



Exploração colaborativa através do compartilhamento de dados de citações do Google Scholar

Collaborative exploitation through citation data sharing of the Google Scholar

Tiago Rodrigo Marçal Murakami*

Sibele Fausto**

Ronaldo Ferreira de Araujo***

RESUMO

A falta de indexação dos títulos de revistas científicas de Ciências Humanas e Sociais em bases de dados comerciais restringe a investigação sobre seu impacto. O Acesso Aberto, ferramentas como o *Google Scholar* (GS) e aplicativos de processamento de dados permitem a busca e a recuperação de citações de artigos, sinalizando uma alternativa para os estudos sobre o impacto da produção científica publicada nessas áreas. Este estudo apresenta um projeto piloto de compartilhamento de dados de citações de periódicos para a investigação colaborativa por parte da comunidade de cienciométrica brasileira com o objetivo de incentivar uma maior utilização do GS para fins bibliométricos.

Palavras-chave: Dados de Citação; Google Acadêmico; Periódicos Científicos; Colaboração.

ABSTRACT

The lack of indexing for titles of scientific journals in the Social Sciences and Humanities in commercial databases makes it difficult to carry out an investigation on their impact. Open Access and tools such as *Google Scholar* (GS) and software for data processing allow search and the recovery of article citations, which can be regarded as an alternative for the studies on the impact of scientific production published in these areas. This study presents a pilot project for sharing citation data from Brazilian journals for further collaborative research by the national scientometrics community with the aim of encouraging greater use of GS for bibliometric purposes.

Keywords: Citation Data; Google Scholar; Sharing; Journals; Scientific Collaboration.

* Graduado em Biblioteconomia e Documentação pela Universidade de São Paulo. Endereço: Departamento Técnico do Sistema Integrado de Bibliotecas da Universidade de São Paulo (DT-SIBI-USP), Rua da Biblioteca s/n – Térreo – Complexo Brasiliana, Cidade Universitária, CEP 05508-500 - São Paulo-SP. Telefone: (11) 3091-1539. Email: trmurakami@gmail.com.

** Mestranda em Ciência da Informação na ECA-USP. Endereço: Programa de Pós-Graduação em Ciência da Informação (Mestrado) da Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP), Rua da Biblioteca s/n – Térreo – Complexo Brasiliana, Cidade Universitária, CEP 05508-500 - São Paulo-SP. Telefone: (11) 3091-1539. Email: sifausto@usp.br.

*** Doutorando em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais (UFMG). Endereço: Universidade Federal do Alagoas (UFAL), Departamento de Biblioteconomia, Rua Coração Eucarístico de Jesus, 287, CEP 30535-460 - Belo Horizonte-MG. Telefone: (31) 2535-5817. Email: ronaldfa@gmail.com

INTRODUÇÃO

O cenário de profundas transformações tecnológicas que permeia a sociedade atual repercute em maior desenvolvimento da produção científica, refletindo a demanda social por educação, conhecimento e soluções técnicas em praticamente todas as áreas, tornando mais complexo o ambiente científico e induzindo novas condições no contexto de pesquisas.

A complexidade verificada fez emergir a necessidade de maximizar o esforço e a alocação de recursos – tanto físicos, materiais como humanos, para a pesquisa, fomentando as colaborações científicas. A colaboração entre indivíduos, para Vanz e Stumpf (2010, p. 43), “[...] é um processo social e de interação humana que pode acontecer de diversas formas e por diferentes motivos”, e na ciência, segundo Mocelin (2009), um determinante para a formação de grupos de colaboração é a livre associação entre pesquisadores pela afinidade temática e a complementação teórico-técnica colaborativa entre diferentes áreas.

Segundo Katz e Martin (1997, apud VANZ; STUMPF, 2010, p. 44) há colaboração entre dois cientistas quando ambos compartilham dados, equipamentos e/ou ideias que produzem experimentos e análises que posteriormente são publicados na literatura científica. As autoras complementam essa definição clássica de colaboração, pontuando que mesmo indivíduos que fornecem materiais ou são responsáveis por experimentos simples, independentemente de análises, também podem ser considerados colaboradores.

O compartilhamento de dados, em especial, caracteriza-se como um aspecto da colaboração científica em franca expansão, carreado pela crescente massa de dados gerados por métodos modernos de extração, manipulação e armazenamento, facilitados pela nova geração de infraestruturas de informação e comunicação e ferramentas baseadas na internet, permitindo ampliar ainda mais o processo de pesquisa criando assim novos formatos e dinâmicas de produção coletiva em ciência, tecnologia e inovação (CT&I). Para David (2006), a colaboração científica e tecnológica é cada vez mais criticamente dependente do acesso e do compartilhamento de dados de pesquisa.

A disponibilização e o compartilhamento foram extremamente facilitados com a expansão do Acesso Aberto (*Open Access – OA*), ampliando-se também a prática de uma ciência mais aberta: além das publicações, considerando outras categorias como os dados, ferramentas, softwares, instrumentos laboratoriais e demais itens que fazem parte de uma estrutura de pesquisa – com o propósito de “tornar a atividade de pesquisa mais transparente, mais colaborativa e mais eficiente” (SAYÃO; SALES, 2014, p. 77). O compartilhamento de dados, para Sayão e Sales (2014, p. 78), “abre a possibilidade de verificação confiável dos resultados dos experimentos e permite pesquisas transversais e inovadoras desenvolvidas sobre informações já existentes, [permitindo] novas formas de interlocução e de socialização no mundo científico”.

Na área dos Estudos Métricos da Informação, cujos propósitos iniciais repousam justamente na necessidade de avaliar para gerenciar a crescente massa da produção de literatura científica pela aplicação de métodos matemáticos e estatísticos que, sistematizados, originaram a Bibliometria, é notável o volume crescente e contínuo de dados gerados pelas bases bibliográficas e outros dispositivos informacionais, que indexam essa produção técnico-científica publicada – principalmente em artigos de revistas.

Para Fausto e Mugnaini (2014), o desenvolvimento dessas bases internacionais no ambiente online facilitou a disponibilidade de dados, aumentando sua popularidade e ampliando os estudos bibliométricos. No entanto, o caráter comercial de tais bases ainda limita a pesquisa bibliométrica pela restrição de acesso aos dados, além de sua indexação limitar também a amplitude dos dados disponíveis, uma vez que a cobertura dessas bases reflete as decisões dos seus compiladores, desconsiderando áreas e publicações fora da chamada *mainstream science*, voltada às ciências “duras”.

De fato, é notória a insuficiência de indexação de títulos de revistas científicas das áreas de Humanidades e Ciências Sociais Aplicadas nas grandes bases de dados comerciais, que privilegiam as áreas de Ciências Biomédicas e Exatas. Essa carência é ainda mais aguda quando se trata de revistas dessas áreas desconsideradas, publicadas em outros idiomas que não o inglês e oriundas de países em desenvolvimento, impondo sérias barreiras à investigação bibliométrica sobre a visibilidade e impacto dos periódicos de Humanidades e Ciências Sociais Aplicadas.

Essa situação restritiva está mudando face às novas possibilidades proporcionadas pelo surgimento do OA, de novas ferramentas como o motor de busca Google Acadêmico (*Google Scholar* – GS) e de aplicativos de tratamento de dados como o *Publish or Perish*. A crescente migração das revistas das áreas de Humanidades e Ciências Sociais Aplicadas – incluindo as revistas da Ciência da Informação (CI) – para o meio *online* e as melhorias na estrutura de seus metadados, facilitando sua recuperação por motores de busca, está incrementando sobremaneira sua visibilidade na web, permitindo buscas minuciosas através do GS e ainda recuperando as citações ao(s) artigo(s) buscado(s), sendo esta última funcionalidade considerada de grande potencial para estudos bibliométricos alternativos às bases de dados tradicionais sobre o impacto da produção científica publicada dessas áreas.

No entanto, esse contexto aparentemente favorável para a investigação bibliométrica de áreas por muito tempo desconsideradas pelas grandes bases de dados ainda enfrenta desafios relacionados à confiabilidade do GS como fonte de dados.

Este estudo discute a questão da avaliação da produtividade e impacto das pesquisas científicas no âmbito de Ciência, Tecnologia e Inovação (CT&I) e o papel dos estudos avaliativos de citações nesse contexto, a situação da indexação das revistas das áreas de Humanidades e de Ciências Sociais Aplicadas; comenta os estudos críticos sobre o uso do GS como fonte para os estudos bibliométricos e propõe o compartilhamento de dados de citação das revistas de CI extraídos por essa ferramenta, seguindo os pressupostos da Ciência Aberta, com o objetivo de facilitar e promover mais investigações pela comunidade de Estudos Métricos da Informação, no esforço de qualificação do Google Acadêmico nas pesquisas bibliométricas – em especial dos periódicos de CI, fomentando uma dinâmica de colaboração científica através do compartilhamento.

A AVALIAÇÃO DA PRODUTIVIDADE E IMPACTO DAS PESQUISAS CIENTÍFICAS E O PAPEL DOS ESTUDOS AVALIATIVOS DE CITAÇÕES

A partir do pós-Guerra as atividades científicas conheceram um crescente desenvolvimento, fomentado por políticas de CT&I direcionando seu planejamento e promoção com aportes de recursos materiais e financeiros. A ideia de que as atividades científicas possibilitam a mensuração de seus resultados condicionou sua avaliação através dos chamados indicadores de produtividade e de impacto.

Tradicionalmente, a produtividade e o impacto da ciência são medidos através de seu produto mais óbvio – as publicações científicas derivadas de pesquisas. Tais publicações, registradas, possibilitam a coleta de dados cuja análise permite a identificação e descrição de uma série de padrões da produção do conhecimento científico. Um desses padrões é o quantitativo do número de publicações, indicando a produtividade, sendo outro as citações, entendidas como reflexo do impacto dessa produção.

A prática científica tem um efeito social proveniente, dentre outras condições, do caráter cumulativo da ciência, sendo inevitável para a produção de novos conhecimentos a expressão, com clareza e precisão, das fontes consultadas para postulação do que se reflete, seja na forma de crítica ou de corroboração, caracterizando as citações. No relacionamento entre os pares, o reconhecimento do pensamento, formulação e produção científica do outro, que serve de lastro para a geração de novas ideias e conhecimentos, se materializa nas citações – e de acordo com Caregnato (2011, p.73) “produtividade e impacto são os dois elementos centrais de interesse nos estudos avaliativos que podem ser levantados a partir do uso dos índices de citação”.

Na década de 1960, a criação do *Institute for Scientific Information* (ISI – atual *ISI-Web of Knowledge* depois da sua aquisição pela empresa Thomson Reuters nos anos 1990), inaugurou um novo ambiente para medir a ciência através de métodos quantitativos e objetivos com o surgimento de novos índices de citação, bem como a ampliação do uso de técnicas e métodos bibliométricos e cientométricos para análise e avaliação da produção científica. Durante um longo tempo apenas os dados produzidos a partir das bases de dados do antigo ISI – *Science Citation Index*, *Social Science Citation Index* e indicadores do *Journal Citation Reports* ofereciam uma noção do impacto das publicações nas comunidades científicas. Porém o cenário vem mudando nos últimos anos. Outras bases de dados que indexam citações têm surgido e estão disputando o espaço da *ISI-Web of Knowledge* (*ISI-WoK*) no ramo de produção de informações para medida de impacto das publicações científicas (MUGNAINI; STREHL, 2008) como o *Scientific Electronic Library Online* (*SciELO*), a base Scopus e o Google Acadêmico (*Google Scholar* – GS).

AS REVISTAS DE HUMANIDADES E CIÊNCIAS SOCIAIS APLICADAS NA INDEXAÇÃO DAS GRANDES BASES DE DADOS

Estudos que discutem a cobertura da literatura entre as diferentes áreas do conhecimento por bases de dados como a *ISI-Web of Knowledge* da Thomson Reuters e a Scopus da Editora Elsevier mostram que as Humanidades e as Ciências Sociais Aplicadas são as áreas com menor presença de títulos indexados (FRANSEN; NICOLAISEN, 2008; NEUHAUS; DANIEL, 2007; HOOD; WILSON, 2001).

Larsen e von Ins (2010), investigando a taxa de crescimento de publicações científicas entre 1907-2007 em uma série de bancos de dados, incluindo o *Science Citation Index* (SCI), o *Social Sciences Citation Index* (SSCI) e o *Arts and Humanities Citation Index* (AHCI), todos da *ISI-WoK*, demonstraram que existem grandes diferenças entre os campos, sendo que a taxa de crescimento de publicações no SSCI apresentou-se notavelmente baixa e em declínio ao longo do tempo.

Quanto ao idioma, Archambault et al (2006) compararam a cobertura linguística das revistas nas bases da *WoK* e as registradas no diretório Ulrich, mostrando que há uma super-representação de 20 a 25% dos periódicos de língua inglesa na *WoK* em

comparação com a lista de periódicos do Ulrich. Essa prevalência do inglês foi confirmada por Archambault e Larivière em 2010, num relatório da UNESCO sobre as Ciências Sociais.

A questão da cobertura do idioma é especialmente sensível no caso das revistas de Humanidades e Ciências Sociais Aplicadas, incluindo a Ciência da Informação, uma vez que frequentemente tais periódicos publicam em seu idioma materno, com orientação nacional ou regional, destacando a relevância do conhecimento local dessas áreas (ARCHAMBAULT; LARIVIÈRE, 2010). Sônia Caregnato (2011) assim resume esses estudos sobre a cobertura das bases de dados comerciais:

[...] destaca-se o número relativamente reduzido de títulos indexados frente ao que se estima haver no mundo, o predomínio da literatura em inglês, o número reduzido de revistas dos países em desenvolvimento, a cobertura limitada de outros tipos de documentos além do artigo de periódico, e a representação exagerada de algumas áreas das ciências, especialmente as duras, em detrimento das ciências sociais e humanidades (CAREGNATO, 2011, p. 74).

Nesse contexto, o surgimento do GS em 2004 despertou grande interesse por parte da comunidade cientométrica, impulsionando estudos variados sobre seus pontos fracos e fortes, conforme comentado na próxima seção.

GOOGLE SCHOLAR: AVALIAÇÕES CRÍTICAS

Após seu lançamento em novembro de 2004, prolifera a literatura internacional explorando o potencial do *Google Scholar* nos estudos bibliométricos, sendo a grande maioria desses estudos de caráter comparativo entre o GS e as bases WoK (*Web of Science* – WoS e *Journal Citation Reports* – JCR) e Scopus, usando diferentes metodologias e enfoques.

Aguillo (2012) e Adriaanse e Rensleigh (2013) elencam tais estudos, dividindo-os entre os favoráveis e os não favoráveis ao uso do GS para objetivos bibliométricos, destacando seus pontos fortes e fracos: como positivos, o fato de ser uma fonte de acesso gratuito em contraste com as dispendiosas bases de dados comerciais; sua ampla cobertura, inclusive no caso das Humanidades e Ciências Sociais Aplicadas, de outros materiais além de artigos – como livros, capítulos de livros, trabalhos de eventos, etc., que normalmente não são cobertos pelas bases comerciais; sua recuperação abrangente de revistas em acesso aberto, de idiomas diversos do inglês e originárias de países emergentes. Os pontos negativos do GS estão relacionados à sua acurácia na contagem de citações, com relatos de duplicações, erros ou supressões no ano da publicação, no nome dos autores e das revistas.

Peter Jacsó (2010; 2008; 2005) figura entre os maiores críticos ao GS, enquanto Anne-Wil Harzing (2013; HARZING; VAN DER WAL, 2008) está entre seus entusiastas, relevando a utilidade do GS para as áreas de Humanidades e Ciências Sociais Aplicadas e também disponibilizando publicamente o *software Publish or Perish*¹ (HARZING, 2007) para ordenar automaticamente as citações recuperadas no buscador.

¹ *Software Publish or Perish*. Disponível em: <http://www.harzing.com/pop.htm>. Acesso em: 27 jan. 2014

Esses estudos críticos são unânimes em afirmar a necessidade de mais pesquisas com a ferramenta, evidenciando assim uma base fundamentada e relevante que aponte realisticamente todo o potencial do *Google Scholar* para fins bibliométricos, em especial nas áreas não cobertas pelas fontes de dados comerciais. No caso brasileiro, Caregnato (2011, p. 76) reforça essa necessidade considerando que grande parte da nossa produção científica nas Humanas, Ciências Sociais e Sociais Aplicadas “[...] não está indexada nas principais bases [...], é veiculada também como tipos de documentos alternativos (por exemplo, trabalhos em eventos) e é registrada em língua portuguesa”. Essa assertiva estimulou a ideia de compartilhamento de dados de citações extraídos do GS, promovendo sua exploração através da colaboração entre a comunidade de Estudos Métricos da Informação.

METODOLOGIA DE EXTRAÇÃO E COMPARTILHAMENTO DE DADOS DE CITAÇÕES DAS REVISTAS BRASILEIRAS DE CI

Para concretizar a ideia de compartilhamento livre e aberto dos dados de citações extraídos do *Google Scholar*, facilitando e estimulando sua exploração colaborativa, foram eleitas as revistas nacionais de Ciência da Informação como piloto, com os seguintes procedimentos:

- a) Levantamento dos nomes dos periódicos de CI, usando como base o WebQualis da CAPES², comparando com listas de periódicos de CI existentes na Internet;
- b) Consulta utilizando o software *Publish or Perish* para Windows, tendo como parâmetro de busca o título do periódico e, quando presente, sua variação de nome ou abreviatura oficial, no período de 28/01/2014 a 03/02/2014;
- c) Exportação dos resultados em planilhas do *Google Drive*, uma para cada título de revista recuperado;
- d) Criação de uma planilha consolidando todas as planilhas, filtrando neste consolidado apenas os trabalhos que receberam ao menos uma citação;
- e) Elaboração de consolidados estatísticos utilizando as tabelas dinâmicas do Excel e o *Tableau Public*³, que é uma ferramenta muito interessante para criação, edição e tratamento de dados.

O *Google Drive* permite o compartilhamento, e as planilhas elaboradas com os dados pré-tratados foram hospedadas de maneira dinâmica e compartilhadas publicamente através do link:

https://docs.google.com/spreadsheets/d/19kcMMnfi_5Ohe60_mev-myFc85FkppqRJy-HhXpfb_Q/edit.

² WebQualis: base de dados de revistas mantida pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), órgão do Ministério da Educação. Disponível em: <http://qualis.capes.gov.br/webqualis/principal.seam>. Acesso em: 20 jan. 2014

³ *Tableau Public*: ferramenta gratuita para criação, edição e visualização de dados. Disponível em: <http://www.tableausoftware.com/public>. Acesso em: 18 jan. 2014

4.1 Resultados preliminares e esperados

A extração de dados do GS com o *Publish or Perish* resultou no total de 24 revistas nacionais da área de Ciência da Informação, sendo todas em acesso aberto. A busca por título recuperou dados imprecisos – neste levantamento tivemos dois casos, o primeiro da revista “Em Questão”, uma vez que existem diversos periódicos que adotam a expressão “em questão” no título; e o segundo da revista “Biblos”, que possui um homônimo publicado em Portugal (<http://dialnet.unirioja.es/servlet/revista?codigo=1764>). Nestes casos, analisamos artigo por artigo e retiramos aqueles fora do escopo da busca.

Os resultados obtidos através da consulta ao *Publish or Perish* são sumarizados no Quadro 1, mostrando os títulos das revistas nas diferentes formas encontradas, o número de artigos recuperados, as citações recebidas e o período dessas citações em anos:

Quadro 1 - Consultas efetuadas e resultados obtidos

Revista	n. Artigos	n. Citações recebidas	n. de Anos das citações
Transinformação	680	1900	28
TPBCI	0	0	0
Tendências da Pesquisa Brasileira em Ciência da Informação	75	148	6
Revista Ibero-americana de Ciência da Informação	71	45	6
Revista Digital de Biblioteconomia e Ciência da Informação	36	151	11
Revista de Biblioteconomia de Brasília	560	1044	41
Revista Brasileira de Biblioteconomia e Documentação	221	470	40
Revista Biblos	12	25	39
Revista ACB	395	440	20
Rev. digit. bibliotecon. cienc. inf.	0	0	0
PontodeAcesso	160	105	7
PG&C	0	0	0
Pesquisa Brasileira em Ciência da Informação e Biblioteconomia	119	83	8

Pesq. Bras. em Ci. da Inf. e Bib.	0	0	0
Perspectivas em Gestão & Conhecimento	149	27	3
Perspectivas em Ciência da Informação	1000	3098	18
Perspect. Ci. Inf.	1	2	3
Liinc em Revista	239	418	20
Informare	20	60	23
Informações@Profissões	0	0	0
Informação & Sociedade: Estudos	628	1858	26
Informação & Informação	351	521	19
Inf. Inf.	5	13	12
Inf. & Soc.:Est.	5	9	7
InCID: Revista	88	37	4
Encontros Bibli	785	963	18
Enc. Bibli.	8	28	10
Em Questão	1000	2290	68
DataGramaZero	172	1691	15
Comunicação & Informação	286	144	16
Ciência da Informação	1000	13289	42
Ci. Inf.	270	3255	42
Brazilian Journal of Information Science	116	68	8
Biblos: Revista	160	94	88
Biblionline	301	79	9
AtoZ: novas	21	1	3

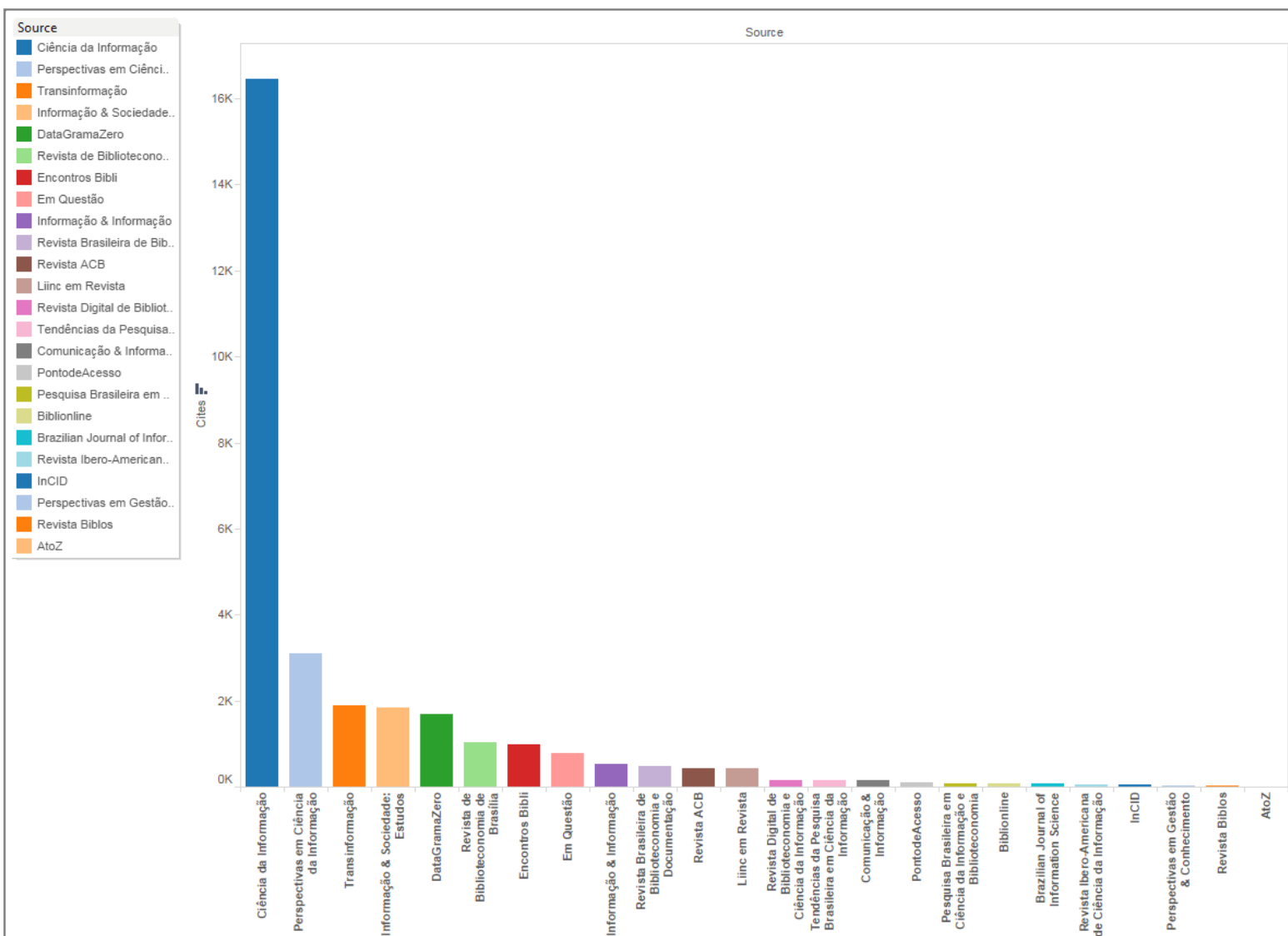
Fonte: Elaborado pelos autores

Os dados obtidos foram dispostos em planilhas dinâmicas do *Tableau Public* através de vários exercícios preliminares, por diferentes categorizações como Citações recebidas por periódico; Citações por ano e periódico e por periódico e ano; Artigos citados, entre outras, permitindo sua exploração neste link compartilhado:

http://public.tableausoftware.com/views/EstudodascitaesrecebidaspoperidicosdaCI/Citaesrecebidaspoperidicos?:embed=y&:display_count=no

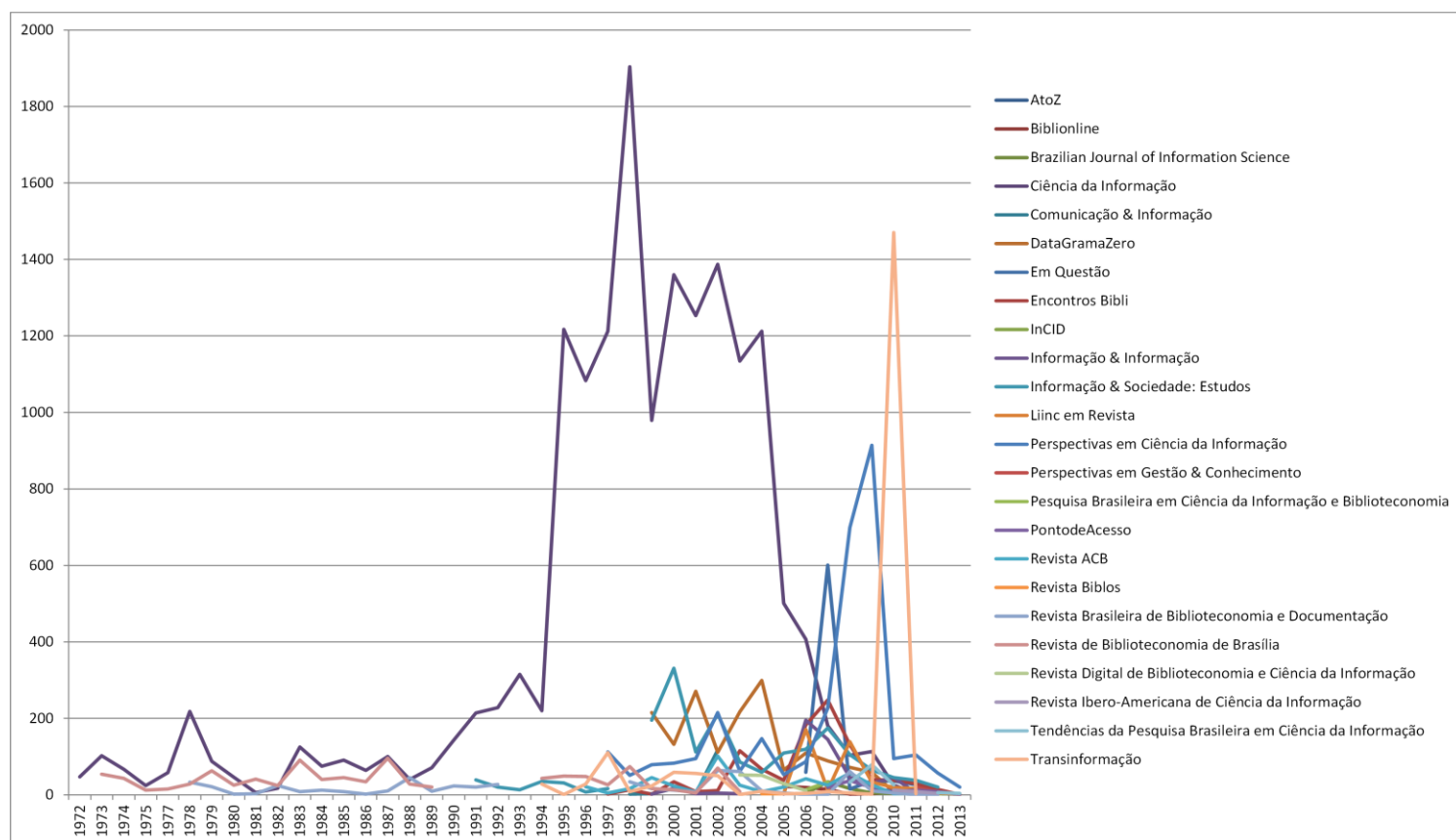
Como exemplo desses exercícios preliminares, a Figura 1 mostra a distribuição das citações obtidas por título de revista, e a Figura 2, a mesma distribuição, por ano.

Figura 1- Distribuição das citações recebidas por título de revista



Fonte: Dados compartilhados no *Tableau Public*

Figura 2 - Distribuição das citações por revista e por ano



Fonte: Dados compartilhados no Tableau Public

A facilidade de extração de dados do GS através do *Publish or Perish* com sua praticidade em elencar e ordenar os resultados, bem como a existência de ferramentas livres e gratuitas para editar e tratar os dados como o *Google Refine* e o *Tableu Public* torna mais amigável a tarefa de explorar e validar os dados obtidos, estimulando análises sobre esses dados de forma colaborativa.

CONSIDERAÇÕES FINAIS

Os estudos de citações são um importante eixo da Bibliometria e Cientometria que até recentemente eram prerrogativas de bases comerciais restritivas e dispendiosas como fonte de dados confiáveis. Ainda assim, tais fontes continuam apresentando inconsistências, largamente discutidas na literatura da área. O *Google Scholar*, surgido há exatos 10 anos, oferece um potencial alternativo para esses estudos, em especial para as áreas de Humanidades e Ciências Sociais Aplicadas, cujas revistas em idiomas diferentes do inglês e oriundas de países emergentes notoriamente não são consideradas pelas grandes bases de dados, criando uma séria barreira aos estudos de avaliação dos periódicos dessas áreas e prejudicando seu desempenho na contribuição para a CT&I.

O surgimento de aplicativos que facilitam a extração e o tratamento dos dados do GS, como o *Publish or Perish* e ferramentas como o *Google Drive* e *Refine* e o *Tableau Public* descomplica e favorece a tarefa de validação desses dados, e acreditamos que o compartilhamento público e aberto de dados pré-tratados de citações, seguindo os pressupostos da Ciência Aberta, estimule mais investigações colaborativas pela

comunidade brasileira de Estudos Métricos da Informação, visando uma desejada qualificação do *Google Scholar* como fonte alternativa nos estudos bibliométricos de revistas nacionais, minimizando a dependência de fontes comerciais restritas e contribuindo para a mensuração de resultados dessas áreas no contexto da avaliação da CT&I no país.

Artigo recebido em 08/07/2014 e aprovado em 12/08/2014

REFERÊNCIAS

ADRIAANSE, L. S.; RENSLEIGH, C. Web of Science, Scopus and Google Scholar: a content comprehensiveness comparison. *The Electronic Library*, v. 31, n. 6, p.727 – 744, 2013.

AGUILLO, I. F. Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics*, v. 91, p. 343-351, 2012.

ARCHAMBAULT, E.; LARIVIÈRE, V. The limits of bibliometrics for the analysis of the social sciences and humanities literature. In: UNESCO. *2010 World Social Science Report: knowledge divides*. Paris: International Social Science Council, 2010. p. 251-254. Disponível em: <http://unesdoc.unesco.org/images/0018/001883/188333e.pdf>. Acesso em: 20 fev. 2014.

ARCHAMBAULT, E. et al. Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, v. 68, p. 329-342, 2006.

CAREGNATO, S. E. Google Acadêmico como ferramenta para os estudos de citações: avaliação da precisão das buscas por autor. *Ponto de Acesso*, v.5, n.3, p. 72-86, dez. 2011.

DAVID, P. A. Towards a cyberinfrastructure for enhanced scientific collaboration: providing its ‘soft’ foundations may be the hardest part. In: KAHIN, B.; FORAY, D. *Advancing knowledge and the knowledge economy*. Cambridge: MIT Press, 2006. p. 431-453.

FAUSTO, S.; MUGNAINI, R. Os rankings como objeto dos estudos métricos da informação. In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 14., Florianópolis, 2013. **Anais ...** Florianópolis: ANCIB; UFSC, 2013. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/2425>. Acesso em: 29 jun. 2014.

FRANDSEN, T. F.; NICOLAISEN, J. Intradisciplinary differences in database coverage and the consequences for bibliometric research. *Journal of the American Society for Information Science and Technology*, v. 59, n. 10, p. 1570-1581, 2008.

HARZING, A.-W. A preliminary test of Google Scholar as a source for citation data: a longitudinal study of Nobel Prize winners. *Scientometrics*, v. 93, n.3, p. 1057-1075, 2013.

_____; VAN DER WAL, R. Google Scholar as a new source for citation analysis? *Ethics in Science and Environmental Politics*, v. 8, n.1, p. 62-71, 2008.

_____. *Publish or perish*. [s.n.] 2007. Disponível em: <http://www.harzing.com/pop.htm>. Acesso em 20 jan. 2014.

HOOD, W. W.; WILSON, C. S. The scatter of documents over databases in different subject domains: how many databases are needed?, *Journal of the American Society for Information Science and Technology*, v.52, n.14, p.1242-1254, 2001.

JACSÓ, P. Metadata mega mess in Google Scholar. *Online Information Review*, v. 34, n.1, p. 175-191, 2010.

_____. Google Scholar revisited. *Online Information Review*, v. 32, n.1, p.102-114, 2008.

_____. As we may search: comparison of major features of Web of Science, Scopus and Google Scholar citation-based and citation-enhanced databases. *Current Science*, v. 89, n. 9, p. 1537-1547, 2005.

LARSEN, P. O.; VON INS, M. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, v. 84, n. 3, p. 575-603, 2010.

MOCELIN, D. G. Formação de grupos de pesquisa e prática de pesquisa em grupo. In: FRANCO, M. E. D. P.; LONGHI, S. M.; RAMOS, M. G. P. (Org.). *Universidade e pesquisa: espaços de produção do conhecimento*. Pelotas: UFPel, 2009.

MUGNAINI, R.; STREHL, L. Recuperação e impacto da produção científica na era Google: uma análise comparativa entre o Google Acadêmico e a Web of Science. *Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.*, n. esp., 1. sem. 2008.

NEUHAUS, C.; DANIEL, H-D. Data sources for performing citation analysis: an overview. *Journal of Documentation*, v. 64, n. 2, p. 193-210, 2007.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. *RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde*, v. 8, n. 2, p.76-92, jun. 2014.

VANZ, S.A.S.; STUMPF, I.R.C. Colaboração científica: revisão teórico-conceitual. *Perspect. Ciênc. Inf.*, v.15, n.2, p. 42-45, 2010.