espaços, o que separa os termos que compõem uma expressão na hora de analisar as *tags*. Assim, alguns termos dentro das *tags* levantadas ficam sem sentido devido à quebra do sintagma terminológico, e por este motivo foram retirados do *corpus* de pesquisa.

Em relação ao *corpus* de pesquisa de 2011, de um total de 979 *tags* atribuídas aos documentos, foram retiradas as *tags* repetidas e restaram 493 *tags* únicas. Já no *corpus* de 2017, de um total de 1.494 *tags*, 651 eram *tags* únicas. A limpeza das *tags* serviu para eliminar expressões com erros de grafia, *tags* que não possuem sentido claro quando empregadas ao assunto Pediatria, ou aquelas que, apesar de estarem no discurso dos especialistas, são muito genéricas e transitam em várias áreas, ou seja, são pouco específicas deste campo. Durante este processo foram retiradas algumas *tags* que possuíam erros de grafia, como “*childrens*” (a grafia correta é *children*) e “*breast\_feeding*” (a redação correta é *breastfeeding*).

Também se identificou que algumas *tags* encontravam-se em outros idiomas, como turco, alemão e português e que tinham relação com a temática dos artigos pesquisados, mas foram retiradas em função das ferramentas de validação (*corpus* textual e vocabulário controlado) estarem no idioma inglês, o que traria resultado inválido. Algumas destas ocorrências estão registradas a seguir: a) *cocuk* – em turco, criança; b) *siber\_zorbalik* – (*siber zorbalik*) em turco, cyber-bullying; c) *epilepsia* – em português; d) *auswirkungen* – em alemão, efeito.

Finalmente, mesmo percebendo que algumas *tags* poderiam ser variantes, como “*neonatal\_thrombocytopenia*” e “*low platelet count*”, optou-se por mantê-las no *corpus* de pesquisa para verificar sua ocorrência nas ferramentas de validação. As siglas encontradas no *corpus* de pesquisa de 2017 foram tratadas à parte, uma vez que as mesmas não foram coletadas e analisadas no *corpus* de 2011. Esta medida foi tomada para não distorcer os resultados comparativos.

Após a limpeza do *corpus* de pesquisa, resultaram **195** *tags* a serem analisadas no *corpus* de 2011 e **282** no *corpus* de 2017. Foi então realizada a normalização das *tags*, corrigindo o uso de hífen e *underline* onde deveria haver espaços (no caso de sintagmas), ou mesmo a

supressão do espaço, devido à limitação do CiteULike descrita anteriormente. São exemplos desta normalização: “*alcohol-abuse*” foi substituído por “*alcohol abuse*”, “*speech\_and\_language\_delay*” foi substituído por “*speech and language delay*” e “*earlyintervention*” foi substituído por “*early intervention*”.

### 3.4. Validação das tags

Nesta etapa realizou-se a validação de ocorrência das *tags* no *corpus* textual através do e-Termos. Verificou-se que, de acordo com a tipologia do termo analisado, é necessário utilizar uma ferramenta diferente em função da maneira como a frequência é calculada em cada uma delas. Por isso, as ferramentas utilizadas foram: “Contador de Frequência” para termos simples, “Consulta Termos” para sintagmas (termos compostos) e “Identificador de Siglas, Acrônimos e Nomes Próprios” para siglas.

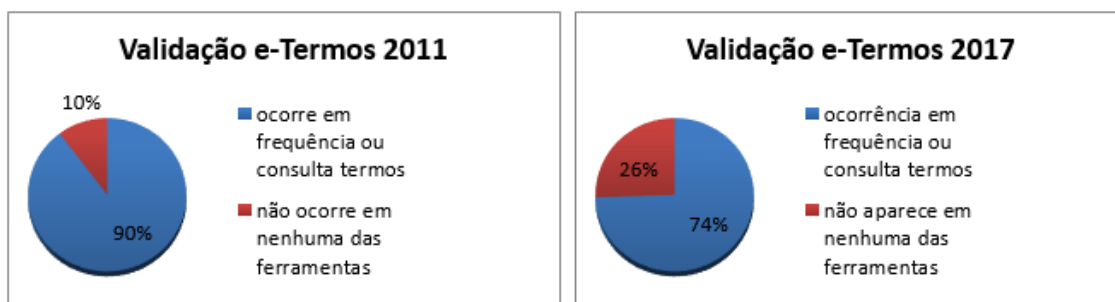
A próxima etapa consistiu na validação das *tags* em vocabulário controlado da área de Medicina. Para isso, utilizou-se a ferramenta Descritores em Ciências da Saúde (DeCS)<sup>8</sup>. Para o resultado do DeCS foram adotadas as seguintes categorias: (a) encontrado: quando a *tag* foi encontrada como descritor com exatamente a mesma grafia; (b) parte de descritor: quando a *tag* foi encontrada como parte de um descritor; (c) variante: quando a *tag* foi encontrada, mas não como descritor e sim variante ou parte de variante no DeCS; e (d) não encontrado: quando a *tag* não pôde ser validada, pois não foi encontrada no DeCS.

A partir do resultado da validação do *corpus* de pesquisa foi possível calcular quanto do *corpus* de pesquisa efetivamente apareceu no *corpus* textual e/ou no vocabulário controlado. A título de ilustração, o **Apêndice A** traz as dez *tags* mais compartilhadas nos *corpora* de 2011 e 2017. A próxima seção refere-se à análise dos dados.

## 4. Análise e resultados da pesquisa

No *corpus* de pesquisa de 2011, das 195 *tags* selecionadas para análise, 175 foram validadas no *corpus* textual (e-Termos), aparecendo pelo menos uma vez. Já no *corpus* de pesquisa de 2017, das 282 *tags* selecionadas para análise, 210 foram validadas no *corpus* textual (e-Termos). Na Figura 1 pode-se visualizar os percentuais de ocorrência em cada um dos períodos.

**Figura 1 – Validação comparada das tags no corpus textual**



Fonte: dados da pesquisa

Segundo os dados da pesquisa apresentados na Figura 1, 90% das *tags* de 2011 foram validadas no *corpus* textual, e em 2017 a validação foi de 74%. Na análise transversal, pode-se observar a diferença de 10% para 26% de *tags* que não ocorrem no e-Termos. Isso significa que entre 2011 e 2017 a distância entre a linguagem do usuário (representada nas *tags*) e a linguagem dos especialistas constante nos textos aumentou.

Como a ferramenta onde os documentos são marcados pela inserção de *tags* é de interesse de especialistas, pode-se inferir que esse afastamento se deve a escolhas linguísticas/terminológicas dos usuários. Sabe-se que a linguagem utilizada para a escrita de artigos científicos é mais formal por exigência do próprio gênero textual. A linguagem utilizada para a marcação do conteúdo de documentos em plataformas sociais de compartilhamento, por outro lado, é livre e informal, apresentando em alguns casos *tags* de interesse particular dos usuários que não representam o conteúdo do documento marcado (como as *tags* “*my\_phd*” ou “*literature\_search*”). Mesmo que *tags* com essa particularidade tenham sido retiradas dos *corpora* de pesquisa durante a limpeza, percebe-se que o usuário faz uso do seu repertório linguístico individual muitas vezes, pois essa é a característica proeminente da folksonomia. Outra possibilidade para justificar esse afastamento é que as *tags* não encontradas sejam neologismos, e nesse caso é natural que não ocorram no *corpus* textual, por serem termos que ainda estão começando a fazer parte da linguagem especializada no campo, sem terem se consolidado em glossários e vocabulários controlados da área.

Em vocabulários controlados, a ausência de variantes causa dificuldades de recuperação de informação pelos usuários. Quando um documento está indexado com um termo, mas o usuário realiza a busca utilizando outro, se esta variante não estiver registrada a recuperação é prejudicada. Faulstich (2002, p. 70) afirma que os termos, no meio linguístico e social, “são



entidades passíveis de variação e mudança”. Entende-se que os profissionais da informação devem estar atentos a estas variações e que seus produtos estejam alinhados com as expressões de busca.

Por isso, além de validar o *corpus* de pesquisa no *corpus* textual, neste trabalho realizou-se a validação das *tags* também no vocabulário controlado. Nesta etapa, reuniram-se as *tags* que apareceram de alguma forma no DeCS, seja como um descritor exatamente com a mesma grafia, seja encontrado como variante ou parte de variante de um descritor ou então como parte de um descritor composto por mais de uma palavra. A Figura 2 representa os resultados obtidos em 2011 e 2017.

**Figura 2 – Validação comparada das tags no vocabulário controlado**

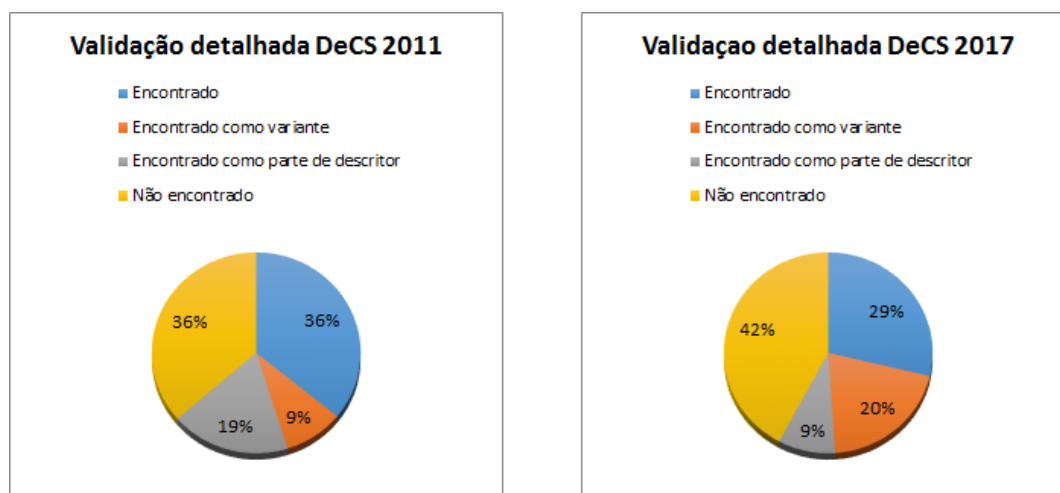


Fonte: dados da pesquisa

Na análise transversal pode-se observar a diferença de 36% para 42% de *tags* que não ocorrem no DeCS. Isso significa que entre 2011 e 2017 a distância entre a linguagem do usuário (representada nas *tags*) e a linguagem controlada (representada pelos descritores e variantes do DeCS) também aumentou, assim como ocorreu com a validação das *tags* no *corpus* textual. Como os descritores do DeCS são utilizados por bibliotecários para a indexação de obras da área da saúde, pode-se dizer então que observamos um maior afastamento entre a linguagem dos usuários e a linguagem usada para indexação pelos bibliotecários da área da saúde.

Por outro lado, embora tenhamos identificado esse distanciamento, percebe-se, também, que a presença de variantes terminológicas, encontradas no DeCS em 2017, aumentou em relação às variantes encontradas em 2011, indo de 9% para 20%. Esse resultado é positivo, pois pode indicar que os gestores do DeCs estão cientes da importância das variantes para a recuperação da informação. A Figura 3 apresenta o detalhamento da validação realizada, na qual é possível identificar como as *tags* foram encontradas (se como descritores ou variantes).

**Figura 3 – Validação comparada das tags no vocabulário controlado**

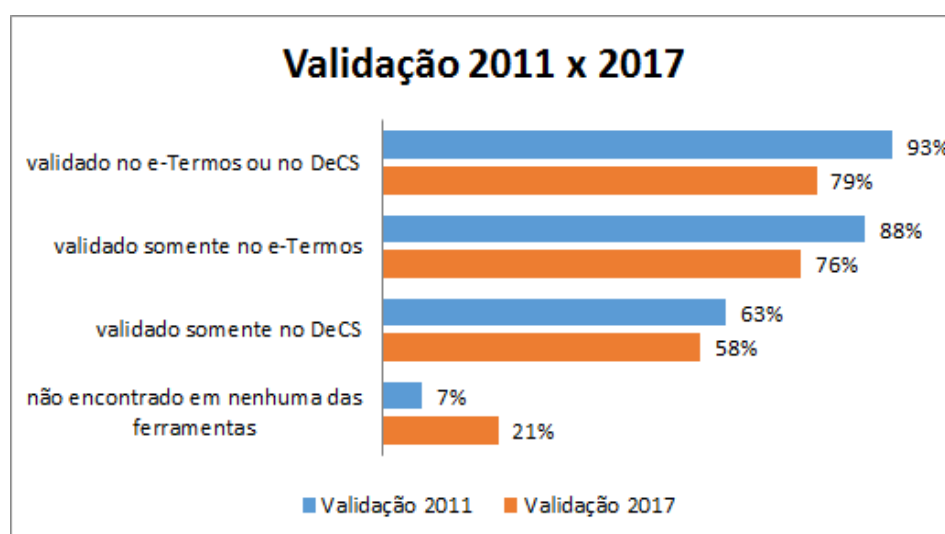


Fonte: dados da pesquisa

Outro aspecto importante é que nenhuma das variáveis (*corpus* textual ou DeCS) **isoladamente** é responsável pelo distanciamento geral da linguagem do usuário em relação à linguagem expressa nos textos (Figura 1) e à linguagem do vocabulário controlado (Figura 2). O percentual de *tags* não encontradas poderia ter sido o mesmo se apenas uma das linguagens (do especialista ou do bibliotecário) estivesse muito distante das *tags* atribuídas, mas isso não ocorre. Tanto a validação no e-Termos (que apresentou 88% em 2011 e 76% em 2017) quanto a validação no DeCS (que apresentou 63% em 2011 e 58% em 2017) contribuem para o aumento do percentual de *tags* não validadas em 2017.

Pode-se verificar, na Figura 4, que a diferença de validação entre 2011 e 2017 foi de 12% no *corpus* textual (e-Termos) e 5% no DeCS (vocabulário controlado) respectivamente.

**Figura 4 – Validação 2011 x 2017**



Fonte: dados da pesquisa

Para entender as possíveis causas desse resultado, analisaram-se as *tags* do *corpus* de pesquisa que não foram validadas em nenhuma das ferramentas. Estas *tags*, que não aparecem nem no discurso dos especialistas (validação no *corpus* textual através do e-Termos) tampouco no vocabulário controlado (validação através do DeCS), somam 14 no *corpus* de 2011 e 59 no *corpus* de 2017. Na Tabela 1 encontram-se as 10 *tags* não validadas com maior frequência no CiteULike.

**Tabela 1 – Tabela de tags sem ocorrência no corpus textual ou no DeCS**

2011		2017	
Tag normalizada	Freq. CiteULike	Tag normalizada	Freq. CiteULike
neurological impairments	3	autism asd	24
Pulm	3	attention deficit hyperactivity disorder adhd	18
abuse teen	2	neurological impairments	5
blood brain barrier	2	nicu nbn	4
constraint induced movement therapy	2	otitis media with effusion ome	4
genetic screening newborn	2	measles mumps rubella mmr vaccine	3
Hypoalbuminaemia	2	abuse teen	2
neurology physical therapy	2	development disorder pervasive	2
autism asd	1	developmental coordination disorder dcd	2
infeccion urinaria	1	gluten free casein free diet	2
Lactância	1	joint hypermobility syndrome jhs	2

Fonte: dados da pesquisa.

Nessa análise observamos que, apesar de não validadas no e-Termos nem no DeCS, essas *tags* não estão desconectadas do contexto da Pediatria, o que indica a possibilidade de serem novos termos (novos demais para já terem sido registrados em vocabulários controlados). A comparação constante de *tags* em relação a descritores e variantes confere dinamismo ao sistema de recuperação da informação e alinha buscas e resultados.

No *corpus* de 2017, percebeu-se também outra característica. Aparentemente, o motivo pelo qual a maior parte das *tags* não foi validada foi o uso concomitante de termo e sigla, como nas *tags* “*developmental coordination disorder dcd*”, “*joint hypermobility syndrome jhs*” e “*attention deficit hyperactivity disorder adhd*”. Ao analisar estas ocorrências pode-se afirmar

que, apesar de serem possíveis bons representantes do conteúdo dos artigos, esta forma de representação dos usuários impede sua validação, pois se as siglas tivessem sido registradas separadamente do termo composto, tanto o termo quanto a sigla teriam sido validados.

Conforme já dito anteriormente, as siglas não foram analisadas no *corpus* de 2011 e por isso não foram incluídas na análise transversal aqui apresentada (juntamente com o *corpus* de 2017). No entanto, na coleta de 2017 identificamos um uso significativo de siglas para representação dos conteúdos dos artigos científicos compartilhados no CiteULike (46 *tags*), sendo exemplos “*adhd*” (frequência 35 CiteULike), “*cpoe*” (12 ocorrências) e “*ome*” (7 ocorrências).

Verificamos que estas siglas se referem a termos da área pediátrica e, analisando-as separadamente, validamos 28 delas no *corpus* textual e 14 no vocabulário controlado (12 como variante ou parte de variante; 1 como parte de descritor e 1 encontrado como descritor). Desta análise pode-se concluir que as siglas são muito usadas pelos usuários e especialistas e por isso, é importante que sejam inseridas pelos bibliotecários como remissivas nos vocabulários controlados.

Adicionalmente, observou-se que alguns conceitos foram representados pelos usuários com variação de numeral, como, por exemplo, “*newborn / newborns*” “*immunization / immunizations*”. Em alguns casos essa variação foi validada, em outros não. Esta é uma característica da folksonomia, com a atribuição livre de etiquetas, que difere da linguagem normalizada dos vocabulários controlados.

## 5. Conclusão

A partir das análises apresentadas no decorrer desse artigo, conclui-se que as *tags* atribuídas pelos usuários do CiteULike a documentos compartilhados no ambiente podem contribuir, enquanto fontes de coleta de candidatos a termos, para o enriquecimento da representação temática de documentos, seja através das remissivas em vocabulários controlados ou em catálogos de autoridades.

A análise transversal dos *corpora* (2011 e 2017) indica que houve distanciamento ao longo do tempo entre a linguagem dos usuários, especialistas e bibliotecários, pois houve um decréscimo tanto na quantidade de *tags* validadas no DeCS quanto no *corpus* textual,

comparando-se 2011 e 2017. Apesar disto, observou-se um crescimento no percentual de variantes encontradas no DeCS de 2011 para 2017, o que indica um esforço de aproximação com a linguagem dos usuários por parte dos bibliotecários responsáveis pelo vocabulário controlado.

Conclui-se, ainda, que as *tags* que não apareceram no *corpus* textual ou no vocabulário controlado podem ser neologismos, e, portanto, devem ser observadas ao longo do tempo na literatura especializada devido à possibilidade de virem a se tornar termos.

Considerou-se alto o número de siglas encontradas entre as *tags* em 2017, e por este motivo, sugere-se que as mesmas sejam incluídas nos vocabulários controlados como remissivas, para melhorar a recuperação da informação. Para pesquisas futuras, sugere-se que novos trabalhos sejam realizados para análise da linguagem dos usuários, incluindo ambientes de compartilhamento de literatura científica como o próprio CiteULike e o Academia.edu.

Observa-se, por fim, que uma limitação do estudo deve-se a uma característica da ferramenta CiteULike, que separa as *tags* pelo espaço e, portanto, não admite termos compostos como etiquetas. Por este motivo, muitas *tags* precisaram ser retiradas dos *corpora* de pesquisa, pois perdem o sentido quando separadas durante a contagem de frequência.

## Referências Bibliográficas

---

- AMARAL, A., & AQUINO, M. C. (2008). Práticas de folksonomia e social tagging no Last.fm. In: Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais, 8, 2008. *Anais...* Paraná: PUC. Recuperado de <http://www.din.uem.br/gsii/downloads/waihews/Praticas-Folksonomia-Social-TaggingLastfm.pdf>.
- CISCO (2013). *Índice mundial sobre entornos de nube de Cisco: previsión y metodología, 2012–2017*. San José: Cisco. Recuperado de <http://docplayer.es/1754500-Indice-mundial-sobre-entornos-de-nube-de-cisco-prevision-y-metodologia-2012-2017.html>
- FAULSTICH, E. [1998] Termo e variação: tendências no português do Brasil. In: *Socioterminologia*. (Excerto, parte II). Brasília: UnB.
- FUJITA, M. S. L., RUBI, M. P. & BOCCATO, V. R. C. (2009) As diferentes perspectivas teóricas e metodológicas sobre indexação e catalogação de assuntos. In: Fujita, M. S. L., Boccato, V. R. C., Rubi, M. P. & Gonçalves, M. C. *A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias*. Um estudo de observação do contexto sociocognitivo com protocolos verbais. São Paulo: SciELO - Editora Unesp. Recuperado de <https://books.google.com.br/books?id=KAyV8MVA88YC&>
- GOMES, H. E., MOTTA, D. F. & CAMPOS, M. L. A. (2006). Princípios normativos. In: \_\_\_\_\_. *Revisitando Ranganathan: a classificação na rede*. Rio de Janeiro. Recuperado de <http://www.conexao.org/bitstream/handle/123456789/123456789/revisitando/revisitando.htm>
- HILBERT, M., & LÓPEZ, P. (2011). The world's technological capacity to store, communicate, and compute information. *science*, 332(6025), 60-65. Recuperado de <http://science.sciencemag.org/content/332/6025/60>
- VANDER WAL, T. (2007) *Folksonomy definition and Wikipedia*. Recuperado de <http://www.vanderwal.net/random/entrysel.php?blog=1750>
- VAN DER LAAN, R. H., FERREIRA, G. I. S., BONOTTO, M. E. K. K.; NEVES, I. C. B., & GASPERIN, I. M. (2004). *Avaliação de descritores relativos às ciências da informação: relato de pesquisa*. Em *Questão*, Porto Alegre, 10(2), 337-347.

## 6. APÊNDICE A - Tags mais compartilhadas

**Tabela 1 – Dez tags mais compartilhadas no corpus de pesquisa de 2011 (CiteULike)**

<b>Termo</b>	<b>Freq. CiteULike</b>	<b>Freq. e-Termos</b>	<b>Ferramenta utilizada</b>	<b>Resultado DeCS</b>
<b>children</b>	16	4391	<i>frequência</i>	encontrado como parte de descritor
<b>asthma</b>	12	558	<i>frequência</i>	encontrado
<b>neonatal</b>	10	148	<i>frequência</i>	encontrado como parte de descritor
<b>adolescent</b>	8	199	<i>frequência</i>	encontrado
<b>analgesia</b>	8	48	<i>frequência</i>	encontrado
<b>prevention</b>	8	199	<i>frequência</i>	encontrado como parte de descritor
<b>childhood</b>	6	296	<i>frequência</i>	encontrado como parte de descritor
<b>cognitive</b>	6	231	<i>frequência</i>	encontrado como parte de descritor
<b>visual motor</b>	6	12	<i>consulta termos</i>	não encontrado
<b>child</b>	5	1269	<i>frequência</i>	encontrado

Fonte: dados da pesquisa.

**Tabela 2 – Dez tags mais compartilhadas no corpus de pesquisa de 2017 (CiteULike)**

<b>Termo</b>	<b>Freq. CiteULike</b>	<b>Freq. E-Termos</b>	<b>Ferramenta utilizada</b>	<b>Resultado DeCS</b>
<b>children</b>	24	5436	<i>frequência</i>	encontrado como variante
<b>autism asd</b>	24	0	<i>consulta termo</i>	não encontrado
<b>autism</b>	22	389	<i>frequência</i>	encontrado como variante
<b>attention deficit hyperactivity disorder adhd</b>	18	0	<i>consulta termo</i>	não encontrado
<b>early intervention</b>	12	54	<i>consulta termo</i>	encontrado como variante
<b>child</b>	10	1783	<i>frequência</i>	encontrado
<b>screening</b>	9	981	<i>frequência</i>	encontrado como variante
<b>epidemiology</b>	8	28	<i>frequência</i>	encontrado
<b>abuse</b>	8	329	<i>frequência</i>	encontrado como parte de descritor
<b>medical</b>	8	691	<i>frequência</i>	encontrado como parte de descritor

Fonte: dados da pesquisa

## Notas

---

<sup>1</sup> O presente trabalho foi realizado com o apoio da Pró-Reitoria de Pesquisa - UFRGS – Brasil.

<sup>2</sup> CUTTER, Charles A. Rules for a dictionary catalogue. Washington: Government Printing Office, 1889.

Disponível em < <http://babel.hathitrust.org/cgi/pt?id=wu.89101448975> >. Acesso em 25 jul. 2013.

<sup>3</sup> <http://www.citeulike.com/>

<sup>4</sup> <http://www.periodicos.capes.gov.br/>

<sup>5</sup> <http://scholar.google.com.br/>

<sup>6</sup> A título de informação, destes 185 artigos, 71 já constavam no corpus textual de 2011.

<sup>7</sup> <http://www.etermos.cnptia.embrapa.br/>

<sup>8</sup> <http://decs.bvs.br/>



# Usos do Arquivamento da Web na Comunicação Científica

*Uses of Web Archiving in Scientific Communication*

**Lisiane Braga Ferreira**

Universidade Federal do Rio Grande do Sul  
[lisianebf@gmail.com](mailto:lisianebf@gmail.com)

**Marina Rodrigues Martins**

Universidade Federal do Rio Grande do Sul  
[mrodriguesmartins@gmail.com](mailto:mrodriguesmartins@gmail.com)

**Moisés Rockembach**

Universidade Federal do Rio Grande do Sul  
[moises.rockembach@ufrgs.br](mailto:moises.rockembach@ufrgs.br)

## Resumo

Esta investigação analisa o ambiente web e as informações nele produzidas, procurando enquadrar, como objetos de estudo, o arquivamento da web, como fonte de dados de pesquisa, e a comunicação científica, como prática de disseminação do conhecimento produzido nas universidades. A metodologia delimitou-se pela pesquisa exploratória, a partir de revisão bibliográfica internacional sobre o tema e análise das iniciativas participantes do Consórcio Internacional de Preservação da Internet (IIPC) vinculadas a Universidades. Realizou-se análise qualitativa dos objetivos e projetos desenvolvidos por estas iniciativas. Conclui que o arquivamento da web é um campo ainda pouco explorado internacionalmente, principalmente dentro das universidades. E observa a carência de pesquisas na América Latina, principalmente no Brasil.

**Palavras-chave:** Arquivamento da web. Comunicação Científica. Ciência da Informação.

## Abstract

*This research analyzes the web environment and the information produced in this medium, aiming to configure web archiving as an object of study, as a source of research data, along with scientific communication, as a practice of disseminating knowledge produced in universities. The methodology was delimited as exploratory research, based on an international bibliographic review on the subject, and analysis of the Initiatives of the International Consortium for the Preservation of the Internet (IIPC) related with Universities. It uses qualitative analysis of the objectives and projects developed by these initiatives. It concludes that the Web archiving is a field still not explored enough, namely inside the universities, and it observes the lack of research in Latin America context, especially in Brazil.*

**Keywords:** Web archiving. Scientific Communication. Information Science.

## 1. Introdução

Esta investigação procura observar e analisar o ambiente web e as informações nele produzidas, a partir da perspectiva da coleta, preservação e recuperação dos websites e outros objetos digitais também produzidos neste meio, metodologia conhecida como arquivamento da web, e os possíveis usos de seu potencial informacional e probatório.

Procuramos enquadrar, como objetos de estudo, o arquivamento da web, como fonte de dados de pesquisa, e a comunicação científica, como prática para validação e disseminação do conhecimento produzido nas universidades. Como procedimentos metodológicos, a investigação se configura como uma pesquisa exploratória e descritiva, com revisão bibliográfica e análise efetuadas a partir dos estudos de caso levantados sobre o arquivamento da web, realizados por universidades vinculadas ao Consórcio Internacional de Preservação da Internet (*International Internet Preservation Consortium - IIPC*), e sua relação com a comunicação científica.

A comunicação científica não possui uma data específica de origem, estudos apontam que ela é proveniente dos povos gregos de Atenas quando grupos se reuniam para discutir questões filosóficas entre os séculos IV e V a.c. (MEADOWS, 1999). O que de fato se pode compreender é que a ciência só existe se ela for comunicada, e sua comunicação é tão importante quanto a própria pesquisa, portanto, o processo de comunicar vem legitimar a ciência, pois possibilita que a mesma seja analisada pelos pares. (ZIMAN, 1979).

No Brasil, Pinheiro (2012) relata que a comunicação científica é entendida como uma subárea da Ciência da Informação e que, conforme Mueller (2007), tem seu princípio relacionado à necessidade de assegurar o acesso ao crescente volume de publicações científicas. Este tipo de comunicação despontou no país através da inserção de disciplinas e professores estrangeiros nos programas de pós-graduação do Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT, pioneiro na América Latina. Entre os nomes internacionais que colaboraram com esta inclusão através de suas publicações traduzidas em português estão Derek Solla Price, John Ziman e Jack Meadows. Outros professores renomados como Derek Langridge, John Joseph Eyre, Suman Datta e Jack Mills da Polytechnic também contribuíram com seus conhecimentos, ministrando aulas no país na década de 70.

Para um melhor contorno sobre o tema, podemos definir a comunicação científica como um processo que reúne uma série de atividades, incluindo produção, disseminação e uso da informação, desde a concepção do problema de pesquisa até que os resultados sejam aceitos como componentes do conhecimento científico, conforme Pinheiro (2012) baseada em Garvey (1979). Estas atividades geram uma quantidade demasiada de informação como referenciado por Meadows (1999), não apenas pela quantidade de produções e publicações como também pela variedade de temáticas pesquisadas nos diferentes campos científicos. Por causa disso, e da importância de perpetuar o conhecimento para futuras gerações, deve-se resguardar o que se deixou de herança, principalmente quando se fala na era da informação e do ambiente web. Este último se trata de um meio de comunicação formado sobre a rede de computadores (Internet) e que a cada dia se concretiza como o meio mais utilizado para publicações da sociedade moderna (Gomes, 2010).

A ideia de preservar o que é disponibilizado pela web ainda não é consenso entre os pesquisadores que estão estudando o arquivamento da mesma por três motivos principais, como diz Masanès (2006). O primeiro é a qualidade dos dados encontrados que não correspondem aos padrões de preservação. Um dos pontos desta visão acredita que seria adequada uma seleção manual dos conteúdos, mas isso é incompatível considerando a amplitude de dados na web. Podemos também, relacionar este ponto de vista com o crescimento exponencial das publicações científicas. O segundo entende que a web se autopreserva e por isso não há necessidade deste trabalho e o terceiro acredita de modo claro que não é possível arquivá-la.

De modo geral, já existem diversas iniciativas ao redor do mundo que visam o arquivamento da web, basicamente o processo envolve a identificação dos dados de interesse, a captura e o devido armazenamento para possível acesso destas informações por seus diferentes públicos. São, pelo menos, quatro continentes envolvidos com este processo, sendo os primeiros registros de ações do ano de 1996.

A partir dos estudos de caso delimitados na pesquisa, foi possível identificar os projetos em desenvolvimento de algumas universidades. Em destaque as Universidades de Harvard, Stanford e Norte do Texas que possuem portais independentes do *Archive-It*, fato que acarreta a necessidade de maiores recursos tecnológicos, de pessoal e financeiros. Ainda se

destaca o projeto *Memento - Time Travel for the Web* desenvolvido pela Biblioteca de Pesquisa do Laboratório Nacional de Los Alamos em colaboração com o Departamento de Ciência da Computação da Old Dominion, bem como o projeto *Hiberlink*, também desenvolvido pela Biblioteca de Pesquisa do Laboratório Nacional de Los Alamos.

## 2. As Comunidades e Sociedades Científicas

As denominadas sociedades ou comunidades tiveram um papel fundamental no desenvolvimento da comunicação científica. Formadas essencialmente por membros sócios tinham como objetivo demonstrar e comprovar suas próprias investigações e, posteriormente, abrir este conhecimento para a sociedade em geral para assim repassá-lo às futuras gerações. Conforme a obra de Meadows (1999) era através de um grupo específico de pessoas que as informações eram difundidas de modo ágil em reuniões que ocorriam regularmente. Mueller (2007) explica as comunidades científicas parafraseando Ziman (1984), definindo-as como grupos de indivíduos ligados a instituições formais, como universidades, institutos de pesquisa, sociedades científicas e também redes informais de colaboração e comunicação, como os colégios invisíveis, formados por pesquisadores que por um determinado momento estão interessados e envolvidos num mesmo problema de pesquisa.

Inicialmente existiu uma diferenciação de nomenclatura entre “academia” e “sociedade” que influenciou na periodicidade da comunicação destes grupos, porém, na contemporaneidade, como afirma Meadows (1999), todas seguem a mesma missão. Entre as primeiras fundadas estão a *Académie Française* (1635), a *Royale de Peinture et de Sculpture* (1648) e a *Royale des Inscriptions et Belles-Lettres* (1663), em Paris; e a *Royal Society* (1662), em Londres. Posteriormente, novos objetivos como controle e fiscalização profissional se agregaram às sociedades e associações, expandindo também à atividade da comunicação.

Meadows (1999, p. 11) afirma que “hoje em dia, a maioria dos sócios tem acesso a acervos adequados, ao alcance da mão, graças às bibliotecas de suas instituições”. Este histórico auxilia a entender como ocorreu todo o processo da comunicação científica desde suas origens, pois estas comunidades e a forma como se organizavam ainda refletem na forma de acesso aos resultados científicos de diferentes áreas.

## 2.1. Formas, Tipos e Canais de Comunicação Científica

É importante identificar qual a natureza da comunidade científica a qual está se direcionando a informação, pois ela orienta a escolha do tipo de forma e meio adequado para a comunicação. A fala e a escrita são as formas mais antigas e importantes de comunicar a ciência, uma complementa a outra e ambas são utilizadas, desde os primórdios da comunicação científica, pelos gregos. A tecnologia da prensa potencializou a difusão da informação, mantendo o formato manuscrito destinado a um público reduzido ainda nos séculos XVII e XVIII, por meio de cartas, por exemplo. Segundo Meadows (1999) isso acontecia devido à possível censura que a ciência poderia sofrer, deste modo, a circulação de informações manuscritas se tornou estratégica para obtenção de prova e testes dos pares, para que posteriormente estes dados fossem difundidos para um público maior. A partir disto, surgem as revistas científicas, os primeiros registros se deram em 1660 e se tornaram regulares com a formalização da *Royal Society* em 1662.

A bibliografia estudada indica que a comunicação científica se divide em dois tipos, a informal e a formal. O primeiro é caracterizado pela troca de informações entre pesquisadores por canais e meios não oficiais. Entre eles estão “conversas pessoais face a face, por telefone ou carta, aulas e palestras, e circulação de preprints (manuscritos ainda não publicados sobre uma pesquisa), trabalhos apresentados em reuniões” (Mueller, 2007, p. 130). O segundo é caracterizado pela troca via canais reconhecidos como oficiais, onde se incluem capítulos e edição de livros, teses, dissertações, anais de eventos científicos, artigos publicados em revistas científicas, entre outros. A autora observa que o avanço da tecnologia e dos repositórios digitais tornou essa fronteira um pouco turva, porém a divisão ainda permanece válida.

Como afirma Meadows (1999), alguns canais de comunicação são óbvios, como uma conversa face a face em um encontro. e eles também se dividem entre formais e informais, embasando o apresentado por Mueller (2007). As editoras são os principais canais formais se tratando dos impressos em papel, seguidas pelas bibliotecas e unidades de informação, que são os mais importantes compradores destas produções. Só na década de 1990, no período de um ano, as bibliotecas do Reino Unido adquiriram mais de 2 milhões de livros e por volta de 600 mil assinaturas de periódicos.

As bibliotecas são entendidas como depositários de informações passadas e presentes (que ainda estão sendo editadas) e o crescimento desta literatura científica demonstra a necessidade de se pensar o armazenamento destas para acesso futuro. Congressos e conferências se enquadram em canais informais e tem, como principal forma de interação, a fala. Como resultado destes encontros se obtém publicações (anais, livros e periódicos), que são considerados meios formais. A qualidade destes está relacionada ao processo de avaliação e pode variar conforme a orientação dos envolvidos, de acordo com o apresentado por Meadows (1999).

A tecnologia não só ampliou as formas e os canais de comunicação científica como também o modo de processar os dados por parte das editoras e bibliotecas; a consulta, a percepção e a absorção do conhecimento por parte do público, como explicam Meadows (1999) e Mueller (2007). Estes avanços trouxeram alguns benefícios como a agilidade, a amplitude e a comodidade do acesso remoto na disponibilização e na busca das informações.

No entanto, ao passo que o conhecimento científico é reescrito, novos experimentos vão sendo realizados e bibliotecas científicas ou unidades de informação são memórias em constante mutação e crescimento, como diz Ziman (1979). Esta ideia também vai ao encontro do que Meadows (1999) apresenta sobre quantidade exponencial de dados em circulação. Não apenas devido à variedade de produção, mas também a diversidade de temáticas que hoje são pesquisadas, as quais atraem cada vez mais acadêmicos e entusiastas.

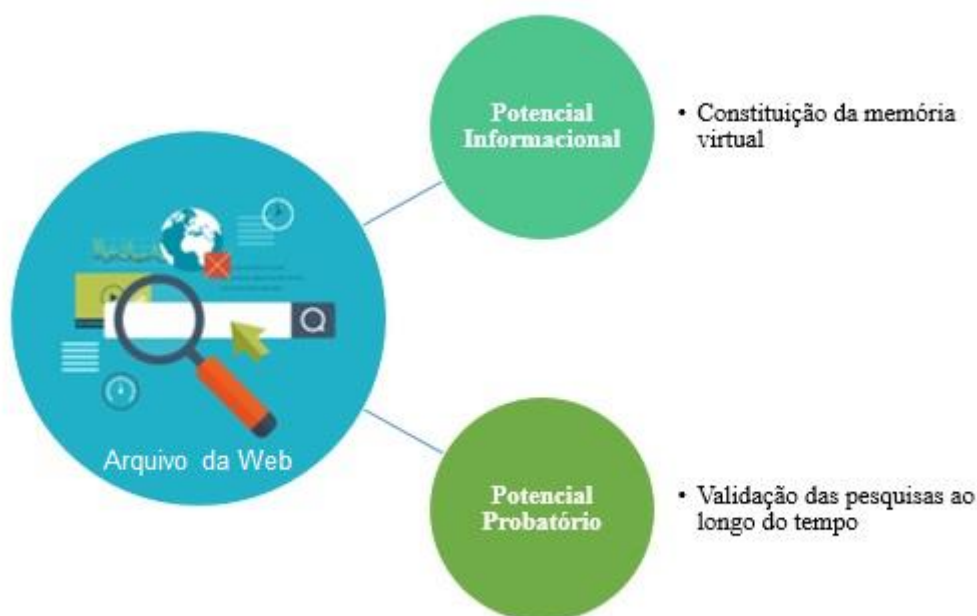
Corroborando com Klein (2014), refletimos sobre os desafios dos usos da Internet que afetam fundamentalmente a maioria dos aspectos da comunicação da informação, incluindo a comunicação científica. E, portanto, o imediatismo que caracteriza a publicação de informações na web, bem como o acesso, permite um aumento dramático na velocidade de disseminação do conhecimento acadêmico. Assim, referenciar fontes - como uma parte fundamental do discurso acadêmico - e a expectativa de que as fontes referenciadas possam e devam ser verificadas por outros, para permitir uma interpretação correta das informações que estão sendo comunicadas e para apoiar a reprodutibilidade dos resultados, passa a ser um problema recorrente no ambiente virtual.

Por este motivo, a discussão sobre o registro e o armazenamento destas informações é importante para acesso futuro e devida comunicação científica, assim, entendemos o

arquivamento da web como uma das práticas que se propõem a garantir o potencial informacional e probatório das referências que dão subsídios e sustentam as pesquisas científicas.

As iniciativas de arquivamento da web dão origem a referências persistentes, como representamos na **Figura 1**, que constituem uma memória virtual, garantindo o seu potencial informacional, bem como possibilitam a validação ao longo do tempo, garantindo o seu potencial probatório.

**Figura 1: Arquivamento da web e as funções das referências persistentes na Comunicação Científica**



Fonte: autores, 2017

Ao nos referirmos ao potencial probatório de uma referência, estamos incorporando não apenas a possibilidade de acesso para a validação da mesma, mas também sobre a preservação de seu conteúdo, tal qual era no ato da coleta dos dados e informações para uma determinada pesquisa.

Neste sentido, e buscando novas possibilidades para o acesso e a validação da comunicação científica, universidades de diversos países vêm trabalhando no desenvolvimento de projetos de arquivamento da web.

### 3. Contexto do Arquivamento da Web

Dentro do contexto desta pesquisa o arquivamento da web é significativo e virá a preservar conteúdos pertinentes para o acesso e desenvolvimento de gerações futuras, pois se destaca a grande quantidade de dados disponibilizados na rede, convergindo com a ideia apresentada por Gomes (2010) citada na introdução deste artigo. Neste caso, é importante esclarecer que o arquivamento da web é um processo, uma ação contínua e prolongada que demanda regularidade em todas suas etapas. De modo simplificado, trata-se de um procedimento que identifica a informação, captura e preserva o conteúdo original das páginas, sendo fiel ao que foi postado pelos editores oficiais, que podem ser qualquer pessoa ou instituição que tenha acesso à web.

Compreende-se que para que o arquivamento da web seja executado com eficiência, eficácia e efetividade, deve-se respeitar as etapas que contemplam desde o perfil de páginas/conteúdo a ser arquivado, *hiperlinks* contidos, período e frequência de arquivamento até questões éticas e políticas envolvidas. Isto se constata importante, pois como explanam Costa, Gomes e Silva (2016) 80% do conteúdo disponibilizado na rede é alterado após o período de um ano, não mantendo seu formato original e 13% das referências *on-line* utilizadas por estudantes em suas pesquisas desaparecem em pouco mais de dois anos.

A partir da descrição acima, é perceptível o quão complexo e estratégico é o processo de arquivamento da web, porém não impraticável. Por este motivo é conveniente destacar que não é apenas uma única iniciativa que irá suprir a necessidade de arquivamento da rede, devido ao já mencionado crescimento exponencial da informação, como observam Gomes (2010) e Masanès (2006). Os autores defendem que somente a união das iniciativas existentes permitirá que os usuários usufruam dos reais benefícios do processo como um todo.

Embasados num *survey* realizado em 2010, Gomes, Miranda e Costa (2011) registraram 42 iniciativas ao redor do mundo, sobretudo concentradas na América do Norte, Europa, Oceania e Ásia. Os dados apresentados demonstraram que 24 destas mantinham finalidade nacional para a coleta de conteúdo, e os primeiros registros são datados de 1996 na Austrália, Suécia e nos Estados Unidos. Os demais escopos variam entre interesses regionais, audiovisuais, institucionais e de literatura, por exemplo, caracterizando uma perspectiva heterogênea. Realizando um levantamento a partir dos dados disponibilizados no site do *International*



*Internet Preservation Consortium* (2017), verifica-se que atualmente existem 52 membros, incluindo uma iniciativa no Chile, a primeira na América Latina. Em Portugal foi identificado o Projeto de Arquivo da Web Portuguesa (AWP), um serviço prestado pela Fundação para a Computação Científica Nacional, que tem como missão capturar, armazenar e preservar a informação que interessa aos portugueses ou a quem se interessar pelo que foi publicado no país.

Numa perspectiva mundial surge, em 2003, pela parceria de 12 instituições com hospedagem oficial na Biblioteca Nacional da França, o Consórcio Internacional de Preservação da Internet (do inglês *International Internet Preservation Consortium* - IIPC). Atualmente, ele conta com a participação de mais de 45 países e tem como membros bibliotecas, museus, arquivos nacionais, universitários, regionais e instituições de património cultural. Fazem parte do IIPC bibliotecas de países como Áustria, Suíça, Estados Unidos, Finlândia, Los Angeles, Portugal e Reino Unido. O Consórcio tem como missão unir organizações para coletar, preservar e tornar acessível o conhecimento da web com escopo global, nesta busca estão incluídas publicações acadêmicas, obras de arte, como também documentos governamentais. Para ser sócio é necessário demonstrar interesse ou experiência no campo de arquivamento da web. O trabalho é executado de modo colaborativo a fim de compartilhar conhecimento e continuar a desenvolver *softwares* e ferramentas adequadas para a melhor operacionalização da missão do IIPC.

Fundamentado na ideia e na atuação do IIPC se percebe a efetividade da união de forças e como isto torna o arquivamento da web praticável e com mais potencialidade de abrangência. Este ponto de vista se fortalece quando se observa uma das tecnologias utilizadas para o processo de arquivamento da rede pelas instituições que fazem parte do Consórcio, o chamado *WayBack Machine*, disponibilizada pelo *Internet Archive*. Esta última se trata de uma organização sem fins lucrativos, fundada em 1996 que fornece acesso universal ao que é coletado e utiliza o programa *Archive-It* como base para identificar as páginas importantes para captura. Conforme informações disponibilizadas no site do *Internet Archive* (2017) a organização possui em seus arquivos 279 bilhões de páginas da web, 11 milhões de livros e textos, 4 milhões de gravações de áudio (incluindo 160.000 concertos ao vivo), 3 milhões de vídeos (incluindo 1 milhão de programas de televisão), 1 milhão de imagens e 100.000 programas de software. Atualmente o programa é capaz de digitalizar e disponibilizar 1.000

livros acadêmicos diariamente e pode ser acessado gratuitamente por qualquer pessoa no mundo.

A partir desta metodologia de preservação de informações da web, abordaremos a seguir nove iniciativas que a utilizam na academia, visando a comunicação científica e o acesso amplo às informações produzidas em suas pesquisas.

## **4. O Arquivamento da Web na Comunicação Científica**

Alguns projetos de arquivamento da web estão sendo tratados por universidades como extensões de suporte oferecidos por suas bibliotecas universitárias, neste contexto, delimitamos a pesquisa a partir da lista de membros que compõem o IIPC, sendo identificadas nove iniciativas desenvolvidas em universidades – oito localizadas nos Estados Unidos e uma na Eslováquia, as quais trazemos abaixo.

Compreendendo que a Internet tem sido, cada vez mais, o ponto de origem de um grande volume de informações, pesquisas e publicações científicas - muitas universidades voltaram seus recursos tecnológicos e seu conhecimento técnico para a preservação de conteúdos e comunicações, que vêm se perdendo com o passar dos anos. Descreveremos a seguir como algumas dessas universidades têm desenvolvido projetos em cooperação para que o conhecimento científico digital não se perca em sua totalidade.

### **4.1. Bibliotecas da Universidade do Norte do Texas**

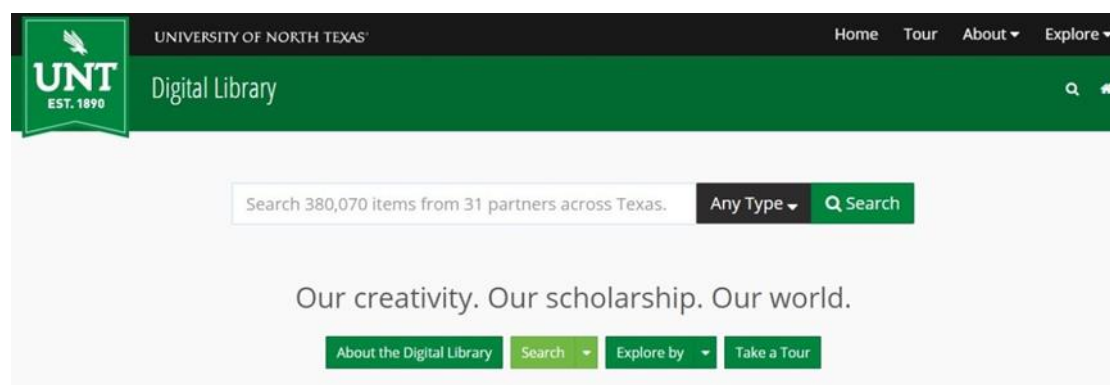
Segundo a *University of North Texas Libraries* (2017), as bibliotecas da instituição assumiram um papel ativo na preservação da Internet, coletando e arquivando domínios da web UNT, sites governamentais e outros conteúdos da web.

A Universidade do Norte do Texas (*International Internet Preservation Consortium*, 2017) foi uma das primeiras instituições acadêmicas dos Estados Unidos a arquivar sites, começando em 1997 com o *CyberCemetery*, um arquivo de sites governamentais que cessaram sua operação. Esta coleção (*University North Texas*, 2017) apresenta uma variedade de tópicos indicativos da natureza ampla das informações do governo e, em particular, sites que abordam temas que apoiam o currículo da universidade e pontos específicos do programa.

Outra coleção em destaque é a *End of Therm Web Archive*, que surgiu da colaboração das Bibliotecas da Universidade do Norte do Texas com a Biblioteca do Congresso, o Arquivo da Internet (*Internet Archive*), a Biblioteca Digital da Califórnia e o Escritório de Impressão do Governo dos EUA visando colher sites que mudariam rapidamente no final dos termos presidenciais em 2008 e 2012.

Apresentamos na **Figura 2**, o portal da Biblioteca Digital da Universidade do Norte do Texas (UNT), onde observamos que o layout da página é bem acessível e oferece a possibilidade de pesquisas, em tipos bem distintos de informações, desde dissertações e artigos até websites.

**Figura 2: Página inicial da Biblioteca Digital da Universidade do Norte do Texas**



Fonte: UNIVERSISITY OF NORTH TEXAS, 2017

Um website arquivado (*snapshot*) dentro da coleção *CyberCemetery* representa uma fonte de pesquisa e referência com grande potencial informacional, que visa adicionar valor a diversas áreas do conhecimento, como Política, Sociologia, Ciência da Informação, Direito, entre outros. Mas também fornece um potencial probatório, o qual possibilitará a validação de uma determinada pesquisa ao longo do tempo. Ao utilizarmos fontes de referência, nascidas em meio digital, na comunicação científica, principalmente as redes sociais, blogs e websites, devemos ter consciência de que as mesmas dificilmente poderão ser validadas daqui a dez anos a partir de suas fontes originais.

A proposta de unificar todas as fontes de pesquisa científica em um único repositório, proporciona maior eficiência na pesquisa e satisfação dos usuários, além de cruzar fontes e tipos distintos sobre uma mesma temática. Consideramos a Iniciativa da UNT um modelo bem planejado e estruturado, que busca atender às distintas demandas do pesquisador de maneira objetiva e prática.

## 4.2. Biblioteca Digital da Califórnia

A *California Digital Library* - CDL também foi uma das pioneiras desta lista que se envolveram no arquivamento da web a partir de 2003, segundo o International Internet Preserving Consortium (2017), com o Projeto Web em Risco (do inglês *Web-at-Risk*) - no qual a CDL desenvolveu e operou o seu Serviço de Arquivamento da Web (WAS).

Em 2010, o Projeto Web em Risco estabeleceu alguns objetivos a serem explorados junto aos pesquisadores e bibliotecários envolvidos no projeto, os quais consistiam em identificar como as necessidades da comunidade para o arquivamento da web poderiam ter mudado desde o início do trabalho; identificar como as necessidades de arquivamento da web da comunidade de pesquisa em larga escala poderiam ser abordadas; identificar/analisar como o WAS, em particular, poderia ser melhorado para melhor ajudar os estudiosos e bibliotecários e mensurar os custos futuros e o crescimento do serviço, além de pensar nas possíveis abordagens para se investir em sustentabilidade (*California Digital Library*, 2017).

A *California Digital Library* (2017), manteve o WAS por oito anos e, em 2015, optou por um trabalho colaborativo com o serviço de arquivamento da web do Archive-It, assim, migrou todos os seus clientes para esta plataforma. A equipe do WAS percebeu que a complexidade e a constante mudança da web representavam desafios significativos para o conjunto de ferramentas de arquivamento da web atual e exigia atualizações frequentes para se manter à frente. Este custo afetaria os objetivos de definir as necessidades técnicas, bem como a estrutura organizacional que poderiam garantir a criação de novas ferramentas e serviços e torná-los amplamente disponíveis em toda a comunidade.

Atualmente, o CDL está explorando oportunidades com Harvard, MIT, Stanford, UCLA e outros para trabalhar em colaboração com o *Archive-It* para criar uma lista expandida de ferramentas e serviços de valor agregado. Esta iniciativa fomenta a convergência dos serviços de arquivamento da web para uma única plataforma, enquanto individualmente cada universidade poderia trabalhar no aprimoramento dos mecanismos de pesquisa e das coleções, entre outros objetivos mais específicos.

### 4.3. Biblioteca de Harvard

Conforme a *Harvard Library* (2017), foi lançado em 2006 um projeto piloto de seu serviço de coleções de arquivos da web, financiado pela Iniciativa da Biblioteca Digital da Universidade (do inglês *University's Library Digital Initiative* - LDI). Este projeto foi o primeiro do LDI especificamente orientado para preservar o material "nato-digital" e, em 2009, foi lançada a interface pública do WAX.

Os gerentes de coleções, Harvard (2017), que trabalham no ambiente on-line, tem como objetivo adquirir o conteúdo que eles sempre coletaram fisicamente. Como os blogs que substituem diários, o correio eletrônico substituindo a correspondência tradicional e os materiais HTML substituindo muitas formas de garantia impressa, assim, os gerentes de coleções estão cada vez mais preocupados com possíveis lacunas na documentação do patrimônio cultural de Harvard.

Portanto, o WAX foi desenvolvido como uma resposta inicial e parcial a essas e outras preocupações, que vão desde a viabilidade técnica até implicações legais e financeiras. O piloto concentrou-se na colheita de conteúdo da superfície da web - conteúdo que é descoberto para os motores de busca por meio de rastreadores da web, em oposição ao conteúdo escondido dos rastreadores da web em um banco de dados ou restrito por senha ou proteção de *login*.

O serviço de arquivamento WAX possui cinco grandes coleções, (Harvard, 2017), que estão divididas em: "*H-Sites: Harvard life and learning*", "*SL Sites: Archived Websites from Schlesinger Library Collections*", "*Capturing Women's Voices on the Web*", "*Constitutional Revision in Japan Research Project*" e "*A-Sites: Archived Harvard Websites*".

A "*H-Sites: Harvard life and learning*", foi criada com a proposta de coletar os materiais, agora criados em sites, que complementam os arquivos pessoais de indivíduos e registros de organizações afiliadas à Universidade. Visa preservar os interesses intelectuais e sociais de um segmento da comunidade de pessoas que vivem, trabalham e aprendem em Harvard. Para a Universidade, as vidas dentro e fora das salas de aula e escritórios são parte integrante da cultura e da história de Harvard.

A *“SL Sites: Archived Websites from Schlesinger Library Collections”* é uma coleção de sites criados e mantidos por organizações e indivíduos cujas coleções são mantidas na Biblioteca Schlesinger. Estes sites completam as coleções baseadas em papel e representam documentação adicional sobre as atividades e contribuições importantes dessas organizações e indivíduos.

A *“Capturing Women's Voices on the Web”* é uma coleção com a missão de capturar as vozes das mulheres cujos pontos de vista não podem ser encontrados em outros lugares, bem como, documentar o uso de blogs e outras formas de publicação na web pelas mulheres americanas no início do século XXI.

A *“Constitutional Revision in Japan Research Project”*, coleção que originou-se a partir do Projeto de Pesquisa de Revisão Constitucional do Japão, criado em 2005, onde são realizadas reuniões para discutir, analisar e documentar o processo de revisão constitucional no Japão. Além das reuniões, são arquivados materiais digitais relevantes de vastas fontes relacionadas à revisão constitucional e importantes para o projeto. Uma vez que a informação sobre as atividades atuais de indivíduos e grupos envolvidos na questão é principalmente gerada em meio digital, uma seleção de cerca de oitenta sites relacionados é coletada periodicamente para garantir que o debate e o processo de revisão constitucional sejam preservados e disponibilizados para os estudiosos. Este projeto, considerando a história entre os países, serve como exemplo de como é possível desenvolver o diálogo entre culturas distintas para se chegar a um ponto de equilíbrio e consenso e é um exemplo de arquivamento que contribui para a pesquisa em diversas áreas da Ciência.

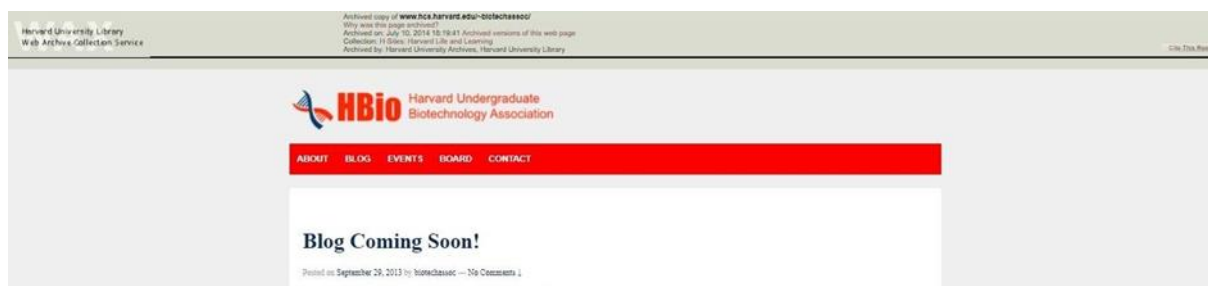
A coleção *“A-Sites: Archived Harvard Websites”*, criada a partir de 2007, consiste na preservação de sites de concessão de diplomas de departamentos e comitês da Faculdade de Artes e Ciências. Estas informações, antes preservadas apenas em formato físico pelos Arquivos da Universidade, agora coletam sites de toda a Instituição, com periodicidades mensais ou anuais. Assim, a lista de sites coletados continua a crescer, proporcionando uma imagem mais completa do espaço web de Harvard.

O serviço de arquivamento da Biblioteca de Harvard (SWAP) é disponibilizado para toda a comunidade acadêmica e não acadêmica de forma aberta e gratuita. As coleções descritas são formadas por listas de sites arquivados, podendo ter mais de uma versão arquivada para cada

site. Estes sites são fontes de informação e referências que se não fossem preservados poderiam prejudicar a validação das pesquisas que as utilizaram, uma vez que a prova delas tende a se perder com o tempo.

E, como mostra a **Figura 3**, ao selecionar uma *snapshot* específica o site arquivado é carregado com as funcionalidades originais da época em que a captura foi realizada. No cabeçalho do site estão descritos alguns dados sobre a captura e a qual coleção pertence.

**Figura 3: Lista de sites arquivados da coleção H-Sites: Harvard life and learning**



Fonte: HARVARD LIBRARY/WAX, 2017.

A coleção possibilita a pesquisa por palavras-chave e termos dentro dos sites antes de acessar qualquer versão e ainda disponibiliza o link atual da página para acesso.

#### 4.4. Bibliotecas da Universidade de Stanford

A *Stanford University Libraries - SUL* (2017), em colaboração com bibliotecários, professores, pesquisadores e outros funcionários da Universidade de Stanford, visa identificar os conteúdos selecionados da web para realizar o seu arquivamento. Os objetivos da SUL consistem em armazenar os arquivos da web no Repositório Digital de Stanford, fornecer descobertas através do *SearchWorks* e habilitar a navegação através de uma instância local da plataforma do *Wayback Machine*.

Segundo o *International Internet Preservation Consortium* (2017), a SUL iniciou seu envolvimento com o arquivamento da web em 2007, focando em informações governamentais e institucionais da web. E após uma pesquisa realizada na Universidade, entre 2011 e 2012, sobre o potencial uso deste recurso, em 2013 começou a ser construído um programa de arquivamento da web mais abrangente e a iniciativa passou a fazer parte do IIPC.

A *Stanford University Libraries* (2017) lista alguns dos fatores que motivaram o seu envolvimento no arquivamento da web, entre eles: sites de campanha de candidatos políticos, que estão disponíveis apenas durante a temporada eleitoral; sites de projetos financiados, que ao deixarem de receber os subsídios, são retirados da web, mesmo com o projeto em andamento; discurso político dissidente sujeito à censura governamental; notícias on-line relacionadas a eventos rápidos, que da mesma forma são rapidamente alterados e submersos; presença na web de personagens importantes falecidos; entre outros.

Na página inicial do Portal de Arquivamento da Web de Stanford (*Stanford Web Archive Portal* - SWAP), a pesquisa é realizada a partir da URL de interesse. E, na **Figura 4**, podemos visualizar como são apresentados os resultados da pesquisa por URL, onde, a partir de um calendário, é possível selecionar por ano, mês e dia a *snapshot* desejada.

**Figura 4: Página inicial da Stanford Web Archive Portal - SWAP**



Fonte: STANFORD UNIVERSITY LIBRARIES/SWAP, 2017.

Assim como no WAX de Harvard, o site arquivado é apresentado com um cabeçalho contendo a data de arquivamento. No caso do SWAP, é possível navegar pelos demais *snapshots* nesta mesma barra/cabeçalho.

Segundo a SUL (2017), assegurar a continuidade da capacidade de acesso ao conteúdo da web que desapareceu ou foi substituído, está diretamente relacionado com objetivos diversos, de pesquisa, ensino, construção de coleções de bibliotecas, legado institucional, conformidade legal e administração de informações governamentais.



#### 4.5. Biblioteca de Pesquisa do Laboratório Nacional de Los Alamos e Departamento de Ciência da Computação da Old Dominion

A equipe de prototipagem da *Los Alamos National Laboratory Research Library* - LANL (2017), explora vários aspectos da comunicação acadêmica na era digital, com foco principal na infraestrutura e interoperabilidade da informação e persistência a longo prazo do registro acadêmico. Dois projetos associados com o arquivamento da web ganharam destaque dentro da iniciativa: o *Hiberlink* e o *Memento - Time Travel for the Web*.

O Projeto *Hiberlink*, Klein (2014), utiliza o termo “*reference rot*” para denotar a combinação de dois problemas envolvidos no uso de referências URI, que se relacionam com a natureza dinâmica e efêmera da web: o ***link rot***, onde o recurso identificado por um URI pode deixar de existir e, portanto, uma referência URI para esse recurso deixará de fornecer acesso ao conteúdo referenciado; e o ***content drift***, onde o recurso identificado por um URI pode mudar ao longo do tempo e, portanto, o conteúdo no final do URI pode evoluir, até mesmo deixando de ser representativo do conteúdo originalmente referenciado.

O Projeto *Memento - Time Travel for the Web* (2017) tem sido desenvolvido em colaboração com o *Old Dominion Department of Computer Science*. Este projeto consiste em um protocolo, que adiciona uma dimensão de tempo ao protocolo HTTP. Inspirado na negociação de conteúdo HTTP, o protocolo introduz a noção de negociação de data e hora que permite que um cliente solicite a versão de um recurso tal como existia em um tempo específico no passado.

#### 4.6. Bibliotecas da Universidade da Columbia e Biblioteca de Pesquisa da UCLA

As Iniciativas da *Columbia University Libraries* e *UCLA Research Library* são parceiras na utilização da plataforma e do serviço do *Archive-It* para coletar e arquivar suas coleções. Contudo, nos sites institucionais existem poucas informações a respeito de suas coleções - sendo necessária a pesquisa direta no *Archive-It* pelo nome da instituição.

#### **4.7. Biblioteca da Universidade de Bratislava**

A Iniciativa mais recente a incorporar o IIPC (2017), a *Univerzitná knižnica v Bratislave*, a partir de 2015, começou a executar o arquivamento de domínios nacionais da web e de documentos nascidos no ambiente digital, mas ainda não se encontra nenhuma plataforma disponível para acesso aberto, bem como a localização de nenhuma coleção no *Archive-It*.

Em uma análise geral, percebemos que as iniciativas de arquivamento da web preocupam-se tanto com os aspectos informacionais, os quais são definidos como a formação da memória virtual - este muito mais evidente nos projetos, quanto com os aspectos probatórios desta informação que se origina, tramita e é comunicada na web e pode vir a ter o seu acesso interrompido ou alterado.

### **5. Considerações Finais**

A sociedade da informação, as tecnologias, a variedade de produção e de temáticas de pesquisa colaboram para o crescimento contínuo das publicações científicas na web fazendo a comunicação da ciência atingir um novo patamar em sua abrangência e disseminação. Estes são alguns dos motivos que despertaram o interesse de estudiosos da Ciência da Informação acerca de novas fontes de referência, como é o caso do arquivamento da web para a comunicação científica.

Iniciativas nesta perspectiva são realidade há pouco mais de duas décadas e demonstram a importância de se pensar o armazenamento dos dados publicados neste meio de comunicação. Estas ações propõem - pouco a pouco - coletar partes criteriosamente selecionadas de informações formando uma memória coletiva, o mais fiel e completa possível, de uma determinada comunidade, a fim de preservar e garantir o conhecimento para gerações futuras.

No entanto, observa-se que é necessário rever e ampliar os métodos do processo de arquivamento da web para a ciência, a fim de garantir a validação das pesquisas atuais, onde as fontes por vezes se limitam ao ambiente virtual. Para tanto, a aferição das referências deve ser assegurada pelos mecanismos já existentes de preservação da web para que, no futuro, tais investigações possam seguir válidas ou postas a prova em função da evolução das fontes e dados e não por falta de acesso ou por inconsistência dos links.

Assim, se constata que o arquivamento da web é um campo ainda pouco explorado internacionalmente, principalmente dentro das universidades e do campo científico. E ainda se observa a carência de pesquisas na América Latina, principalmente no Brasil sobre esta temática.

## Referências Bibliográficas

---

- BUENO, W. C. (2010). Comunicação científica e divulgação científica: aproximações e rupturas conceituais. *Informação & Informação*, 15, 1-12. doi: 10.5433/1981-8920
- COSTA, M; GOMES, D; SILVA, M.J. (2016). The evolution of web archiving. *International Journal on Digital Libraries*, 1-15. doi: 10.1007/s00799-016-0171-9
- GARVEY, W. D. (1979) *Communication: essence of science; facilitating information exchange among librarians, scientists, engineers and students*. Oxford: Pergamon Press.
- GOMES, D.(2010.) Preservar a Web: um desafio ao alcance de todos. In: *Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*. (pp. 1-9). Lisboa, POR: Retirado de <http://sobre.arquivo.pt/wp-content/uploads/PreservarAWebBADFormat-v.14.pdf>
- GOMES, D.; MIRANDA, J. & COSTA, M. (2011). A survey on web archiving initiatives. In: *International Conference on Theory and Practice of Digital Libraries*. (pp. 408-420). Lisboa, POR: Springer Berlin Heidelberg. Retirado de [https://link.springer.com/content/pdf/10.1007%2F978-3-642-24469-8\\_41.pdf](https://link.springer.com/content/pdf/10.1007%2F978-3-642-24469-8_41.pdf)
- MASANÈS, J. (2006). Web Archiving: Issues and Methods. In Julien Masanès, *Web Archiving* (pp. 1-46). Paris, FRA: Springer-Verlag Berlin Heidelberg.
- MEADOWS, A. J. (1999). *A Comunicação Científica*. Brasília, DF: Briquet de Lemos.
- MUELLER, S. P. M. (2007). Literatura Científica, Comunicação Científica e Ciência da Informação. In: TOUTAIN, L. B. (Org.). *Para entender a Ciência da Informação*. (pp. 125-144). Salvador, BA: Editora da Universidade Federal da Bahia. Retirado de: <https://repositorio.ufba.br/ri/bitstream/ufba/145/1/Para%20entender%20a%20ciencia%20da%20informacao.pdf>
- PINHEIRO, L. V. R. (2012). Constituição epistemológica e social da comunicação científica no Brasil. In: PINHEIRO, L. V. R.; OLIVEIRA, E. da C. P. de (Orgs.). *Múltiplas facetas da comunicação e divulgação científicas: transformações em cinco séculos*. (pp. 115-148) Brasília, DF: Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT. Retirado de: <http://livroaberto.ibict.br/handle/1/711>
- ZIMAN, J. (1979). *Conhecimento público*. Belo Horizonte, MG: Itatiaia.
- ZIMAN, J. (1984). *An introduction to science studies: The philosophical and social aspects of science and technology*. Cambridge, GB: Cambridge University Press.

KLEIN, M.; SOMPEL, H. V. de; SANDERSON, R. et al. (2014). *Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot*. Retirado de: <https://doi.org/10.1371/journal.pone.0115253>

## Hiperlinks

ARQUIVO.PT. Retirado de: <http://arquivo.pt/> Versão de referência arquivada no Archive.is: <http://archive.is/CIHTO>

HARVARD LIBRARY. WEB ARCHIVE COLLECTION SERVICE (WAX). Retirado de: <http://wax.lib.harvard.edu/collections/about.do?kind=about&lang=eng> Versão de referência arquivada no Archive.is: <https://archive.is/O6aev>

HARVARD LIBRARY. Constitutional Revision in Japan Research Project. Retirado de: <http://wax.lib.harvard.edu/collections/collection.do?coll=101&lang=eng> Versão de referência arquivada no Archive.is: <http://archive.is/uOCGf>

IMAGEM REFERÊNCIAS, FIGURA 1. Retirado de: <https://comunidade.rockcontent.com/o-que-sao-referencias-na-hora-de-escrever-um-texto/> Versão de referência arquivada no Archive.is: <https://archive.is/ZFf84>

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM (IIPC). Membros. Retirado de: <http://netpreserve.org/about-us/members/> Versão de referência arquivada no Archive.is: <http://archive.is/RLaVv>

INTERNET ARCHIVE. About. Retirado de: <http://archive.org/about/> Versão de referência arquivada no Archive.is: <http://archive.is/xokXX>

INTERNET ARCHIVE. Wayback Machine. Retirado de: <https://archive.org/web/> Versão de referência arquivada no Archive.is: <http://archive.is/xokXX>

LOS ALAMOS NATIONAL LABORATORY. Digital Library Research and Prototyping. Retirado de: <http://www.lanl.gov/library/about/research-prototyping.php> Versão de referência arquivada no Archive.is: <http://archive.is/8cdqA>

MEMENTO - TIME TRAVEL SERVICE. Retirado de: <http://timetravel.mementoweb.org/about/> Versão de referência arquivada no Archive.is: <http://archive.is/4LvJ9>

STANFORD UNIVERSITY LIBRARIES. Stanford Web Archive Portal. Retirado de: <https://swap.stanford.edu/> Versão de referência arquivada no Archive.is: <http://archive.is/YcObD>

STANFORD UNIVERSITY LIBRARIES. Web Archiving. Retirado de: <https://library.stanford.edu/projects/web-archiving> Versão de referência arquivada no Archive.is: <http://archive.is/oVFbZ>

UNIVERSITY OF CALIFORNIA. California Digital Library – CDL. The Web-At-Risk: Preserving Our Nations's Digital Cultural Heritage. Retirado de: <http://www.cdlib.org/services/uc3/partners/webatrisk.html> Versão de referência arquivada no Archive.is: <http://archive.is/geMrM>

UNIVERSITY OF CALIFORNIA. California Digital Library – CDL. Announcing a New Partnership: California Digital Library, UC Libraries, and Internet Archive's Archive-It Service. Retirado de: <http://www.cdlib.org/cdlinfo/2015/01/14/announcing-a-new-partnership-california-digital-library-uc-libraries-and-internet-archives-archive-it-service/> Versão de referência arquivada no Archive.is: <http://archive.is/bLpkE>

UNIVERSITY OF NORTH TEXAS. Website. Retirado de: <https://digital.library.unt.edu/> Versão de referência arquivada no Archive.is: <http://archive.is/6fncP>

UNIVERSITY OF NORTH TEXAS. CyberCemetery. Retirado de: <https://digital.library.unt.edu/explore/collections/GDCC/> Versão de referência arquivada no Archive.is: <http://archive.is/vzYxP>

WEB ARCHIVING SERVICE. Website. Retirado de: <https://archive-it.org/> Versão de referência arquivada no Archive.is: <http://archive.is/RMAQK>

# A relação entre referências e menções: estudo exploratório em artigos na base de dados BRAPCI<sup>1,2</sup>

*The relation between references and mentions: exploratory study on articles in the BRAPCI database*

**Rodrigo Aquino de Carvalho**

Universidade Federal do Rio Grande do Sul  
[racfurg@gmail.com](mailto:racfurg@gmail.com)

**Sônia Caregnato**

Universidade Federal do Rio Grande do Sul  
[sonia.caregnato@ufrgs.br](mailto:sonia.caregnato@ufrgs.br)

**Maria de Fátima Santos Maia**

Universidade Federal do Rio Grande  
[mafas.maia@gmail.com](mailto:mafas.maia@gmail.com)

**Raquel Oroski**

Colégio São Carlos, Caxias do Sul – RS  
[queloroski@gmail.com](mailto:queloroski@gmail.com)

## Resumo

Apresenta uma perspectiva para os estudos de citação a partir da relação entre a lista de referências e as menções desses documentos no texto. Os objetivos são: i) identificar os padrões existentes na relação entre as referências e as menções na área da Recuperação da Informação, através da análise de artigos publicados em periódicos acadêmicos indexados na base de dados BRAPCI; e ii) comparar um ranking de citação tradicional de autores pessoais com um ranking de menções. Trata-se de um estudo cientométrico e exploratório, para o qual foram utilizados 66 artigos, publicados entre 2008 e 2012, indexados na base de dados BRAPCI e

## Abstract

*It presents a perspective to citation studies through the relation between references and mentions to those documents in the text. The objectives are: i) to identify the existing patterns in the relationship between references and mentions in the area of Information Retrieval through articles published in academic journals indexed in the BRAPCI database; and ii) to compare a traditional citation ranking of personal authors with a ranking of mentions. It is a scientometric and exploratory study, to which were used 66 articles, published between 2008 and 2012, indexed in the BRAPCI database and retrieved using 12 terms related to "information retrieval". The articles*

recuperados a partir de 12 termos relacionados à “recuperação da informação”. Os artigos apresentaram 1468 referências, sendo que 60% (n=877) são mencionadas uma única vez, 30% (n=451) são mencionadas duas ou mais vezes e 10% (n=136) não são mencionadas. O número de menções é de 2246, não considerando 139 menções não referenciadas. O ranking elaborado pelo estudo mostrou que dos 30 autores mais citados (sete ou mais citações), 20 estão entre os mais mencionados. Conclui-se que há um núcleo de autores que são mais mencionados e que são basicamente os mais citados, e isso valida os estudos de citação a partir da lista de referências. Sugere-se a ampliação do recorte temporal desse teste, assim como a sua aplicação em outras áreas do conhecimento.

**Palavras-chave:** Análise de citação. Cientometria. Comunicação científica.

*presented 1468 references, being 60% (n=877) mentioned once, 30% (n=451) mentioned two or more times and 10% (n=139) not mentioned. The number of mentions is 2240, not considering 139 mentions not referenced. The ranking prepared by the study showed that of the 30 most cited authors (seven or more citations), 20 are also the most frequently mentioned. The conclusion is that there is a core of authors who are most frequently mentioned and cited, and this evidence validates citation studies through the reference list. It suggests expanding the time range and the test application in other areas of knowledge.*

**Keywords:** Citation analysis, Scientometrics, Scientific communication.

## 1. Introdução

Os estudos de citação são necessários para entender o comportamento da produção e do uso de informação na ciência. Utilizar outros documentos para gerar conhecimento e apresentar uma lista desse material é elemento indispensável na publicação de pesquisas.

Tradicionalmente realizados a partir da lista de referências, os estudos de citação e os seus derivados (como análises de cocitação) buscam, cada vez mais, entender o contexto da citação, no sentido de diminuir as limitações dos estudos tradicionais. Assim, pesquisas utilizando o texto completo têm surgido com mais frequência, como apresentam Ding et al (2013), Liu, Zhang e Guo (2013), Song e Kim (2013), Strotmann e Bleier (2013), Zhang, Ding e Milojevic (2013), Jeonga, Songa e Ding (2014), Silveira e Caregnato (2014), para citar alguns exemplos.

O presente trabalho tem como tema os estudos de citação e busca responder dois objetivos, que relacionam as referências bibliográficas apresentadas na lista e a menção desses documentos no texto completo. A questão norteadora é muito simples: existe a tendência de mencionar um documento referenciado mais de uma vez no texto completo? Para tanto, os objetivos são:

- i) Identificar os padrões existentes na relação entre as referências e as menções na área da Recuperação da Informação, através da análise de artigos publicados em periódicos acadêmicos indexados na base de dados BRAPCI; e
- ii) Comparar um *ranking* de citação tradicional de autores pessoais com um *ranking* de menções.

Os objetivos apresentados serão atendidos a partir de duas unidades de contagem, a referência bibliográfica, que representa o documento, e os autores pessoais, respectivamente. Assim, referências com entradas por título, entidade coletiva e jurisdição não serão aproveitadas para atender o segundo objetivo do trabalho. A contagem dos autores inclui todas as autorias das referências e isso implica que as relações de coautoria apresentem uma contagem que não reflète o número de documentos referenciados.

Vale salientar ainda que “menção” nesse trabalho é entendida como o uso efetivo, no texto de um artigo, de um documento presente na sua lista de referências, ou seja, um documento citado (ou referenciado) deve ser mencionado no texto ao menos uma vez para caracterizar a citação. Sendo esta identificada, principalmente no texto completo, pela junção da entrada da referência (p.ex. o sobrenome do autor pessoal) e da data de publicação do documento. Essa forma de nomear a relação citação/referência com citação/menção é utilizada por Silveira e Caregnato (2014), já Ding et al (2013)<sup>3</sup> utilizam citação para uma menção e recitação para outras menções da mesma referência. Nenhum dos trabalhos pretendeu discutir questões terminológicas que envolvem a análise de citação, assim como esse texto.

## 2. Aspectos conceituais e trabalhos relacionados

A citação é uma das formas de fazer ciência e um indicador que ajuda na sua avaliação. Os estudos mais comuns de análise de citação utilizam os autores como elemento de contagem (documentos e periódicos são outros exemplos), indicando que os mais citados são os mais reconhecidos, independente da forma como o contexto do documento citante aborda essa relação. Reitz (c2014) e Cunha e Cavalcanti (2008) definem a citação como uma menção à opinião ou argumento de uma autoridade sobre algum assunto. Nesse sentido, a citação deve, também, indicar o “caminho” para a verificação dessa menção: a referência bibliográfica. Assim, pode-se afirmar que a citação é algo mais amplo que a referência bibliográfica, mas é medida principalmente a partir dessa indicação.



Romancini (2010) diz que a dimensão básica e fundamental da citação é relacionar dois participantes de um circuito científico, “o par ‘autor (texto) citado/autor (texto) citante” (ROMANCINI, 2010, p. 20). Essa relação se mostra muitas vezes política (social) e assim “forçada” e conveniente, mas também se apresenta com caráter cognitivo (epistemológico), fazendo com que as citações sejam utilizadas para comparar métodos e resultados de pesquisa, posições teóricas e de pioneirismo em determinado campo científico etc. Até a não citação é um aspecto que pode ser estudado e que mostra o caráter competitivo e político da ciência.

O certo é que, somente estudando as listas de referências é mais difícil perceber algumas facetas que caracterizam o processo de citar, como as motivações, a relação com outros documentos citados, se o contexto menção é para comparar resultados ou uma discussão teórica, por exemplo. Nesse sentido, alguns estudos merecem destaque.

Brambilla, Vanz e Stumpf (2006) mapearam, através da análise de conteúdo, as razões de 24 citações de um artigo de um pesquisador da Universidade Federal do Rio Grande do Sul – UFRGS e concluíram que todas foram focadas no uso de métodos práticos ou teóricos para a solução de problemas.

Ding et al (2013) apresentam resultados sobre a distribuição das menções (eles utilizam citação) das referências na estrutura textual dos artigos de um periódico, pois questionam que a citação pela lista de referência não deixa claro o crédito ou reconhecimento ao documento citado, além de igualar a contribuição de todos os documentos listados nas referências. Os autores (DING et al, 2013) indicam que a maior parte das menções ocorre nas seções de introdução, revisão de literatura e metodologia dos artigos e que documentos recitados são mais significativos para os documentos citantes.

Jeonga, Songa e Ding (2014) apresentam uma análise de cocitação de autores (ACA) baseado em conteúdo, mostrando que os pares formados a partir das menções são mais específicos e favorecem uma análise mais precisa, pois há maior detalhamento de subdomínios do que os resultados oriundos de uma ACA tradicional.

Silveira e Caregnato (2014) propõem um estudo aplicado para a classificação das razões que levam a citação em comunicações orais e indicam que 93% das menções expressam a dimensão conceitual da citação, ou seja, “as razões das citações [...] se concentram,

essencialmente, para fins de sustentação e exemplificação da argumentação necessária para justificar as contribuições, como também para a construção do arcabouço teórico e metodológico” (SILVEIRA; CAREGNATO, 2014, p. 3627). Esse trabalho conta apenas as menções e não faz uma comparação com dados de lista de referência.

Zhao, Cappello e Johnston (2017) apresentam um estudo que, assim como esse, compara a relação quantitativa de referências e menções (mas não usa esse termo). O objetivo do texto é ponderar a contagem de citações, identificando documentos não essenciais em artigos citantes, mas a partir da localização da menção no texto, já que concluem que essa abordagem é mais eficiente do que a análise de recitação (menção) e remoção de menções únicas (ZHAO; CAPPELLO; JOHNSTON, 2017).

É evidente que os estudos de citação a partir das referências são necessários e devem continuar a ser realizados, pois, como indicam Vanz e Caregnato (2003), possibilitam identificar características e mapear a comunicação científica, além de servirem para atividades de gestão para políticas científicas. Mas o avanço em técnicas que transpõem essa tendência é uma realidade, ainda que muitos estudos sejam de caráter metodológico (JEONGA; SONGA; DING, 2014; SILVEIRA; CAREGNATO, 2014; ZHAO; CAPPELLO; JOHNSTON, 2017) e experimental, inclusive esse.

Assim, o trabalho quer mostrar se há diferenças quantitativas e significativas entre referências e menções e ampliar as possibilidades para *rankings* de citações feitos a partir das referências, como também, deixar à disposição da comunidade dados empíricos para futuras comparações.

### 3. Procedimentos metodológicos

O trabalho se caracteriza por ser empírico, cientométrico e exploratório. O corpus do estudo foi constituído de 66 artigos publicados entre 2008 e 2012 (cinco anos), recuperados na base BRACPI, nos meses de fevereiro e março de 2015. . Documentos publicados entre 2013 e 2014 não foram incluídos, pois apresentaram poucos resultados, indicando que números atrasados de alguns periódicos ainda não estavam na base.

As expressões que seguem foram utilizadas, como frase (entre aspas), para realizar as buscas com ocorrência dos termos no “título” e nas “palavras-chave”: a) RECUPERAÇÃO DA INFORMAÇÃO; b) RECUPERAÇÃO DE INFORMAÇÃO; c) RECUPERAÇÃO INTELIGENTE DA INFORMAÇÃO; d) RECUPERAÇÃO INTELIGENTE DE INFORMAÇÃO; e) RECUPERAÇÃO DE DADOS; f) ARMAZENAMENTO DE DADOS; g) ARMAZENAGEM DE DADOS; h) COLETA DE DADOS; i) BUSCA DE INFORMAÇÃO; j) BUSCA DA INFORMAÇÃO; k) INFORMATION RETRIEVAL; e, l) RETRIEVAL OF INFORMATION.

Definido o *corpus* foi gerada uma planilha de dados com os indicadores de caracterização e os primeiros resultados da pesquisa. A matriz foi formada por 66 casos (artigos) e as seguintes variáveis: ano de publicação, periódico, tipo de artigo, metadado (título e palavra-chave) de ocorrência do termo de busca, número de autores, número de referências (separando as com entradas de autoria pessoal das outras e identificando as que tinham apenas um autor das que tinham dois ou mais), número total de menções no corpo do texto, número de referências com uma menção, número de referências com duas menções, número de referências com três ou mais menções, referências não mencionadas e menções não referenciadas. A tabela 1 apresenta os principais indicadores de caracterização do *corpus*.

**Tabela 1. Corpus da pesquisa, por ano de publicação e título do periódico**

REVISTAS	ANO DE PUBLICAÇÃO					Total por revista (%)
	2008	2009	2010	2011	2012	
Informação e Sociedade	01	-	03	01	-	05 7.58
AtoZ: novas práticas em informação e conhecimento	-	-	-	01	-	01 1.52
Biblionline	-	-	03	-	03	06 9.09
Biblos	-	02	01	01	-	04 6.06
Brazilian Journal of Information Science	-	01	01	-	-	02 3.03
Ciência da Informação	01	01	-	-	-	02 3.03
DataGramaZero	-	03	-	03	01	07 10.61
Em Questão	-	-	-	01	-	01 1.52
Encontros Bibli	03	-	-	02	02	07 10.61
InCID: Revista de Ciência da Informação e Documentação	-	-	-	02	01	03 4.55
Informação & Informação	01	01	01	01	-	04 6.06
Perspectivas em Ciência da Informação	01	02	-	03	01	07 10.61
Perspectivas em Gestão & Conhecimento	-	-	-	01	-	01 1.52
Pontodeacesso	-	-	01	-	02	03 4.55
Revista Brasileira de Biblioteconomia e Documentação	01	-	-	-	-	01 1.52
Revista Digital de Biblioteconomia e Ciência da Informação	-	-	01	-	-	01 1.52
Revista ACB	02	01	01	-	01	05 7.58
Revista Ibero-Americana de Ciência da Informação	-	-	-	01	-	01 1.52
TransInformação	-	-	03	01	01	05 7.58
Total por ano (%)	10 15.15	11 16.67	15 22.73	18 27.27	12 18.18	66 100.00

Fonte: Dados da pesquisa, 2015

Depois da recuperação dos artigos, as referências foram manualmente colocadas em arquivos de texto apenas com os autores (ou entradas das referências) e a data de publicação. A partir desses dados foi realizada a contagem manual das menções e elaborado o *ranking* de citação de autores do *corpus* (299 autores com duas ou mais citações). Não foi possível localizar a fonte de duas referências elaboradas em coautoria, com a indicação “et al”, e assim apenas o primeiro autor entrou no *ranking*. Os dados foram normalizados com a verificação de autores com o mesmo sobrenome e iniciais próximas. A figura 1 exemplifica a forma de organização dos dados das referências.

A contagem de menções dos primeiros e únicos autores, das referências com entradas por título, entidade coletiva e jurisdição atendem o primeiro objetivo específico do trabalho e os resultados são apresentados na Tabela 2 e na Figura 2.

Figura 1. **Exemplo de organização dos dados retirados de cada artigo**

Autoria	Menção
1. AUTOR, A	02
AUTOR, B	
AUTOR, C	
AUTOR, D	
AUTOR, E	
2. AUTOR, C	01
3. ENTIDADE COLETIVA A	01
4. JURISDIÇÃO A	06
Total de menções no ARTIGO A	10

Fonte: Dados da pesquisa, 2015.

Além do indicador tradicional de citação de autores e do número de menções de cada autor, mais dois indicadores foram elaborados para ajudar na comparação, como apresentado no Quadro 1, que considera as primeiras 30 posições do *ranking* elaborado. Segue abaixo a descrição dos quatro indicadores:

- Número de citações: ocorrência do autor na lista de referências (com e sem autocitação);
- Número de menções: ocorrência do autor no corpo do texto (a menção é pelo documento, que mencionado duas vezes e com dois autores, p.ex., acarreta duas menções para cada autor);
- Número de artigos citantes: documentos que citam os autores, no caso desse trabalho são os 66 artigos do corpus (esse dado ajuda a entender a distribuição das citações e menções, uma vez que há casos de autores muito citados ou mencionados, mas em poucos artigos citantes); e
- Pico de menção de um único documento referenciado por autor: o maior número de menção de um único documento de um autor em todo o corpus (informação que ajuda a entender a distribuição das menções nos artigos citantes).

Foram medidos os coeficientes de correlação de Spearman<sup>4</sup> e Pearson<sup>5</sup> para as quatro categorias no software SPSS<sup>6</sup>, considerando nível de significância igual a 0,05. Esses testes servem para testar a significância da relação entre variáveis intervalares (BECKER, 2015).

## 4. Resultados e discussão

Os resultados iniciais mostram que existe o comportamento de mencionar mais de uma vez um documento referenciado (Tabela 2), ainda que a maioria das referências seja mencionada uma única vez (n=877), dados proporcionalmente parecidos com os apresentados por Zhao, Cappello e Johnston (2017, p.60). As 1468 referências geraram 2240 menções, com média de

33,94, um pouco abaixo dos dados apresentados por Silveira e Caregnato (2014). Mas é interessante observar que, de 451 referências, as mencionadas duas ou mais vezes correspondem à 1363 menções.

Os dados discrepantes são as “menções não referenciadas” (n=63) e as “referências não mencionadas” (n=139), mas esses resultados não parecem ser uma tendência, uma vez que apenas a mediana das referências não mencionadas foi diferente de zero. A tabela abaixo apresenta os principais dados descritivos das variáveis relativas às referências.

**Tabela 2. Distribuição do número de referências e menções no corpus da pesquisa (N=66)**

CATEGORIAS	Totais	Medidas de centralidade			Medidas de dispersão		
		Média	Mediana	Moda	Desvio padrão	Mínimo	Máximo
Número de referências	1468	22.24	19	17	11.85	01	54
Referências com entrada de autoria pessoal*	1348	20.74	19	14	10.99	01	53
Referências com entrada por título, entidade coletiva ou jurisdição	120	1.82	00	00	3.16	00	16
Referências com UMA menção**	877	13.29	11.50	06	8.16	00	38
Referências com DUAS menções	239	3.62	03	02	3.19	00	13
Referências com TRÊS ou MAIS menções	212	3.21	2.50	00	3.07	00	13
Referências NÃO MENCIONADAS	139	2.11	01	00	4.06	00	21
Menções no texto completo***	2240	33.94	29.50	18	21.21	02	91
Menções não referenciadas	63	0.95	00	00	1.88	00	10

Obs: \* referências utilizadas para elaborar os rankings do segundo objetivo do trabalho | \*\* uma referência foi desconsiderada para a contagem das menções | \*\*\* não inclui as menções não referenciadas.

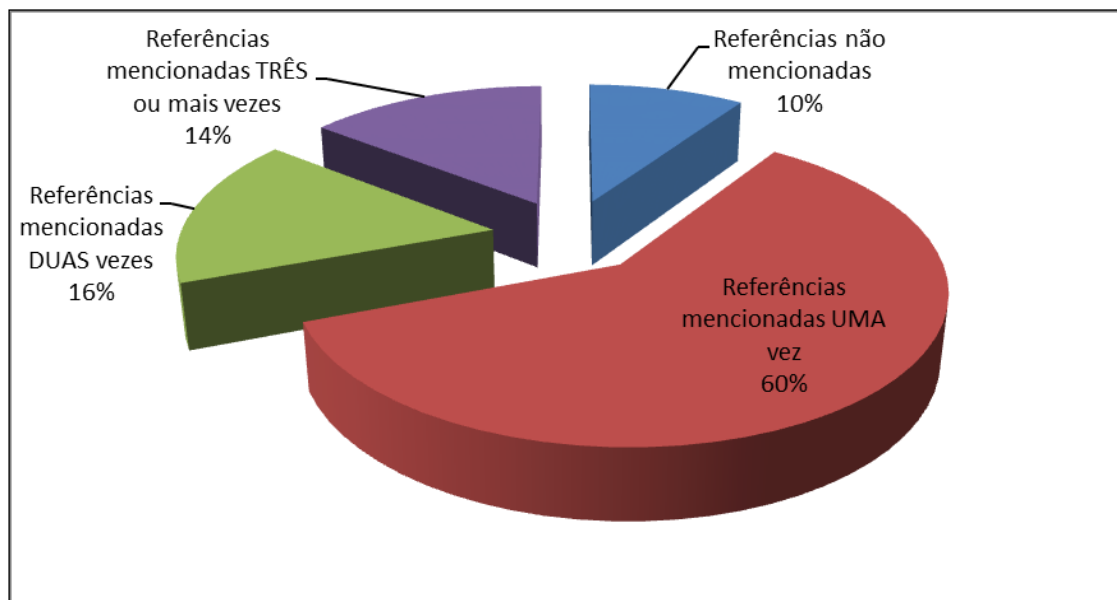
Fonte: Dados da pesquisa, 2015

A Figura 2 mostra que 30% das referências foram mencionadas mais de uma vez. Parece necessário estudar outros contextos (áreas de conhecimento e tipos de publicação) para verificar se isso é um comportamento. Voos e Dagaev<sup>7</sup> (1976 apud DING et al, 2013) sugerem que o número de menções de uma referência indica a relevância para o assunto do documento.

Os 10% de referências não mencionadas se apresentam como uma prova empírica que Romancini (2010) reivindica quando comenta a afirmação de Wouters sobre muitas referências não serem citadas. Essa proporção pode ser um caso isolado, mas indica que os editores precisam estar atentos para diminuir essa prática, ainda mais que os dados

descritivos apresentaram um desvio padrão de 4,06 e um valor absoluto máximo igual a 21. As boas práticas na publicação de pesquisas são fundamentais, além de ser uma questão ética. A responsabilidade é principalmente dos autores, mas também de editores e pareceristas.

**Figura 2. Proporção entre referências mencionadas**



Fonte: dados da pesquisa, 2015

Para atender o segundo objetivo do trabalho foi gerado um ranking com os pesquisadores mais citados e mais mencionados. A essas duas categorias juntaram-se “número de artigos citantes” e “pico de menção de um único documento referenciado”, como salientado na seção dos procedimentos metodológicos.

O ranking ficou com 299 autores (com duas ou mais citações). A Tabela 3 apresenta os coeficientes de correlação de Spearman e Pearson das variáveis. Ding e Cronin (2011) utilizaram o coeficiente de Spearman para mostrar a relação de três *rankings* em um estudo de citação e o coeficiente de Pearson foi realizado por ser o mais comum e robusto. A análise apresentou valor de significância igual à zero para todos os pares de correlação em ambos os coeficientes, o que, segundo Field (2009), é aceitável visto que precisa ser  $< 0,05$ .

Em relação às categorias, “número de citação” apresentou correlação alta com “número de artigos citantes” e “número de menções” (as três mais importantes). A categoria “pico de menções” apresentou correlação baixa com “número de citação” e “número de artigos citantes”, mas alta com “número de menções”. O comportamento foi parecido com os dois

coeficientes utilizados e mostrou que a variável “pico de menção” é a mais afastada, ainda que diretamente relacionada com o “número de menções”.

**Tabela 3. Distribuição das taxas de coeficiente de correlação de Spearman e de Pearson**

CATEGORIAS	Nº Citação	Nº Artigos citantes	Nº Menções	Pico de menções	
Nº Citação	1,000	0,858	0,873	0,521	Correlação de Pearson
Nº Artigos citantes	0,742	1,000	0,682	0,388	
Nº Menções	0,735	0,577	1,000	0,784	
Pico de menções	0,412	0,341	0,860	1,000	
Correlação de Spearman					
O valor de significância de todos os pares da correlação foi igual a zero (0,000), ou seja, menor que 0,05.					

Fonte: dados da pesquisa, 2017

A Tabela 4 apresenta os dados descritivos do *ranking* de autores considerando as três principais variáveis, e os valores de menções se destacam, indicando um padrão que merece estudos mais apropriados, levando em consideração os autores como unidade de contagem.

**Tabela 4. Medidas descritivas das variáveis: citação, artigos citantes e menção (n=299)**

Medidas	Nº Citação	Nº Artigos citantes	Nº Menções
Média	34.55	26.82	56.19
Mediana	20	20	30
Moda	02	02	02
Desvio padrão	2.75	1.78	6.87
Mínimo	02	01	00
Máximo	22	13	66

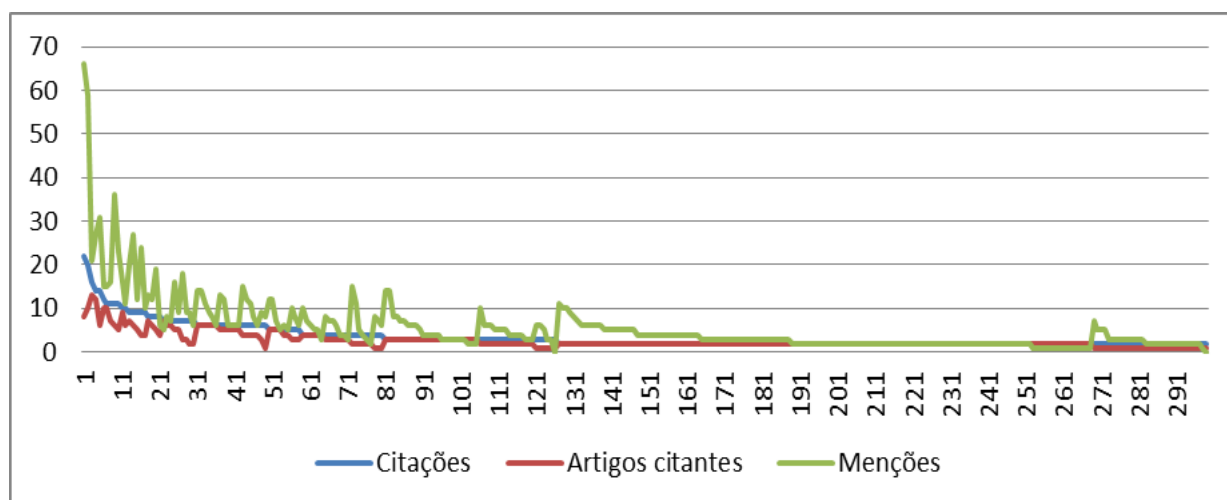
Fonte: Dados da pesquisa, 2015

Os valores “mínimos” baixos da Tabela 4 indicam a necessidade da escolha mais criteriosa de periódicos para futuros estudos. Não fazia sentido para esse trabalho remover os documentos que apresentam esses dados, uma vez que o ponto de partida era a base de dados BRAPCI e não os periódicos. Tratar os dados como *outliers* também não pareceu apropriado, uma vez que não são duvidosos (BECKER, 2015).

A Figura 3 mostra uma distribuição de frequência das três principais categorias relacionadas aos autores. A menção apresenta os maiores valores, e há picos, mesmo quando a citação e o número de documentos citantes são baixos. As referências não mencionadas deixam a linha da menção abaixo da linha dos artigos citantes, mostrando um ruído nos dados.



**Figura 3. Distribuição comparativa das variáveis: citação, artigos citantes e menção**



Fonte: Dados da pesquisa, 2017

No Quadro 1 são apresentados os autores mais citados do ranking elaborado (primeira coluna de nomes), com sete ou mais citações. As demais colunas apresentam os resultados das outras categorias. Os números apresentados após o nome dos autores indicam a posição em cada categoria. Assim, o autor “Saracevic [3,1,9,42]” foi o terceiro mais citado, o primeiro em número de artigos citantes, o nono em número de menções e 42º na categoria pico de menção.

Optou-se por colocar os autores nas 30 primeiras posições de cada categoria, ao invés de uma única lista de autores e suas posições em cada categoria, para evidenciar as diferenças. Autores que aparecem na coluna da categoria “pico de menção”, como “Moura [134,117,63,2]”, “Taylor [135,118,41,6]”, “Almeida Júnior [270,269,68,10]”, “Gadamer [136,119,42,16]” e “Stambaugh [137,120,48,17]” acabariam não aparecendo no quadro e isso diminuiria a percepção do alcance das menções. Na categoria menções, autores como “Heidegger [73,96,19,14]”, “Cox [81,69,22,4]” e “Hall [82,70,23,5]” exemplificam essa percepção.

**Quadro 1. Distribuição dos rankings de autores pessoais (30 posições)**

Número de citações (referências) – sem e com autocitação		Número de artigos citantes		Número de menções		Pico de menção de um único documento referenciado			
01	<u>Hjorland [1,7,1,1]</u>	22	22	Saracevic [3,1,9,42]	13	<u>Hjorland [1,7,1,1]</u>	66	Hjorland [1,7,1,1]	19
02	<u>Kuhlthau [2,3,2,8]</u>	20	20	Lancaster [4,2,5,19]	12	<u>Kuhlthau [2,3,2,8]</u>	59	Moura [134,117,63,2]	10
03	<u>Saracevic [3,1,9,42]</u>	16	16	Kuhlthau [2,3,2,8]	10	<u>Ellis [9,13,3,3]</u>	36	Ellis [9,13,3,3]	09
04	<u>Lancaster [4,2,5,19]</u>	14	14	Kobashi [6,4,16,103]	10	<u>Wilson [5,12,4,9]</u>	31	Cox [81,69,22,4]	09
05	<u>Wilson [5,12,4,9]</u>	14	14	Barreto [7,5,17,43]	10	<u>Lancaster [4,2,5,19]</u>	27	Hall [82,70,23,5]	09
06	<u>Kobashi [6,4,16,103]</u>	12	12	Souza [11,6,14,20]	09	<u>Capurro [14,15,6,7]</u>	27	Taylor [135,118,41,6]	09
07	<u>Barreto [7,5,17,43]</u>	11	11	Hjorland [1,7,1,1]	08	<u>Ingwersen [16,40,7,21]</u>	24	Capurro [14,15,6,7]	08
08	<u>Spink [8,8,13,44]</u>	11	11	Spink [8,8,13,44]	07	<u>Fujita [10,25,8,11]</u>	23	Kuhlthau [2,3,2,8]	07
09	<u>Ellis [9,13,3,3]</u>	11	11	Belkin [13,9,10,45]	07	<u>Saracevic [3,1,9,42]</u>	21	Wilson [5,12,4,9]	07
10	<u>Fujita [10,25,8,11]</u>	11	11	Alvarenga [18,10,24,12]	07	<u>Belkin [13,9,10,45]</u>	20	Almeida Júnior [270,269,68,10]	07
11	<u>Souza [11,6,14,20]</u>	10	11	Cunha [22,11,96,182]	07	<u>Ferreira [20,27,11,13]</u>	19	Fujita [10,25,8,11]	06
12	<u>Lima [12,14,32,104]</u>	10	10	Wilson [5,12,4,9]	06	<u>Järvelin [27,56,12,23]</u>	18	Alvarenga [18,10,24,12]	06
13	<u>Belkin [13,9,10,45]</u>	09	09	Ellis [9,13,3,3]	06	<u>Spink [8,8,13,44]</u>	16	Ferreira [20,27,11,13]	06
14	<u>Capurro [14,15,6,7]</u>	09	09	Lima [12,14,32,104]	06	<u>Souza [11,6,14,20]</u>	16	Heidegger [73,96,19,14]	06
15	<u>Crespo [15,26,26,60]</u>	09	09	Capurro [14,15,6,7]	06	<u>Choo [25,28,15,22]</u>	16	Rubi [74,97,35,15]	06
16	<u>Ingwersen [16,40,7,21]</u>	09	09	Jansen [19,16,27,46]	06	<u>Kobashi [6,4,16,103]</u>	15	Gadamer [136,119,42,16]	06
17	Campos [17,41,37,105]	09	09	Castells [23,17,49,107]	06	<u>Barreto [7,5,17,43]</u>	15	Stambaugh [137,120,48,17]	06
18	<u>Alvarenga [18,10,24,12]</u>	08	09	Marcondes [24,18,59,183]	06	Guarino [43,43,18,25]	15	Tibbo [139,122,67,18]	06
19	<u>Jansen [19,16,27,46]</u>	08	08	Baeza-yates [31,19,20,48]	06	Heidegger [73,96,19,14]	15	Lancaster [4,2,5,19]	05
20	<u>Ferreira [20,27,11,13]</u>	08	08	Ribeiro-neto [32,20,21,49]	06	Baeza-yates [31,19,20,48]	14	Souza [11,6,14,20]	05
21	Jardim [21,42,69,106]	08	08	Ferneda [33,21,33,61]	06	Ribeiro-neto [32,20,21,49]	14	Ingwersen [16,40,7,21]	05
22	Cunha [22,11,96,182]	07	07	Gil [35,22,46,24]	06	Cox [81,69,22,4]	14	Choo [25,28,15,22]	05
23	Castells [23,17,49,107]	07	07	Lara [32,23,50,62]	06	Hall [82,70,23,5]	14	Järvelin [27,56,12,23]	05
24	Marcondes [24,18,59,183]	07	07	Le Coadic [36,24,71,111]	06	<u>Alvarenga [18,10,24,12]</u>	13	Gil [35,22,46,24]	05
25	<u>Choo [25,28,15,22]</u>	07	07	Fujita [10,25,8,11]	05	Figueiredo [37,30,25,50]	13	Guarino [43,43,18,25]	05
26	Lévy [26,29,43,47]	07	07	Crespo [15,26,26,60]	05	<u>Crespo [15,26,26,60]</u>	12	Almeida [44,44,29,26]	05
27	<u>Järvelin [27,56,12,23]</u>	07	07	Ferreira [20,27,11,13]	05	<u>Jansen [19,16,27,46]</u>	12	Rowley [59,50,39,27]	05
28	McGuinness [28,57,44,108]	07	07	Choo [25,28,15,22]	05	Campello [38,31,28,63]	12	Buckland [65,62,54,28]	05
29	Maturana [29,94,45,109]	07	07	Lévy [26,29,43,47]	05	Almeida [44,44,29,26]	12	Sousa [78,260,55,29]	05
30	Gomes [30,95,70,110]	07	07	Figueiredo [37,30,25,50]	05	Baptista [50,36,30,51]	12	Mathes [83,71,56,30]	05

Fonte: dados da pesquisa, 2015

Os resultados mostram que dos 30 autores mais citados, 20 estão entre os mais mencionados (destaques em itálico e sublinhado nas colunas um e três). Esse resultado mostra uma evidência da importância e da confiabilidade dos estudos de citação realizados a partir das referências. Ou seja, estudar as menções de autores altamente citados parece relevante para identificar o que Zhao, Cappello e Johnston (2017) chamam de documentos essenciais, mas vale salientar que o número de documentos citantes melhora essa visualização.

O autor “Hjorland [1,7,1,1]”, p.ex., que é o mais citado no *corpus*, tem 22 citações de oito documentos citantes e 66 menções, e isso pode indicar que ele é essencial para alguns dos artigos que o citam. Mas o autor “Saracevic [3,1,9,42]” que tem 16 citações, em 13 documentos citantes, e 21 menções pode não ser essencial para muitos documentos citantes, apesar de ter um alcance maior de uso no *corpus* da pesquisa. Assim, estudar as menções dos autores em mais documentos citantes pode ser uma alternativa.

Os resultados dos autores “Moura [134,117,63,2]”, “Taylor [135,118,41,6]” e “Almeida Júnior [270,269,68,10]”, na categoria pico de menção, evidenciam a importância de relacionar as categorias, uma vez que os valores são altos, mas oriundos de apenas um documento citante. Ou seja, os autores apareceriam em um *ranking* de menções mais completo, mas sem alcance no *corpus* da pesquisa.

Os resultados dos autores “Baeza-Yates [31,19,20,48]” e “Ribeiro-Neto [32,20,21,49]” que aparecem em duas categorias no Quadro 1 (nº de artigos citantes e menção) evidenciam o cuidado que se deve ter com a unidade de contagem, pois esses valores são oriundos de um mesmo documento. Assim, fazer estudos de menções de documentos, identificar os documentos dos citados e a relação de coautoria dos autores melhor posicionados no *ranking* pode melhorar a visibilidade dos dados.

As menções podem oferecer perspectivas mais amplas ou perspectivas que as referências nunca tiveram efetivamente a função de esclarecer, independentemente da unidade contagem. Mas afinal, o que leva um autor a mencionar uma referência 19, 10 ou nove vezes, como mostram as três primeiras posições da quarta coluna do Quadro 1? Este é um questionamento importante em estudos de contexto de citação.

## 5. Considerações finais

Na produção brasileira sobre Recuperação da Informação (2008-2012) há o comportamento de mencionar um documento mais de uma vez, ainda que a maioria das referências seja mencionada uma única vez. As referências não mencionadas e menções não referenciadas aparecem como um complicador, que distorce levemente os dados e que pode ser resolvido com uma maior atenção das equipes que gerenciam as revistas da área da Ciência da Informação no Brasil.

Em relação aos autores pode-se afirmar que os mais mencionados são basicamente os mais citados, o que implica em uma das limitações do estudo, já que essa comparação não pode ser baseada apenas em dados absolutos. Entende-se que as categorias adicionadas, artigos citantes e pico de menção, melhoram a visualização dos dados.

As limitações tecnológicas parecem ser um impedimento para estudos com menções em larga escala, pois automatizar a extração em textos completos é mais complexo do que lidar com a lista de referências. Mas de qualquer forma, estudos com menção sugerem contexto de citação e métodos qualitativos para inferências mais precisas, o que exige *corpus* menor para análises mais qualificadas e precisas.

### *Limitações do trabalho*

O trabalho apresenta algumas limitações que merecem destaque:

- a utilização de duas unidades de contagem (documento e autoria) impede o aprofundamento maior de algumas questões, mas se justifica pelo caráter exploratório da pesquisa; e
- o número de artigos do *corpus* é pequeno, mas o trabalho manual exigiu essa condição, e lidar com citações amplifica muito o trabalho de análise.

### *Sugestões para futuras pesquisas*

Segue abaixo algumas sugestões para futuras pesquisas:

- um estudo teórico terminológico parece necessário para definir e consolidar alguns termos, como: menção, citação, referência e recitação;
- a ampliação do recorte temporal e um número maior de documentos pode apresentar dados mais robustos, ainda que a pesquisa seja replicada; e

- o estudo de aspectos éticos relacionados a boas práticas de produção do conhecimento, tanto pelos autores como pelos periódicos é necessário.

Espera-se que dados possam colaborar em pesquisas no campo da Cientometria, principalmente no contexto brasileiro.

## Referências Bibliográficas

---

- BRAMBILLA, S. D. S., VANZ, S. A. De S., STUMPF, I. R. C. (2006). Mapeamento de um artigo produzido na UFRGS: razões das citações recebidas. *Encontros Bibli: R. Eletr. Bibliotecon. Ci. Inf.*, 195-209. Retrieved from <https://goo.gl/frk4Zp>
- BECKER, J. L. (2015). *Estatística Básica: Transformando Dados em Informação*. Porto Alegre: Bookman.
- CUNHA, M. B., CAVALCANTI, C. R. de O. (2008). *Dicionário de Biblioteconomia e Arquivologia*. Briquet de Lemos: Brasília, DF.
- DING, Y., CRONIN, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. *Information Processing and Management*, 47, 80-86. doi: 10.1016/j.ipm.2010.01.002
- DING, Y., LIU, X., GUO, Chun, CRONIN, B. (2013) The distribution of references across texts: some implications for citation analysis. *Journal of Informetrics*, 7, 583-592. doi: 10.1016/j.joi.2013.03.003
- FIELD, A. (2009). *Descobrimo a estatística usando o SPSS*. (2nd ed). Porto Alegre: Artmed.
- JEONGA, Y., SONGA, M., DING, Y. (2014). Content-based author co-citation analysis. *Journal of Infor-metrics*, 8, 197-211. doi: 10.1016/j.joi.2013.12.001
- LIU, X., ZHANG, J., GUO, C. (2013). Full-Text Citation Analysis: A New Method to Enhance Scholarly Networks. *Journal Of The American Society For Information Science And Technology*, 64(9):1852–1863. doi: 10.1002/asi.22883
- REITZ, J. M. (2014). *ODLIS - Online Dictionary for Library and Information Science*. Retrieved from [http://www.abc-clio.com/ODLIS/odlis\\_c.aspx](http://www.abc-clio.com/ODLIS/odlis_c.aspx)
- ROMANCINI, R. (2010). O que é uma citação? Análise de citações na ciência. *Intexto*, 23(2), 20-35. Retrieved from <http://goo.gl/Jgeidh>
- SILVEIRA, M. A. A., CAREGNATO, S. E. (2014). Estudo das razões das citações na Ciência da Informação: proposta de classificação. *Proceedings of the Encontro Nacional de Pesquisa em Ciência da Informação*, Brazil, 15. Retrieved from <http://goo.gl/Q5MfKA>

- SONG, Min, KIM, S. Y. (2013). Detecting the knowledge structure of bioinformatics by mining full-text collections. *Scientometrics*, 96(1), 183–201. doi: 10.1007/s11192-012-0900-9
- STROTMANN, A., BLEIER, A. (2013). Author Name Co-Mention Analysis: Testing a Poor Man's Author Co-Citation Analysis Method. *Proceedings of the International Society of Scientometrics and Informetrics Conference*, USA, 14. Retrieved from <https://arxiv.org/abs/1309.5256>
- VANZ, S. A. de S., CAREGNATO, S. E. (2003). Estudos de Citação: uma ferramenta para entender a comunicação científica. *Em Questão*, 9(2), 295-307. Retrieved from <http://goo.gl/HdRvbB>
- ZHANG, G., DING, Y., MILOJEVIC, S. (2013). Citation Content Analysis (CCA): A Framework for Syntactic and Semantic Analysis of Citation Content. *Journal Of The American Society For Information Science And Technology*, 64(7), 1490–1503. doi: 10.1002/asi.22850
- ZHAO, D., CAPPELLO, A., JOHNSTON, L. (2017). Functions of Uni- and Multi-citations: Implications for Weighted Citation Analysis. *Journal of Data and Information Science*, 2(1), 51–69. doi: 10.1515/jdis-2017-0003

## Notas

---

<sup>1</sup> 1 Base de Dados em Ciência da Informação | Acervo de Publicações Brasileiras em Ciência da Informação.

<sup>2</sup> Texto originalmente publicado nos anais do XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB), no Brasil, em 2015, que passou por revisão em ampliação dos dados.

<sup>3</sup> “[...] cited/re-cited in the body of the text [...]”, p.591.

<sup>4</sup> “[...] medida padronizada da força do relacionamento entre duas variáveis que não depende das hipóteses de um teste paramétrico” (FIELD, 2009, p.643).

<sup>5</sup> “O coeficiente mede o grau de dependência linear entre duas variáveis” (BECKER, 2015, p.98).

<sup>6</sup> Statistical Package for the Social Sciences for Windows.

<sup>7</sup> Voos, H., & Dagaev, K. S. (1976). Are all citations equal? Or Did we op. cit. your idem? *J. of Acad. Librarianship*, 1, 20–21. [Artigo não localizado na Web e esse número do periódico não consta em nenhuma Biblioteca que faz parte do Catálogo Coletivo Nacional – CCN/IBICT]

# Altmetria no Brasil: estudo de citação e cocitação na base de dados BRAPCI

*Altmetrics in Brazil: study of citation and co-citation in BRAPCI database*

**Fernanda Bochi dos Santos**

Universidade Federal do Rio Grande do Sul  
[nandabochi@gmail.com](mailto:nandabochi@gmail.com)

**Gonzalo Rubén Alvarez**

Universidade Federal do Rio Grande do Sul  
[gonzalorubenalvarez@gmail.com](mailto:gonzalorubenalvarez@gmail.com)

**Rene Faustino Gabriel Junior**

Universidade Federal do Rio Grande do Sul  
[rene.gabriel@ufrgs.br](mailto:rene.gabriel@ufrgs.br)

**Ana Maria Mielniczuk de Moura**

Universidade Federal do Rio Grande do Sul  
[ana.moura@ufrgs.br](mailto:ana.moura@ufrgs.br)

## Resumo

O presente artigo aborda a influência dos autores nacionais e internacionais na literatura científica brasileira na temática 'altmetria', através de análise de citação e cocitação das publicações indexadas na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI). Com base em indicadores bibliométricos, a pesquisa analisou 14 artigos com 112 referências e 92 autores citados. Os resultados sugerem (1) predomínio dos pesquisadores americanos e europeus na rede de coautoria das obras citadas e na rede de cocitação dos autores citados, (2) baixa presença de pesquisadores nacionais no ranking de autores mais citados, (3) proximidade temática significativa na rede de cocitação advinda de coautorias e ligações conceituais entre autores nacionais e internacionais citados, com destaque para J. Priem (95%

## Abstract

*This paper discusses the influence of national and international authors in the Brazilian scientific literature on altmetrics, through the analysis of citation and co-citation of publications indexed in the Reference Database of Journals in Information Science (BRAPCI). Based on bibliometric indicators, the present study analyzed 14 articles with 112 references and 92 cited authors. The results suggest (1) the supremacy of American and European researchers in the network of co-authorship of the works cited and in the co-citation network of the authors cited, (2) the low presence of national researchers in the ranking of the most cited authors, (3) significant thematic proximity in the network of co-citation coming from co-authorships and conceptual links between cited national and international authors, with emphasis on J. Priem (95% of possible co-citations).*

das cocitações possíveis). Conclui que a pesquisa brasileira em altmetria indexada na BRAPCI é influenciada, majoritariamente, pelas correntes teóricas e metodológicas de autores internacionais pioneiros no estudo das métricas alternativas. *It concludes that the Brazilian research on altmetrics indexed in BRAPCI is influenced, mainly, by theoretical and methodological currents of pioneering international authors in the study of alternative metrics.*

**Palavras-chave:** Comunicação Científica, Estudos Métricos, Análise de Citação e Cocitação, Altmetria. **Keywords:** *Scientific Communication. Metric Studies. Citation and Co-citation Analysis. Altmetrics.*

## 1. Introdução

A comunidade científica, para o melhor desenvolvimento da pesquisa, faz uso de diferentes documentos com a finalidade de registrar e validar o que foi ou está sendo pesquisado. Esses documentos propiciam fundamentação teórica aos estudos e proporcionam aos cientistas a disseminação das suas ideias, que de acordo com Grácio, Oliveira e Matos (2009), é comandada pela sua capacidade de produção para desenvolvimento das pesquisas. O processo de construção do conhecimento científico realizado pelos pesquisadores, utilizando como base as pesquisas anteriores, é considerado subjetivo, dotado muitas vezes de motivos pessoais. Macias-Chapula (1998) reafirma isso ao dizer que as razões que levam um autor a citar outro, são muitas vezes para prestar homenagem, para apresentar a comunidade científica que conhece e domina a literatura da área, ou até mesmo, para ratificar o próprio trabalho, fazendo uma espécie de autopromoção.

Na percepção de Grácio (2016), as listas de referências que aparecem no final dos trabalhos publicados podem ser analisadas como reflexo de uma comunidade científica discursiva, caracterizada pelo diálogo e pela proximidade temática entre os autores que citam e os autores que são citados. Para isso, Grácio e Oliveira (2013) consideram necessário o estudo de citação e cocitação, uma vez que o primeiro proporciona visualizar mais claramente as temáticas abordadas e os autores que abordam tais assuntos, e o segundo permite visualizar a frequência com que esses autores aparecem concomitantes em determinada pesquisa.

Partindo das argumentações anteriores, observa-se que os estudos métricos de citação e cocitação apresentam claramente uma familiaridade entre pesquisador e temática, o que Grácio, Oliveira e Matos (2009, p. 81) chamam de “. . . similaridade de assuntos . . .” entre os citados, mostrando assim a proximidade de ideias desses autores e suas relações. Small (1973)



afirma que, para ser cocitado, um grande número de autores deve citar as obras anteriores. Sendo assim, o autor reforça a ideia de que a cocitação é uma relação que é estabelecida pelos autores citados. Small (1973) ainda diz que, ao medir a cocitação, medimos o grau de relacionamento ou associação entre os autores, ou até mesmo entre as instituições.

Neste novo cenário da pesquisa científica, onde o diálogo entre os pesquisadores ou entre a comunidade acadêmica e a sociedade é imediato, novas propostas de estudos métricos foram surgindo. Priem e outros colaboradores (2010) apresentam pela primeira vez o termo altmetria, também chamada de métricas alternativas, que tem como proposta avaliar e medir as atividades científicas em meio virtual. Apesar de ser uma temática relativamente nova, a altmetria, de acordo com Nascimento e Oddone (2015) tem sido adotada intensamente por diversos estudiosos da comunicação científica, uma vez que permite visualizar o impacto da pesquisa em redes sociais como Twitter, Facebook, Blogs, entre outros. Segundo Barros (2015) essa métrica objetiva compreender as características e o comportamento dos pesquisadores na web.

Para Priem e colaboradores (2010), os estudos métricos tradicionais negligenciam o impacto que os estudos causam fora da academia, e também ignoram o contexto e os motivos da citação. Tendo em vista que os cientistas estão deslocando os seus trabalhos para o mundo da web, a altmetria surge como uma alternativa para avaliar, validar e recomendar determinados estudos em tempo real. Contudo, segundo Nascimento e Oddone (2015), devido a contemporaneidade dessa temática ainda estar em desenvolvimento, destaca-se que são necessários mais estudos que mostrem as vantagens e desvantagens de seu uso.

Com o objetivo de ampliar o conhecimento que se tem sobre comunicação científica, e com auxílio de indicadores bibliométricos, a proposta deste estudo é realizar uma análise de citação e cocitação sobre a temática 'altmetria', na área da Ciência da Informação, a partir dos documentos indexados pela Base de Dados em Ciência da Informação (BRAPCI).

## **2. Referencial Teórico**

Na comunidade científica citar é imprescindível para ratificar a pesquisa realizada. Além de incluir parte da ideia de alguém, é fundamental que seja incluída a fonte de onde essa informação foi extraída, respeitando assim os direitos autorais. Gil (2010) legitima essa

afirmativa ao dizer que as ideias dos autores, quando mencionadas em um documento, devem ser seguidas da autenticação do autor que as produziu, assegurando os direitos legais do seu executor. Em função disso, a análise de citação torna-se uma ferramenta importante para a avaliação da produção científica. Contudo, Meadows (1999) lembra que as citações são excelentes medidas de qualidade quando corretamente construídas, do contrário, podem se tornar um problema sério para a pesquisa. Por sua vez, Ravichandra Rao (1986, p. 200) declara que os essenciais propósitos da análise de citação são: “. . . avaliar e interpretar as citações de artigos, autores, instituições e outros agregados das atividades científicas.” Freitas (1998) acrescenta ainda que a análise de citação tem a pretensão de medir a qualidade dos trabalhos científicos via dados quantitativos, baseando-se no impacto ou importância dada pela comunidade científica a alguns autores e seus trabalhos.

Vanz (2004) afirma que a análise de citação permite visualizar a relação citante e citado, proporcionando transparência nos assuntos abordados nas diversas áreas do conhecimento. Da mesma forma, Leydesdorff (1998) entende que as citações são referências para outros elementos textuais e que elas estabelecem relações entre pares (citados e citantes). O autor explica que uma rede de citações é modelada a cada momento e reproduzida ao longo do tempo. A partir da repetição do processo de citação, a rede adquire uma estrutura com condições suficientes para revelar características de uma comunidade científica.

Weinstock (1971) destaca quinze razões para uso da citação: 1) homenagear os pioneiros; 2) dar crédito para trabalhos relacionados; 3) identificar metodologia, equipamento etc.; 4) prover leitura de referência; 5) corrigir o próprio trabalho; 6) corrigir o trabalho de outros; 7) analisar criticamente estudos anteriores; 8) reivindicar publicações; 9) alertar investigadores para futuros trabalhos; 10) dar relevância para investigações pouco disseminadas, inapropriadamente indexadas ou não citadas; 11) autenticar informações e categorias de fatos, constantes físicas etc.; 12) identificar artigos originais nos quais uma ideia ou conceito foi discutido; 13) identificar publicações originais que descrevem termos epônimos como, por exemplo, Lei de Pareto, Doença de Hodgkin etc.; 14) contrapor ideias de pares; 15) discutir manifestações prioritárias de outros. Como observado, o número de citações depende de uma série de fatores sociais e hábitos dos pesquisadores, portanto, deve ser apreendido como um valor estimado e parcial do impacto e/ou da qualidade de uma publicação (Leta & Cruz, 2003).

Bavelas (1978) aprecia que o momento da escolha de trabalhos para citar envolve aspectos relacionados com o impacto acadêmico (uso do método, paradigma ou teoria do autor citado) e com aspectos sociopsicológicos (pressões sociais, interesses pessoais etc.).

Alguns autores como MacRoberts e MacRoberts (1989) relacionam uma série de problemas com relação aos estudos de citação em termos de influências formais e informais não citadas; citação tendenciosa; autocitação; diferentes tipos de citações; variações na taxa de citação relacionadas ao tipo de publicação, nacionalidade, período de tempo, tamanho e tipo de especialidade; limitações técnicas dos índices de citação e bibliografias (autoria múltipla, sinônimos, homônimos, erros de escrita, cobertura da literatura). Outro aspecto abordado por Garfield (1979) é o fenômeno da obliteração, que ocorre quando o trabalho de um cientista se torna tão integrado no corpo de conhecimentos do campo, que as pessoas frequentemente negligenciam a citá-lo explicitamente. Apesar dessas limitações, análises de citação são eficientes ferramentas para estimar a contribuição que os resultados de atividade ou produção científica tiveram para um campo ou área de conhecimento (Romancini, 2010). Vanz e Caregnato (2003) manifestam que o uso de indicadores bibliométricos baseados na contagem de citações permite conhecer como acontece a comunicação dentro de um campo científico e mapear teorias e metodologias consolidadas.

Da mesma maneira, Grácio e Oliveira (2013), em seus estudos, mostram que a análise de cocitação de autores é uma ferramenta importante para compreender e visualizar a frequência com que temáticas e autores aparecem em determinadas áreas do conhecimento. As autoras afirmam que tal estudo propicia compreender o grau de proximidade dos autores citados, a partir da similitude dos assuntos por eles abordados. Sendo assim, quanto maior for a frequência de cocitação, maior a relação temática dos citados.

Garfield (2001) afirma que a análise de cocitação tem um papel fundamental não apenas na área de Ciência da Informação, como também em outras áreas do conhecimento, uma vez que permite aos estudiosos mapear suas disciplinas, além das suas especialidades. Para o autor, esse estudo baseia-se principalmente na identificação de um conjunto de trabalhos citados simultaneamente. Spinak (1996), um dos pioneiros nesse estudo, reforça essas afirmações ao dizer que a análise de cocitação estuda as relações e a frequência com que os pares de documentos são conjuntamente citados em um terceiro documento. Ele ainda diz

que as redes de cocitação mais próximas podem ser descritas em mapas ou grafos, onde os nós denotam os documentos e as linhas representam as relações de cocitação.

Entretanto, analisar o desempenho e o desenvolvimento da ciência não é algo fácil. Por essa razão, pesquisadores fazem uso de diferentes ferramentas e metodologias, visando compreender como e onde estão sendo abordados determinados assuntos nas diversas áreas de conhecimento. Santos e Kobashi (2009) afirmam que esse desafio de avaliar qualitativamente a informação, foi contrabalanceado com a criação de métodos e técnicas de tratamento, análise e visualização dessa informação, fundamentado em princípios estatísticos como: a bibliometria, a cientometria, a informetria e webometria. Embora esses termos sejam congêneres, como afirmam os autores Bufrem e Prates (2005), Job (2006) e Vanti (2002), suas aplicabilidades são diferentes, uma vez que cada um propõe aferir a disseminação do conhecimento sobre perspectivas diferentes. Os estudos métricos são ferramentas que têm colaborado para o aperfeiçoamento das produções científicas nas diferentes áreas. Porém, Nascimento (2016) aponta que, com as novas formas de disseminação da informação científica, foi necessário pensar e criar novos métodos para avaliar essas produções, visto que, com as tecnologias, antes mesmo de serem citados, os trabalhos científicos já foram lidos, curtidos, compartilhados e disseminados.

Priem et al (2010), na tentativa de analisar o desempenho da ciência na web, tornam-se os primeiros a abordar os estudos métricos voltados para a web social com o termo altmetrics. Priem e os demais autores (2010) asseveram que a contagem de citação, o índice h e a revisão por pares, embora úteis, não conseguem mensurar o impacto que a literatura científica está causando fora do ambiente acadêmico e de pesquisa. Dessa forma, surgem, intrinsecamente, os estudos métricos voltados para as mídias sociais.

A altmetrics, também denominada de métricas alternativas ou simplesmente altmetria, vem complementar os estudos métricos já existentes, avaliando a disseminação da produção científica nos diferentes suportes midiáticos. Alguns autores a descrevem como um subconjunto da Cientometria e Webometria, uma vez que trabalha com dados em ambientes online. Porém, as 'métricas alternativas' trouxeram a proposta de um novo olhar para as revisões por pares e fator de impacto, permitindo à pesquisa científica percorrer um ecossistema mais diversificado, saindo do âmbito dos filtros tradicionais e trabalhando com

indicadores fundamentados na web social. Souza e Almeida (2013) afirmam que, com a altmetria é possível medir quantas vezes determinados documentos foram citados em redes como Facebook, Twitter, Blogs, entre outros.

No entendimento de Haustein, Bowman e Costas (2016), a altmetria, como disciplina, pode ser compreendida como um novo conjunto de métricas baseadas em eventos que acontecem nas diferentes mídias sociais e que estão relacionados com a comunicação de resultados acadêmicos. Métricas com base em “vestígios” de uso e produção de resultados derivados de investigações científicas em plataformas de mídia social podem contribuir notoriamente para o desenvolvimento de instrumentos de avaliação da pesquisa mais eficientes (Priem et al., 2010). Nesse contexto, a inclusão das mídias sociais, na avaliação da ciência produzida nas instituições acadêmicas e centros de pesquisa por meio de indicadores de impacto, abriu um novo canal de comunicação e discussão de resultados científicos entre pesquisadores. Os estudos altmétricos podem ser considerados uma ferramenta importante para analisar a comunicação e as relações entre pesquisadores e pesquisadores-sociedade em geral. Por outro lado, como ainda é uma abordagem recente no Brasil, torna-se imprescindível compreender como essa temática vem sendo explorada em nosso país. Para isso, é necessário utilizar as métricas tradicionais, tais como análise de citação e cocitação, já que elas permitem visualizar onde, como, quem e com que frequência tal temática está sendo investigada.

### **3. Metodologia**

A pesquisa desenvolvida caracteriza-se como um estudo bibliométrico, com foco em uma análise de citação e cocitação de autores no domínio temático “altmetria” a partir da perspectiva brasileira. A coleta de dados foi realizada em 07 de julho de 2017 na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), utilizando uma estratégia de busca parametrizada pelo termo “altmet”, sem delimitação temporal.

O corpus referente à produção científica indexada na BRAPCI, com a temática “altmetria”, constitui-se de 14 registros bibliográficos. Após a coleta da totalidade dos documentos, organizaram-se as referências extraídas de cada um dos registros recuperados em ordem alfabética de autores. Inicialmente, foram identificadas e organizadas em uma planilha Excel, 377 referências dos 14 documentos recuperados. A partir desse procedimento, foram selecionados todos os trabalhos que abordavam o assunto “altmetria”, restando 112

referências para análise de citação. Da mesma maneira, para a análise de cocitação foram identificados todos os autores citados/referenciados nos 14 artigos analisados, considerando-se somente as citações/referências que tinham como foco temático ‘altmetria’. No total, identificaram-se 92 pesquisadores referenciados no campo da ‘altmetria’ no conjunto de documentos para análise, considerando todos os coautores dos artigos. Além das ocorrências dos artigos de origem, foram relacionadas também as coautorias nos trabalhos citados. Com base nessas informações, gerou-se a matriz de cocitação quadrada simétrica 92 x 92, com as frequências absolutas de cocitação entre os autores citados. Para facilitar a identificação das cocitações, a matriz construída em uma planilha Excel foi padronizada, indicando com 1 quando houve cocitação e com 0 quando não houve cocitação.

Com relação ao tratamento dos dados, foram utilizados os softwares Microsoft Excel para organização das informações e geração de tabelas para as análises descritivas, Pajek para geração das redes de coautoria das obras citadas e Ucinet para geração da rede de cocitação dos autores referenciados.

#### **4. Análise dos Resultados e Discussão**

A produção científica no domínio temático “altmetria” indexada na BRAPCI é constituída por 14 artigos. A Tabela 1 apresenta a relação dos 24 autores que foram mais citados nos documentos, com destaque para o pesquisador J. Priem dos Estados Unidos, que recebeu 34 citações, 18,28% da totalidade do número de citações (n=24).

O pesquisador Jason Priem da University of North Carolina at Chapel Hill dos EUA foi quem recebeu o maior número de citações nos artigos da BRAPCI, sendo um dos defensores da ‘altmetria’ como disciplina (Priem et al. 2010). Na publicação *Altmetrics: a manifesto* de 2010 é cunhado pela primeira vez o termo ‘altmetrics’, estando relacionado com a comunicação científica e as métricas alternativas, como um complemento às tradicionais métricas de citações para avaliar o impacto da pesquisa. Priem (2014) defende a ideia de que com a altmetric, aspectos invisíveis na comunicação – como ler, discutir e sugerir publicações científicas – deixam vestígios que podem ser coletados de maneira mais fácil e rápida do que com as citações, fornecendo assim uma alternativa para as métricas tradicionais. O autor mais citado tem diversos trabalhos em coautoria com o terceiro pesquisador mais citado nos

artigos indexados pela BRAPCI, B. Hemminger, do School of Information and Library Science da University of North Carolina at Chapel Hill dos EUA.

**Tabela 1 - Frequência de citações dos autores referenciados nas publicações sobre 'altmetria' indexadas na BRAPCI**

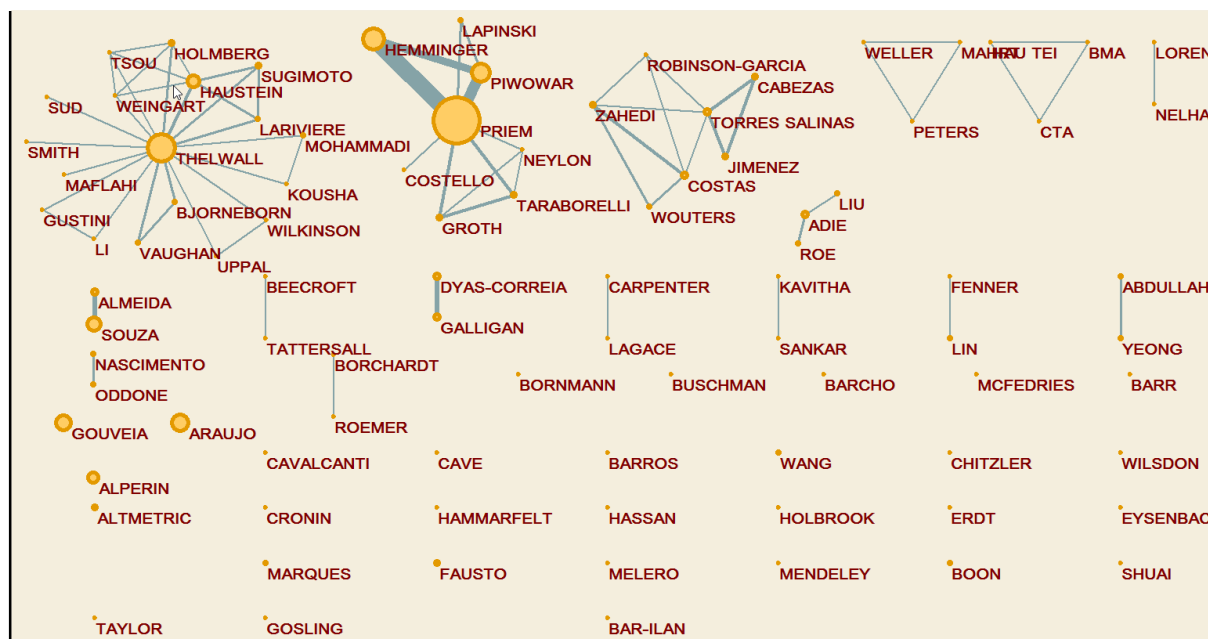
Autores mais citados	Nº de citações	%	Autores mais citados	Nº de citações	%
Priem, J. (EUA)	34	18,28	Adie, E. (Reino Unido)	4	2,15
Thelwall, M. (Reino Unido)	20	10,75	Galligan, F. (Reino Unido)	4	2,15
Hemminger, B. (EUA)	16	8,60	Almeida, C. (Brasil)	4	2,15
Piwowar, H. (EUA)	14	7,53	Cabezas, A. (Espanha)	3	1,61
Araújo, R. (Brasil)	12	6,45	Sugimoto, C. (EUA)	3	1,61
Gouveia, F. (Brasil)	11	5,91	Fausto, S. (Brasil)	3	1,61
Souza, I. (Brasil)	10	5,38	Jimenez, E. (Espanha)	3	1,61
Haustein, S. (Canadá)	9	4,84	Taraborelli, D. (EUA)	3	1,61
Alperin, J. (EUA)	8	4,30	Groth, P. (Holanda)	3	1,61
Torres Salinas, D. (Espanha)	5	2,69	Holmberg, K. (Finlândia)	3	1,61
Costas, R. (Holanda)	4	2,15	Altmetric	3	1,61
Dyas-Correia, S. (Canadá)	4	2,15	Zahedi, Z. (Holanda)	3	1,61

Mike Thelwall, com 20 citações recebidas, é o segundo autor mais citado nas publicações indexadas pela BRAPCI na temática 'altmetria'. Atualmente encontra-se filiado à instituição acadêmica School of Technology da University of Wolverhampton da Inglaterra, sendo um dos pesquisadores mais produtivos em assuntos relacionados com as métricas alternativas. H. Piwowar, da National Evolutionary Synthesis Center (NESCent) dos Estados Unidos, é o quarto colocado no ranking de autores mais citados (Tabela 1). Ronaldo Ferreira de Araújo, com 12 citações, é o pesquisador brasileiro que apresenta maior influência nessa temática no Brasil. Filiado à Universidade Federal de Alagoas, o professor Araújo tem se destacado pela contínua participação em eventos como, por exemplo, o Encontro Brasileiro em Bibliometria e Cientometria (EBBC), sendo um dos pesquisadores pioneiros na temática 'altmetria' no Brasil. O sexto lugar no ranking de autores mais citados pertence a outro pesquisador brasileiro, Fabio Castro Gouveia da Fundação Oswaldo Cruz do Rio de Janeiro, destacando-se pela sua

participação como coordenador de apresentações temáticas no âmbito das métricas alternativas em eventos como o IV EBBC de São Paulo. Iara Souza, da Universidade Federal do Rio de Janeiro, com dez citações, é a terceira pesquisadora brasileira mais influente na temática altmetria.

À continuação, na Figura 1 apresenta-se a rede de coautoria das obras citadas pelas publicações indexadas pela BRAPCI. Na rede de colaboração, os nós caracterizam uma comunidade científica representada pelos autores citados pelos documentos na BRAPCI com o assunto 'altmetria' e as ligações os trabalhos publicados em coautoria. De maneira a evitar a sobreposição, alguns nós foram moderadamente reposicionados dentro do grafo principal. A estrutura decorrente das ações e interações sociais entre 'atores sociais' está constituída por 92 autores, com destaque para o autor Jason Priem dos Estados Unidos como o mais popular dentro da rede de colaboração dos autores citados.

**Figura 1. Rede de coautoria das obras citadas pelas publicações sobre 'altmetria' indexadas na BRAPCI**



A respeito da densidade da rede de coautoria das obras citadas, observam-se três grupos (clusters) principais de indivíduos altamente relacionados, e diversos nós com apenas uma ou nenhuma ligação. Um dos clusters é formado por pesquisadores altamente citados e filiados a instituições norte-americanas (J. Priem, B. Hemminger, H. Piowar). Os dois primeiros pertencem à mesma instituição (University of North Carolina at Chapel Hill), confirmando a teoria de que a proximidade geográfica propicia a interação entre dois ou mais parceiros e

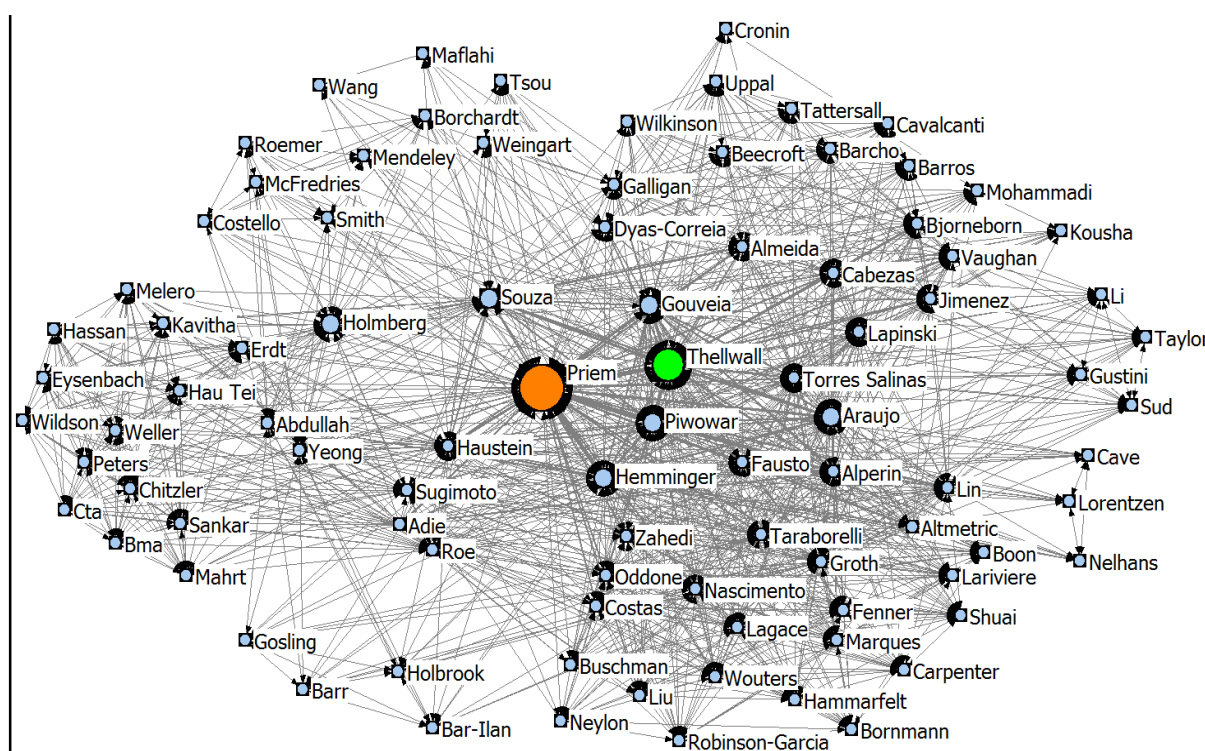


eleva as chances de publicar em coautoria (Katz & Martin, 1997). A cultura, a língua, os costumes e as questões sociopolíticas são fatores que afetam o nível de colaboração (Katz, 1994). Outro dos clusters detém o segundo autor mais citado nas publicações sobre altmetria indexadas na BRAPCI, M. Thelwall do Reino Unido, com um número considerável de conexões (coautores), destacando o grau de centralidade desse autor. O nível de agrupamento neste cluster possibilita o acesso à informação e o desenvolvimento das comunidades científicas. O posicionamento do pesquisador Thelwall é fundamental para a circulação e comunicação de resultados de pesquisas científicas e para o estabelecimento de parcerias com pesquisadores de diversas instituições interessados pelos mesmos assuntos: C. Sugimoto do School of Information and Library Science, Indiana University Bloomington (EUA), S. Haustein da École de bibliothéconomie et des sciences de l'information da Université de Montréal (Canadá). O terceiro agrupamento mais importante no grafo principal é constituído por pesquisadores filiados a instituições espanholas e holandesas com um número moderado de citações (Tabela 1). Nesse cluster destacam-se os autores D. Torres Salinas (Universidad de Navarra) e E. Jimenez, A. Cabezas (Universidad de Granada) da Espanha; R. Costas, Z. Zahedi, P. Wouters (CWTS-Leiden University) da Holanda. Pesquisadores brasileiros como R. Araújo e F. Gouveia com um número considerável de citações, 12 e 11 respectivamente, destacam-se por publicarem isoladamente seus trabalhos na temática 'altmetria', inferindo que as denominadas 'métricas alternativas' são um assunto incipiente no Brasil, não havendo ainda um grupo de pesquisa consolidado em nível nacional. A predominância de autores estrangeiros na rede de coautoria das obras citadas denota uma maior influência da pesquisa internacional na produção científica, com a temática 'altmetria', que se encontra indexada na base de dados BRAPCI.

A seguir na Figura 2, apresenta-se a rede de cocitação entre os 92 pesquisadores citados nas publicações sobre altmetria indexadas na BRAPCI. Mediante análise da rede gerada a partir das frequências absolutas de cocitação, percebeu-se uma maior proximidade temática entre aqueles autores que foram mais citados, os quais se encontram posicionados no centro do grafo. A composição estrutural da rede revela quatro grupos de autores citados, diferenciados pela totalidade do número de cocitações de cada um. O primeiro grupo identificado detém apenas um pesquisador, Jason Priem dos Estados Unidos, que tem cocitações com praticamente todos os demais autores citados, 86 de 91 cocitações possíveis. Desde o

surgimento do termo 'altmetria' em 2010, Jason Priem e outros colaboradores se preocuparam com a elaboração de fundamentos teóricos que sejam capazes de consolidar o campo altmetria como métrica alternativa no âmbito da comunicação científica. A importância da validação de novas métricas fundamenta-se nas deficiências detectadas pelo sistema tradicional de comunicação científica que determina a qualidade da ciência: lentidão da avaliação por pares, limitação da contagem de citações (restritas a publicações formais), fator de impacto suscetível a manipulações e distorções (Priem et al. 2010).

**Figura 2. Rede de cocitações entre os 92 pesquisadores citados nas publicações sobre 'altmetria' indexadas na BRAPCI**



O segundo grupo que se destacou pela quantidade de cocitações é composto por pesquisadores norte-americanos e ingleses: H. Piowar (66), B. Hemminger (64) e M. Thellwall (64). Os dois primeiros integrantes apresentam, especialmente com J. Priem, proximidades temáticas advindas de coautorias entre autores. Por outro lado, M. Thellwall destaca-se pela alta produtividade em temas relacionados com as métricas alternativas. Desponta também um terceiro grupo de cocitação, que inclui pesquisadores brasileiros precursores no assunto altmetria no âmbito local como F. Gouveia (51), R. Araújo (50) e I. Souza (50). S. Haustein (54) do Canadá e D. Torres Salinas (53) da Espanha completam esse seleto grupo. Finalmente, o

quarto grupo, constituído por indivíduos de diferentes instituições brasileiras e estrangeiras, apresenta-se com frequência menor que 50 ocorrências, distanciando-se dos outros grupos.

## 5. Considerações Finais

A proposta deste estudo baseou-se em uma análise de citação e cocitação a partir das publicações com a temática 'altmetria' indexadas pela Base de Dados em Ciência da Informação (BRAPCI). No ranking de autores mais citados, destaca-se a presença de pesquisadores norte-americanos e europeus pioneiros nos estudos altmétricos, com destaque para J. Priem, M. Thelwall, B. Hemminger, H. Piwowar. Esse indicador ressalta a interlocução e a aproximação temática entre os autores brasileiros que publicam sobre o assunto e os pesquisadores mais influentes e reconhecidos no âmbito internacional.

A baixa presença de pesquisadores nacionais no ranking de autores mais citados (R. Araújo, F. Gouveia, I. Souza) denota a incipiência do campo da altmetria no Brasil. Os três autores brasileiros mais citados (e também mais produtivos) são filiados a instituições públicas federais de ensino superior dos estados do Rio de Janeiro e Alagoas. Em relação à rede de cocitação, percebe-se uma maior proximidade temática advinda de coautorias e de ligações conceituais entre aqueles autores que foram mais citados no domínio da altmetria. Os autores posicionados no centro da rede são considerados, pelos autores citantes, os pesquisadores mais influentes e renomados da área, com destaque para J. Priem, que se relaciona, por meio das cocitações, com quase todos os demais autores citados, 95% das cocitações possíveis. A proximidade temática entre autores nacionais e internacionais observada na rede de cocitação revela também a forte similaridade de conteúdos produzidos, assim como o fortalecimento da altmetria como disciplina na comunidade científica identificada com os estudos métricos da ciência. Conclui-se que a pesquisa brasileira em altmetria indexada na BRAPCI é influenciada, majoritariamente, pelas correntes teóricas e metodológicas de autores internacionais pioneiros no estudo das métricas alternativas.

Esta pesquisa contribui para a identificação dos pesquisadores nacionais e internacionais referenciados nas publicações indexadas na BRAPCI, que constituem os fundamentos teóricos e conceituais para a pesquisa da comunidade brasileira no domínio da altmetria. Espera-se que este estudo bibliométrico colabore com subsídios elementares para um maior desenvolvimento das investigações vinculadas às métricas alternativas como campo de

estudo no Brasil, contribuindo para a visibilidade daqueles pesquisadores que têm se empenhado em compreender, teorizar e aplicar as técnicas altmétricas na avaliação do impacto da atividade científica fora da academia. Da mesma maneira, a análise de cocitação tornou possível observar como os autores, nacionais e internacionais, relacionam-se nessa temática.

Sugere-se, para futuras pesquisas, a inclusão de bases internacionais, como Web of Science (WoS) ou Scopus, para coleta de dados, visando uma compreensão mais ampla acerca dos expoentes teóricos e metodológicos referenciados pela literatura internacional e que suportam a pesquisa científica no contexto da altmetria.

## Referências Bibliográficas

---

- BARROS, M. (2015). Altmetrics: métricas alternativas de impacto científico com base em redes sociais. *Perspectivas em Ciência da Informação*, 20(2), 19-37. Doi: 10.1590/1981-5344/1782
- BAVELAS, J. (1978). The social psychology of citations. *Canadian Psychological Review*, 19(2), 158-163. Doi: 10.1037/h0081472
- BUFREM, L., & PRATES, Y. (2005). O saber científico registrado e as práticas de mensuração da informação. *Ciência da Informação*, 34(2), 9-25. Doi: 10.18225/ci.inf..v34i2.1086
- FREITAS, M. H. A. (1998). Avaliação da produção científica: considerações sobre alguns critérios. *Psicologia Escolar e Educação*, 2(3), 211-228, Recuperado de [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-85571998000300002&lng=en&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-85571998000300002&lng=en&nrm=iso&tlng=pt)
- GARFIELD, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, 1(4), 359-375. Doi: 10.1007/BF02019306
- GARFIELD, E. (2001). From bibliographic coupling to co-citation analysis via algorithmic historio-bibliography. Recuperado de <http://www.garfield.library.upenn.edu/papers/drexelbelvergriffith92001.pdf>
- GIL, A. C. (2010). *Métodos e técnicas de pesquisa social*. (6a. ed.). São Paulo: Atlas.
- GRÁCIO, M. C. C. (2016). Acoplamento bibliográfico e análise de cocitação: revisão teórico-conceitual. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 21(47), 82-99. Doi: 10.5007/1518-2924.2016v21n47p82

- GRÁCIO, M. C. C., OLIVEIRA, E. F. T., & MATOS, G. I. (2009). Visibilidade dos pesquisadores no tema Estudos Métricos: análise de citação e cocitação nos periódicos do SciELO. *IBERSID: revista de sistemas de información y documentación*, 3, 75-80. Recuperado de: <http://www.ibersid.eu/ojs/index.php/ibersid/article/view/3727/3488>
- HAUSTEIN, S., BOWMAN, T. D., & COSTAS, R. (2016). Interpreting “altmetrics”: viewing acts on social media through the lens of citation and social theories. In C. R. Sugimoto (Ed.), *Theories of informetrics and scholarly communication* (pp. 372-406). Frankfurt: De Gruyter.
- JOB, I. (2006). Análise bibliométrica das teses de uma comunidade científica em educação física com um uso do método indiciário. *Revista Brasileira de Ciência do Esporte*, 28 (1), 201-216. Recuperado de: <http://www.oldarchive.rbceonline.org.br/index.php/RBCE/article/view/47/55>
- KATZ, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31-43. Doi: 10.1007/BF02018100
- KATZ, J. S., & MARTIN, B. R. (1997). What is research collaboration? *Research Policy*, (26), 1-18. Doi: 10.1016/S0048-7333(96)00917-1
- LETA, J., & CRUZ, C. H. B. (2003). A produção científica brasileira. In: E. B. Viotti & M. M. Macedo. (Orgs). *Indicadores de Ciência, Tecnologia e Inovação no Brasil* (pp. 121-168). Campinas: Unicamp.
- LEYDESDORFF, L. (1998). Theories of citation? *Scientometrics*, 43(1), 5-25. Doi: 10.1007/BF02458391
- MACROBERTS, M. H.; & MACROBERTS, B. R. (1989). Problems of citation analysis: a critical review. *Journal of the American Society for Information Science*, 40(5), 342-349. Doi: 10.1002/(SICI)1097-4571(198909)40:5<342::AID-ASI7>3.0.CO;2-U
- MEADOWS, A. J. (1999). *A Comunicação Científica*. Brasília: Briquet de Lemos.
- NASCIMENTO, A. G. (2016). *Altmetria para bibliotecários: guia prático de métricas alternativas para avaliação da produção científica*. Porto Alegre: Revolução eBook.
- NASCIMENTO, A. G.; & ODDONE, N. E. (2015). Uso de altmetrics para avaliação de periódicos científicos brasileiros em ciência da informação. *Ciência da Informação em Revista*, 2(1), 3-12. Recuperado de: <http://www.seer.ufal.br/index.php/cir/article/view/1745/1260>
- PRIEM, J. Altmetrics. (2014). In B. Cronin & C. R. Sugimoto (Eds.). *Beyond bibliometrics: harnessing multidimensional indicators of performance* (pp. 263-287). Cambridge, MA: MIT Press.
- PRIEM, J., TARABORELLI, D., GROTH, P., & NEYLON, C. (2010). Altmetrics: A manifesto, 26 October 2010. Recuperado de: <http://altmetrics.org/manifesto>

- RAVICHANDRA RAO, I. K. (1986) Métodos quantitativos em biblioteconomia e ciência da informação. Brasília, DF: ABDF.
- ROMANCINI, R. (2010). O que é uma citação? a análise de citações na ciência. *In* *Texto*, 2(23), 20-35. Doi: 10.19132/1807-8583201023.5-17
- SANTOS, R. N. M.; & KOBASHI, N. Y. (2009) Bibliometria, cientometria, infometria: conceitos e aplicações. *Ciência da Informação*, 2(1), 155-172. Recuperado de: <http://inseer.ibict.br/ancib/index.php/tpbci/article/view/21/43>
- SMALL, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265-269. Recuperado de: [http://polaris.gseis.ucla.edu/gleazer/296\\_readings/small.pdf](http://polaris.gseis.ucla.edu/gleazer/296_readings/small.pdf)
- SOUZA, I. V. P., & ALMEIDA, C. H. M. (2013) Introdução à altmetria: métricas alternativas da comunicação científica. In: 14º Encontro Nacional de Pesquisa em Ciência da Informação. Florianópolis, SC.
- SPINAK, Ernesto. (1996). *Diccionario enciclopédico de Bibliometria, Cienciometria e Infometria*. Venezuela: UNESCO.
- VANTI, N. A. P. (2002). Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. *Ciência da Informação*, 31(2), 152-162. Doi: 10.1590/S0100-19652002000200016.
- VANZ, S. A. S. (2004). A produção discente em comunicação: análise das citações das dissertações defendidas nos programas de pós-graduação do Rio Grande do Sul. (Dissertação de Mestrado). Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- VANZ, S. A. S.; & CAREGNATO, S. E. (2003) Estudos de citação: uma ferramenta para entender a comunicação científica. *Em Questão*, 9(2), 295-307.
- WEINSTOCK, M. (1971). Citation index. In A. Kent & H. Lancour (Eds.), *Encyclopedia of Library and Information Science* (pp.19). (Vol. 5). New York: M. Dekker.

# Preservação de acervos digitais: um estudo bibliométrico<sup>i</sup>

*Preservation of digital archives: a bibliometric study*

**Raquel Oroski**

Colégio São Carlos, Caxias do Sul – RS  
[queloroski@gmail.com](mailto:queloroski@gmail.com)

**Maria de Fátima Santos Maia**

Universidade Federal do Rio Grande  
[mafas.maia@gmail.com](mailto:mafas.maia@gmail.com)

**Rodrigo Aquino de Carvalho**

Universidade Federal do Rio Grande do Sul  
[racfurg@gmail.com](mailto:racfurg@gmail.com)

## Resumo

Apresenta um estudo bibliométrico que indica características da produção científica sobre o tema preservação digital. Os dados foram coletados na base de dados LISA - Library and Information Science Abstracts, sendo identificadas 1.671 referências. Destas, foram identificados os autores mais produtivos assim como os seus países de origem e instituições de vínculo, a tipologia documental predominante e os periódicos onde se concentram maior número de referências. Os trabalhos identificados cobriram uma abrangência temporal de 30 anos (1986 – 2015) e 2009 foi o ano com maior número de publicações sobre preservação digital. É possível afirmar que o tema da preservação digital ganhou importância e destaque na produção acadêmica da área da ciência da informação a partir de 2007. A autora com maior número de registros foi Judy Duke, editora de uma revista de divulgação indexada na LISA, que trata de tecnologias aplicadas em bibliotecas. O país com maior quantidade de autores foi os EUA. No Brasil não houve destaque para autores. Os autores da base LISA têm média colaborativa de 1,6. O periódico

## Abstract

*This article presents a bibliometric study that indicates characteristics of the scientific production about digital preservation. The data were collected in the LISA (Library and Information Science Abstracts) database, in which 1.671 references have been identified. From these, the more productive authors have been identified, as well as their countries of origin and institutions of contact, the predominant documentary typology and the journals in which there is a higher number of references. The identified papers covered 30 years (1986-2015) and 2009 was the year with the highest number of publications about digital preservation. It can be affirmed it was in 2007 that this issue has been given importance on the academic production about Information Science. The author with the highest number of registrations was Judy Duke, an editor from a magazine indexed to LISA, which deals with technologies applied in libraries. USA was the country with the highest number of authors while Brazil had no emphasized authors. The ones from the LISA database have a collaborative average of 1,6. *Liber Quarterly: Journal of European Research Libraries* has*

científico que mais se destacou pelo número de referências sobre preservação digital foi o *Liber Quarterly: Journal of European Research Libraries*. A tipologia documental encontrada nas análises foi *Journal Article* (1.662), *Conference paper* (8) e *Magazine* (1). Verificou-se que a palavra “libraries” está presente em 967 trabalhos, destacando o tema da área da Biblioteconomia. Considera-se que este trabalho foi uma primeira imersão sobre o tema da produção científica sobre preservação digital, sendo válido dar continuidade a estas análises, incluindo buscas em outras fontes de informação da área da ciência da informação.

*been the most outstanding journal due to its number of references about digital preservation. The documentary typology found on the analysis was Journal Article (1.662), Conference paper (8) and Magazine (1). It was verified that the word “libraries” is present in 967 papers, giving emphasis to Library Science. The present article can be considered a first immersion about the scientific production about digital preservation and it is relevant to continue these analyses, including researching in other sources of Information Science.*

**Palavras-chave:** Preservação digital. Comunicação científica. Produção científica. Bibliometria.

**Keywords:** *Digital preservation. Scientific communication. Scientific production. Bibliometry.*

## 1. Introdução

Durante o século XX, o desenvolvimento tecnológico transformou significativamente a maneira das pessoas compartilharem informações, sendo que a criação da internet, no final da década de 90, está no centro destas mudanças. Sobre estas alterações, Wolton (2003, p.107) destaca que o rádio e a televisão impactaram nas relações sociais, entretanto estas mídias influenciaram principalmente a vida privada, enquanto a internet provocou transformações em todos os setores da sociedade, incluindo, entre outros, o lazer, os serviços, a educação e a busca de conhecimentos. Conforme Miranda e Farias (2009, p.3) “a internet representa hoje o maior repositório de informações disponíveis para qualquer pessoa que a acesse de qualquer parte do mundo”.

Conseqüentemente, estas transformações também têm impactado no crescimento da produção de documentos em formato digital que, de forma acelerada, começam a predominar os espaços dos acervos e arquivos pessoais, assim como repositórios informacionais de instituições e empresas.

Acrescenta-se ainda que, atualmente, as informações geradas de atividades cotidianas ou institucionais são, cada vez mais, armazenadas em formato digital, tornando real e importante a preocupação com a preservação desses materiais.



Além disso, a preservação destes acervos digitais exige mais atenção do que o formato em papel, pois estão “constantemente ameaçados pela fragilidade do suporte e pela obsolescência da tecnologia” (RONDINELLI, 2002 apud INNARELLI, 2012, p.12). Portanto, a preocupação com a preservação das informações disponíveis em formato digital é um assunto relevante, e que merece atenção destacada entre profissionais da ciência da informação.

Documento digital pode ser definido como toda a informação registrada, codificada em formato de dígitos binários, acessível e interpretável por computador. E a preservação digital consiste no conjunto de cuidados e procedimentos que asseguram a acessibilidade e integridade de documentos em formato digital, através da preservação física adequada do suporte; atualização tecnológica do formato e preservação da integridade dos seus dados e informações (BRASIL, 2007; CONARQ, 2004 *apud* FRANCO, 2008).

Partindo deste contexto, este trabalho investigou as características da produção científica sobre preservação digital no âmbito da área da ciência da informação, tais como, se a quantidade de publicações aumentou ou diminuiu ao longo do tempo; identificação dos principais autores que publicam sobre o tema; principais fontes; e também, se há mais trabalhos publicados na área de biblioteconomia ou arquivologia.

## **2. Contexto teórico**

A seguir apresentamos alguns tópicos considerados importantes para melhor contextualizar os resultados das análises.

### **2.1. Evolução dos registros de informações**

Até o predomínio da utilização do livro impresso em papel como principal suporte de informação, outras formas de armazenagem e difusão foram produzidas, tais como as tabuletas de barro, os rolos de papiro ou pergaminhos. O aumento na produção de livros impressos ocorreu, especialmente, a partir do século XVI, quando a quantidade de prensas mecânicas e tipografias se proliferaram na Europa (LYONS, 2011). Febvre e Martin (1992) afirmam que em 1501, na cidade de Paris, foram impressos 88 livros e em 1549 já ultrapassavam 300. Durante os séculos XVII e XVIII, uma tipografia europeia publicava, aproximadamente, 2.000 livros por ano, sendo que este movimento se ampliou sem cessar

(FEBVRE; MARTIN, 1992). Dados mais atuais apontam que, em 2016, o total de livros produzidos no mundo já ultrapassava 134 milhões.

Ao longo da história, os diversos formatos de materiais utilizados para registrar informações trouxeram diferentes formas para evitar problemas relacionados com a preservação. Neste contexto, destacam-se as informações em formato digital, que se não forem transferidas constantemente para formatos mais atualizados, podem ser perdidas pela inexistência de computadores e softwares adequados para acessá-las. Portanto, umas das principais preocupações dos pesquisadores da área é a questão da obsolescência tecnológica e a fragilidade das mídias, como afirma Rondinelli (2002).

Partindo deste contexto, é possível afirmar que os arquivos digitais são mais difíceis de preservar do que os em papel, pois as tecnologias de registro de informações se transformam rapidamente e os equipamentos, programas e redes precisam ser atualizados a intervalos frequentes (MEADOWS, 1999, p. 111).

No âmbito das informações científicas, desde a criação dos primeiros periódicos acadêmicos em 1665, o crescimento de publicações só aumentou. Além disso, a transferência do conteúdo para o formato digital se ampliou de maneira acelerada, pois o compartilhamento de resultados de pesquisas com maior amplitude e rapidez são uma grande vantagem em relação ao formato impresso.

Portanto, a preservação de informações em formato digital deveria ocupar posição central nos estudos e pesquisas da área da ciência da informação.

## **2.2. Preservação de documentos em formato digital**

Documento digital pode ser definido como toda informação registrada, codificada em dígitos binários, acessíveis e interpretáveis por computadores (SANTOS; FLORES, 2015). No meio digital, o suporte e o conteúdo são perfeitamente separáveis, o que permite que as informações sejam migradas para mídias atuais, sendo que essa migração só é possível quando o conjunto de dígitos binários é mantido exatamente igual (INNARELLI, 2012, p.23).

A preservação digital consiste na capacidade de garantir que a informação digital permaneça acessível, e com qualidades de autenticidade suficientes, para que possa ser interpretada no

futuro recorrendo a uma plataforma tecnológica diferente da utilizada no momento da sua criação. Segundo Hedstrom (1996 *apud* MARDERO ARELLANO, 2004), a preservação digital é um processo distribuído que envolve “o planeamento, alocação de recursos e aplicação de métodos e tecnologias para assegurar que a informação digital de valor contínuo permaneça acessível e utilizável”.

Os documentos digitais são perdidos com a mesma facilidade com que são produzidos, sendo que sua vulnerabilidade pode ser atribuída a diversos fatores, tais como os suportes usados para o registro da informação, que estão em constante evolução, ou pelos equipamentos e programas de computador utilizados. Innarelli (2012) destaca alguns problemas que podem acontecer no momento de migração, tais como, perda de dados, alterações da estrutura original dos documentos. O mesmo autor também aponta que é importante que haja controle de acesso aos documentos, permitindo somente usuários autorizados; que os metadados de preservação garantam que as informações funcionais do documento digital sejam preservadas; que sejam utilizadas ferramentas de migração adequadas e que as instituições estabeleçam rotinas eficientes (INNARELLI, 2012). Além disso, dependendo da estratégia escolhida pela unidade de informação, é importante que sejam considerados os aspectos econômicos e a compatibilidade com as necessidades da instituição.

Para um melhor entendimento do tema, a tabela abaixo mostra alguns conceitos considerados importantes sobre as diferentes estratégias de preservação de documentos em formato digital, suas características principais assim como suas desvantagens.

**Quadro 1. Características das estratégias de preservação digital**

ESTRATÉGIA	DEFINIÇÃO	DESvantagem
Migração	Transferência de um documento digital de um suporte obsoleto para outro mais atualizado.	Pode provocar alteração na cadeia de bits.
Emulação	Usa recursos computacionais para fazer uma tecnologia atual funcionar com as características de uma obsoleta.	O emulador também sofrerá obsolescência.
Encapsulamento	Reunião necessária de toda informação para a correta representação do documento.	Demanda maior espaço lógico de armazenamento.
Preservação da Tecnologia	Manutenção de todo o contexto tecnológico no qual o documento foi criado e armazenado.	Estratégia cara a longo prazo.

Refrescamento	Transferência da informação de um suporte físico de armazenamento considerado antigo para outro mais atual, sendo realizado antes da obsolescência do mesmo.	Deve ser auxiliado por outras estratégias de preservação.
---------------	--	---

**Fonte:** Adaptado de Santos (2015)

É importante ressaltar que a preservação digital não é resolvida pela própria tecnologia, e sim por estabelecimentos de políticas que auxiliem no processo de preservar. Tornando-se assim um processo complexo e recente, devendo ser tratado institucionalmente e de forma interdisciplinar, devendo ser de responsabilidade dos profissionais da informação garantir a preservação e manutenção do documento digital.

Assim, partindo deste breve contexto histórico e dos aspectos importantes que envolvem o tema da preservação de documentos digitais, apresentamos a seguir os procedimentos metodológicos utilizados para a realização das análises.

### 3. Procedimentos metodológicos

Como esta pesquisa é resultado de um trabalho de conclusão de curso de graduação, o tempo disponível é reduzido. Portanto, a primeira etapa consistiu em decidir qual a melhor fonte de dados para a realização da pesquisa que consistiu em realizar uma busca prévia sobre preservação digital em três diferentes bases de dados bibliográficas que indexam publicações da área da ciência da informação: WoS (Web of Science), SCOPUS e LISA (Library and Information Science Abstracts). Com este procedimento foi possível verificar a fonte que recuperaria maior número de referências, assim, utilizando a expressão “digital preservation” e limitando por estudos da área da ciência da informação, foram encontradas 311 referências na WoS; 568 na Scopus e 950 na LISA. A base de dados escolhida (LISA) reúne o conteúdo de mais de 440 periódicos da área da ciência da informação, publicados em mais de 60 países, incluindo o Brasil.

Após a definição da fonte de dados, consultou-se o tesauro da própria base para definição da sintaxe de busca que consistiu em utilizar “digitization” OR “preservation” OR “digital preservation”. Através desta estratégia, foram recuperadas 1.671 referências, imediatamente transferidas para o software gerenciador de referências bibliográficas EndNote X5. O uso desta ferramenta facilita na identificação de problemas de grafia, padronização, e permite

também, agrupar referências com características em comum, como a fonte ou ano de publicação.

Os campos utilizados nas análises e seus respectivos códigos foram: autores (AUT), ano das publicações (ANO), o vínculo institucional dos autores (VIN), o país informado no endereço de contato do primeiro autor (PAI), o idioma da publicação (IDI) e os títulos dos periódicos utilizados (PER); tipo de publicação (TIP); temática (TEM).

**Quadro 2. Descrição dos campos de análise conforme os objetivos específicos**

OBJETIVO	CAMPO	DESCRIÇÃO
Analisar se o tema da preservação digital tem aumentado ou diminuído na literatura da ciência da informação.	COD; ANO	Quantidade de publicações por ano
Identificar os autores mais produtivos no campo de estudos da preservação digital	COD; AUT	Total de trabalhos publicados
Identificar os autores brasileiros que mais publicaram sobre o assunto.	COD; PAI; AUT	Autores do Brasil que publicaram maior quantidade de trabalhos.
Verificar os países com maior e menor produção científica sobre o tema.	COD; PAI	Total de publicações por país
Identificar as instituições de vínculo dos autores mais prolíficos.	COD; AUT; VIN	Vínculo institucional dos autores com mais publicações.
Analisar a principal tipologia documental utilizada para divulgar estudos sobre a temática.	COD; TEM.	Total de trabalhos publicados conforme a tipologia (livros, resumos de eventos, etc.).
Examinar os periódicos que reúnem maior número de trabalhos sobre preservação digital.	COD; PER	Quantidade de registros conforme o título dos periódicos.
Examinar qual disciplina da ciência da informação - arquivologia, biblioteconomia - produz mais sobre o assunto.	COD; TEM	Quantidade de publicações conforme disciplina.

Fonte: Elaborado pelos autores.

Para cada registro há um campo identificador (COD) que permite calcular as frequências de ocorrência das características que se pretende analisar.

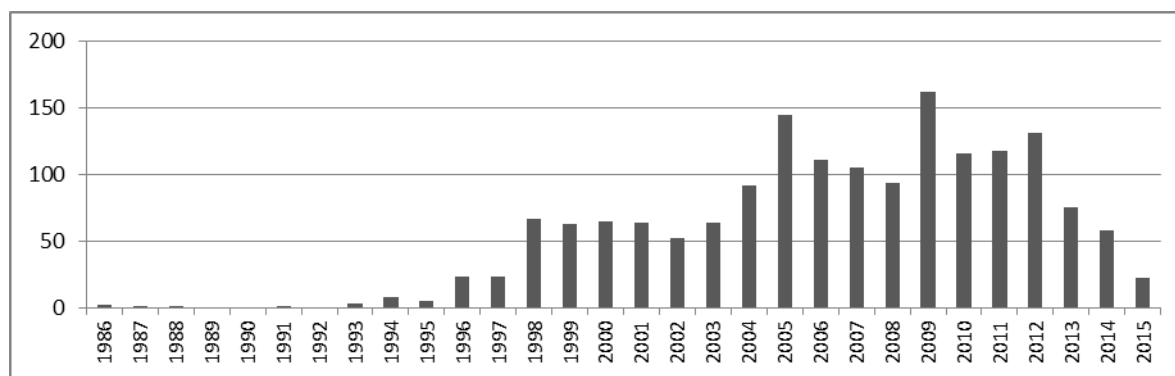
A busca foi realizada em junho de 2016, portanto as referências deste ano foram descartadas porque não representaria a totalidade do ano.

## 4. Resultados e considerações

As 1.671 referências recuperadas abrangeram um período de 30 anos (1986-2015). A primeira análise consistiu em verificar a quantidade de publicações, na qual o ano de 2009 se destacou. Na Figura 1 é possível observar que o interesse pelo tema cresceu, especialmente a partir de 2004. Provavelmente, este resultado tem relação com o aumento da produção de documentos em formato eletrônico, implicando em reflexões sobre a importância de preservá-los.

A Tabela 1 complementa a análise representada na Figura 1. Nela é possível observar, através do percentual acumulado, que nos primeiros 14 anos (1986 - 1999) foram identificadas 11,8% do total de todas as publicações recuperadas. Por outro lado, nos últimos nove anos, entre 2007 e 2015, concentram-se a maior parte das publicações, sendo possível afirmar que o tema da preservação digital ganhou importância e destaque, na produção acadêmica da área da ciência da informação, a partir de 2007. Talvez por demora na indexação dos registros na base, nos últimos anos se observa um decréscimo no número de trabalhos. Este resultado fortalece a necessidade de dar continuidade ao trabalho, acrescentando outras fontes.

**Figura 1. Distribuição das publicações sobre preservação digital por ano (LISA, n=1.671)**



Fonte: Dados da pesquisa, 2016.

**Tabela 1. Distribuição das publicações sobre preservação digital por ano (LISA, n=1.671)**

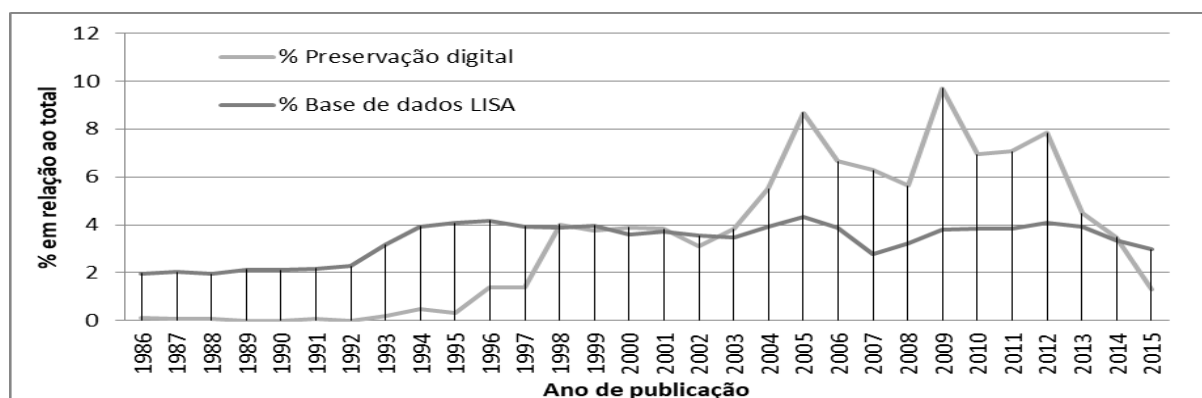
Ano	Nº publicações	%	% Acumulado
1986	02	0,1	0,1
1987	01	0,1	0,2
1988	01	0,1	0,2
1989	00	0,0	0,2
1990	00	0,0	0,2
1991	01	0,1	0,3

1992	00	0,0	0,3
1993	03	0,2	0,5
1994	08	0,5	0,9
1995	05	0,3	1,2
1996	23	1,4	2,6
1997	23	1,4	4,0
1998	67	4,0	8,0
1999	63	3,8	11,8
2000	65	3,9	15,7
2001	64	3,8	19,5
2002	52	3,1	22,6
2003	64	3,8	26,4
2004	92	5,5	31,9
2005	145	8,7	40,6
2006	111	6,6	47,3
2007	105	6,3	53,5
2008	94	5,6	59,2
2009	162	9,7	68,9
2010	116	6,9	75,8
2011	118	7,1	82,9
2012	131	7,8	90,7
2013	75	4,5	95,2
2014	58	3,5	98,7
2015	22	1,3	100,0
TOTAL	1671	100,0	-----

Fonte: Dados da pesquisa, 2016.

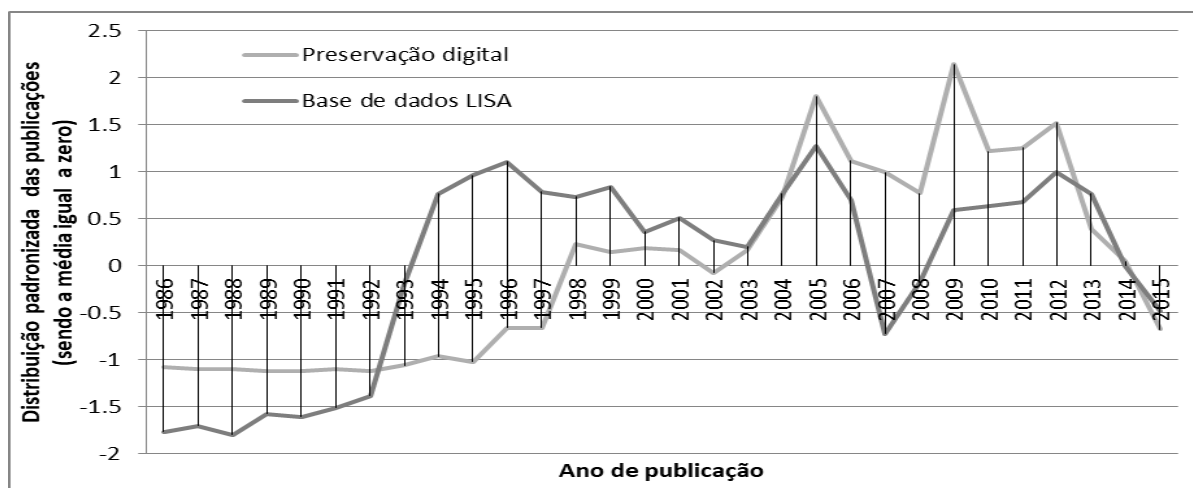
Para conferir se o aumento do número de publicações mostrado nos resultados acima não estava sendo influenciado pelo crescimento geral de qualquer assunto na base de dados LISA, optou-se em verificar o número de registros por ano. Nas Figuras 2 e 3 são apresentadas as tendências de crescimento do número de publicações gerais na base e das sobre preservação digital. Verifica-se que o aumento proporcional nos registros sobre o assunto, tanto pelo percentual como pela padronização dos dados<sup>ii</sup> (BECKER, 2015), foi maior entre os anos de 2004 e 2012 e que, a partir de 1998, o tema da preservação digital começa a apresentar um aumento significativo no número de documentos indexados na base. Observa-se também que, em 2015, houve uma diminuição no número de publicações, talvez influenciada pela dinâmica de trabalho de inserção de dados, pois a coleta foi realizada em maio de 2016 e talvez ainda houvesse registros de 2015 sendo inseridos na base.

**Figura 2. Distribuição proporcional, em %, de publicações na base LISA e sobre preservação digital**



Fonte: Dados da pesquisa, 2016.

**Figura 3. Distribuição das publicações padronizando os dados absolutos (média = 0)**



Fonte: Dados da pesquisa, 2016.

O próximo passo contemplou a verificação dos autores mais produtivos sobre o tema da preservação digital. Antes de calcular a frequência, foi necessário padronizar a grafia dos nomes, por exemplo, Allen, N. S. e Allen, Nancy S. que foi deixado de uma só forma. Além disso, foram identificados alguns registros sem autoria, tais como informações sobre relatórios institucionais publicados em periódicos. Estas referências foram descartadas nesta análise e, sendo assim, o total de registros sobre a produtividade foi de 1.671 e os de autoria 1.591.



**Tabela 2. Autores com cinco ou mais publicações entre 1986 – 2015 (LISA, n = 1.591)**

Autor	Nº de publicações	Instituição de vínculo
Duke, Judy	14	Millwood Group Corp. (EUA)
Conway, P.	08	University of Michigan (EUA)
Kashimura, Masaaki	06	HUMI Project, Keio University. (Japão)
Lavoie, Brian F.	06	OCLC Online Computer Library Center (EUA)
Ashling, Jim	05	Ashling Consulting. (Reino Unido)
Beagrie, N.	05	University of Edinburgh. (Reino Unido)
Charlton, John	05	Information Today inc. (EUA)
Dorr, M.	05	Não informado.
Gertz, J.	05	Columbia University. (EUA)
Jankovic, L.	05	Slovenska Narodna Kniznica. (Eslováquia)
Kenney, A. R.	05	Cornell University. (EUA)

**Fonte:** Dados da pesquisa, 2016.

Na Tabela 2, estão dispostos os 11 autores que totalizaram mais de cinco publicações sobre preservação digital. A maior quantidade de referências são autoria de Judy Duke, com 14 publicações. Verificou-se que esta autora se insere neste cenário temático em 2009, permanecendo até 2014. Todos os artigos de Duke foram publicados na revista *Advanced Technology Libraries*. Judy Duke também utiliza o nome Judith Duke. É interessante apontar a necessidade de investigar de maneira mais detalhada as publicações desta autora, pois se verificou que ela também foi editora desta publicação, caracterizando-se assim um cenário de endogenia. Não se sabe se é um padrão o editor contribuir em todos os fascículos ou se existe um propósito de aumentar a produtividade através deste tipo de estratégia.

Também na Tabela 2, pode-se verificar as instituições que os autores informam nas publicações indexadas pela base LISA. Como alguns autores indicaram diferentes instituições em suas publicações, optou-se pela informada mais recentemente. Somente um autor não indicou a sua instituição de vínculo. Ainda sobre a Tabela 2, nota-se que seis dos 11 autores que totalizaram mais de cinco trabalhos são dos Estados Unidos da América (EUA). Portanto, é possível dizer que os EUA são líderes na produção científica sobre o tema preservação digital.

As análises também mostraram que 1.038 diferentes pesquisadores colaboraram na elaboração dos 1.671 artigos, resultando em uma média geral de 1,6 autores por artigo. Esse

resultado está em conformidade com trabalhos anteriores que já mostraram que na área da ciência da informação os autores publicam mais individualmente ou no máximo com mais um colaborador (COSTA; VANZ, 2012). É válido destacar que, segundo Katz e Martin (1997 apud VANZ, 2010), a colaboração científica pode ser definida como, “[...] o trabalho conjunto de pesquisadores para atingir um objetivo comum de produzir novos conhecimentos científicos”.

Foram encontrados 20 autores brasileiros, com oito artigos no total. Entre os brasileiros, a primeira publicação sobre o assunto registrada na base foi em 1999 e a última em 2012. Entre estes oito registros, cinco são de artigos publicados no periódico *Ciência da Informação*. Este resultado evidencia uma carência de estudos sobre o tema no Brasil.

A tabela abaixo mostra a distribuição de número de artigos por título, sendo possível observar que *Advanced Technology Libraries* foi o periódico que reuniu maior quantidade de trabalhos sobre preservação digital. Entretanto, ao analisar o site da revista, verificou-se que não se trata de um periódico acadêmico cujos artigos passam por revisão por pares e sim uma fonte de divulgação de novos produtos e sistemas relacionados à tecnologia aplicada aos serviços bibliotecários. Este resultado também está relacionado com a fonte de dados, isto é, a base de dados LISA contempla vários tipos de publicações e não somente periódicos acadêmicos. Além disso, esta revista é publicada mensalmente, explicando a maior frequência de registros. Para confirmar qual o periódico acadêmico que possui avaliação por pares e que reúne maior quantidade de artigos sobre preservação digital, identificou-se o *Liber Quarterly: The Journal of European Research Libraries* como o periódico de cunho acadêmico mais utilizado. Portanto, este resultado aponta que se fosse necessário priorizar uma fonte de dados para atender, por exemplo, uma demanda de um curso, este seria o periódico científico mais indicado.

O cálculo do percentual acumulado mostrou que 43 periódicos concentram 50,4% do total de publicações, e o restante (49,6%) se distribuiu entre os 298 diferentes títulos, mostrando bastante dispersão de fontes de informação.

Para identificação da tipologia documental, utilizou-se o campo “tipo de referência” (reference type) de cada registro. O resultado mostrou que o artigo de periódico (journal article) foi o mais prevalente, com 1.662 registros. Entretanto, conforme já mencionado, *Advanced Technology Libraries*, que apresentou maior número de registros, é uma publicação

que se caracteriza mais como uma revista (magazine) e não um periódico. Portanto, é preciso estar atento com este problema na base de dados LISA.

**Tabela 3. Periódicos com mais publicações sobre preservação digital (LISA, n=1.671)**

Título do periódico	Nº Artigos	%	%A
Advanced Technology Libraries	58	3,5	3,5
International Preservation News	52	3,1	6,6
D-Lib Magazine	51	3,1	9,7
Microform & Imaging review	51	3,1	12,8
Computers in Libraries	30	1,8	14,6
Bulletin des Bibliothèques de France	29	1,7	16,3
Nordisk Arkivnyt	29	1,7	18,1
IASA Journal	28	1,7	19,7
Ariadne	27	1,6	21,4
Microform & Digitization Review	26	1,6	22,9
Liber Quarterly: The Journal of European Research Libraries	25	1,5	24,2
Against the Grain	24	1,4	25,9
Informatie Professional	24	1,4	27,3
Knjiznica	23	1,4	28,7
Library Hi Tech	23	1,4	30,1
Library Trends	22	1,3	31,4
Biblioteche Oggi	17	1,0	32,4
IFLA Journal	16	1,0	33,4
Information Today	16	1,0	34,4
Kniznica	16	1,0	35,3
Serials Librarian	15	0,9	36,2
Vjesnik Bibliotekara Hrvatske	15	0,9	37,1
Zeitschrift für Bibliothekswesen und Bibliographie	15	0,9	38,0
Biuletyn EBIB	14	0,8	38,9
Collection Management	12	0,7	39,6
Serials	12	0,7	40,3
Tudományos és Muszaki Tajekoztató	12	0,7	41,0
Information World Review	11	0,7	41,7
Journal of Agricultural & Food Information	11	0,7	42,4
Profesional de la Información	11	0,7	43,0
Journal of Information Science and Technology Association (Joho no Kagaku to Gijutsu)	10	0,6	43,6
Library Hi Tech News	10	0,6	44,2
Library Philosophy and Practice	10	0,6	44,8
META: tijdschrift voor bibliotheek & archief	10	0,6	45,4
Research Information	10	0,6	46,0
Vine	10	0,6	46,6
ESARBICA Journal: Journal of the Eastern and Southern Africa Regional Branch of the International Council on Archives	09	0,5	47,2
First Monday	09	0,5	47,7
Konyvtari Figyelo	09	0,5	48,3
OCLC Systems & Services: International Digital Library Perspectives	09	0,5	48,8
RLG DigiNews	09	0,5	49,3
Toshokan Zasshi/The Library Journal	09	0,5	49,9
American Libraries	08	0,5	50,4
Outros títulos	825	49,6	100,0
TOTAL	1662	100,0	-

Legenda: %A – percentual acumulado.

Fonte: Dados da pesquisa, 2015.

Para verificar qual disciplina da ciência da informação produz mais sobre o assunto preservação digital, foram analisados todos os resumos das referências, utilizando o software Atlas.Ti6. Após uma limpeza minuciosa dos dados para retirar and, or, library e libraries, digitization e digitisation, verificou-se as palavras mais prevalentes. Na Tabela 4 se verifica que libraries tem maior número de ocorrência, pronomes, sinônimos ou diferentes grafias para um mesmo termo.

**Tabela 4. Palavras que apresentaram frequência igual ou superior a 500, nos resumos dos artigos sobre preservação digital indexados na base de dados bibliográfica LISA (n = 1.671)**

Palavras	Total de registros
DIGITAL	2.305
PRESERVATION	1.612
DIGITIZATION	1.314
PROJECT	1.032
LIBRARIES	967
SOURCE	878
INFORMATION	821
COLLECTIONS	767
ACCESS	740
NATIONAL	642
RESEARCH	620
UNIVERSITY	565
MATERIALS	551

**Fonte:** Dados da pesquisa, 2016.

Verificou-se que a palavra “libraries” está presente em 967 trabalhos. Este resultado pode estar associado com o maior destaque do tema na área da biblioteconomia do que da

arquivologia. É importante destacar a necessidade de aprofundar esta análise para evidenciar melhor esta característica.

## 5. Conclusões

Este trabalho teve como objetivo a realização de um estudo bibliométrico sobre preservação digital, utilizando da base bibliográfica LISA como fonte para coleta dos dados. Através dos resultados foi possível revelar um panorama sobre o assunto pesquisado. É importante ressaltar que neste estudo foi utilizada apenas uma fonte de informação (LISA) e para uma maior compreensão da situação geral é necessário dar continuidade ao estudo, incluindo outras fontes que permitam alcançar maior abrangência da produção científica na temática pesquisada.

Verificou-se que nos últimos 30 anos o tema se expandiu, especialmente a partir de 2004, sendo possível dizer que a preocupação com a preservação de documentos digitais é um assunto atual.

Conclui-se que as publicações brasileiras sobre o tema ainda são poucas, evidenciando a necessidade de mais investigações sobre o mesmo pelos pesquisadores da área da ciência da informação do Brasil.

## Referências Bibliográficas

---

- BECKER, J. L. (2015). *Estatística Básica: Transformando Dados em Informação*. Porto Alegre: Bookman.
- COSTA, J. G., VANZ, S. A. de S. (2012). Indicadores da produção científica e co-autoria: análise do Departamento de Ciências da Informação da UFRGS. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, 33, 97-115. doi: 10.5007/1518-2924.2012v17n33p97
- CUNHA, M. B., CAVALCANTI, C. R. O. (2008). *Dicionário de biblioteconomia e arquivologia*. Brasília: Briquet de Lemos.
- CUNHA, M. R. (2007). *Gestão de documentos eletrônicos e preservação digital*.
- FEBVRE, L., MARTIN, L. (1992). *O aparecimento do livro*. São Paulo: UNESP; Hucitec.

- FRANCO, S. A. O. (2008). Digitalização de documentos: acesso à informação com preservação de acervo. In M. Cassares & A. P. H. Tanaka, *Preservação de acervos bibliográficos: homenagem a Guita Mindlin*. São Paulo: Associação Brasileira de Encadernação e Restauro.
- INNARELLI, H. C. (2012). Preservação digital e seus dez mandamentos. In V. B. dos Santos, *Arquivística: temas contemporâneos: classificação, preservação digital, gestão do conhecimento*. (3rd ed.). Brasília: SENAC.
- LYONS, M. (2011). *Livro: uma história viva*. São Paulo: SENAC.
- MARDERO ARELLANO, M. A. (2004). Preservação de documentos digitais. *Ciência da Informação*, 33(2), 15-27. Retrieved from [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652004000200002](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652004000200002)
- MEADOWS, A. J. (1999). *A comunicação científica*. Brasília: Briquet de Lemos.
- MIRANDA, L. M., FARIAS, S. F. (2009). *A internet enquanto ferramenta de orientação de atividade física na promoção da saúde do idoso*.
- NORONHA, D. P., MARICATO, J. de M. (2008). Estudos métricos da informação: primeiras aproximações. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, 116-128. doi: 10.5007/1518-2924.2008v13nesp1p116
- RONDINELLI, R. C. (2002). *Gerenciamento arquivístico de documentos eletrônicos: uma abordagem teórica da diplomática arquivística contemporânea*. Rio de Janeiro: FVG.
- SANTOS, H. M. dos, FLORES, D. (2015). Preservação de documentos arquivísticos digitais: reflexões sobre as estratégias de emulação. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 20(43), 3-19. doi: 10.5007/1518-2924.2015v20n43p3
- VANZ, S. A. De S., STUMP, I. R. C. (2010). Colaboração científica: revisão teórico-conceitual. *Perspectivas em Ciência da Informação*, 15(2), 42-55. Retrieved from <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1105>
- WOLTON, D. (2003). *Internet, e depois? Uma teoria crítica das novas mídias*. Porto Alegre: Sulina.

## Notas

---

i Artigo adaptado de um Trabalho de conclusão de curso – TCC, apresentado ao curso de Biblioteconomia da Universidade Federal do Rio Grande – FURG, em 2016.

ii A padronização dos dados consiste em transformar os dados utilizando a média e o desvio padrão como pontos de referência e serve para ajudar na comparação de medidas com escalas diferentes, mas que focam um mesmo fenômeno (BECKER, 2015).

# Questões éticas na comunicação científica

*Ethical issues in scientific communication*

---

**Karen Ribeiro de Freitas Irizaga**

Universidade Federal do Rio Grande do Sul, Brasil  
[karenirizaga@ufrgs.br](mailto:karenirizaga@ufrgs.br)

**Luciana Monteiro Krebs**

Universidade Federal do Rio Grande do Sul, Brasil  
[luciana.monteiro@ufrgs.br](mailto:luciana.monteiro@ufrgs.br)

**Maiara Bettio**

Universidade Federal do Rio Grande do Sul, Brasil  
[maiara.bettio@ufrgs.br](mailto:maiara.bettio@ufrgs.br)

**Moisés Rockembach**

Universidade Federal do Rio Grande do Sul, Brasil  
[moises.rockembach@ufrgs.br](mailto:moises.rockembach@ufrgs.br)

## Resumo

O artigo aborda a relação entre ética e comunicação científica, trazendo dilemas que se situam nas práticas de pesquisa e publicação, com o objetivo de fomentar a discussão ética sobre práticas científicas e no sentido de evitar a má conduta de investigação, que pode vir a prejudicar todo um campo de conhecimento. O enfoque sobre as questões éticas envolve a ciência aberta, o acesso aberto aos dados de pesquisa versus a privacidade dos dados, a múltipla afiliação institucional e os impactos na avaliação da produtividade científica dos autores e instituições, assim como a prática de fatiar os resultados em várias publicações, conhecida como *salami slicing*. Conclui-se que as discussões destes enfoques são fundamentais para a manutenção do status científico das pesquisas e desenvolvimento do conhecimento como um todo.

## Abstract

*The paper discusses the relationship between ethics and scientific communication, bringing dilemmas that are placed in the practices of research and publication, with the objective of fomenting the ethical discussion about scientific practices and in the sense of avoiding research misconduct, that can harm a whole field of knowledge. The focus on ethical issues involves open science, open access to research data versus data privacy, multiple institutional affiliation and impacts on the assessment of authors and institutions scientific productivity, as well as the practice of slicing results in various publications, known as salami slicing. It is concluded that the discussions of these approaches are fundamental for the maintenance of the research scientific status and development of knowledge as a whole.*

**Palavras-chave:** Comunicação científica. Ética. **Keywords:** *Scientific communication. Ethics. Open Access aberto. Múltipla filiação institucional. Salami access. Multiple institutional affiliation. Salami slicing. slicing.*

## 1. Introdução

Define-se a comunicação científica como o intercâmbio de informações entre membros que compõem a comunidade científica, sendo consideradas também as atividades associadas à produção de ciência, a fim de propagar a informação, desde o momento no qual se concebe uma ideia até a divulgação dos resultados de uma pesquisa. Portanto, a literatura científica e a comunicação científica são importantes elementos para fazer-se ciência, bem como disseminar o conhecimento científico. A comunicação científica, conforme afirma Meadows (1999, p.7), “[...] situa-se no próprio coração da ciência. É para ela tão vital quanto a própria pesquisa.”. Acredita-se que uma publicação científica que siga os preceitos éticos gera credibilidade e maiores chances de ser aceita pela comunidade científica. Ou seja, além de ser reconhecida e citada, também traz prestígio para o autor, e ainda possibilita financiamentos para os projetos de pesquisa a ela associados.

A ética, por sua vez, como campo de estudo filosófico, possui múltiplas vertentes a serem exploradas, aplicadas a diversos dilemas humanos. Na busca de uma eticidade hegeliana, refletindo sobre as subjetividades da moral e procurando a realização do bem, com o suporte de instituições que o garantam – representado aqui pela própria comunidade científica – temos muitos desafios na prática da investigação acadêmica, sendo importante abordar e discutir possíveis questões éticas enfrentadas.

A relevância da comunicação científica é reforçada com a observância de que, além de simplesmente divulgar resultados de pesquisa, publique-se de forma ética. Esta ética não está ligada somente a uma moralidade individual ou ainda a um código deontológico profissional, mas acima de tudo a uma “Ciência da conduta” (Abbagnano, 2007), que nos leva à reflexão do impacto que o conhecimento científico produz na sociedade contemporânea. Aqui entra em jogo não apenas a ética aplicada aos profissionais da informação, mas também a ética da informação, onde questões envolvendo privacidade, sigilo, segurança, acesso e autoria tornam-se essenciais para a compreensão da informação enquanto fenômeno.



A partir destas reflexões, a responsabilidade do cientista cresce proporcionalmente ao seu papel transformador nesta Sociedade da Informação. O cerne deste artigo é apontar três desafios éticos, em que as relações entre o acesso aberto e a privacidade, a autoria institucional e a forma de “fatição” na publicação dos resultados tomam papel fundamental.

## **2. Questões éticas na comunicação científica**

Muitas podem ser as questões éticas envolvendo a comunicação científica. Marcovitch et al. (2010) relatam desde plágio, publicação duplicada, mudanças não divulgadas nos protocolos pré-pesquisa ou comportamento ético duvidoso. Este espectro que vai de um “erro honesto”, ou seja, sem intenção de fraude, até a fraude total, é representado por uma série de más condutas científicas. Por um lado, deste espectro, temos aquelas que se aproximam do erro não-intencional e de difícil detecção, como uma observação ou análise errada, arquivamento inadequado, publicações duplas e fatiadas e ainda a ação de ignorar o trabalho anterior de outros. Enquanto do outro, temos más condutas que se aproximam da fraude, de forma intencional e de detecção mais fácil, como pesquisas sem consentimento informado, plágio, falsificação ou fabricação de dados e experiências humanas ilegais (Marcovitch et al., 2010).

A produtividade acadêmica é exigência dos órgãos reguladores da pós-graduação no país e das agências de fomento, e talvez por isso os autores estejam preocupados, majoritariamente, com a prevenção de más condutas no processo de publicação, “[...] tais como casos de plágio, fraude, falsificação de dados, duplicidade e segmentação de resultados, autoria indevida, conflito de interesses, etc.” (Shinkai, 2011, p. 2). Das diversas questões éticas possíveis, trazemos uma discussão sobre a perspectiva da ciência aberta e o dilema entre dados abertos de pesquisa versus privacidade dos dados, a representação e vinculação institucional do pesquisador nas publicações científicas na relação com múltiplas instituições e a prática de dividir ou fatiar publicações científicas (*salami slicing*).

### **2.1. A ciência aberta e o dilema entre dados abertos de pesquisa versus privacidade**

A coleta de dados é algo intrínseco ao fazer científico, pois é a partir de dados coletados, organizados e analisados que a ciência se desenvolve. Davis e Vickery (2007) afirmam que os dados primários tornaram-se a principal moeda da ciência. A Organização para a Cooperação

e Desenvolvimento Econômico (OECD) define os dados de pesquisa como “[...] registros factuais [...] utilizados como fontes primárias para pesquisa científica e que são comumente aceitos na comunidade científica como necessários para validar os resultados da pesquisa.” (OECD, 2007, p. 13, tradução nossa).

O processo de coleta e tratamento dos dados envolve um custo. No intuito de minimizar o risco de perda e maximizar a acessibilidade aos dados, formou-se o *World Data Centre* (WDC). Na ocasião, o *International Council for Science* (ICS)<sup>i</sup> estabeleceu vários centros de dados mundiais e recomendou que os dados coletados de pesquisas observacionais fossem disponibilizados, ainda em 1955, em formato legível por máquina (*National Centers for Environmental Information*, [20--?], documento eletrônico). De lá para cá, muitas iniciativas têm colocado o Acesso Aberto a Dados de Pesquisa (AADP) em pauta, tendo um crescimento especialmente alto nos últimos anos.

Chama a atenção, no entanto, que a emergência da ciência aberta e o movimento de AADP tragam consigo uma série de decorrências éticas e sociais ainda pouco discutidas na literatura especializada. Uma busca na plataforma *Web of Science*<sup>ii</sup> (WoS) por “*research data*” e “*data sharing*” recupera 24.079 documentos. Ao acrescentar o termo “*privacy*”, o número cai para 877<sup>iii</sup>. No Portal de Periódicos da CAPES<sup>iv</sup>, a busca por “*research data sharing*” e “*privacy*” retorna apenas 161 documentos.

A sociedade pode beneficiar-se do acesso aberto, como é o caso da transparência para contas públicas, mas também pode ser prejudicada, no caso de tratamento inadequado de informações pessoais. O direito à privacidade é uma das liberdades fundamentais garantidas pela Declaração Universal dos Direitos Humanos (Organização das Nações Unidas [ONU], 1948). Torna-se então responsabilidade das instituições garantir que informações pessoais dos indivíduos, como seus hábitos, propriedades materiais, histórico de saúde, entre outras, não sejam publicizadas para terceiros como familiares, comunidade ou empresas.

Já vigoram algumas medidas no sentido de proteger os indivíduos participantes de pesquisas científicas como o direito de desistência (deixar de participar da pesquisa em qualquer tempo) e a avaliação da proposta de pesquisa por um comitê apropriado (no Brasil, um comitê de ética), antes de a pesquisa ser iniciada. Porém, essas medidas são derivadas dos Princípios de Nuremberg<sup>v</sup>, e foram pensadas para proteger os indivíduos participantes de pesquisa de

danos físicos, não sendo suficientes em um contexto de ciência globalizada, em que o fluxo de dados circula internacionalmente através da colaboração científica (Kaye, 2012).

Então, se o fluxo de dados a partir dos processos de colaboração científica permite o compartilhamento e reuso de dados de pesquisa, a aprovação de um projeto de pesquisa por um comitê pode ser questionado. Afinal, a avaliação considerou o método, as políticas de gestão de dados e os limites de pesquisa para um objetivo específico, em um determinado tempo, elaborado por um pesquisador (ou grupo) definido previamente. A partir do momento em que os dados estão disponíveis para novos projetos, esses limites são extrapolados e os dados passam a servir a outros objetivos, sob a égide de novos olhares e podendo inclusive ser cruzados com outros conjuntos de dados de fontes heterogêneas.

Mesmo o termo de consentimento informado, cujo objetivo é prover ao participante a autonomia de tomar decisões sobre os riscos que a pesquisa oferece à sua privacidade, não é capaz de informar os possíveis usos posteriores dos dados, simplesmente porque um dos princípios da ciência aberta repousa justamente sobre o livre (e, portanto, imprevisível) intercâmbio destes, pois considera o conhecimento científico como um bem da humanidade. Assim, os mecanismos atuais para assegurar a proteção da privacidade dos indivíduos participantes das pesquisas científicas mostram-se insuficientes. É necessário discutir e pensar soluções que garantam a devida privacidade dos participantes, mas que, ao mesmo tempo, não inviabilizem o avanço da ciência e impeçam a colaboração científica.

Uma técnica para garantir a privacidade dos participantes em um conjunto de dados é a anonimização. Este processo consiste em ocultar (ou embaralhar, ou substituir) o nome do participante da pesquisa, mantendo os demais dados relativos à pessoa disponíveis. No entanto, Mayer-Schonberger e Cukier (2013, p. 108) alertam para o fato de que a anonimização funciona em um pequeno universo de dados. Segundo os autores, com o aumento de quantidade e variedade de informações, fica mais fácil reidentificar os participantes. Infere-se, assim, que apenas ocultar nomes ou IPs pode não ser mais suficiente, devido à triangulação que se torna possível a partir do uso secundário e terciário dos dados de pesquisa (cruzamento destes com outras bases para gerar novas informações).

Conhecer as percepções dos cientistas torna-se extremamente relevante quando se pensa em políticas de incentivo para acesso aberto a dados de pesquisa. Isso porque são aspectos que

podem motivar ou inibir a adesão dos pesquisadores aos repositórios desta tipologia de dados. Mapeando a opinião de pesquisadores<sup>vi</sup> a respeito de reuso de dados de pesquisa em acesso aberto, Curty (2016) identificou que um dos riscos percebidos pelos pesquisadores é o medo de infringir códigos de ética (em caso de confidencialidade, por exemplo, ou uso secundário não previsto em acordos prévios).

Um estudo das opiniões das partes interessadas na pesquisa em saúde pública com o compartilhamento de dados no Quênia destaca que a preocupação com a privacidade é importante, pois sua perda gera riscos de estigmatização e desrespeito quando associada a dados confidenciais. Essas informações podem incluir dados clínicos, diagnóstico individual de HIV, dados genéticos e sobre comportamento sexual (Jao et al., 2015). Além da perda de privacidade e possível estigmatização, outros desafios apontados na pesquisa são o impacto na autonomia e concorrência desleal, inclusive através de “uso indevido” (intencional ou não) dos dados.

Também há que se considerar, naturalmente, a opinião dos indivíduos participantes em relação às práticas de uso e compartilhamento de dados de pesquisa das quais eles fizeram parte. Estudos realizados no âmbito das ciências médicas (compartilhamento de dados genéticos) apontam que a grande maioria dos participantes deseja ser consultada sobre a disponibilização dos seus dados e que muitos deles consideram inaceitável a consulta ocorrer após o compartilhamento (Ludman et al., 2010; McGuire et. al, 2008).

Kaye (2012) acredita que uma solução possível para este dilema reside no que chamou de *participant-centric interfaces* (PCIs), ou seja, interfaces centradas no participante. A abordagem das PCIs possui “[...] um modelo de consentimento ‘dinâmico’ e não ‘informado’, que permite a obtenção do consentimento quando necessário, em tempo real, como parte de um processo bidirecional, contínuo e interativo entre pacientes e pesquisadores [...]” (Kaye, 2012, p. 426).

Esse tipo de interface pressupõe uma nova forma de lidar com os participantes da pesquisa, dando-lhes autonomia para tomar decisões, pois os considera parceiros no processo científico, ao invés de provedores passivos de informação. Isto porque permite que os indivíduos “[...] mudem de ideia e de preferências ao longo do tempo, revoguem suas escolhas quando apropriado, rastreiem e auditem as mudanças feitas e escolham quando e como eles

são contatados.” (Kaye, 2012, p. 426). Outra opção viável é a desidentificação, que desvincula totalmente os dados resultantes da pesquisa a quem o produziu/forneceu, incluindo todos os dados pessoais (e não apenas o nome).

No que diz respeito a publicações da Ciência da Informação, Sayão e Sales (2014) sinalizam a necessidade de se pensar diferentes instâncias envolvidas no âmbito dos dados abertos de pesquisa, incluindo a "Instância social, legal e ética". No entanto, os autores não chegam a dissertar sobre os impactos, desafios e possíveis encaminhamentos para essas questões.

No Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã, o IBICT menciona que cabe aos pesquisadores “[...] adotar padrões éticos, legais, especialmente em relação à privacidade e à confidencialidade [...]” (Instituto Brasileiro de Informação em Ciência e Tecnologia [IBICT], 2016, documento eletrônico). O Manifesto prevê ainda que gestores e executores de programas e projetos de dados de pesquisa devam “[...] adotar medidas para proteger a confidencialidade dos dados sensíveis, como anonimização, adotar padrões éticos e observar os direitos associados aos dados [...]” (IBICT, 2016, documento eletrônico). Porém, não especifica outras medidas cabíveis para garantia da confidencialidade além da anonimização.

Percebe-se uma oportunidade para os pesquisadores da Ciência da Informação discutirem não apenas a privacidade, mas outros pontos envolvendo o acesso aberto a dados de pesquisa, levando em conta não só as implicações legais da jurisdição, mas também o contexto global no qual a pesquisa invariavelmente se insere atualmente. O respeito a todos os participantes envolvidos torna-se fundamental para garantia de uma ciência justa que promove a cidadania e a paz.

## **2.2. Múltiplas filiações institucionais**

Para a ciência, a comunicação é tão vital quanto a própria pesquisa, razão pela qual os pesquisadores veiculam seu trabalho para a comunidade científica e a sociedade em geral, principalmente através de revistas científicas (Meadows, 1999). Para tanto, são estabelecidas diretrizes para a publicação que vão além do formato acadêmico do trabalho desenvolvido, e que requerem uma temática original, a apresentação de resultados confiáveis, de

contribuições relevantes para a ciência e a humanidade, e que seja assinado por pesquisadores íntegros, capazes tecnicamente e honestos eticamente (Krokosz, 2015).

Entre as diretrizes estabelecidas pelas revistas científicas, por comitês de ética internacionais e órgãos profissionais estão aquelas que dizem respeito às concepções de autoria e às relações entre os pesquisadores, no intuito de garantir a correta atribuição dos créditos aos autores envolvidos. Podemos citar como exemplos desses esforços as publicações dos comitês de ética em publicação científica *International Committee of Medical Journals Editors* (ICMJE, 2016) - *Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals* e *Committee on Publication Ethics* (COPE, 2011) - *Code of conduct and best practice guidelines for journal editors*, além da *American Psychological Association* (APA, 2017) - *Ethical principles of psychologists and code of conduct/Section 8: research and publication*, e, no contexto nacional do Brasil, o Conselho Federal de Odontologia (CFO, 2012) - Código de Ética profissional odontológica.

Tais documentos, apesar de trazerem definições de autoria e de condutas para a publicação de pesquisas quanto a uma vasta gama de dilemas éticos importantes, não apresentam diretrizes que tratem da atribuição da vinculação institucional dos autores, seja ela única ou múltipla. O docente que atua em duas universidades, por exemplo, não encontra nesses códigos nenhuma informação sobre como atribuir os créditos a cada instituição com a qual possui vínculo. Qual das instituições deve aparecer no artigo? Apenas aquela que forneceu os recursos financeiros para o desenvolvimento da pesquisa? Ou aquela com a qual ele possui contrato formal de trabalho? E quanto à instituição que forneceu outros tipos de recursos importantes para a pesquisa, tais como a bibliografia utilizada no referencial teórico, os serviços de assessoria editorial, as salas de estudo na biblioteca? Não existiriam aí outros tipos de investimentos que deveriam ser considerados quando da atribuição dos créditos institucionais da pesquisa? Ambas as instituições poderiam ou não ser indicadas na publicação?

Tais questões se tornam ainda mais relevantes se observarmos que as avaliações realizadas pelas universidades, assim como por agências de fomento à pesquisa, avaliam quantitativamente a produção científica dos docentes. Para isso, são utilizadas as informações atribuídas pelos autores quanto à sua filiação institucional, revelando, portanto, o impacto de

tal informação na futura representação da produtividade das instituições. Na base de dados multidisciplinar de abrangência e cobertura internacional *Web of Science* (WoS), amplamente utilizada em estudos que visam à avaliação da produção científica de instituições (Packer & Meneghini, 2006) e referência em estudos bibliométricos, o campo Endereço (C1) é utilizado para reunir todos os documentos de uma instituição, indexados na base de dados.

As indicações de filiações institucionais dos pesquisadores, apresentadas nas publicações, podem afetar também as medidas de desempenho das universidades em *rankings*, uma vez que a sua representatividade na pesquisa científica mundial costuma ser aferida através das bases de dados como a WoS. Os *rankings* universitários, tais como o *Shanghai Jiao Tong University* e o Ranking Universitário Folha, se constituem em sistemas de classificações de instituições de ensino superior com credibilidade para comparar o desempenho institucional no ensino e na pesquisa (Morosini, 2009).

Outra questão ética importante e que merece destaque diz respeito ao critério da produtividade como elemento para a progressão do docente na carreira, no âmbito das universidades. De acordo com Zabalza (2004), os mecanismos básicos da promoção estão vinculados à produtividade científica, muitas vezes utilizando como critérios de mérito pessoal as atividades de pesquisa ou de produção científica. A múltipla filiação institucional indicada pelo autor em uma publicação pode, neste caso, servir como forma de estabelecer sua vinculação a mais de uma instituição, tanto para fins de avaliação interna e externa de produtividade, quanto para a sua ascensão na carreira docente.

No artigo intitulado *A first look at multiple institutional affiliations: a study of authors in Germany, Japan and the UK* (Um primeiro olhar sobre a múltipla filiação institucional: um estudo de autores na Alemanha, no Japão e no Reino Unido, tradução nossa), publicado pelas pesquisadoras Hanna Hottenrott e Cornelia Lawson (2017), são analisados artigos das áreas de Biologia, Química e Engenharia, concluindo que o número de publicações com autores de múltipla filiação institucional mais que dobrou nos últimos dois anos. Elas destacam a importância dos vínculos institucionais na organização e governança da ciência, capazes de alterar o quadro tradicional de emprego e de carreira acadêmica. Outro ponto destacado diz respeito à avaliação institucional realizada com base nos resultados de pesquisa. A múltipla

filiação institucional teria, portanto, implicações tanto na economia da ciência quanto na avaliação da produtividade científica dos autores e das próprias instituições.

A questão da múltipla filiação institucional associada à produção intelectual (PI) dos docentes é tratada pela Universidade Federal do Rio Grande do Sul (UFRGS, 2017) no *site* Document@, que reúne e disponibiliza informações e documentação relativa ao Sistema de Automação de Bibliotecas (SABi) e ao Repositório Digital da UFRGS (Lume). Os registros bibliográficos das publicações de autores vinculados à universidade alimentam o cálculo dos índices departamentais de produtividade, que definiam a alocação de vagas docentes em anos anteriores e, hoje, definem a progressão funcional, além de relatórios individuais de PI. No Capítulo 2 do Document@ constam diretrizes para inclusão de publicações de docentes com dupla vinculação institucional, além da previsão para outros casos semelhantes, demonstrando um entendimento, por parte da universidade, de que a múltipla vinculação institucional identificada nas publicações de seus docentes é prática recorrente e considerada eticamente aceitável para fins de registro e reunião de sua produção intelectual, bem como para fins avaliativos que observam a sua produtividade.

### **2.3. *Salami slicing* e duplicidade de informação**

Em um primeiro momento definimos os termos conhecidos por *salami publication*, *salami slicing* ou *salciping*, em inglês; publicação em salame ou corte em salame, em português, conforme encontrados na literatura científica. *Salami slicing* trata de uma prática de publicação utilizada por pesquisadores na ciência como forma de publicar os dados de uma pesquisa. Este modo de publicação foi definido por Smolčić (2013) como a publicação de dois ou mais artigos derivados de um único estudo. Por exemplo, dados coletados de um único estudo são divididos em vários segmentos e publicados em diferentes artigos, o de tamanho suficiente apenas para obterem resultados e conclusões razoáveis, também conhecidos como “unidade editável mínima”.

Devido à definição acima, é importante expor que a publicação em salame pode ser aceita pela comunidade científica, pois muitas vezes faz-se necessário. Para exemplificar essa situação podemos citar o exemplo de uma pesquisa na qual o conjunto de pesquisadores é oriundo de diferentes áreas do conhecimento. É provável que cada um deles utilize os resultados da sua investigação científica para publicar em sua área do saber o seu olhar sob



um aspecto da pesquisa, diferente do publicado no artigo original. Além deste motivo, existem pesquisas extensas e que exigem acompanhamento durante anos, sendo seus resultados parciais publicados inicialmente em um artigo e posteriormente usados em um novo trabalho completo.

No entanto, nesses casos, os autores devem fornecer essas informações impreterivelmente ao editor, a fim de que ele avalie a justificativa para a publicação. Sendo assim, os autores devem informar claramente quais resultados já foram publicados anteriormente, além de fornecer a fonte completa e o consentimento do autor original quando necessário. Citar o artigo original apenas nas referências não seria suficiente, pois o valor agregado do novo artigo deve ser evidente, à medida que dois artigos similares podem ser facilmente identificados. Além do exposto, é interessante dizer também que um artigo publicado em mais de um idioma, desde que autores e editores estejam cientes desta publicação, também não é considerado publicação em salame, pois segundo o *Committee on Publication Ethics* (COPE, 1999) a publicação de um artigo em outra língua é aceitável, desde que exista uma divulgação completa e proeminente de sua fonte original no momento da submissão.

Todavia, a publicação em salame ainda é considerada um problema de má conduta. Em outras palavras, é antiético “fatiar” os resultados de uma mesma pesquisa em vários artigos, sem necessidade, apenas para mascarar produtividade, tornando-se uma forma de publicação redundante que aumenta dissimuladamente o número de trabalhos científicos de um autor. Tal prática pode gerar um benefício imérito a autores tais como o avanço da carreira acadêmica ou o alcance de financiamentos de projetos de pesquisa. A editora Elsevier (2015) considera que uma publicação fatiada torna-se um problema quando segmenta um grande estudo em duas ou mais publicações. Esses segmentos são referidos como “fatias” de um estudo que compartilham os mesmos pressupostos, população e métodos e, portanto, considerada uma prática inaceitável, visto que a mesma “fatia” nunca deve ser publicada mais de uma vez.

Ressalta-se a dificuldade por parte dos editores de periódicos científicos em reconhecer uma publicação em salame, pois não há um *software* ou algoritmo criado para detectar este tipo de publicação. A identificação desta prática torna-se complexa uma vez que as publicações em salame não incluem frequentemente o plágio textual. Editores e avaliadores podem

desconfiar de uma publicação em salame. Embora não existam formas objetivas de detectar esse tipo de publicação redundante, os artigos suspeitos geralmente relatam tamanho, hipótese, metodologia e resultados de pesquisa idênticos ou similares, e muitas vezes incluem os mesmos autores. Devido a este contexto, a editora Elsevier apresentou instruções aos pesquisadores a fim de prevenir publicações em salame eticamente questionáveis (Quadro 1).

**Quadro 1 – Como prevenir problemas éticos na publicação**

<b>Ação</b>	<b>O que é isso?</b>	<b>Isto é antiético?</b>	<b>O que você deve fazer?</b>
Romper ou segmentar dados de um único estudo e criar diferentes manuscritos para publicação.	Publicação de pequenas "fatias" de uma pesquisa em vários artigos diferentes é chamada de "publicação em salame".	Sim. O corte em salame pode resultar em uma distorção da literatura, levando os leitores desavisados a acreditar que os dados apresentados em cada "fatia" são derivados de uma amostra de assunto diferente.	Evite separar os dados de um único estudo em dois ou mais artigos.  Ao enviar um documento, seja transparente. Envie cópias de qualquer artigo intimamente relacionado ao artigo referido.  Isso inclui todos os manuscritos publicados, recentemente submetidos e já aceitos.

Fonte: Elsevier (2015).

Com este quadro explicativo, a editora Elsevier aposta na diminuição dos problemas de má conduta por parte dos autores e procura evitar que uma publicação em salame chegue às mãos do avaliador ou editor de periódico. Além disso, procura evitar o desperdício de tempo de que a equipe editorial dispõe para avaliação dos trabalhos. Vale salientar que decisão final pela publicação ou não do artigo recai sobre o editor do periódico científico. No entanto, frisamos que a honestidade e a transparência dos autores são essenciais para uma comunicação científica eticamente responsável.

Segundo o *Committee on Publication Ethics* (1999) a publicação em salame e a publicação redundante constituem-se em dois ou mais artigos compartilhando as mesmas hipóteses, dados e pontos de discussão. As seguintes diretrizes são apresentadas em seu guia de boas práticas para a publicação:

- a) estudos publicados não precisam ser repetidos a menos que seja necessária alguma informação adicional;
- b) a publicação prévia de um resumo durante os anais de evento não exclui a submissão subsequente para publicação em periódicos, mas a divulgação completa do artigo deve ser feita no momento da submissão;
- c) a republicação de um artigo em outra língua é aceitável, desde que exista uma divulgação completa e proeminente de sua fonte original no momento da submissão;
- d) no momento da submissão, os autores devem divulgar detalhes de artigos relacionados, mesmo que em uma linguagem diferente, e artigos similares na imprensa.

Iniciativas como estas visam a garantir uma publicação dentro dos preceitos éticos com a finalidade de aperfeiçoar a *expertise* das publicações, visto que elas reforçam a concepção de produzir uma ciência de qualidade, na qual as descobertas são publicadas e divulgadas à comunidade somente quando se obtêm resultados relevantes para pesquisa científica.

### 3. Considerações finais

Convém chamarmos a atenção para o sentido geral da ética, como ramo próprio da filosofia, e que é possível aplicá-la a diversos setores da atividade humana, incluindo à atividade infocomunicacional, bem como à atividade científica e de pesquisa. Das muitas questões éticas encontradas, optamos por abordar três pontos que demandam importantes discussões e reflexões.

A ciência aberta traz uma série de possibilidades sem precedentes na história e pode garantir o melhor aproveitamento dos investimentos feitos em pesquisa, uma distribuição mais igualitária do conhecimento gerado e a aceleração da disseminação da informação em diferentes esferas da sociedade. Porém, o acesso aberto a dados de pesquisa demanda cuidados em relação à preservação da privacidade dos participantes das pesquisas. Os riscos envolvidos neste contexto foram apontados em diversas pesquisas de mapeamento de percepção sobre acesso aberto a dados de pesquisa de diferentes partes envolvidas (tanto pesquisadores quanto participantes e investidores). Identificou-se a necessidade de discussão dessa e outras questões éticas pelos pesquisadores e, em especial, uma oportunidade para os cientistas da informação.

As indicações de filiações institucionais apresentadas pelos próprios pesquisadores nas publicações refletem-se nas medidas de desempenho das universidades em *rankings*, uma vez que a importância de uma instituição na pesquisa científica mundial é verificada através de sua representatividade em bases de dados internacionais. O fenômeno da múltipla filiação, nesse contexto, levanta uma série de questões para as quais faltam respostas na literatura e nos códigos de ética de editores científicos e de órgãos profissionais. Apesar de uma ampla bibliografia direcionada a autores e editores quanto a dilemas decorrentes da atividade e da publicação dos resultados de pesquisa, o tema da múltipla filiação institucional ainda não figura entre os principais temas discutidos na comunicação científica. Destacam-se, no entanto, as diretrizes apresentadas pela Universidade Federal do Rio Grande do Sul, que regulamentam o registro da produção intelectual de seus docentes quando do envolvimento profissional com outras universidades.

Demonstramos que em alguns casos a publicação em salame é admitida pela comunidade científica e até mesmo necessária, devido à complexidade de algumas pesquisas e à dificuldade de arranjar os resultados obtidos em apenas um artigo científico. Entretanto, em outros casos a publicação em salame não é aceita, tornando-se antiética quando utilizada para mascarar a produtividade dos pesquisadores, além da ausência de efetiva contribuição à ciência. Assim sendo, considera-se pertinente que projetos de pesquisa sejam claramente delimitados desde a sua concepção, com o devido cuidado no planejamento e na execução do processo, no intuito de evitar condutas eticamente questionáveis de publicação.

A relação entre ética e comunicação científica foi aqui abordada, trazendo dilemas que se situam nas práticas de pesquisa e de publicação, no intuito de fomentar a discussão ética sobre práticas científicas e no sentido de evitar a má conduta de investigação, que pode vir a prejudicar todo um campo de conhecimento. As discussões destes enfoques são consideradas fundamentais para a manutenção do status científico das pesquisas e para o desenvolvimento da ciência como um todo.

## Referências Bibliográficas

---

- ABBAGNANO, N. (2007). *Dicionário de filosofia* (5. ed.). São Paulo: Martins Fontes.
- AMERICAN PSYCHOLOGICAL ASSOCIATION [APA] (2017). *Ethical principles of psychologists and code of conduct*. Washington, DC: APA. Recuperado de <http://www.apa.org/ethics/code/index.aspx#Teaching>
- COMMITTEE ON PUBLICATION ETHICS [COPE] (1999). *Guidelines on good publication practice*. Recuperado de <https://publicationethics.org/files/u7141/1999pdf13.pdf>
- COMMITTEE ON PUBLICATION ETHICS [COPE] (2011). *Code of conduct and best practice guidelines for journal editors*. Harleston: COPE. Recuperado de [https://publicationethics.org/files/Code%20of%20Conduct\\_2.pdf](https://publicationethics.org/files/Code%20of%20Conduct_2.pdf)
- CONSELHO FEDERAL DE ODONTOLOGIA [CFO] (2012). *Código de ética odontológica*. Brasília: CFO. Recuperado de [http://cfo.org.br/wp-content/uploads/2009/09/codigo\\_etica.pdf](http://cfo.org.br/wp-content/uploads/2009/09/codigo_etica.pdf)
- CURTY, R. G. (2016). Factors influencing research data reuse in the social sciences: An exploratory study. *International Journal of Digital Curation (IJDC)*, 11(1), 96-117. Recuperado de <http://www.ijdc.net/index.php/ijdc/article/view/11.1.96>
- DAVIS, H. M., & VICKERY, J. N. (2007). Datasets, a shift in the currency of scholarly communication: Implications for library collections and acquisitions. *Serials Review*, 33(1), 26-32. Recuperado de <http://www.tandfonline.com/doi/abs/10.1080/00987913.2007.10765089>
- ELSEVIER (2015). *Ethics in research & publication*. Recuperado de [https://www.publishingcampus.elsevier.com/websites/elsevier\\_publishingcampus/files/Guides/Brochure\\_Ethics\\_2\\_web.pdf](https://www.publishingcampus.elsevier.com/websites/elsevier_publishingcampus/files/Guides/Brochure_Ethics_2_web.pdf)
- HOTTENROTT, H., & LAWSON, C. (2017). A first look at multiple institutional affiliations: A study of authors in Germany, Japan and the UK. *Scientometrics*, 111(1), 285-295.
- INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA [IBICT] (2016). *Ibict lança Manifesto de Acesso Aberto a Dados de Pesquisa Brasileira para Ciência Cidadã*. Brasília: IBICT. Recuperado de <http://www.ibict.br/Sala-de-Imprensa/noticias/2016/ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada>
- INTERNATIONAL COMMITTEE OF MEDICAL JOURNALS EDITORS [ICMJE] (2016). *Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals*. Recuperado de <http://www.icmje.org/icmje-recommendations.pdf>

- JAO, I., KOMBE, F.; MWALUKORE, S., BULL, S., PARKER, M., KAMUYA, D., MARSH, V. (2015). Research stakeholders' views on benefits and challenges for public health research data sharing in Kenya: The importance of trust and social relations. *PLOS ONE*, 10(9), e0135545. DOI:10.1371/journal.pone.0135545.
- KAYE, J. (2012). The tension between data sharing and the protection of privacy in genomics research. *Annual Review of Genomics and Human Genetics*, 13, 415-431. Recuperado de <http://www.annualreviews.org/doi/full/10.1146/annurev-genom-082410-101454>
- KROKOSZ, M. (2015). A autoria na redação científica. *Informação & Informação*, 20(1), 319-333.
- LUDMAN, E. J., FULLERTON, S. M., SPANGLER, L., TRINIDAD, S. B., FUJII, M. M., JARVIK, G. P., BURKE, W. (2010). Glad you asked: Participants' opinions of re-consent for dbGap data submission. *Journal of Empirical Research on Human Research Ethics*, 5(3), 9-16. Recuperado de <https://www.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&retmod=e=ref&cmd=prlinks&id=20831417>
- MARCOVITCH, H., BARBOUR, V., BORRELL, C., BOSCH, F., FERNÁNDEZ, E., MACDONALD, H., NYLENNA, M. (2010). Conflict of interest in science communication: More than a financial issue. *Croatian Medical Journal*, 51(1), 7-15.
- MAYER-SCHONBERGER, V. & CUKIER, K. (2013). Riscos. In: V. Mayer-Schonberger & K. Cukier, *Big Data: Como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana* (pp. 105-118). Rio de Janeiro: Elsevier.
- MCGUIRE, A. L., HAMILTON, J. A., LUNSTROTH, R., MCMULLOUGH, L. B., & GOLDMAN, A. (2008). DNA data sharing: research participants' perspectives. *Genetics in Medicine*, 10(1), 46-53. Recuperado de <http://www.nature.com/gim/journal/v10/n1/full/gim20087a.html?foxtrotcallback=true>
- MEADOWS, A. J. (1999), *A comunicação científica*. Brasília: Briquet de Lemos Livros.
- MOROSINI, M. (2009). Qualidade na educação superior: Tendências do século. *Estudos em avaliação educacional*, 20(43), 165-186.
- NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION (20--?). *World Data Centers*. Recuperado de <https://www.ncdc.noaa.gov/customer-support/world-data-centers>
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT [OECD] (2007). *OECD principles and guidelines for access to research data from public funding*. Danvers: OECD. Recuperado de <http://www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>

- ORGANIZAÇÃO DAS NAÇÕES UNIDAS [ONU] (1948). *Declaração Universal dos Direitos Humanos*. Paris: ONU. Recuperado de <http://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=por>
- PACKER, A. L. & MENEGHINI, R. (2006). Visibilidade da produção científica. In: D. A. Población, G. P. Witter, & J. F. M. Silva, *Comunicação e produção científica: Contexto, indicadores e avaliação* (pp. 235-259). São Paulo: Angellara.
- PORTELLA JR., J. C. (2016). Princípios de Nuremberg e a justiça penal internacional. *Canal Ciências Criminais*, Seção Direito Penal Internacional. Recuperado de <https://canalcienciascriminais.com.br/principios-de-nuremberg-justica-penal/>
- SAYÃO, L. F. & SALES, L. F. (2014). Dados abertos de pesquisa: ampliando o conceito de acesso livre. *RECIIS*, 8(2), 76-92. Recuperado de <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/viewFile/611/1252>
- SMOLČIĆ, V. S. (2013). Salami publication: Definitions and examples. *Biochemia Medica*, 23(3), 237-241. Recuperado de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900084/>
- UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL [UFRGS] (2017). *Document@: Dúvidas mais frequentes na identificação da PI*. Porto Alegre: UFRGS. Recuperado de <http://www.ufrgs.br/documenta/manuais-sabi/politicas-e-procedimentos/registro-de-producao-intelectual/duvidas-mais-frequentes-na-identificacao-da-pi>
- ZABALZA, M. A. (2004). *O ensino universitário: Seu cenário e seus protagonistas*. Porto Alegre: Artmed.

## Notas

<sup>i</sup> Site do ICS: <https://www.icsu.org/about-us>.

<sup>ii</sup> [www.webofknowledge.com](http://www.webofknowledge.com)

<sup>iii</sup> Pesquisas realizadas em 28/08/2017, na Principal Coleção do *Web of Science*, sem restrição de anos nem área do conhecimento, e usando os índices SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI. Na primeira consulta buscou-se pelos tópicos (*research data*) AND (*data sharing*). Na segunda, buscou-se os tópicos (*research data*) AND (*data sharing*) AND (*privacy*).

<sup>iv</sup> [www.periodicos.capes.gov.br](http://www.periodicos.capes.gov.br).

<sup>v</sup> “Os princípios de Nuremberg advieram do estatuto do Tribunal de Nuremberg, criado pelo Acordo de Londres de 1945 e que se destinou ao julgamento dos oficiais nazistas por crimes cometidos contra a paz e a segurança da Humanidade durante a Segunda Guerra Mundial.” (Portella Jr., 2016, documento eletrônico).

<sup>vi</sup> Curty (2016) entrevistou treze pesquisadores norte-americanos das Ciências Sociais.