

Registros valiosos

Dados compilados da Plataforma Lattes abastecem estudos sobre a ciência no país e revelam tendências

A Plataforma Lattes, que reúne mais de 4 milhões de currículos acadêmicos, tornou-se fonte de informações para um número crescente de pesquisadores que buscam dados sobre a ciência brasileira a fim de estudar seus fenômenos e tendências. Criado em 1999 pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), o sistema de currículos registra a trajetória e a contribuição de cada estudante, técnico e pesquisador do Brasil e fornece ao governo e a agências de fomento informações sobre produção científica, participação em projetos e orientações e supervisões, entre outros. Sua utilidade, porém, há tempos extrapolou a esfera da gestão, ajudando pesquisadores a produzir conhecimento original. “Raros países dispõem de uma plataforma com as atividades de sua comunidade científica como um todo”, diz Rogério Mugnaini, professor da Escola de Comunicações e Artes (ECA), da Universidade de São Paulo (USP).

Mugnaini participa de um projeto de pesquisa apoiado pelo CNPq e a USP que busca desenvolver ferramentas para combinar informações da Plataforma Lattes com dados sobre publicações científicas

de fontes nacionais e internacionais. No mês passado, Mugnaini e colegas publicaram um artigo na revista *PLoS One* analisando a relação entre pesquisadores sêniores na área de ciências exatas e da Terra e estudantes de pós-graduação orientados por eles, utilizando produções bibliográficas de 1981 a 2010 registradas na Plataforma Lattes. Constatou-se que, quanto mais tempo eles trabalham juntos, maior é a produtividade do jovem pesquisador orientado. “Observamos que parte dos orientadores deixou de ter uma linha de pesquisa própria, passando a declarar como sua produção apenas artigos de alunos”, afirma. Outro estudo do grupo analisou até que ponto as informações do Lattes estão atualizadas ao comparar dados dos currículos com os obtidos em relatórios de programas de pós-graduação. A conclusão é de que a falta de atualização pode alcançar até 20% dos artigos publicados nos três anos anteriores. As áreas de engenharias e ciências agrárias são as que mais sofrem com o problema.

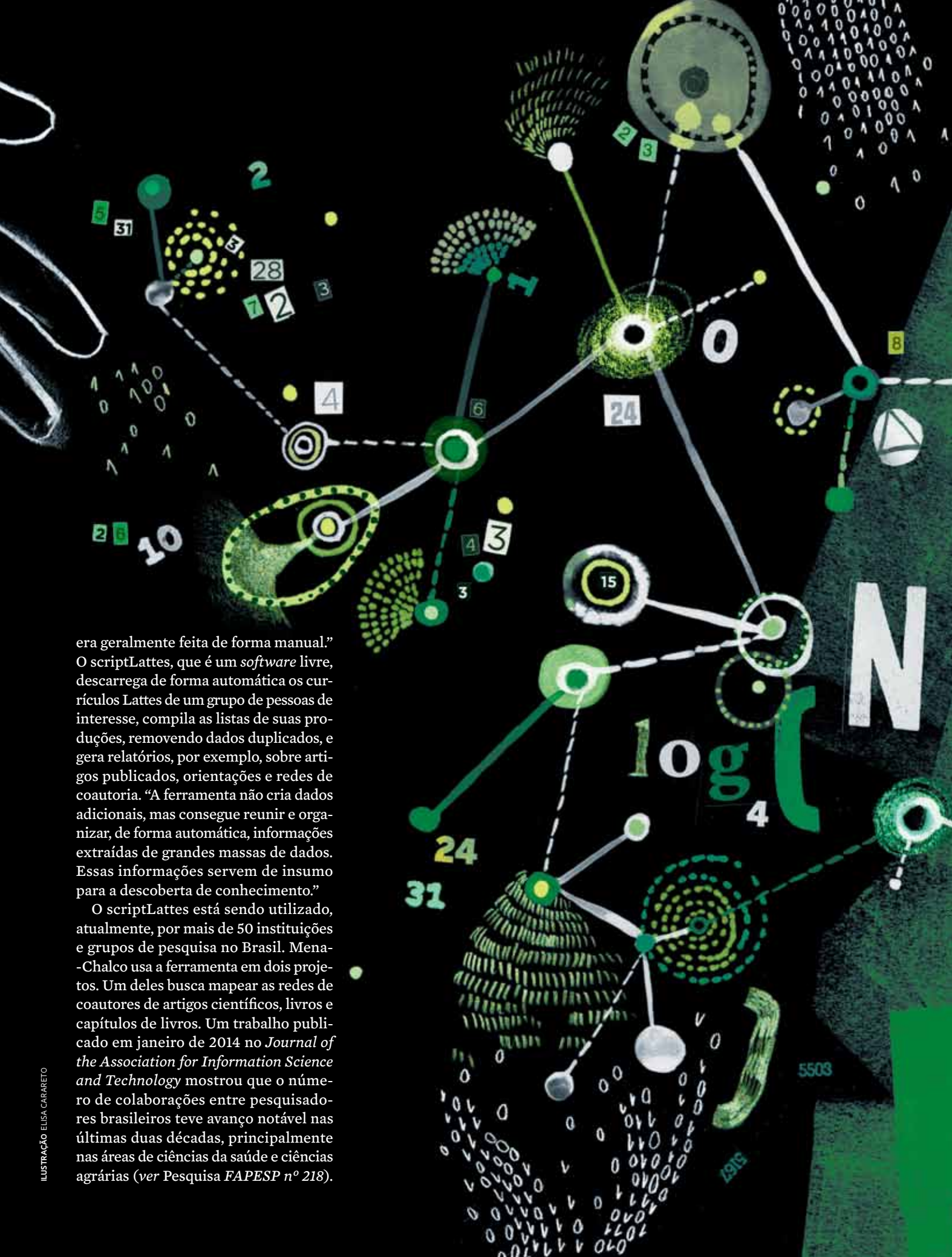
Mugnaini acredita que seja viável criar indicadores específicos baseados no Lattes. “Nas referências bibliográficas dos currículos Lattes, há informação sobre

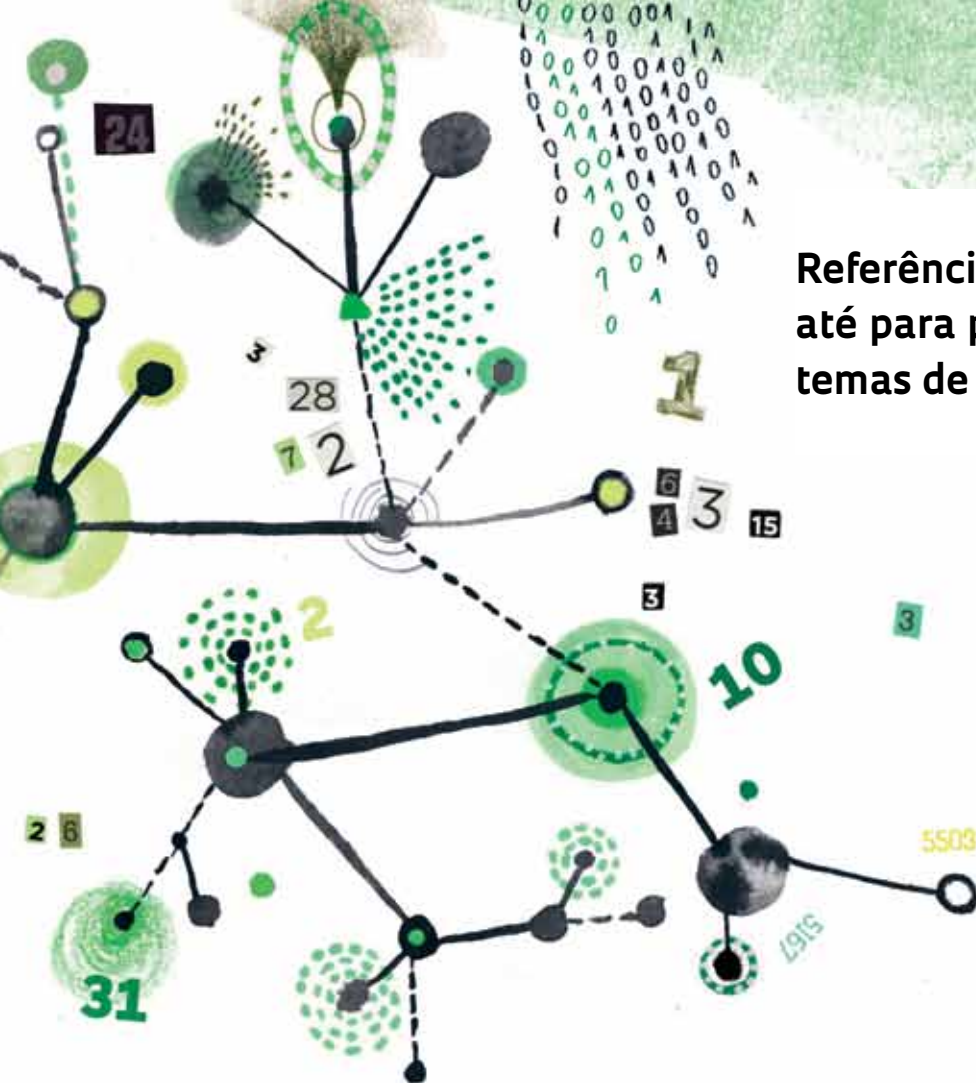
a produção científica em revistas não indexadas nas bases de dados de revistas tradicionais, além de teses, livros e outros documentos. O Lattes representa uma visão completa do conjunto da produção científica brasileira e poderia municiar um sistema de avaliação mais fidedigno”, afirma.

O advento de uma ferramenta que ajuda a extrair e organizar grandes quantidades de dados do Lattes disponíveis na internet foi fundamental para pavimentar o trabalho dos pesquisadores. Desde 2005, está disponível o scriptLattes, desenvolvido pelo professor do Instituto de Matemática e Estatística (IME) da USP, Roberto M. Cesar-Jr., e seu então aluno de doutorado, Jesús Mena-Chalco, hoje professor da Universidade Federal do ABC (Ufabc). Mena-Chalco também integra o grupo de pesquisa de Mugnaini, que é coordenado por Luciano Digiampietri, professor da Escola de Artes, Ciências e Humanidades (EACH) da USP. “Nossa pretensão era criar uma ferramenta para uso doméstico do IME, mas acabou se tornando útil para muito mais gente”, diz Mena-Chalco. “Antes do scriptLattes, a coleta de dados do Lattes

era geralmente feita de forma manual.” O scriptLattes, que é um *software* livre, descarrega de forma automática os currículos Lattes de um grupo de pessoas de interesse, compila as listas de suas produções, removendo dados duplicados, e gera relatórios, por exemplo, sobre artigos publicados, orientações e redes de coautoria. “A ferramenta não cria dados adicionais, mas consegue reunir e organizar, de forma automática, informações extraídas de grandes massas de dados. Essas informações servem de insumo para a descoberta de conhecimento.”

O scriptLattes está sendo utilizado, atualmente, por mais de 50 instituições e grupos de pesquisa no Brasil. Mena-Chalco usa a ferramenta em dois projetos. Um deles busca mapear as redes de coautores de artigos científicos, livros e capítulos de livros. Um trabalho publicado em janeiro de 2014 no *Journal of the Association for Information Science and Technology* mostrou que o número de colaborações entre pesquisadores brasileiros teve avanço notável nas últimas duas décadas, principalmente nas áreas de ciências da saúde e ciências agrárias (ver Pesquisa FAPESP nº 218).





Referências da plataforma servem até para predizer quais serão os temas de pesquisa mais estudados

seriam os assuntos mais estudados no futuro próximo. Em 2012 a expressão que mais se destacou segundo a metodologia foi “serviços *web*”, mas está previsto, tanto para 2015 como para 2020, que o campo mais quente será o de redes neurais. O pesquisador da EACH-USP também planeja criar uma ferramenta capaz de sugerir a um pesquisador artigos recém-publicados que possam lhe interessar, com base nos temas a que se dedica. “Comecei a trabalhar com esses assuntos há cinco anos. Não tinham a ver com o que fiz no doutorado, mas me fascinaram”, diz Digiampietri.

FENÔMENOS

Estudos com base nos dados da Plataforma Lattes estão mostrando fenômenos pouco conhecidos ou movimentos ainda não registrados em indicadores oficiais. Fabio Mascarenhas, professor do Departamento de Ciência da Informação da Universidade Federal de Pernambuco (UFPE), acaba de orientar a dissertação de mestrado do estudante Guilherme Alves de Santana, que, em meio a dados compilados sobre colaboração científica, encontrou uma peculiaridade sobre a formação de grupos de pesquisa no Brasil. “Foi analisada a produção científica dentro de grupos de pesquisa. Esperávamos encontrar mais artigos assinados em coautoria entre os membros dos grupos. Mas observamos que há mais colaboração com pesquisadores de fora dos grupos do que entre eles”, afirmou. O assunto será investigado em profundidade no doutorado de Santana.

Outra curiosidade, essa envolvendo a produção científica brasileira sobre medicina tropical, foi observada na dissertação de mestrado de Natanael Vitor Sobral, também defendida neste ano. Com base em dados do Programa de

Outro projeto busca construir árvores genealógicas de cientistas, analisando as relações de orientação e supervisão. O pesquisador e seus colaboradores planejam criar uma plataforma nacional, na qual seja possível verificar a contribuição de cada pesquisador na formação de outros. “Já havia iniciativas para levantar árvores genealógicas em campos do conhecimento, como a matemática, a física e a neurociência, mas não sobre a comunidade científica de um país”, afirma o pesquisador. “A análise das informações sobre relações de orientação acadêmica é uma tentativa de mensurar a notoriedade de um pesquisador com base em sua repercussão em outras gerações.”

Já Luciano Digiampietri, professor do bacharelado em sistemas da informação da EACH-USP, que já publicou mais de uma dezena de artigos baseados em dados do Lattes, também utiliza as informações da plataforma para desenvolver algoritmos talhados para antecipar tendências. Uma de suas pesquisas, em parceria com um aluno de mestrado,

William Maruyama, busca prever as colaborações futuras de um pesquisador. O estudo utilizou dados sobre coautoria de trabalhos científicos de pesquisadores da área de ciência da computação registrados na Plataforma Lattes entre 1970 e 2010. Os dados entre 1970 e 2000 serviram para representar os padrões do passado. As informações entre 2001 e 2005 foram usadas para representar o presente. O algoritmo comparou os dois intervalos e tentou predizer com quem os pesquisadores iriam se relacionar no futuro. Para validar o algoritmo, seus resultados foram comparados com dados registrados entre 2006 e 2010. A representação do futuro feita pelo algoritmo e o que aconteceu de verdade entre 2006 e 2010 teve índice de coincidência de 97%.

Outra linha de pesquisa é a análise de tendências em determinados campos do conhecimento. Utilizando palavras-chave extraídas dos títulos de publicações científicas de pesquisadores da ciência da computação registradas no Lattes ao longo do tempo, buscou-se predizer quais



Pós-graduação em Medicina Tropical da UFPE, constatou-se que parte significativa dos artigos publicados na área não tem como foco de forma direta as doenças endêmicas no Brasil, como dengue, malária ou esquistossomose, mas de moléstias que despertam mais interesse de revistas científicas internacionais, como a Aids. “A explicação é que grandes periódicos científicos têm pouca abertura para doenças de países pobres se não estiverem associadas a temas mais universais”, diz Mascarenhas.

Alberto Laender, professor titular do Departamento de Ciência da Computação da Universidade Federal de Minas Gerais (UFMG) e membro da Academia Brasileira de Ciências, coletou dados de mais de 4 milhões de currículos e, com a ajuda de Thiago Magela Rodrigues Dias, aluno de doutorado do Centro Federal de Educação Tecnológica de Minas Gerais, do qual é coorientador, está analisando a produção científica dos mais de 220 mil doutores com currículos cadastrados na plataforma, de 64 mil docentes de programas de pós-graduação e de 15 mil bolsistas de produtividade do CNPq, pesquisadores considerados mais produtivos pela agência de fomento federal. “Vamos analisar a evolução da

“A abertura integral das informações do Lattes seria importante para a comunidade científica”, diz Alberto Laender

produção nesses três grupos e em sete grandes áreas do conhecimento”, diz Laender, que integra o Instituto Nacional de Ciência e Tecnologia para a Web (InWeb). Dados preliminares sugerem que, de modo geral, começou a haver em 2012 uma redução no número total de publicações pela comunidade científica brasileira. Isso, diz Laender, pode ter duas origens. A primeira é o aumento da carga de trabalho didático dos dou-

tores contratados como professores em universidades federais. “É possível que eles não estejam conseguindo publicar o mesmo volume de artigos do tempo em que faziam o doutorado”, afirma. Outra seria a expansão de vagas de docentes em institutos federais de ensino, onde há pouco espaço para os doutores fazerem pesquisa. A queda na produção científica, porém, não é observada nos docentes de programas de pós-graduação nem entre os bolsistas de produtividade.

PRODUÇÃO COMPARADA

Laender e seu grupo têm uma extensa produção de artigos baseada em dados do currículo Lattes. Já publicaram *papers* sobre o perfil dos pesquisadores de ciência da computação e sua produção científica em comparação com a de colegas da América do Norte e da Europa. Também utilizaram dados do Lattes para organizar um portal na internet (www.cienciabrasil.org.br) onde é possível vislumbrar a produção dos Institutos Nacionais de Ciência e Tecnologia (INCTs).

Uma queixa frequente de pesquisadores que trabalham com informações do Lattes é a dificuldade de obter diretamente do CNPq dados brutos da plataforma. Em abril, tornou-se mais complicado extrair informações do Lattes na internet, pois o CNPq introduziu um código de confirmação para cada consulta a currículos. Isso para evitar o crescente compartilhamento de dados da plataforma por *sites* comerciais. O CNPq tem como política fornecer a cada instituição os dados consolidados de seu corpo de pesquisadores, professores e alunos, mas o conjunto de informações da plataforma não é facilmente franqueado. “A abertura integral dos dados do Lattes seria importante para a comunidade científica. Não me refiro a currículos individuais, mas ao conjunto de dados e à atualização deles para que possamos tratá-los com mais facilidade”, diz Alberto Laender, da UFMG.

Mônica Ramalho, analista de ciência e tecnologia e inovação do CNPq que atua na assessoria de planejamento, coordenação de estatísticas e indicadores do órgão, alega que há um caminho institucional a ser seguido para obter tais informações. “Para conseguir acesso mais direto, é preciso enviar um pedido ao CNPq explicando os motivos pelos quais o pesquisador precisa dos dados do Lattes”, afirma. ■

Fabício Marques