

- I Universidade do Estado do Rio de Janeiro, Rio de Janeiro - RJ, Brasil  
lascampos@iesp.uerj.br  
<https://orcid.org/0000-0003-2153-547X>
- II Fundação Getulio Vargas, Rio de Janeiro - RJ, Brasil  
matheus.pestana@iesp.uerj.br  
<https://orcid.org/0000-0002-5388-8950>
- III Universidade do Estado do Rio de Janeiro, Rio de Janeiro - RJ, Brasil  
andre.felix@iesp.uerj.br  
<https://orcid.org/0000-0002-9480-0126>
- IV Instituto Universitário de Lisboa, Lisboa, Portugal  
marcia.candido@iscte-iul.pt  
<https://orcid.org/0000-0003-3466-000X>
- V Universidade Estadual de Campinas, Campinas - SP, Brasil  
chaguri@unicamp.br  
<https://orcid.org/0000-0002-5201-360X>

Luiz Augusto Campos<sup>I</sup>  
Matheus Cavalcanti Pestana<sup>II</sup>  
André Costantin Felix de Souza<sup>III</sup>  
Marcia Rangel Candido<sup>IV</sup>  
Mariana Miggiolaro Chaguri<sup>V</sup>

## ATLAS DAS CIÊNCIAS SOCIAIS<sup>1</sup>

### INTRODUÇÃO

As informações encontradas nos currículos cadastrados na Plataforma Lattes constituem uma fonte de dados relevante sobre a produção acadêmica nacional. Seus registros já foram utilizados em análises sobre o conhecimento científico por diversas abordagens, tais como a identificação de redes de coautoria (Ferraz, Quoniam & Alvares, 2014; Mena-Chalco & Digiampietri et al., 2012); a avaliação da produção de programas de pós-graduação (Ferraz, Quoniam & Maccari, 2014; Nigro et al., 2015); a detecção do estado do conhecimento sobre um dado assunto (Brito, Quoniam & Mena-Chalco, 2016; Nigro, 2016); e as desigualdades de gênero em perspectiva comparada (Beigel et al., 2023).

A maioria desses trabalhos utilizaram o programa de computador *scriptLattes* (Mena-Chalco & Cesar Junior, 2009) que automatizou o tratamento em lote dos currículos. Todavia, a implementação do dispositivo de segurança CAPTCHA<sup>2</sup> no site da plataforma inviabilizou o funcionamento do *scriptLattes* (Corrêa et al., 2017) o que criou dificuldades consideráveis para a aquisição de informações na quantidade necessária para realização deste tipo de estudo. Neste registro de pesquisa, apresentamos uma alternativa para a superação do obstáculo CAPTCHA, bem como uma nova forma de utilização da base Lattes: o delineamento da produção de todos/as os/as docentes de uma área inteira do conhecimento, as Ciências Sociais.

Qual o perfil de pesquisadores(as) que atuam nas Ciências Sociais no Brasil? Foi com o intuito de mapear nossos(as) acadêmicos(as), entendendo suas diferenças de gênero, seu ritmo de publicações, suas origens e seus destinos, bem como os vínculos entre orientadores(as) e orientandos(as),

que consolidamos uma base de dados. Tal conjunto de conteúdos é oriundo de consultas à Plataforma Sucupira, da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), e à Plataforma Lattes, do CNPq, de onde agregamos informações da produção dos(as) docentes de pós-graduação do Brasil. Trabalhamos com as seguintes áreas de avaliação da CAPES: a Sociologia; a Ciência Política/Relações Internacionais; e a Antropologia/Arqueologia. O processo de obtenção dos dados, tratamento das bases e criação de painel para visualização dos resultados serão discriminados na sequência, sendo esses os objetivos principais do presente texto.

Os dados foram utilizados no *Atlas Digital das Ciências Sociais*<sup>3</sup>, parte do projeto “Democracia e Resistências Acadêmicas: as Ciências Sociais na atual conjuntura social, política e cultural brasileira” cujo intuito é promover a coleta e análise sistemática de informações quantitativas, com a produção de estatísticas e indicadores sociais sobre a organização, funcionamento e impacto acadêmico e público desta grande área no país<sup>4</sup>.

## MÉTODOS E RESULTADOS

A Plataforma Lattes não disponibiliza a área disciplinar de pertencimento dos docentes ali registrados, dado gerado apenas pela Capes. Por isso, antes de efetuar a raspagem das informações dos CVs Lattes, definimos o recorte a partir dos nomes dos(as) docentes dos programas de pós-graduação, obtidos na área de dados abertos do site da Capes<sup>5</sup>. Nesta página são disponibilizados para descarregamento arquivos no formato de planilha eletrônica por triênio de avaliação dos programas de pós-graduação. Para além das razões técnicas, a pesquisa buscou estabelecer as características dos pesquisadores brasileiros e de seus programas de pós-graduação. Embora esta seleção não esgote toda a ciência produzida no Brasil, sabemos que a maior parte dela é feita em tais espaços. Recortamos deste arquivo<sup>6</sup> todos os registros referentes às áreas das Ciências Sociais, como já mencionadas: Antropologia/Arqueologia (código=35), Ciência Política/Relações Internacionais (código=39) e Sociologia (código=34)<sup>7</sup>. Na sequência, padronizamos os valores das variáveis “Nome Docente” e “Instituição de Titulação” (NM\_DOCENTE e NM\_IES\_TITULACAO), convertendo todos os caracteres para minúsculos e removendo todos os sinais diacríticos.

Para construir a base de dados de produção acadêmica seguimos três etapas: começamos por obter os possíveis identificadores Lattes dos(a) docentes; descarregamos os currículos no formato XML8; e convertemos os arquivos XML para o formato de tabela. Na primeira etapa, executamos um script na linguagem Python que realiza, para cada nome de docente extraído do site da CAPES, a busca no site da Plataforma Lattes<sup>9</sup> e grava um arquivo texto com todos os identificadores Lattes resultantes da busca.

A segunda etapa, como mencionado, compreendeu o descarregamento dos arquivos no formato XML de cada identificador Lattes encontrado na etapa anterior. A superação do mecanismo CAPTCHA foi possível por meio da

contratação de um serviço de terceirização de tarefas via internet (*crowdsourcing*) especializado na solução deste problema. O tempo de resolução foi, em média, de cinco segundos e o custo de cerca de dois dólares para cada mil CAPTCHAs resolvidos.

A partir da leitura e interpretação de cada arquivo XML, estruturamos dois conjuntos de dados. O primeiro utilizamos para estabelecer a relação com os registros da Plataforma Sucupira, contendo as informações Identificador Lattes, Nome Docente, Instituição de Titulação e Ano de Titulação. Na base de dados, os nomes destas variáveis são, respectivamente, “NUMERO-IDENTIFICADOR”, “NOME-COMPLETO”, “NOME-INSTITUICAO-DOUT” e “ANO-DE-OBTENCAO-DO-TITULO”. Da mesma forma que os dados da CAPES, convertimos os valores das variáveis textuais para caracteres minúsculos sem acentuação. O segundo arquivo contém a publicação de artigos, livros e capítulos de livros. Ambos incluem a variável “NUMERO-IDENTIFICADOR”, o que permite a combinação das informações entre as duas bases, isto é, a associação entre produção bibliográfica de cada docente.

Realizamos a relação entre as bases CAPES e Lattes pelas variáveis “NUMERO-IDENTIFICADOR”, “NOME-COMPLETO”, “NOME-INSTITUICAO-DOUT” e “ANO-DE-OBTENCAO-DO-TITULO”. Fizemos tal combinação em sete rodadas sucessivas, partindo de um critério mais rigoroso para um critério mais frouxo. O maior rigor consistiu na comparação dos valores exatos, em caracteres minúsculos sem acentuação, das variáveis. No afrouxamento, empregamos a medida de similaridade de palavras denominada *Distância de Levenshtein*<sup>10</sup>.

Nas rodadas um e dois, comparamos os valores exatos de “NOME-COMPLETO” e “NOME-INSTITUICAO-DOUT”, sendo que a primeira incluiu o “ANO-DE-OBTENCAO-DO-TITULO”. O resultado foi de 70% de registros coincidentes. Na terceira rodada, a variável “NOME-INSTITUICAO-DOUT” foi comparada de maneira aproximada e as variáveis “NOME-COMPLETO” e “ANO-DE-OBTENCAO-DO-TITULO” de maneira exata. A comparação na quarta rodada foi feita pela variável “NOME-COMPLETO” de forma exata e “NOME-INSTITUICAO-DOUT” de forma aproximada. Ao fim destas quatro rodadas, alcançamos a conformidade de 90% dos registros das duas tabelas. Nas três últimas rodadas, utilizamos o primeiro nome do(a) docente como critério base de combinação, e as demais variáveis foram comparadas frouxamente, sendo selecionados os registros com maior semelhança. É importante notar que essas três últimas rodadas geraram falso-positivos. Descartamos, então, os registros não coincidentes.

O resultado desse processamento é um arquivo que relaciona as variáveis identificadoras de cada plataforma, “ID\_PESSOA” e “NUMERO-IDENTIFICADOR”, da CAPES e dos currículos Lattes, respectivamente. Com essa tabela de relação, é possível, por exemplo, atribuir a produção acadêmica de docentes obtida na Plataforma Lattes aos programas de pós-graduação listados do site da CAPES. Como exemplo prático dos possíveis usos dessas informações combinadas, apresentamos a seguir o Atlas das Ciências Sociais.

### TRATAMENTO E PROCESSAMENTO DOS DADOS

Com os dados descarregados da internet, criamos as bases Acadêmicos e Produções relacionáveis entre si por meio da variável “NUMERO-IDENTIFICADOR”. A base Acadêmicos é um recorte do arquivo baixado da CAPES, com a inclusão da variável “genero”. Tal variável foi atribuída a cada registro de docente a partir de um modelo probabilístico<sup>11</sup> que utiliza o primeiro nome do docente para estimar o gênero segundo informações de registros de nascimento disponibilizados no sítio eletrônico do Instituto Brasileiro de Geografia e Estatística (IBGE)<sup>12</sup>. Vale notar que os dados do IBGE não permitem ir além da classificação dicotômica entre “feminino” e “masculino”. É importante ressaltar que essa abordagem é problemática, uma vez que não captura adequadamente a diversidade de identidades de gênero que existem, desconsiderando pessoas não-binárias, de gênero fluido e outras identidades de gênero que não se encaixam estritamente nessa dicotomia tradicional. Nos casos de nomes que não constam nesta base do IBGE, o dado foi imputado de forma manual pelos(as) pesquisadores(as), seguindo a dicotomia feminino e masculino. A base contém 2.757 registros de docentes vinculados(as) a 241 instituições, com as seguintes informações:

**Quadro 1** – Variáveis da base de dados Acadêmicos

Variável	Descrição
<i>an_base</i>	Ano base da coleta dos dados
<i>an_nascimento_docente</i>	Ano de nascimento do pesquisador
<i>codigo_instituicao</i>	Código da instituição
<i>ds_faixa_etaria</i>	Faixa etária
<i>ds_tipo_nacionalidade_docente</i>	Nacionalidade
<i>ds_tipo_vinculo_docente_ies</i>	Tipo de vínculo do docente
<i>genero</i>	Gênero
<i>nm_area_avaliacao</i>	Nome da área de avaliação
<i>nm_docente</i>	Nome do docente
<i>nm_entidade_ensino</i>	Nome da entidade de ensino ao qual está vinculado
<i>nm_ies_titulacao</i>	Nome da instituição na qual houve titulação
<i>nm_municipio_programa_ies</i>	Nome do município do programa ao qual está associado
<i>nm_pais_ies_titulacao</i>	País da instituição na qual houve titulação
<i>nm_pais_nacionalidade_docente</i>	País da nacionalidade do docente
<i>nm_regiao</i>	Região geográfica da entidade de ensino ao qual está vinculado
<i>nome_completo</i>	Nome completo
<i>nome_instituicao</i>	Nome da instituição
<i>sg_entidade_ensino</i>	Sigla da entidade de ensino
<i>sg_ies_titulacao</i>	Sigla da entidade de titulação
<i>sg_uf_programa</i>	Sigla da UF do programa

Fonte: elaborado pelos autores (2022).

A segunda base, intitulada Produção, contém a produção científica dos docentes extraída dos arquivos XML descarregados da Plataforma Lattes. Nesta base, incluímos a variável “id\_pessoa” a partir da tabela de relacionamento

criada anteriormente. Com essa informação, é possível combinar as duas bases e, por exemplo, totalizar a produção científica por gênero. A base de dados contém 203.357 publicações com as variáveis descritas no Quadro 2, abaixo.

**Quadro 2** – Variáveis da base de dados Publicações

Variável	Descrição
<i>ano</i>	Ano da publicação
<i>id_pessoa</i>	Identificador único do docente na CAPES
<i>idioma</i>	Idioma da produção
<i>país</i>	País da produção
<i>tipo</i>	Tipo da produção: artigo, capítulo de livro e livro

Fonte: elaborado pelos autores (2022).

Com essas duas bases, criamos o painel interativo no serviço em nuvem Google Looker Studio<sup>13</sup>, que será tratado na seção seguinte.

## DESENVOLVIMENTO DE PAINÉIS E OUTRAS VISUALIZAÇÕES

A construção do painel com os dados foi feita a partir do Google Looker Studio, que permite uma conexão com um ou mais bancos de dados, que podem estar disponíveis em diversos formatos, a exemplo do CSV, Microsoft Excel, SQL, entre outros, como também no BigQuery. Com a conexão foi possível desenvolver os gráficos interativos que listamos abaixo, que permitem ao usuário explorar, filtrar, e, assim, desenvolver análises próprias sobre os dados.

O painel foi dividido em 3 setores: docentes/pesquisadores, produção bibliográfica e origens e destinos. O setor de docentes contém os seguintes gráficos:

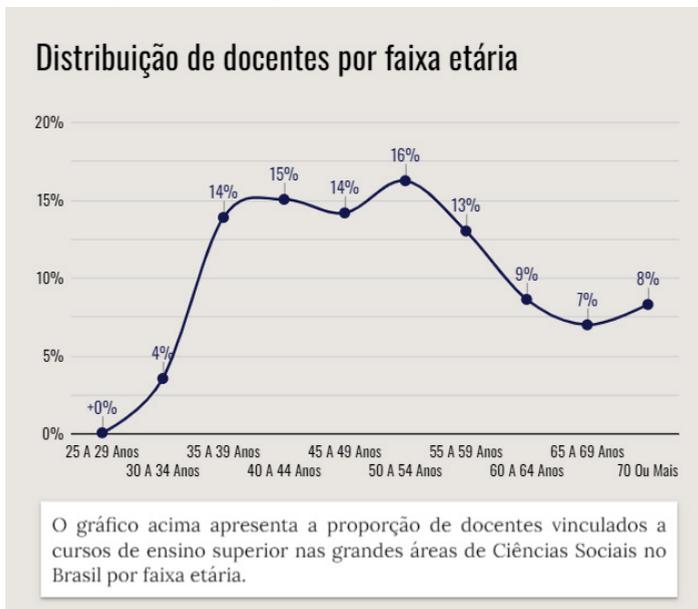
- Distribuição de docentes por área;
- Distribuição de docentes por região;
- Distribuição de docentes por estado;
- Quantidade de programas por área;
- Distribuição da origem dos docentes;
- Distribuição dos docentes por faixa etária;
- Distribuição de docentes por disciplina;
- Distribuição de docentes por região e área.

No setor de produção bibliográfica, estão presentes os gráficos:

- Produção de artigos por idioma;
- Artigos por idioma estrangeiro na última década;
- Produção por categoria;
- Produção por categoria na última década;
- Produção total por gênero.

Ainda que o objetivo da plataforma não seja fornecer análises mais aprofundadas sobre o tema, alguns dados já apresentam tendências que merecem nota. Em termos de gênero, há uma leve sub-representação de mulheres no grupo, já que 57,9% dos docentes são homens contra 42,1% de mulheres. Na área de Ciência Política e Relações Internacionais, contudo, o desequilíbrio é maior, com 65,5% de homens contra 34,5% de mulheres. Em termos regionais, há programas de pós-graduação em quase todos os estados brasileiros, sendo o Acre e o Tocantins as únicas duas exceções. Isso não anula a forte concentração no Sudeste, que possui 42% dos programas. No que concerne à idade média dos docentes, um dos gráficos da plataforma, abaixo reproduzido, evidencia a existência de duas ondas de incorporação de docentes, com um pico nos docentes entre 40 e 44 anos (que somam 15% do total) e nos docentes entre 50 e 54 anos (que somam 16%). Esses dois picos provavelmente refletem momentos de expansão do sistema de ensino superior, um próximo da década de 1990 e outro mais recente. Em termos de produção, há uma tendência crescente na publicação de artigos acadêmicos e um crescimento em ritmo mais lento da publicação de capítulos e livros. Em 2010, os(as) pesquisadores(as) incluídos no recorte publicaram 2.925 artigos, 2.162 capítulos e 707 livros. Em 2020, o número de artigos saltou para 4.379, enquanto o de capítulos foi 2.561 e o de livros 866. Note-se, no entanto, que somente se contabilizou a produção dos professores credenciados no ano do levantamento, o que deve matizar essa análise longitudinal.

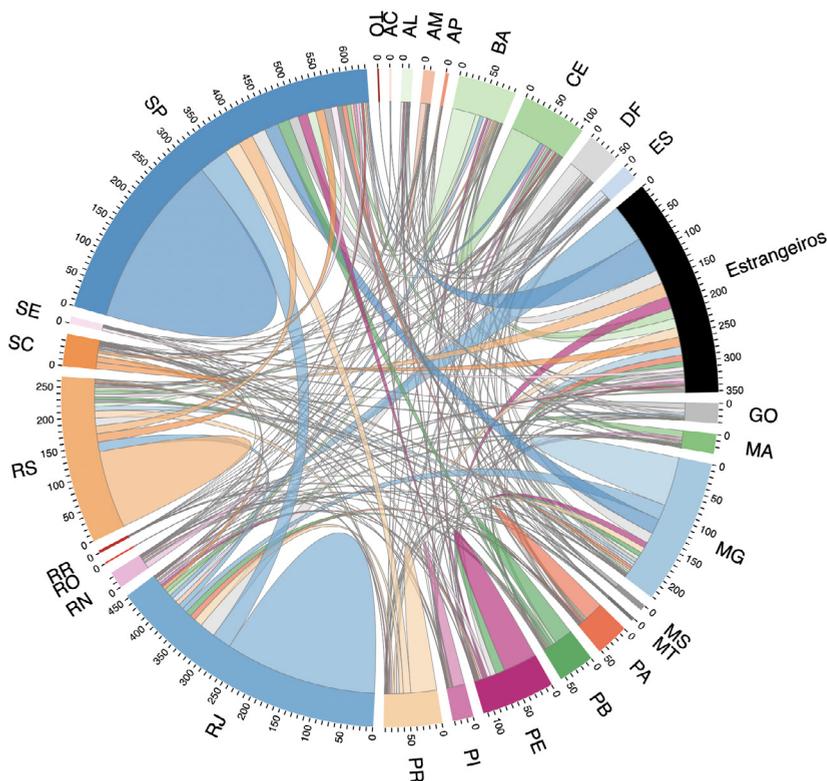
Figura 1 – Distribuição de docentes por faixa etária



Fonte: Atlas Digital das Ciências Sociais.

Por fim, o setor de Origens e Destinos mostra, a partir de um *chord diagram* onde se formaram e para onde foram os pesquisadores(as) em Ciências Sociais no Brasil. O gráfico, que pode ser visto abaixo, permite apresentar de forma interativa e inovadora a movimentação de pesquisadores(as) no país.

Gráfico 1 – Diagrama Chord de Origens e Destinos



Fonte: elaborado pelos autores (2022).

## CONCLUSÃO

Este registro de pesquisa explicitou o funcionamento e as escolhas que conformaram a raspagem de dados do projeto *Atlas Digital das Ciências Sociais*. A iniciativa está em acesso aberto<sup>14</sup>, bem como todas as rotinas que a originaram<sup>15</sup>. Além de fornecer transparência ao seu processo de construção, buscamos também incentivar a aplicação do algoritmo em pesquisas futuras sobre as Ciências Sociais ou as demais áreas do conhecimento no Brasil. O grande volume de dados disponível à comunidade acadêmica no país foi conquistado ao longo dos anos por muito esforço de cientistas em articulação

com gestores políticos favoráveis ao progresso científico. É preciso que continuemos avançando para que mais informações estejam abertas a reflexões e ao debate público, especialmente àquelas que tocam em questões urgentes, como as desigualdades. Esperamos que o presente trabalho colabore para isso.

Editora responsável: Thays Monticelli

Recebido em 21/03/2024 | Revisado em 16/05/2024 | Aprovado em 02/07/2024

**Luiz Augusto Campos** é professor no Instituto de Estudos Sociais e Políticos da Universidade do Estado do Rio de Janeiro, e doutor pela mesma instituição. Coordena o Grupo de Estudos Multidisciplinares da Ação Afirmativa (GEMAA) e o Observatório das Ciências Sociais (OCS). É autor de *Em Busca do Público: a controvérsia das cotas na Imprensa* (EdUERJ) e co-autor de *Raça e Eleições no Brasil* e *Ação Afirmativa: conceito, história e debates* (EdUERJ). É também editor-chefe de DADOS.

**Matheus Cavalcanti Pestana** é professor na Escola de Comunicação, Mídia e Informação da Fundação Getúlio Vargas (FGV-ECMI), doutorando em Ciência Política no Instituto de Estudos Sociais e Políticos da Universidade do Estado do Rio de Janeiro e pesquisador no Instituto de Estudos da Religião.

**André Costantin Felix de Souza** é mestre em Ciência Política pelo Instituto de Estudos Sociais e Políticos da Universidade do Estado do Rio de Janeiro. Pesquisador vinculado aos grupos de pesquisa GEMAA (Grupo de Estudos Multidisciplinar da Ação Afirmativa) e LEMEP (Laboratório de Estudos de Mídia e Esfera Pública) do IESP/UERJ. Tem experiência na área de Ciência da Computação, com ênfase em Linguagem de Programação PHP para as Ciências Sociais por meio da automação de rotinas utilizando scripts e análise estatística de dados nas linguagens R e Python.

**Marcia Rangel Candido** é Doutora em Ciência Política pelo Instituto de Estudos Sociais e Políticos da Universidade do Estado do Rio de Janeiro, onde também foi bolsista de pós-doutorado Faperj Nota 10 do Programa de Pós-Graduação em Sociologia em projeto sobre desigualdades na ciência. Atualmente, é investigadora integrada do Centro de Investigação e Estudos de Sociologia do Instituto Universitário de Lisboa (CIES-Iscte) e editora adjunta de DADOS. É autora do livro *Dois gêneros, duas histórias? A fundação da ciência política no Brasil* (EdUERJ, 2024).

**Mariana Miggiolaro Chaguri** é doutora em Sociologia, professora do Departamento de Sociologia e do Programa de Pós-Graduação em Sociologia da Unicamp. Bolsista de produtividade em pesquisa CNPQ, nível 2. Atua em temas como: pensamento social, gênero e feminismos. Publicou recentemente (2024) “A carreira acadêmica como profissão de mulheres: algumas histórias a partir da FFCL-USP”, revista *Tempo Social* (36/1), em coautoria com Bárbara Pires.

**NOTAS**

- 1 Agradecemos à Fundação Ford pelo financiamento deste estudo e à FAPERJ -Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro pelo financiamento parcial com bolsa de pós-doutorado nota 10 concedida a Marcia Candido e supervisionada por Luiz Augusto Campos, Processo SEI 260003/019653/2022. Agradecemos também aos revisores anônimos pelas sugestões que ajudaram a aprimorar este texto.
- 2 Acrônimo da expressão *Completely Automated Public Turing test to tell Computers and Humans Apart* (Teste de Turing Público Completamente Automatizado para Diferenciar Computadores de Humanos). Um desafio de segurança digital projetado para distinguir entre seres humanos e programas de computador, visando evitar o acesso automatizado a sistemas ou recursos online. Geralmente, os CAPTCHAs apresentam tarefas fáceis para humanos, como identificar objetos em imagens, mas desafiadoras para máquinas, ajudando a proteger contra abuso e atividades maliciosas na internet.
- 3 Anpocs. Atlas Digital das Ciências Sociais, 2021. Disponível em: <http://atlas.anpocs.com>. Acesso em: 15 dez. 2023.
- 4 Projeto conduzido no âmbito da Articulação das Ciências Sociais, que reúne a Associação Nacional de Pós-Graduação e Pesquisa em Ciências Sociais (Anpocs), a Associação Brasileira de Antropologia (ABA), a Associação Brasileira de Ciência Política (ABCP) e a Sociedade Brasileira de Sociologia (SBS).
- 5 Disponível em: <https://dadosabertos.capes.gov.br/dataset?organization=diretoria-de-avaliacao>. Acesso em: 15 dez. 2023.
- 6 Título do arquivo: br-capes-colsucup-docente-2019-2021-03-01.csv
- 7 Sobre as áreas de avaliação da Capes, cf. <https://www.gov.br/capes/pt-br/aceso-a-informacao/acoes-e-programas/avaliacao/sobre-a-avaliacao/areas-avaliacao/sobre-as-areas-de-avaliacao/colégio-de-humanidades/ciencias-humanas/ciencia-politica-e-relacoes-Internacionais>. Acesso em: 15 dez. 2023.
- 8 Linguagem de Marcação Extensível (XML) é um formato de texto simples e altamente flexível derivado do SGML (ISO 8879). Originalmente projetado para atender às demandas da publicação eletrônica em grande escala, o XML também desempenha um papel cada vez mais

- importante na troca de uma ampla variedade de dados na Web e em outros lugares. <https://www.w3.org/XML/>
- 9 Disponível em: <http://buscatextual.cnpq.br/buscatextual/busca.do>. Acesso em: 15 dez. 2023.
- 10 Foi utilizado o pacote FuzzyWuzzy da linguagem Python.
- 11 Disponível em: <https://github.com/meirelesff/genderBR/blob/master/README.md>. Acesso em: 15 fev. 2024.
- 12 Disponível em: <https://censo2010.ibge.gov.br/nomes/#/search>. Acesso em: 15 fev. 2024.
- 13 Disponível em: <https://lookerstudio.google.com/>. Acesso em: 15 dez. 2023.
- 14 Como já citado, disponível em: <http://atlas.anpocs.com/>. Acesso em: 21 mar. 2024.
- 15 Disponível em: <https://github.com/AndreFelix74/lattes-capes-integrator>. Acesso em: 21 mar. 2024.
- 16 ANPOCS. Atlas Digital das Ciências Sociais, 2021. Disponível em: <http://atlas.anpocs.com>. Acesso em: 15 fev. 2024.

## REFERÊNCIAS

- Beigel, Fernanda; Almeida, Ana María; Gallardo, Osvaldo; Digiampietri, Luciano; Gomez, Soledad; Candido, Marcia Rangel; Ciriza, Alejandra; Rossomando, Pia; Moschkovich, Marilia, & Pecheny, Mario. (2023). Scientific Production and Gender Inequalities in Two Academic Elites: Brazil and Argentina. *Revue D'Histoire des Sciences Humaines*, 42.
- Brito, Aline; Quoniam, Luc, & Mena-Chalco, Jesús P. (2016). Exploração da Plataforma Lattes por assunto: proposta de metodologia. *Transinformação*, 28/1, p. 77-86.
- Corrêa, Tiago; Suzuki, Manuella; Cintra, Paulo, & Costa, Luzia. (2017). O fim do scriptLattes? Uma análise de suas funcionalidades, alternativas para o presente e perspectivas para o futuro. *Revista do EDICC*, 3, p. 138-148.
- Ferraz, Renato; Quoniam, Luc, & Maccari, Emerson. (2014). A utilização da ferramenta Scriptlattes para extração e disponibilização on-line da produção acadêmica de um programa de pós-graduação stricto sensu em Administração. *Revista Brasileira de Pós-Graduação*, 11/24, p. 361-389.
- Ferraz, Renato; Quoniam, Luc, & Alvares, Lilian. (2014). Avaliação de redes multidisciplinares com a ferramenta scriptlattes: os casos da nanotecnologia, da dengue e de um programa de pós-graduação Stricto Sensu em Admi-

nistração. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, 19/40, p. 67-98.

Mena-Chalco, Jesús, & Cesar Junior, Roberto. (2009). ScriptLattes: an open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, 15/4, p. 31-39.

Mena-Chalco, Jesús Pascual; Digiampietri, Luciano Antonio, & César Júnior, Roberto Marcondes. (2012). Caracterizando as redes de coautoria de Currículos Lattes. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING – BRASNAM, 2012, Curitiba. Anais [...]. Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, 2012. Disponível em: <http://www.lbd.dcc.ufmg.br/bdbcomp/servelet/Trabalho?id=11999>. Acesso em: 10 fev. 2025.

Nigro, Carolina. (2016). *Uso das ferramentas computacionais scriptlattes, scriptgp e patent2net para análise da produção bibliográfica e tecnológica sobre a dengue*. 152 f. Master – Universidade Nove de Julho, São Paulo.

Nigro, Carolina; Ferraz, Renato; Quoniam, Luc; Alves, Wonder, & Messais, Luis. (2015). Prestação de contas anual e quadrienal à Capes por um programa de Pós-Graduação stricto sensu em Engenharia de Produção: utilização da ferramenta computacional Scriptlattes-Scriptsupira. *Prisma*, 29, p. 3-26.

## ATLAS DAS CIÊNCIAS SOCIAIS

### Palavras-chave

Ciências sociais;  
Indicadores sociais;  
Programação;  
Plataforma Lattes;  
Plataforma Sucupira.

### Resumo

A Plataforma Lattes, que disponibiliza virtualmente currículos acadêmicos, do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), é uma valiosa fonte de informações sobre a ciência brasileira. Contudo, as pesquisas a partir da plataforma vêm se tornando mais complexas por conta das crescentes imposições para acessar os dados, seja pela exigência de autenticação humana no processo, seja porque o material não é disponibilizado pelo CNPq. Neste registro de pesquisa, propomos um método para coleta de um grande volume de currículos, sua conversão em uma base estruturada e combinação com a base da Plataforma Sucupira. Como aplicação prática deste método, apresentaremos o *Atlas das Ciências Sociais*<sup>1</sup>, demonstrando, ainda, como orientar a produção de indicadores sociais com recorte de gênero.

## ATLAS OF BRAZILIAN SOCIAL SCIENCES

### Keywords

Social sciences;  
Social indicators;  
Programming;  
Plataforma Lattes;  
Plataforma Sucupira.

### Abstract

The Plataforma Lattes, which makes academic CVs available virtually from the National Council for Scientific and Technological Development (CNPq), is a valuable source of information on Brazilian science. However, research using the platform has become more complex due to the increasing requirements for accessing the data, either because of the need for human authentication in the process or because the material is not made available by the CNPq. In this research report, we propose a method for collecting a large volume of CVs, converting them into a structured database and combining them with the Plataforma Sucupira database. As a practical application of this method, we will present the Atlas of the Social Sciences, and demonstrate how to guide the production of gender-specific social indicators.