

O fim do scriptLattes? Uma análise de suas funcionalidades, alternativas para o presente e perspectivas para o futuro

TIAGO SILVA CORRÊA ¹

MANUELLA BARBOSA SUZUKI ²

PAULO ROBERTO CINTRA ³

LUZIA SIGOLI FERNANDES COSTA ⁴

RESUMO

A Plataforma Lattes, desenvolvida pelo CNPq, ofereceu acesso e padronização de registros acadêmicos e profissionais dos pesquisadores atuantes em território brasileiro. É uma ferramenta na qual os usuários podem se conectar de maneira simples, com a possibilidade de recuperar dados. Existem restrições para se analisar, quantificar e compilar a pesquisa científica nacional como um todo, e os currículos lattes podem ser utilizados para essas análises, visto que os cientistas buscam manter seus currículos atualizados na plataforma. Este trabalho buscou apresentar e discutir as vantagens em se utilizar a ferramenta scriptLattes para extração e visualização dos dados obtidos a partir dos currículos. Entre essas vantagens estão facilidade e diminuição de tempo para a análise de uma grande quantidade de currículos, geração de indicadores bibliométricos e verificação das redes de interação entre autores e coautores. O trabalho também abordou os recentes mecanismos de segurança implantados pelo CNPq que evitam a extração de informações e sua publicação indevida por sites não autorizados, o que consequentemente prejudica o funcionamento do scriptLattes.

Palavras-chave: Plataforma Lattes; scriptLattes; Plataforma Sucupira; bibliometria; extração de dados.

ABSTRACT

Lattes Platform, developed by CNPq, offered access and standardization of academic and professionals records about active researchers in Brazil. It is a tool in which users can connect simply and whose data recovery is possible. There are restrictions to analyze, quantify and compile the national scientific research as a whole, so Resumes Lattes can be used for these analysis as scientists seek to update their resumes on the Platform. This study aimed to present and discuss the advantages of using the scriptLattes tool for extraction and visualization of data

¹ Universidade Federal de São Carlos, Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, Brasil. Mestrando.
E-mail: tscorrea@ufscar.br.

² Universidade Federal de São Carlos, Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, Brasil. Mestranda.
E-mail: manuh_suzuki@hotmail.com.

³ Universidade Federal de São Carlos, Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, Brasil. Mestrando.
E-mail: paulocintra89@gmail.com.

⁴ Universidade Federal de São Carlo, Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, Brasil. Docente.
E-mail: luziasigoli@gmail.com.

obtained from the curriculum; among these benefits are the ease and decrease time to analyze a lot of resumes, generation of bibliometric indicators and verification of interaction networks between the authors and co-authors. The work also addressed the recent security mechanisms implemented by CNPq to avoid the extraction of information and its improper publication by unauthorized sites, which consequently impairs the operation of scriptLattes.

Keywords: *Lattes Platform, scriptLattes, Sucupira Platform, Bibliometrics, Data Extraction.*

1. Introdução

O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) criou a Plataforma Lattes, o maior banco de dados curriculares de estudantes e pesquisadores do País. O foco deste artigo é a ferramenta de extração de dados scriptLattes, de autoria de J. P. Mena-Chalco e R. M. Cesar Junior, que utiliza os dados cadastrados nos currículos lattes. O objetivo do artigo é apresentar as vantagens de se utilizar o scriptLattes para extração dos dados obtidos dos currículos lattes.

O presente trabalho é um estudo descritivo e de análise do scriptLattes. Para compilação de currículos lattes, foram extraídos os currículos lattes dos docentes do Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade (PPGCTS) da Universidade Federal de São Carlos (UFSCar), no ano de 2014. Foi realizada também a extração manual dos currículos via Plataforma Lattes e Plataforma Sucupira (PS), para comparação da eficiência no tempo de extração e compilação do scriptLattes, que se mostrou o método mais eficiente. O texto também aborda os recentes mecanismos de segurança implantados pelo CNPq. Neste novo e atual panorama, são apresentadas alternativas para o presente, e perspectivas para o futuro.

2. A Plataforma Lattes

A criação da plataforma padronizou os currículos dos acadêmicos e estudantes brasileiros, e atualmente ela é utilizada por diversas instituições, como: o próprio CNPq, Ministério da Ciência e Tecnologia, Universidades, grupos de pesquisa etc. Além disso, a plataforma tornou-se estratégica não só para as atividades de planejamento e gestão, mas também para a formulação das políticas do Ministério de Ciência e Tecnologia e de outros órgãos governamentais da área de ciência, tecnologia e inovação (PLATAFORMA LATTES, 2015).

Os dados cadastrados na Plataforma também são utilizados para a avaliação da competência de candidatos à obtenção de bolsas e auxílios, seleção de consultores, membros de comitês e de grupos assessores, além de dar subsídio à avaliação da pesquisa e da pós-graduação brasileiras (AMORIN, 2003).

Atualmente estão cadastrados 3.098.215 currículos na plataforma, separados por nível acadêmico, o que confirma a sua importância e abrangência territorial nacional.

Em relação à compilação de dados e estatística, a plataforma oferece o serviço Lattestats, que fornece informações qualificadas e atualizadas sobre a produção em ciência, tecnologia e inovação (CTI) cadastrada por intermédio do currículo. Cabe ressaltar que os dados são separados por região, instituição, sexo e idade (PLATAFORMA LATTES, 2015). Ademais, também é possível extrair dados dos currículos, os quais são disponibilizados em duas modalidades: espelhamento e extração de CV; e grupos da instituição. Em ambas as modalidades, é necessário

enviar um formulário para a presidência do CNPq, exceto quando a instituição já possui o protocolo de cooperação técnica firmada com eles (CNPq), sendo necessário, nesse caso, apenas o envio de e-mail à presidência do conselho.

3. ScriptLattes e a bibliometria

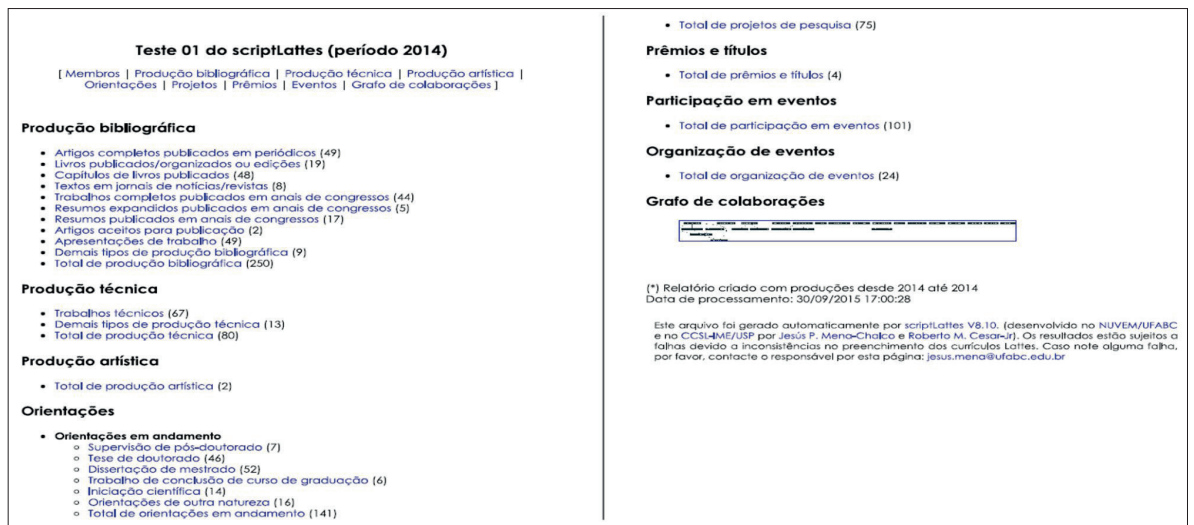
Conforme indicam Alves, Yanasse e Soma (2010) e Farias, Vargas e Borges (2012), há grande dificuldade em se analisar a pesquisa científica como um todo, pois existem trabalhos desenvolvidos pelos cientistas que não estão indexados em bases bibliográficas como Web of Science, Scopus, SciELO e Google Scholar. Além do mais, em alguns casos, podem existir dados duplicados acerca de uma mesma publicação, seja por erros de digitação não corrigidos, seja por homônimos não diferenciados, que acabam causando distorções no resultado final da análise. Por esses motivos, os autores julgam a Plataforma Lattes, do CNPq, como uma fonte apropriada para análise da produção científica brasileira, haja vista que os pesquisadores buscam manter seus currículos sempre atualizados; pois, conforme discutido anteriormente, é por meio desse sistema que algumas agências de fomento fazem a classificação para avaliar as propostas de pesquisa.

De acordo com Gläser e Laudel (2007), as análises bibliométricas tornaram-se um método válido e confiável de se legitimar a tomada de decisão acerca das políticas científicas e tecnológicas, pois elas auxiliam no delineamento da produção científica. Assim, para os autores, a coleta, a análise e a construção de indicadores bibliométricos podem contribuir na avaliação da qualidade das pesquisas realizadas por um pesquisador ou por uma instituição. Nesse contexto, o scriptLattes aparece como uma importante ferramenta, uma vez que facilita a prospecção de informações contidas sobre os pesquisadores que estão cadastrados na Plataforma Lattes. Entre os usos para os dados compilados pelo scriptLattes, Mena-Chalco e Cesar Junior (2011) citam:

- Criação de relatórios de produção acadêmica: os dados obtidos podem ser utilizados para determinação do estado da arte de um campo de pesquisa específico ou explorado em conjunto com indicadores de produção como o Qualis e Fator de Impacto dos periódicos.
- Criação de grafos de colaboração acadêmica: por meio dos relatórios gerados pelo scriptLattes, é possível verificar a existência de redes de colaboração entre os pesquisadores e suas características, como densidade, grau de aproximação e índice de centralização.
- Criação de árvores de genealogia acadêmica: como os currículos lattes contêm informações acerca das orientações recebidas e realizadas por um profissional, é possível verificar a ascendência e descendência da orientação acadêmica.
- Análise da distribuição geográfica de pesquisadores: nesse tipo de análise, busca-se identificar o impacto que a localização geográfica exerce na formação acadêmica de um grupo. Por

FIGURA 1

Resultados obtidos e análises disponíveis no scriptLattes



Fonte: Elaboração dos autores via scripLattes.

fim, Mena-Chalco e Cesar Junior (2011) salientam que os estudos não precisam seguir essas orientações estritas, mas elas podem ser modificadas e utilizadas conjuntamente, a fim de serem estabelecidas as relações que caracterizem a produção científica brasileira.

3.1 – INSTALAÇÃO, EXECUÇÃO E TESTES DO SCRIPTLATTES

O scriptLattes está disponível para download em <http://scriptlattes.sourceforge.net/>, e, conforme exposto no site, é um projeto de software livre que está sendo desenvolvido no CMC-C-UFABC (Centro de Matemática, Computação e Cognição da Universidade Federal do ABC) e no CCSL-Ime/USP (Centro de Competência em Software Livre - Instituto de Matemática e Estatística da Universidade de São Paulo) por Jesús P. Mena-Chalco e Roberto M. Cesar Junior. O site oferece diversos tutoriais, além de links, um FAQ (perguntas frequentes), e também opções para instalação nos ambientes Windows, Linux e MacOS.

O script oferece dois exemplos de execução: o teste-01, que apresenta extração e compilação automáticas de: (a) produções bibliográficas, (b) produções técnicas, (c) produções artísticas, (d) orientações, (e) projetos de pesquisa, (f) prêmios e títulos, (g) grafo de colaborações, (h) mapa de geolocalização e (i) coautoria e internacionalização; e o teste-02, que oferece as mesmas informações, mas com o arranjo da produção pela classificação Qualis. Caso o usuário domine a linguagem de programação Python, as possibilidades são infinitas, pois é possível aprimorar e personalizar o script. Neste caso, configuramos o período para 2014 e delimitamos o grupo escolhido para docentes do Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade / UFSCar. Na Figura 1 temos os resultados obtidos para o teste-01.

FIGURA 2

Artigos completos publicados em periódicos (teste 02)



Fonte: scriptLattes.

Na figura 1, observamos todos os campos que foram extraídos e analisados. Cada nome em azul traz ao seu lado a quantidade total, e clicando-se nesses nomes abrirá uma aba com a análise completa.

Na execução do teste-02 do scriptLattes, destacamos principalmente a distribuição da produção pela Qualis. O Programa de Pós-Graduação em Ciência, Tec-

nologia e Sociedade é abrangido pela área Interdisciplinar, e a Figura 2 (inserimos apenas a 1ª página) mostra a análise dos artigos completos publicados em periódicos no ano de 2014 (49 artigos) pela distribuição na Qualis. Assim temos: (A1: 1, A2: 4, B1: 15, B2: 13, B3: 7, B4: 0, B5: 0, C: 0, Qualis não identificado: 9).

3.2 – PROBLEMAS NA EXECUÇÃO DO SCRIPTLATTES

Em maio de 2015, o CNPq implantou um mecanismo de segurança (captcha) que evita a extração em lotes de informação e sua publicação indevida por sites não autorizados. O mecanismo funciona, basicamente, exigindo a digitação de um código apresentado para cada acesso a um determinado currículo. Tal fato impossibilitou o funcionamento do scriptLattes por um período. Após algumas semanas, entretanto, foi possível rodar o script novamente, mas a extração não era mais automática: foi necessário acessar cada currículo e baixar uma cópia em html. É verdade que inúmeras pesquisas foram produzidas utilizando-se a ferramenta scriptLattes, como mostrado no início do artigo, e que o não funcionamento do script paralisa outras tantas ferramentas. Neste contexto, abordamos algumas alternativas não tão eficientes e práticas para o presente, como a Plataforma Sucupira, além de extração e análise manuais dos dados.

3.3 – PLATAFORMA SUCUPIRA

Lançada em 23 de março de 2014, a Plataforma Sucupira é uma nova ferramenta com a qual o Sistema Nacional de Pós-Graduação (SNPG) realiza a coleta de informações e executa análises e avaliações acerca dos cursos de pós-graduação brasileiros. Foi desenvolvida pela

parceria da CAPES com a Universidade Federal do Rio Grande do Norte (UFRN), e recebeu esse nome em homenagem ao professor Newton Sucupira, que em 1965 normatizou o modelo-base da pós-graduação atual, através do conhecido “Parecer Sucupira”. Ademais, a proposta do sistema é integrar em um único banco de dados informações provenientes de outros sistemas internos e externos, como o GeoCapes, o Banco de Teses, a Receita Federal, o ISSN Register, entre outros (ARCANJO, 2014).

Paralelamente, entre os objetivos da Plataforma Sucupira estão: disponibilizar em tempo real e com clareza os dados do SNPG; facilitar e reduzir a inexatidão nos processos de acompanhamento de avaliação dos cursos; e permitir melhor controle de gestão, confiabilidade, segurança e transparência acerca das informações. A ferramenta também possibilita maior colaboração dos coordenadores e das pró-reitorias dos programas de pós-graduação, e permite o envio ininterrupto de dados ao longo do ano ao mesmo tempo em que fornece as informações armazenadas da instituição. Com isso, os cursos de pós-graduação têm à disposição os dados necessários para elaborar metas futuras, políticas institucionais e ações que visem contribuir para os planos de desenvolvimento (PLATAFORMA SUCUPIRA, 2014).

As produções intelectuais são importadas automaticamente da Plataforma Lattes. Entretanto, o usuário pode excluir, editar e acrescentar artigos manualmente. Somente são importadas publicações dos docentes credenciados no programa. Interessou-nos a consulta de Artigo em Periódico pela Plataforma Sucupira. O Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade tem 42 artigos publicados em periódicos. Como a plataforma não oferece a classificação dos artigos pela Qualis, os resultados foram transferidos para uma planilha de Excel para análise.

4. Comparativo entre os resultados obtidos pelo scriptLattes, Plataforma Sucupira e análise manual

A análise manual consistiu em entrar nos currículos lattes (CL) dos docentes credenciados no programa. Para visualizar cada um dos CL, foi necessário digitar um *captcha* diferente a cada nova consulta, e copiar em uma planilha de Excel os artigos publicados no ano de 2014. Em seguida, os artigos foram distribuídos manualmente pela Qualis da área Interdisciplinar. O método para a análise via Plataforma Sucupira foi muito semelhante, a única diferença se deu pelo fato de que a lista de artigos já estava disponível. Observamos que o tempo gasto para organizar e analisar as informações é enorme e, além disso, a tarefa demanda grande atenção e domínio de uma planilha de texto.

Resultados obtidos: a Plataforma Sucupira retornou 42 artigos, o scriptLattes 49 e a análise manual 50. Coincidiram, ou seja, apareceram nas três análises, 39 artigos. A Plataforma

Sucupira mostrou três artigos que não foram identificados pelos outros dois métodos; o scriptLattes retornou 10 artigos que não constam na Plataforma Sucupira, mas como 2 foram eliminados por não pertencerem à Qualis Interdisciplinar, concluímos que 8 artigos não constaram na Plataforma; a análise manual identificou 12 artigos que não constam na Plataforma Sucupira, e destes eliminamos 2 por não pertencerem à Qualis Interdisciplinar. Portanto, são 10 artigos que não constam na Plataforma. É importante ressaltar que os artigos excluídos têm uma excelente classificação pela Qualis.

A partir dessas informações, observamos que a falta de alguns artigos na Plataforma Sucupira pode ter ocorrido por diversos motivos, como: erro do sistema na extração dos dados da Plataforma Lattes, não atualização dos currículos dos docentes ou atualização posterior à data em que os dados foram baixados etc. A falta desses artigos tem grande impacto na produção e, posteriormente, na avaliação dos programas, pois há classificações Qualis que variam de B4 até A2.

Nesse sentido, o scripLattes seria a ferramenta ideal, pois gera relatórios automáticos e em pouco tempo. Entretanto, os problemas atuais na utilização desse recurso são preocupantes, pois tornam necessária a análise manual que, como já indicado neste artigo, demanda muito tempo e esforço, inviabilizando um avanço já conquistado. A seguir, temos duas tabelas comparativas entre as três formas de verificar os artigos em periódicos em relação à Qualis.

Tabela 1

Comparativo entre a Plataforma Sucupira x scriptLattes x análise manual

Comparativo	Plataforma Sucupira	ScriptLattes	Análise manual
Tempo de geração dos dados	1 minuto	5 minutos	4 horas
Tempo de análise dos dados	4 horas	5 minutos	2 horas
Gera compilação automática dos dados	Não	Sim	Não
Permite escolha de qualquer membro?	Não (só mostra docentes credenciados em um PPG)	Sim	Sim
É possível verificar qualquer intervalo de tempo?	Não (só mostra individualmente os anos de 2013 e 2014)	Sim	Sim
Quantidade de artigos em periódicos localizados	42	49	50
Artigos coincidentes	39	39	39
Conflitos	3 artigos constam somente na Plataforma Sucupira,	10 artigos não constam na Plataforma Sucupira. Destes, eliminamos 2 artigos por não pertencerem à Qualis Interdisciplinar. Assim, não constam 8.	12 artigos não constam na Plataforma Sucupira. Destes, eliminamos 2 artigos por não pertencerem à Qualis Interdisciplinar. Assim, não constam 10.
Fonte: Elaboração dos autores.			

4.1 – ALTERNATIVAS PARA O PRESENTE E PERSPECTIVAS PARA O FUTURO

Além da possibilidade da demorada análise manual, a própria Plataforma Lattes trata da extração de dados e oferece algumas possibilidades para extração de dados dos currículos, já descritas: espelhamento e extração de CV; e grupos da instituição. Essas modalidades são restritas a poucos grupos, pois dependem de autorização.

A própria Plataforma Sucupira pode ser uma alternativa no futuro para a geração automática de informações. Em 2015, os arquivos de Qualis puderam ser visualizados na Plataforma Sucupira (anteriormente eram disponibilizados no site da Capes), o que sugere que futuramente os artigos sejam distribuídos automaticamente pelas respectivas Qualis, porém essa análise ficará restrita a docentes credenciados em programas de mestrado e doutorado.

5. Considerações finais

O uso do scriptLattes pode ter grande utilidade em pesquisa científica, bibliometria e nos cursos de pós-graduação no Brasil, considerando o modelo de avaliação anual dos cursos implantado pela Capes, e denominado Coleta Capes: os responsáveis pelos cursos devem preencher a Plataforma Sucupira durante todo o ano com dados de discentes, docentes, disciplinas, e concluir inserindo diversos dados qualitativos e quantitativos referentes à proposta do programa. O uso do scriptLattes pode servir como ferramenta de conferência para esse processo, já que com ela é possível ter uma completa análise da produção do corpo docente do programa e sua distribuição conforme as Qualis existentes para a área acadêmica (que também podem ser inseridas na proposta do programa).

Outra possível forma de aplicação do scriptLattes em programas de pós-graduação ou outros órgãos educacionais e institucionais é utilizá-lo no levantamento da produção de determinado docente para fins de admissão, credenciamento, recredenciamento e análise de sua produção. Cabe salientar que não pretendemos discutir se tal forma de avaliação é justa ou não; nós apenas identificamos uma alternativa com o uso de uma ferramenta que minimizaria o tempo para tal análise. Além disso, o script também retorna dados referentes ao relacionamento da produção entre determinados docentes (grafos de colaboração), tornando possível a verificação de quais grupos de pesquisadores apresentam maior produção, e quais seus pontos fortes e fracos. O script tem também potencial para ser utilizado na identificação de relações acadêmicas entre candidatos de processos seletivos de docentes, mestrado ou doutorado com o corpo docente de determinado programa – considerem-se os recorrentes problemas jurídicos em tais processos, principalmente em relação à neutralidade da banca e aos recursos que os candidatos impetram.

Por fim, destaca-se que o scriptLattes pode ter grande aplicabilidade em diversos campos de pesquisa que utilizam os dados cadastrados na Plataforma Lattes. Ressaltamos também que

o recente impedimento da utilização eficiente do scriptLattes é um retrocesso enorme e prejudicial para diversas pesquisas e ações que vinham utilizando a ferramenta, cuja solução acordada entre os autores seria o próprio CNPq disponibilizar métodos eficazes de consulta, pois os dados e estatísticas oferecidos no site para consulta pública são de pouca utilidade no presente.

6. Referências

ALVES, A. D.; YANASSE, H. H.; SOMA, N. H. Especificação e implementação de uma linguagem de domínio específico para análise de redes sociais acadêmicas. In: Workshop de Computação Aplicada, 10º, São José dos Campos, 2010. **Anais...** São José dos Campos: INPE, 2010.

AMORIN, C. V. Organização do currículo: plataforma Lattes. **Pesqui. Odontol. Bras.**, São Paulo, v. 17, supl. 1, p. 18-22, maio 2003. Disponível em: <<http://www.scielo.br/pdf/pob/v17s1/a04v17s1.pdf>>. Acesso em: 19 jun. 2015.

ARCANJO, P. Capes lança Plataforma Sucupira para gestão da pós-graduação. **Coordenação de Comunicação Social da Capes**, Brasília, 27 mar. 2014.

FARIAS, L. R. de; VARGAS, A. P.; BORGES, E. N. **Um sistema para análise de redes e pesquisa, baseado na Plataforma Lattes**. 2012. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/erbd/2012/003.pdf>>. Acesso em: 16 jun. 2015.

GLÄSER, J.; LAUDEL, G. The social construction of bibliometric evaluation. In: WHITLEY, R.; GLÄSER, J. **The changing governance of the sciences: the advent of research evaluation systems**. Dordrecht, Springer, 2007. p. 101-126.

FRANCISCO, L. G. **Manual ScriptLattes (UFSCAR)**. 2015. Disponível em: <<http://scriptlattes.sourceforge.net/tutorial-para-o-uso-do-scriptLattes.pdf>>. Acesso em: 10 jun. 2015.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. **Prospecção de dados acadêmicos de currículos Lattes através de ScriptLattes**. 2011. Disponível em: <<http://professor.ufabc.edu.br/~jesus.MENA/publications/pdf/scriptLattes-2011-bibliometria.pdf>>. Acesso em: 17 jun. 2015.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. ScriptLattes: an open-source knowledge extraction from the Lattes platform. **Journal of the Brazilian Computer Society**, v. 15, n. 4, p. 31-39, dez. 2009.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes. In: HAYASHI, M. C. P. I.; MUGNAINI, R.; HAYASHI, C. R. M. (Org.). **Bibliometria e Cientometria: reflexões teóricas e interfaces**. São Carlos: Pedro & João, p. 109-128, 2013.

MOREIRA, M. L.; VELHO, L. Pós-graduação no Brasil: da concepção “ofertista linear” para “novos modos de produção do conhecimento” - implicações para avaliação. **Avaliação** (Campinas) [online], v. 13, n. 3, p. 625-645. 2008. Disponível em: <<http://dx.doi.org/10.1590/S1414-40772008000300002>>. Acesso em: 5 abr. 2016.

PLATAFORMA LATTES. **Sobre a plataforma Lattes**. 2015. Disponível em: <<http://lattes.cnpq.br/>>. Acesso em: 19 jun. 2015.

PLATAFORMA SUCUPIRA. Disponível em: <<https://sucupira.capes.gov.br/sucupira>>. Acesso em: 7 out. 2015.

SCRIPTLATTES. Disponível em: <<http://scriptlattes.sourceforge.net/>>. Acesso em: 10 jun. 2015.

SILVA, F. M.; SMIT, J. W. Organização da informação em sistemas eletrônicos abertos de Informação Científica & Tecnológica: Análise da Plataforma Lattes. **Perspectivas em Ciência da Informação**, Minas Gerais, v. 14, n. 1, p. 77-98, jan./abr. 2009. Disponível em: <<http://portal-deperiodicos.eci.ufmg.br/index.php/pci/article/view/720>>. Acesso em: 19 jun. 2015.