

Extração e visualização de informação em publicações científicas: análise da distribuição de artigos open access com base em currículos de pesquisadores da comunidade brasileira

Information extraction and visualization in scientific publications: analysis of distribution the open access papers based on researchers curriculum vitae of the brazilian community

Alexandre Ribas Semeler

Universidade Federal do Rio Grande do Sul, Geociência
alex.semeler@icloud.com

Adilson Luiz Pinto

Universidade Federal de Santa Catarina
adilson.pinto@ufsc.br

Rosângela Schwarz Rodrigues

Universidade Federal de Santa Catarina
rosangela.rodrigues@ufsc.br

Juarez Vieira do Nascimento

Universidade Federal de Santa Catarina
juarez.nascimento@ufsc.br

RESUMO: Os periódicos de acesso aberto amplificam e oferecem maior visibilidade para a comunicação científica na web. Além de servirem como base para extração e visualização de informações sobre a produtividade de determinada comunidade científica, por meio deles é possível identificar padrões métricos e realizar a efetiva visualização da interação entre redes de colaboração científica. A distribuição da produção científica de países como o Brasil, sob as vias do acesso aberto, ainda é um desafio a ser mapeado. Assim, o objetivo desse estudo é propor um método para extração e visualização da produção científica de pequenos grupos ou comunidades de cientistas. Os aspectos metodológicos subdividem-se em fases: (A) extração de informação; (B) reorganização e coleta de documentos científicos; (C) classificação e processamento dos dados; e, (D) visualização da informação. Como resultado parcial constatou-se que é pequena representatividade de pesquisadores brasileiros em periódicos de modelo comercial. Os testes realizados mostraram-se úteis para elaborar uma resposta sobre a produção de pesquisadores brasileiros em periódicos de acesso aberto, principalmente, quando publique fora do núcleo de periódicos predominante.

PALAVRAS-CHAVE: Periódico Científico. Acesso Aberto. Distribuição de Artigos. Plataforma Lattes

ABSTRACT: The open access journals amplify and provide greater visibility for scientific research on the web. Serve as a basis for information extraction and information visualization on the productivity of particular scientific community. Through them you can identify metric standards and carry out effective view of the interaction of scientific collaboration networks. The distribution of the scientific production of countries like Brazil, under the way of open access is still a challenge to be mapped. The aim of this study is proposing a method for information extract and visualization of scientific production of small groups or communities of scientists. The methodological aspects are subdivided into phases: (A) information extraction; (B) reorganization and collect of scientific papers; (C) classification and processing of data; and (D) information visualization. As a partly result it was found that there is little representation of Brazilian researchers in commercial model journals. The tests proved useful to prepare a response on the production of Brazilian researchers in open access journals, especially when post outside of predominant journals.

KEYWORDS: Scientific journal. Open Access. Paper Distribution. Lattes Platform.

1 Introdução

Os periódicos de acesso aberto e os repositórios amplificam e oferecem maior visibilidade para a comunicação científica na web. Eles consolidam um importante sistema de publicações fundamentadas no acesso ao conhecimento científico. O Movimento Open Access (OA) pode ser uma das vias que proporciona mudanças ao paradigma de Comunicação da Ciência em redes digitais, com base no conceito de Software Livre e de uma Cultura Livre, no qual esse movimento busca a construção de um domínio público para o acesso a Ciência, permitindo a difusão e a reutilização de dados e informações científicas (ABADAL, 2012).

As dimensões do Movimento Open Access extrapolam questões econômicas, éticas, territoriais e tecnológicas. Além dessas dimensões surgem outras necessidades relacionadas ao acesso aberto, como a criação de métodos e indicadores alternativos para avaliação do desenvolvimento e da distribuição das publicações científicas produzida em diferentes partes do mundo.

Nesse sentido, o acesso aberto pode ser utilizado como fonte para visualização da produção científica de regiões do mundo em desenvolvimento e subdesenvolvida. Essas regiões antes esquecidas por não atenderem os padrões de qualidade estabelecidos pelo modelo editorial comercial/proprietário de empresas como Reed Elsevier, John Wiley, Springer, Blackwell, Taylor & Francis, Thonpson Routers, que mais de 350 anos instauraram os moldes dos periódicos científicos,

consolidando-se como carteis da publicação científica.

Assim, como hipótese arriscamos afirmar que a Ciência produzida em países como o Brasil não é visualizada no todo pelos editores comerciais; e que somente parte dos pesquisadores nacionais conseguem atingir os critérios e as regras internacionais desses editores, o que condiciona o sistema de publicação científica ser realizada pela academia e por conseguinte a visibilidade das descobertas científicas brasileiras ficam circulando, na sua maioria, internamente.

Nesse contexto, a questão de pesquisa busca responder: qual a distribuição da produção de artigos científicos de pesquisadores brasileiros depositada em periódicos de acesso aberto e/ou periódicos comerciais? A fim de responder nossa inquietação traçamos como objetivo geral extrair informações de currículos de professores/docentes de programas de pós-graduação para identificar e visualizar a sua distribuição. A ideia foi mapear: onde e em que periódicos foram publicados os artigos e quais os modelos preponderantes de output¹ dos mesmos no que tange os indicadores de: país do editor, Qualis/Capes² do periódico e o tipo de acesso ao artigo.

Como atitude inicial de pesquisa descartamos os metadados referenciais e os serviços de citação extraídos de bases de dados comerciais de países desenvolvidos, como Web of Science Thomson Scientific e Scopus Elsevier. Isso por acreditar que essas fontes não representam a Ciência produzida localmente no Brasil devido a língua, ao modelo de indexação e as exigências adotadas pela Thomson Scientific e Elsevier para inserir um periódico em sua base de dados.

Estes pontos tornam necessária a proposição de soluções alternativas para fontes de dados e para análise cienciométrica do output das publicações científicas brasileiras que estejam fora do núcleo Web of Science e Scopus. Nesse contexto, o estudo utilizou fontes de dados locais brasileiras como: o Web Qualis e a Plataforma Lattes do Conselho Nacional de Pesquisa Brasileiro (CNPq).

A plataforma Lattes permite a coleta automatizada de informações sobre a produção bibliográfica de pesquisadores brasileiros, incluindo o Qualis/Capes dos periódicos. O Lattes possui cerca de 200 mil currículos que são o padrão nacional

1: Como indicadores de input há que se considerar a importância dos insumos necessários ao fortalecimento da comunidade científica de pesquisadores e as condições encontradas para a efetivação de suas investigações. Por outro lado, como indicadores de output, têm-se as medidas dos produtos, isto é, a validação do conhecimento gerado, comprovada com a aceitação pelos pares (disseminação) e pela sociedade (divulgação). Nesta etapa devem-se considerar, ainda, os veículos adequados para a disseminação e divulgação das pesquisas e os recursos que garantam sua ampla visibilidade e formas de acessibilidade. (NORONHA; MARICATO, p. 119, 2008)

2: O Qualis constitui-se num sistema brasileiro de avaliação de periódicos, mantido pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Ele relaciona e classifica os veículos utilizados para a divulgação da produção intelectual dos programas de pós-graduação "stricto sensu" quanto ao âmbito da circulação (local, nacional ou internacional) e à qualidade (A1, A2, B1, B2, B3, B4, B5, C), por área de avaliação. (CAPES, 2015).

brasileiro para a avaliação das atividades científico-acadêmicas/profissionais de pesquisadores de programas de pós-graduação, projetados para mostrar informação pública e individual de cada usuário cadastrado na plataforma (MENA-CHALCO; CESAR-JR, 2013).

Os currículos são Lócus onde pesquisadores brasileiros de pós-graduação são obrigados a depositar informações sobre sua produção científica, o que o transforma na fonte ideal para analisar o output geral de programas de pós-graduação brasileiros. Os pesquisadores e docentes de pós-graduação são sujeitos condicionados por exigência a depositar suas informações pessoais e profissionais relacionadas às suas descobertas científicas na Plataforma Lattes, por razões diversas: de status científico (visibilidade perante sua comunidade de pares); econômica (bolsas de estudos) entre outros fatores que tornam esses sujeitos mais acostumados a depositar sua produção científica em plataformas que avaliam seu output.

Assim, com base nos currículos Lattes dos pesquisadores nacionais propõe-se extrair informações sobre a categoria artigos completos publicados em periódicos para analisar as variáveis entre os indicadores de dispersão em relação ao país do editor, ao Qualis do periódico e o tipo de acesso ao artigo.

O estudo se divide em uma breve explanação sobre os aspectos teóricos do sistema de publicação periódica, citando os modelos preponderantes de negócio comercial, porém focando esforços em produzir subsídios sobre o modelo de acesso aberto a informação científica.

Os aspectos metodológicos de pesquisa foram baseados na realização de testes e observações a respeito do output dos artigos científicos publicados por pesquisadores brasileiros de pós-graduação. Por fim, os procedimentos metodológicos compreenderam a utilização de: (A) métodos de extração de informação; (B) reorganização das informações coletadas na Plataforma Lattes; (C) classificação e processamento dos dados, e; (D) métodos de visualização da informação para analisar a dispersão dos artigos em relação ao país do editor, ao Qualis do periódico, e ao tipo de acesso ao artigo.

2 Aspectos sobre o sistema de publicações periódicas

No que tange os aspectos do scholarly journal publishing, no campo editorial em específico, existem modelos de publicação comerciais, híbridos e de acesso aberto. O surgimento desse último modelo de scholarly journal system

afeta o modelo comercial de negócios. Os modelos comerciais preponderantes são o Australiano, o Britânico, o Dinamarquês, o Holandês e o Estadunidense. No entanto, os modelos alternativos, como author-pays, o auto-arquivamento de artigos publicados já impactam sobre a economia global do scholarly journal system comercial (KING, TENOPIR, 2011), porém ainda existe resistência ao modelo OA (BEALL, 2015), em especial por que existe um monopólio editorial e isso não agrada a alguns países e editoras com suas bases de dados particulares.

As formas de concorrência científica, nacional e internacional, tratam dos processos e estruturas de poder que permeiam a busca pelo controle científico, seguindo o exemplo do campo editorial. Nesse contexto, os periódicos científicos podem ser categorizados conforme algum tipo de escala, na qual estão envolvidos prestígio, autoridade e visibilidade. Assim, a distinção entre nacional e internacional no que tange a estrutura hierárquica de periódicos consiste em categorização de dois níveis: no topo prevalece um único conjunto de títulos; abaixo estão aqueles que desfrutam de diferentes níveis de visibilidade e invisibilidade (GUÉDON, 2010).

Outra questão que o Jean-Claude Guédon destaca é o idioma, ele explica que em um momento pré-guerra o alemão, o francês e o inglês eram o triunvirato da Ciência, sendo o pós-guerra fato crucial em que definiu o Inglês como a língua franca das Ciências. Tal situação impulsionou o surgimento de cartéis científicos e do surgimento de um mercado internacional para as publicações científicas. O que gerou as condições essenciais para empresas como Thomson Scientific e Elsevier no começo dos anos 1950, a configurarem-se como as maiores editoras de conteúdo científico, técnico e médico do mundo (GUÉDON, 2010).

Os cartéis de periódicos, que até os dias de hoje utilizam-se da cooperação para criar uma forma centralizada de Ciência, fundamentam-se solidamente em evidências estatísticas para listar os títulos de seus periódicos, que devem ter credibilidade internacional. Os cartéis permitem o oligopólio ao operar como monopólio a indexação de periódicos. Sendo a Thompson Scientific e Elsevier as empresas que decidem de forma unilateral e obscura quantos títulos de periódicos serão incluídos em suas listagens de periódicos de qualidade (GUÉDON, 2010).

Dessa forma, como alternativa o autor aponta para o movimento de acesso aberto e analisa o centro e a periferia, questionando como a presença ou a visibilidade da ciência periférica é tratada em fóruns centrais e em fóruns periféricos? Há duas tarefas aplicadas às nações em desenvolvimento: (a) a construção de um catálogo

nacional de atividades científicas; e (b) as suas contribuições para ciência nacional, as quais põem em dúvida o que são as contribuições científicas universais. Guédon explica que o universalismo científico, apesar de significar que os resultados científicos são igualmente validados em todos os lugares, onde ele não afirma porém questiona se os resultados são igualmente úteis ou aplicáveis em todos os lugares, tampouco que as pesquisas científicas precisam adotar os assuntos da moda para serem validados (GUÉDON, 2010).

No que se segue, surge a necessidade de se superar a divisão entre ciência predominante e ciência periférica, sendo essa uma função do movimento do acesso aberto. O movimento de acesso aberto (open access) representa uma tendência atual dos estudos sobre a comunicação científica e surge como uma alternativa para divulgação e amplificação da visibilidade das publicações na web. O conceito de acesso aberto foi desenvolvido inicialmente na conferência de Budapeste (2001). Os seus principais veículos de comunicação e distribuição são publicações de revistas científicas e iniciativas de repositórios.

3 Publicações científicas de acesso aberto

O acesso à informação científica revela-se como uma tendência de estudos para sua comunicação. Inúmeras iniciativas consolidam as preocupações da comunidade científica com a filosofia open, ou seja, em direção ao uso de ferramentas, estratégias e metodologias para o desenvolvimento de aplicações de computador (software livre), arquivos abertos (open archive) e acesso aberto (open access) aos conteúdos. Essas novas iniciativas surgiram com a internet e atualmente são muito influentes no que se referem aos processos da comunicação científica (ABADAL, 2012; BRKOVI, 2011; CRAWFORD, 2011; GUÉDON, 2009; HARNAD, 2009; MUELLER, 2006; SUBER, 2004).

O movimento de acesso aberto (open access) representa uma tendência atual dos estudos sobre a comunicação científica, que surge como uma alternativa para divulgação e amplificação da visibilidade das publicações na web. Uma publicação científica OA define-se pela capacidade de estar disponível online, de forma gratuita e livre de direitos autorais ou restrições de licenciamento.

O conceito de acesso aberto foi definido ao longo de três conferências: Budapeste (2001), Bethesda (2003) e Berlim (2003). Nesses eventos definiu-se que publicações open access compreendem toda literatura disponibilizada de forma

livre pela internet e que ela deve permitir a qualquer utilizador ler, fazer download, copiar, distribuir e imprimir completamente o conteúdo dessas publicações. O objetivo do movimento OA é remover as barreiras entre todos os usos legítimos da literatura acadêmica. Ainda, os principais veículos de comunicação e distribuição de publicações são as revistas científicas. Esse veículo pode ser classificado em duas vias (green route e gold route). Esses caminhos ou vias representam os tipos possíveis para sistemas de publicações de acesso aberto (ABADAL, 2012; BRKOVI, 2011; CRAWFORD, 2011; GUÉDON, 2009; HARNAD, 2009; MUELLER, 2006; SUBER, 2004).

As dimensões das publicações científicas Open Access podem ser exemplificadas, conforme o quadro 1:

Quadro 1: Dimensão das Publicações Open Access

Green Route	The author can self-archive at the time of submission of the publication whether the publication is grey literature, a peer-reviewed journal publication, a peer-reviewed conference proceedings paper or a monograph
Golden Route	The author or author institution can pay a fee to the publisher at publication time, the publisher thereafter making the material available 'free' at the point of access.
Preprints	Preprints are articles that are pre-peer-review
Postprints	Postprints are articles that are post-peer-review
eprints	eprints can be either preprints or postprints but in electronic form
White Literature	White literature is peer-reviewed, published articles
Grey Literature	Grey literature is preprints or internal 'know-how' material

Fonte: Jeffrey (2006)

As estratégias da green route, ou via verde, propõe que as universidades, institutos de pesquisa e outras instituições de cunho científico ou tecnológico criem seus próprios repositórios com o propósito de armazenar e dar maior visibilidade para a produção científica de seus pesquisadores. Nas palavras de Leite (2009, p.17), “[...] a via verde significa o sinal verde dos editores científicos para o arquivamento da produção científica pelos autores em repositórios digitais”. Já a gold route, propõe aos editores científicos a criação e/ou conversão das revistas científicas comerciais para revistas científicas de acesso livre. Conforme Leite (2009, p. 7) “[...] a via dourada significa o acesso aberto promovido nos próprios periódicos científicos, de modo que os artigos científicos possam ser disseminados sem restrições de acesso ou uso”. Crawford (2011, p.15, tradução nossa) expande a definição de publicações OA explicando que “[...] contribuições open access incluem também os resultados originais de pesquisas científicas, dados brutos e metadados [...]”.

4 Procedimentos Metodológicos

As questões relacionadas a produção científica em ambientes de comunicação digital estão no cerne das preocupações e atendem em parte ao objetivo desse artigo: extrair e visualizar informação sobre a distribuição da produção científica de pequenos grupos ou comunidades de cientistas brasileiros por meio de métodos e tecnologias alternativas utilizados em projetos de e-Science³ e Ciência da Informação.

Tendo em vista uma abordagem de pesquisa quantitativa que possibilitou a análise métrica e visual de informação científica; buscou-se também a exploração de tecnologias utilizadas para extração, coleta, organização, análise e visualização de informações em fontes brasileiras de informação como a Plataforma Lattes e o Web Qualis/Capes.

No que tange a extração de informação, o corpus investigado por essa proposta foi os currículos de pesquisadores, vistos aqui como um tipo de fonte de dados/informação, onde os próprios pesquisadores indexam informações sobre sua produção. Em outra esfera buscou-se visualizar a distribuição e os tipos de outputs coletados dos currículos.

A operacionalização da proposta inicialmente se fundamentou na exploração de softwares científicos, como ScriptLattes⁴ usado para extração e visualização de informações em currículos. Outro grupo de aplicativos foi destinado para normalização, recuperação e organização da informação digital, como Open Refine⁵, Hermetic Word Frequency Counter Advanced Version (HWC)⁶. Já o terceiro grupo de aplicativos foram softwares de visualização de informações, como Ucinet e NetDraw⁷.

Tais programas compõem o corpus tecnológico explorado e utilizado para simular o método proposto, além de permitirem a argumentação e a instrumentação das seguintes etapas metodológicas: (A) extração de informação; (B) reorganização da coleta; (C) classificação e processamento dos dados, e; (B) visualização da informação.

Por fim, após organizados e replicados em uma base de dados local de um

3: [...] e-Science as the term embracing many of the features commonly associated with how scholarship is conducted in a network environment, utilizing Internet-based tools and involving collaboration among scholars often separated by large distances on a global scale. [...] In list form, these features include: (a) International collaboration among researchers; (B) Increasing use of high-speed interconnected computers, applying Grid architecture; (c) Visualization of data; (d) Development of Internet-based tools and procedures; (e) Construction of virtual organizational structures for conducting research; (f) Electronic distribution and publication of Findings (JANKOWSKI, 2007, p. 553).

4: <http://scriptlattes.sourceforge.net/>

5: <http://openrefine.org/>

6: <http://www.hermetic.ch/>

7: <http://analytictech.com/>

repositório de dados e documentos, foram aplicadas técnicas de análise métrica nos metadados coletados nas plataformas de currículos. Estas técnicas métricas foram dedicadas à identificação de tendências e associações contidas nos dados. A aplicação de métodos de visualização da informação permitiu reflexões sobre a evolução e a coocorrência ao longo do tempo de variáveis de dispersão do país do editor, o Qualis/Capes do periódico e o tipo de acesso ao artigo.

Nesse sentido, se descreveu as fases do método proposto para extração e visualização de informação sobre comunidades de cientistas locais (Local Science), conforme as informações registradas em seus currículos:

- (a) extração de Informação, fase que buscou informação quantitativa sobre os padrões e frequências de publicações;
- (b) reorganização da coleta de registros sobre a produção técnica/científica de documentos científicos, através da criação de uma base de dados local replicada;
- (c) classificação e processamento dos dados, durante esse processo foi necessária a limpeza e normalização dos dados coletados a partir de clusterização, classificação, inclusão, proximidade e equivalência dos registros;
- (d) visualização da informação, estratégia para elaboração, interpretação e visualização dos resultados sobre a distribuição da pesquisa em publicações científicas.

Figura 1: Procedimentos Metodológicos



Fonte: Dados da Pesquisa (2015)

Nesse sentido, surgiu a necessidade de se realizar um teste a tais procedimentos metodológicos.

5 Resultados: estudo sobre o Programa de Pós-Graduação em Educação Física (PPGEF)

Os resultados desse estudo partem da prática das fases do esboço no método proposto. Assim, o corpus de análise prototípica compõem-se de informações extraídas de 28 currículos de pesquisadores/docentes do Programa de Pós-Graduação em Educação Física (PPGEF) da Universidade Federal de Santa Catarina, dentre os quais cinco são bolsistas de produtividade do CNPq. O programa possui Nota 5 e já formou cerca de 300 mestres e 20 doutores; os cursos possuem as seguintes áreas de concentração: (a) - Atividade Física Relacionada à Saúde; (b) - Teoria e Prática Pedagógica em Educação Física, e; (c) - Biodinâmica do Desempenho Humano (PPGEF, 2015).

O corpus de análise representa uma parte do total da produção bibliográfica do PPGEF/UFSC, que são os artigos publicados em periódicos. Com objetivo de avaliar a distribuição dos artigos segundo: o Qualis/Capes, o país do editor do periódico e o tipo de acesso realizamos a sua distribuição, conforme a tabela 1.

Tabela 1: Quantidade total da Produção Bibliográfica Geral e total de artigos

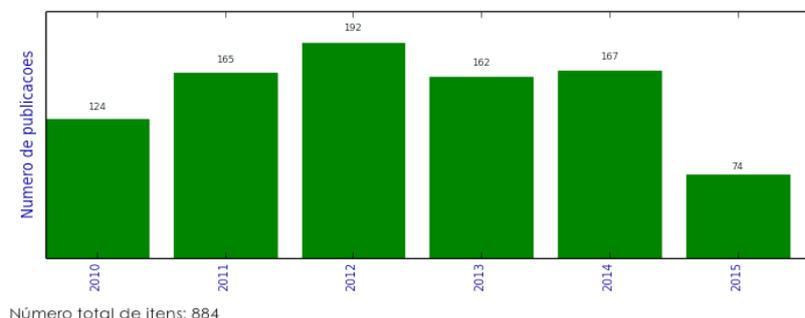
Data	Produção Geral	Artigos em Periódicos
2010	383	124
2011	400	165
2012	408	192
2013	336	162
2014	401	167
2015	211	74
Total	2139	884

Fonte: Dados da Pesquisa, 2015.

O total da Produção Bibliográfica extraído foi de $n=2.174$. No entanto, 35 produções não continham data e, por tanto, não foram analisadas. No que tange a produção intelectual geral do programa, os itens distribuídos quantitativamente em um montante foram de: (a) Artigos completos publicados em periódicos ($n=884$); (b) Livros publicados/organizados ou editados ($n=39$); (c) Capítulos de livros publicados ($n=132$); (d) Textos em jornais de notícias/revistas ($n=13$); (e) Trabalhos completos publicados em anais de congressos ($n=125$); (f) Resumos expandidos publicados em anais de congressos ($n=48$); (g) Resumos publicados em anais de congressos ($n=613$); (h) Artigos aceitos para publicação ($n=52$); (i) Apresentações de trabalho ($n=177$); (j) Demais tipos de produção bibliográfica ($n=52$)⁸.

O distribuição do total de Artigos completos publicados em periódicos ($n=884$) pode ser visualizada no Gráfico 1.

Gráfico 1: Total de Artigos completos publicados em periódicos por ano



Fonte: Dados da Pesquisa, 2015.

8: Os dados foram extraídos automaticamente com scriptLattes V8.10. (desenvolvido no NUVEM/UFABC e no CCSL-IME/USP por Jesús P. Mena-Chalco e Roberto M. Cesar-Jr). Os resultados estão sujeitos a falhas devido a inconsistências no preenchimento dos currículos Lattes. Relatório criado com produções de 2010 até 2015. Data de processamento: 22/09/2015 [16:00:31].

Os artigos publicados em periódicos passam por um rigor de avaliação maior no que tange a qualidade, pois dependem de um processo de submissões a cegas em que há um sistema peer review e se supõe que a revisão feita pelos pares está eticamente amparada pela comunidade científica. Nesse contexto isolou-se os artigos completos publicados em periódicos, os quais correspondem a 51,37% do total da produção intelectual do programa. Outro critério foi a adoção de frequência de publicação de artigo por periódico => 4 no intervalo de tempo entre janeiro de 2010 a setembro de 2015.

Entretanto, é possível que alguns dados extraídos dos currículos da Plataforma Lattes possam ter apresentados falhas na data atual devido a inconsistências no preenchimento dos currículos, como falta de indexação de dados e troca de ordem de autores.

Desta forma, entra em ação toda uma normativa para a realização do tratamento e processamento das informações extraídas. O processo se fundamentou em: (a) Normalização; (b) Limpeza; (C) Conversão; (d) Análise de frequências de artigo por periódico. O aspecto tecnológico dessa etapa centrou-se na operacionalização e uso de softwares como: Open Refine, Hermetic Word Frequency Counter Advanced Version (HWC). Tais programas de computador subsidiam a argumentação, a instrumentação e a interpretação dos resultados.

Após analisar a frequência, os registros duplos e as inconsistências contatou um total de n=632 artigos únicos publicados em periódicos de alta frequência, ou seja, 29,6% do que foi extraído da Plataforma Lattes gerou inconsistências. Depois de realizar a organização e o tratamento dos dados dos currículos, 71,4% da extração foi validada como registro sem inconsistências e em periódico de alta frequência (=> 4). Assim, a distribuição de artigos por ano e Qualis/Capes se deu conforme a tabela 2.

Tabela 2: Artigos por Data e Qualis/Capes dos Periódicos

Data							
Qualis/Capes	2010	2011	2012	2013	2014	2015	Total
A1	6	15	16	13	21	8	79
A2	29	22	30	28	18	8	135
B1	27	49	77	38	63	19	273
B2	6	15	18	17	20	2	78
B3	1	1	1	1	1		5
B4	12	5	8	10	2		37

B5	5	6	3	2	2	4	22
C		2	1				3
Total	86	115	154	109	127	41	632

Fonte: Dados da Pesquisa, 2015.

O critério de frequência de publicação de artigo por periódico => 4 no intervalo de tempo (jan./2010 - set/2015) revelou que 12,5% da produção do programa se concentrou em periódicos A1 e 21,3% em A2. A maior parte da produção do programa está contemplada em periódicos B1 (43,1%). No geral, os n=632 artigos estão distribuídos em 63 títulos de periódicos.

Com relação ao tipo de acesso: (a) OA = periódico de acesso aberto; (b) P = periódico comercial, constatou-se que n=452 do total de artigos (71,5%) são publicados em periódicos de frequência => 4 e estão em acesso aberto. Os outros 29,5% são publicados por instituições de 14 países diferentes, divididos entre as editoras comerciais como Elsevier, Springer, Wiley, Taylor & Francis Group e Thomson Scientific.

Outro ponto a se destacar é o grau de endogenia da produção do PPGEF, onde n=100 artigos (15,8%) foram publicados no período do próprio Centro em que o Programa de Pós-Graduação está inserido.

O total de periódicos por ordem de frequência de publicação de artigo pode ser visualizado na tabela 3.

Tabela 3 Lista de Distribuição Publicações por periódicos (artigos =>4)

Título Periódico	Freq/Artigos	Acesso
Revista Brasileira de Cineantropometria Desempenho Humano (Online)	57	OA
Revista Brasileira de Cineantropometria Desempenho Humano (Impresso)	43	OA
Motriz : Revista de Educação Física (Online)	36	OA
Revista Brasileira de Atividade Física e Saúde	35	OA
Revista da Educação Física (UEM, Online)	26	OA
Revista Brasileira de Ciências do Esporte	23 (34,81%)	P
Revista Paulista de Pediatria (Impresso)	21	OA
Work (Reading, MA)	19	P
Cadernos de Saúde Pública (ENSP, Impresso)	18	OA
Lecturas Educación Física y Deportes (Buenos Aires)	18	OA
Revista Brasileira de Ciência e Movimento	18	OA
Ciência e Saúde Coletiva (Impresso)	14	OA
Journal of Strength and Conditioning Research	14	P
Revista Brasileira de Epidemiologia (Impresso)	14	OA
Revista Brasileira de Medicina do Esporte (Impresso)	14	OA
Revista Brasileira de Educação Física e Esporte (Impresso)	13	OA
Revista Mineira de Educação Física (UFV)	13	OA

Motricidade (Santa Maria da Feira)	12	OA
Pensar a Prática (Online)	12	OA
Human Movement	11 (33,38%)	P
Revista da Educação Física (UEM. Impresso)	9	OA
Revista de Nutrição (Impresso)	9	OA
Revista de Saúde Pública (Impresso)	9	OA
Journal of Physical Activity Health	8	P
Journal of Science and Medicine in Sport	8	P
Neuroscience	8	P
Preventive Medicine (1972. Print)	8	P
Journal of Sports Sciences (Print)	7	P
Plos One	7	OA
Revista de Salud Publica	7	OA
Revista Andaluza de Medicina del Deporte	6	P
Revista Brasileira de Ciências do Esporte (Online)	6	OA
Revista Portuguesa de Ciências do Desporto	6	OA
Estudos Interdisciplinares sobre o Envelhecimento	5	OA
Isokinetics and Exercise Science	5	P
Journal of Community Health	5	P
Movimento (UFRGS. Impresso)	5	OA
Neurotoxicity Research	5	P
Revista Brasileira de Fisiologia do Exercício	5	OA
Revista Brasileira de Geriatria e Gerontologia (UnATI. Impresso)	5	P
Revista da Associação Médica Brasileira (1992. Impresso)	5	P
Clinical Physiology and Functional Imaging (Print)	4	P
European Journal of Sport Science (Print)	4	P
International Journal of Sports Medicine	4	OA
International Journal of Sports Physiology and Performance	4	P
International Sportmed Journal for Fims	4	P
Journal of Sports Science and Medicine	4	P
Licere (Centro de Estudos de Lazer e Recreação. Online)	4	OA
Nutrition Journal	4	OA
Pensar a Prática (Impresso)	4	OA
Revista de Saúde Pública (USP. Impresso)	4	OA
Revistas com =<3	23 (31,81%)	--

Fonte: Dados da Pesquisa, 2015.

Além da distribuição dos artigos por tipo de acesso, também foi possível cruzar o país do editor com o Qualis/Capes do periódico, conforme a tabela 4, onde foi possível identificar alguns cenários de centralização da ciência local por categorias de indexação.

Essa categoria de indexação passa pela atribuição da editora e sua relação direta com a base de dados, como as revistas A1 que estão vinculadas a: John Wiley, Springer, Blackwell, Taylor & Francis e Thonpson Routers indexadas na Web of Science. Já as revistas A2 estão vinculadas Reed Elsevier e, conseqüentemente, sua indexação está articulada a Scopus. A partir das revistas B1 sua articulação está

A figura 2 pode ser utilizada para visualizar a distribuição geográfica do editor do periódico em relação ao Qualis/Capes. A produção do PPGEF de maior relevância (Qualis/Capes A1) concentrou-se em periódicos de Países como: Estados Unidos da América, Inglaterra, França, Austrália, Polônia, Escandinávia, Canadá e Alemanha, que no período de 5 anos somaram um total de n=79 artigos. Os periódicos brasileiros são de acesso aberto, pontuando n=135 artigos e começaram a pontuar a partir do Qualis/Capes A2. Nota-se que os pesquisadores do PPGEF concentram a maior parte de seus artigos (n=431) em revistas B1, brasileiras e de acesso aberto.

6 Conclusões

Após realizarmos a extração de dados dos currículos Lattes foi possível reorganizar os registros sobre os documentos científico em uma base de dados local, com o objetivo de analisar a produção de artigos de 28 pesquisadores/docentes do Programa de Pós-Graduação em Educação Física (PPGEF). O que permitiu constatar que 29,6% do que foi extraído da Plataforma Lattes e do Web Qualis/Capes possuía inconsistências de dados, o que resultou na reorganização, a classificação e o processamento de dados extraídos da Plataforma Lattes.

O nível de inconsistência se dá devido ao preenchimento equivocado dos currículos pelos pesquisadores, no qual também existe falhas na ferramenta ScriptLattes, em especial pelas proibições impostas pelo CNPq em fornecer dados sobre os pesquisadores brasileiros, onde o sistema de CAPCHA, implantado em agosto de 2015, prejudica a extração dos dados e, conseqüentemente, os estudos de análises sobre os dados contidos na Plataforma.

Outro ponto é o resultado sobre o output de artigos completos publicados em periódicos pelo PPGEF/UFSC. A pesquisa constatou que a maioria dos artigos é publicada em acesso aberto e por editores brasileiros, onde o Qualis/Capes do periódico é predominantemente B1. Também foi possível constatar que parte da produção de artigos publicados em periódicos A1 estão contidos em periódicos comerciais de editores como Elsevier, Espringer, Wiley, Taylor & Francis Group e Thomson Scientific.

Em se tratando de revistas do estrato A1, independente da área de

concentração, a tendência é que tais periódicos sejam de editoras comerciais, visto que os parâmetros utilizados no Brasil para a classificação de revistas são as grande bases de dados comerciais e multidisciplinares, como a Web of Science, Scopus, Pro-Quest e Ebscoh.

Os métodos de visualização da informação facilitaram a análise da produção desses pesquisadores. No que tange a sua internacionalização e o Qualis/Capes dos periódicos, os pesquisadores do PPGEF costumam publicar 12,5% da produção do programa em periódicos A1, e que são periódicos de editores: Australianos, Britânicos, Dinamarqueses/Holandeses e Estadunidense, que seguem o modelo comercial.

A distribuição dos artigos em publicações científicas que seguem o modelo de acesso aberto preponderou sob essa comunidade, o que justifica a ideia introdutória sobre a pouca representatividade de pesquisadores brasileiros em editores comerciais, como Reed Elsevier, John Wiley, Springer, Blackwell, Taylor & Francis, Thonpson Routers. O processo se deve ao fato de estas editoras também serem responsáveis pelas bases de dados que o Governo brasileiro assina a partir do Portal de Periódicos da Capes. Assim, esse estudo constatou que 71,5% dos artigos publicados (n=452 dos n=632) estão fora do núcleo Web of Science e Scopus, o que comprova a necessidade se estudar fontes de dados locais para avaliar a produção científica de pesquisadores brasileiros.

Os testes realizados nesse estudo mostraram-se uteis para se elaborar respostas sobre output de pesquisadores Brasileiros no que tange a análise das variáveis investigadas (o país do editor, o Qualis/Capes do periódico e o tipo de acesso ao artigo), principalmente porque é possível indicar o que está sendo publicado fora do núcleo de Ciência Predominante dos editores comerciais. Nesse contexto, o estudo também pode ser útil àqueles que buscam visualizar o output da suposta Ciência periférica.

No que tange os aspectos tecnológicos, ressalta-se a necessidade da utilização de softwares científicos, como o ScriptLattes, Open Refine e Hermetic Word Frequency Counter Advanced Version (HWC) para realizar estudos sobre a produção científica, entre outros possíveis resultados que deriva o método proposto nesse artigo, podendo revelar: (A) Relatórios de produção acadêmica; (B) Criação de grafos de colaboração acadêmica; (C) Exploração de indicadores alternativos; (D) Caracterizar redes sociais e identificar comunidade de colaboração entre grupos de

pesquisa; (E) Estudar e caracterizar a evolução temporal de co-autorias e a dinâmica de colaborações.

Por fim, acredita-se que a Plataforma Lattes é uma fonte confiável, mas que necessita de revisão para o seu uso em análises métricas sobre cientistas brasileiros, principalmente se forem de uma comunidade específica, como um programa de pós-graduação ou uma unidade departamental.

Referências

ABADAL, Ernest. Acceso abierto a la ciencia. Barcelona: Editorial UOC. (Colección El profesional de la información), 2012. Disponível em: <<http://diposit.ub.edu/dspace/bitstream/2445/24542/1/262142.pdf>> Acesso em: 15 ago. 2015.

BEALL, Jeffrey. Is SciELO a publication favela?. In: Scholarly Open Access: Critical analysis of open-access publishing, 2015. Disponível em: <<https://www.emeraldcityjournal.com/2015/07/is-scielo-a-publication-favela/>>. Acesso em: 01 out. 2015.

BERLIM. Berlin Declaration on open access to knowledge in the Sciences and Humanities. 2003. Disponível em: <https://www.fu-berlin.de/sites/open_access/weiteres/Veranstaltungen/oa_berlin/poster/Berlin-Declaration_Simone-Rieger_MPIWG.pdf>. Acesso em: 10 jan. 2018.

BETHESDA. Bethesda Statement on Open Access Publishing. 2003. Disponível em: <<http://www.earlham.edu/%7Eepeters/fos/bethesda.htm>>. Acesso em: 10 jan. 2018.

BRKOVI, M. Academic Authors, Scientific Information and Open Access Publishing. In: KATSIRIKOU, A. Open Access to STM Information: Trends, Models and Strategies for Libraries. Boston: International Federation of Library Associations, 2011.

BUDAPESTE. Budapest Open Access Initiative. 2001. Disponível em: <<http://www.soros.org/openaccess/read.shtml>>. Acesso em: 10 jan. 2018.

CAPES. Relatório de Gestão de Exercício 2015. 2015. Disponível em: <http://www.capes.gov.br/images/stories/download/Contas_Publicas/2015_Relatorio_de_Gestao.pdf>. Acesso em: 22 jan. 2018.

CRAWFORD, W. Open access: what you need to know now. Chicago: ALA Editions,

2011.

GUEDÓN, J.C. Acesso Aberto e divisão entre ciência predominante e ciência periférica. In: FERREIRA, S.M.; TARGINO, M.G. (Orgs.). *Acessibilidade e visibilidade de revistas científicas eletrônicas*. São Paulo: Editora São Paulo, 2010. cap. 1, p. 21-78.

GUEDÓN, J.C. It's a Repository, It's a Depository, It's an Archive...: open access, digital collections and value. *Arbor Ciencia, Pensamiento y Cultura*, CLXXXV, 737, maio-junio, p. 581-595. 2009. Disponível em: <arbor.revistas.csic.es/index.php/arbor/article/download/315/316> Acesso em: 15 ago. de 2015.

HARNAD, S. Open access scientometrics and the UK Research Assessment Exercise. *Scientometrics*, v. 79, n. 1, p. 147-156, 2009.

JANKOWSKI, N. W. Exploring e-Science: An Introduction. *Journal of Computer-Mediated Communication*, v. 12, n. 2, p. 549- 562, 2007.

JEFFERY, T. Open access: an introduction. *ERCIM News*, n. 63, 2006. Disponível em: <http://www.ercim.eu/publication/Ercim_News/enw64/jeffery.html>. Acesso em: 15 ago. de 2015.

KING, D.; TENOPIR, C. Some Economic Aspects of the Scholarly Journal System. *Annual Review of Information Science and Technology*. v. 45, p. 295-366, 2011.

LEITE, F.C.L. Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto. Brasília: IBICT, 2009. Disponível em: <<http://livroaberto.ibict.br/bitstream/1/775/4/Como%20gerenciar%20e%20ampliar%20a%20visibilidade%20da%20informa%C3%A7%C3%A3o%20cient%C3%ADfica%20brasileira.pdf>>. Acesso em: 22 jan. 2018.

MENA-CHALCO, J. P.; CESAR-JR R., M. Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes. Capítulo do livro *Bibliometria e Cientometria: reflexões teóricas e interfaces* São Carlos: Pedro & João, páginas 109-128, 2013.

MUELLER, S.P.M. A comunicação científica e o movimento de acesso livre ao conhecimento. *Ciência da Informação*, v.35, n.2, p.27-38, 2006. Disponível em: <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=925&layout=abstract>>. Acesso em: 3 jul. 2012.

NORONHA, D.; MARICATO, J. Estudos métricos da informação: primeiras

aproximações. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Florianópolis, p. 116-128, abr. 2008. ISSN 1518-2924. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13nesp1p116>>. Acesso em: 28 set. 2015.

SUBER, P. A very brief introduction to open access. 2004. Disponível em: <<http://www.earlham.edu/~peters/fos/brief.htm>>. Acesso em: 15 ago. 2015.