

Identificação e Avaliação da Conexão Científica e Tecnológica entre os Pesquisadores Atuando em Oncologia e Medicina Nuclear Segundo a Base Lattes

Identification and Assessment of the Scientific and Technological Connection Between Researchers Working in Oncology and Nuclear Medicine According to the Lattes Base

Henrique Koch Chaves¹

Fábio Luiz Navarro Marques²

Jesús Pascual Mena Chalco³

Jorge Lima de Magalhães^{1,4,5}

¹Farmanguinhos, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brasil

²Universidade de São Paulo, São Paulo, SP, Brasil

³Universidade Federal do ABC, São Paulo, SP, Brasil

⁴Instituto de Higiene e Medicina Tropical, Global Health and Tropical Medicine, Lisboa, Portugal

⁵Plataforma Internacional de Ciência, Tecnologia e Inovação em Saúde, Aveiro, Portugal

Resumo

Na era do *Big Data*, torna-se premente o compartilhamento aberto de dados em pesquisa, trabalho em rede e a rápida identificação de especialistas em áreas específicas. A Plataforma Lattes é o principal repositório curricular brasileiro, sendo possível extrair diversas informações de pesquisadores, como instituição onde trabalham, linhas de pesquisa, produções científicas, etc. Há limitações na plataforma, como as informações dos especialistas estarem dispostas individualmente e sem um padrão de indexação. Dessa forma, considerando a importância dos radiofármacos para a saúde pública brasileira e mundial, este trabalho objetivou identificar os especialistas mais relevantes em território brasileiro atuando no tema. Utilizou-se como estudo de caso a abordagem quantitativa para análise das informações cadastradas em câncer com especificidade em tecnécio-99m das competências essenciais na área, utilizando a ferramenta computacional *ScriptLattes*. Os resultados mostraram-se eficientes pela geração de conhecimento científico e tecnológico em vários níveis institucionais, podendo ser replicado em diversas áreas da ciência.

Palavras-chave: Plataforma Lattes. ScriptLattes. Câncer.

Abstract

In the era of *Big Data*, the open sharing of data in research, networking and the rapid identification of experts in specific areas becomes urgent. The Lattes Platform is the main Brazilian curriculum repository, making it possible to extract various information from researchers, such as the institution where they work, research lines, scientific productions, etc. There are limitations on the platform, such as the specialists' information being arranged individually and without an indexing standard. Thus, considering the importance of radiopharmaceuticals for Brazilian and global public health, this work aimed to identify the most relevant specialists in Brazilian territory working on the subject. A quantitative approach was used as a case study to analyze the information registered in cancer with specificity in technetium-99m of the essential competences in the area, using the computational tool *ScriptLattes*. The results proved to be efficient in generating scientific and technological knowledge at various institutional levels, which can be replicated in different areas of science.

Keywords: Lattes Database. ScriptLattes. Cancer.

Área Tecnológica: Gestão da Informação. Prospecção Científica. Inovação em Saúde.



1 Introdução

Nas últimas décadas, o aumento exponencial dos volumes de dados disponibilizados na *web* fez surgir, na área de ciência dos dados, o termo *Big Data* (RAJA; MUKHERJEE; SARKAR, 2020). Esse termo pode ser definido de várias formas, por exemplo, pela definição dos “3V’s” de Gartner: alta *velocidade*, alto *volume* e alta *variedade* (GÄRTNER; HIEBL, 2017; WHITE, 2014) Devido à quantidade de novos dados gerados constantemente, torna-se necessário uma análise criteriosa, para que essas informações tenham real aplicabilidade, com base nisso, surge um quarto V, cunhado pela IBM, que é relativo à *veracidade*, ou seja, à confiabilidade dos dados coletados, necessitando, dessa maneira, de ferramentas para a gestão e mineração dos dados incertos (GANDOMI; HAIDER, 2015). Alguns autores ainda propõem a adição de mais duas dimensões: *versatilidade* e *viabilidade*. A combinação dessas seis dimensões levaria a um sétimo V: *valor* (DIAS; DUARTE, 2015). O valor seria gerado por meio da aplicação dos aprendizados obtidos na análise e na interpretação dos dados, de modo que os dados obtidos se convertam em informações útil e utilizável. Por meio do uso das tecnologias do *Big Data*, setores como os da economia, da educação, do transporte e da saúde poderiam ser beneficiados com a utilização sistemática desses dados, gerando *insights* e contribuindo para a proposição de novas soluções.

O aumento expressivo na quantidade de dados gerados se deve ao aumento do uso de tecnologias da informação nos mais diversos locais, como residências, empresas e universidades, podendo ser gerados por vários tipos de fontes, por exemplo, dispositivos móveis, redes sociais, transações bancárias e comerciais, internet das coisas, sensores (dados de localização, dados meteorológicos), satélites, câmeras e sistemas de monitoramento, registros de *softwares*, entre outros (MARTINO *et al.*, 2014). Nesse cenário, o acesso à informação na *web* passa a ser em tempo real (WEB, 2021) e para processar essa vasta quantidade de conteúdo, novas formas computacionais devem ser adotadas para capturar, armazenar, manipular e, assim, extrair valor desses dados, sejam eles estruturados (planilhas, tabelas) ou não estruturados (imagens, blocos de texto, etc.).

Uma vez que o grande volume de dados obtido precisa ser processado em alta velocidade, os métodos tradicionais até então utilizados na área da saúde têm apresentado baixa aplicabilidade, reforçando a necessidade de instrumentos facilitadores para a análise do *Big Data* na área (GALVÃO; VALENTIM, 2019). Tais instrumentos facilitariam a análise desses dados para os gestores e tomadores de decisão em sistemas de pequeno, médio ou grande porte, pois seria possível comparar os dados oriundos de várias fontes, estabelecer correlações e, ainda, prevenir epidemias e doenças (GALVÃO; VALENTIM, 2019). Assim, utilizando as ferramentas do *Big Data* na saúde, além da otimização do tempo, os recursos financeiros poderiam ser melhor administrados, evitando assim desperdícios e gerando melhor prestação de serviços à população (GALVÃO; VALENTIM, 2019).

Um exemplo de *Big Data* no Brasil é o repositório curricular dos pesquisadores, a Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Essa base de dados integra tanto os currículos acadêmicos de profissionais de todas as áreas do conhecimento, principalmente de brasileiros, mas também de pesquisadores estrangeiros que em algum momento desenvolveram projetos em parceria com brasileiros (BRITO; QUONIAM; MENA-CHALCO, 2016). As informações disponibilizadas têm sido essenciais para estratégias técnico-científicas e promovem o estudo e reflexão para o avanço científico e tecnológico do conhecimento em todas as áreas da Ciência, Tecnologia e Inovação (CT&I).

Per si, a plataforma é um banco de dados e o seu sistema de extração está baseado na busca individual por nome da pessoa cadastrada na plataforma. Para permitir a extração de diversos dados simultaneamente, foi desenvolvida uma ferramenta computacional chamada ScriptLattes (BRITO; QUONIAM; MENA-CHALCO, 2016; MAGALHÃES *et al.*, 2020; MENA-CHALCO; JUNIOR, 2009). Esse programa é de código aberto e apresenta, organiza e disponibiliza várias informações dos currículos dos pesquisadores cadastrados, como produções científicas, orientações, colaborações, geolocalização, entre outras (FERRAZ *et al.*, 2018). Assim, essa ferramenta pode permitir que agentes governamentais, sociedades científicas e empresas analisem a distribuição de cientistas e da ciência que é feita em cada região do país.

Nesse cenário, a Plataforma Lattes associada ao programa ScriptLattes foi pensado para ser um instrumento para sumarizar e mapear a produção científica dos pesquisadores cadastrados na plataforma, determinando as regiões do país com capacidade científica instalada e aquelas que deverão ser desenvolvidas, tendo em vista que a atividade científica nem sempre é corretamente mapeada por ministérios e agências governamentais.

Para se ter uma avaliação da viabilidade do uso da Plataforma Lattes e do programa ScriptLattes na gestão das tecnologias do *Big Data*, optou-se por identificar, em território brasileiro, os especialistas de maiores relevâncias na área da medicina nuclear e oncologia, especificamente as competências brasileiras, atuando com o radioisótopo mais utilizado no processo diagnóstico em medicina nuclear, o tecnécio-99m.

A relevância da escolha na área da saúde se fundamenta pelo grande desafio da quimioterapia em entregar seletivamente os fármacos às células tumorais com interação mínima com tecidos saudáveis (WEBSTER *et al.*, 2014). A área da medicina nuclear é responsável pela produção, manipulação, controle de qualidade e fracionamento de radiofármacos (MATHER, 2001). Cabe ressaltar que, segundo dados do Ministério da Saúde (MS) e do Instituto Nacional do Câncer (INCA), o número de mortes causadas por neoplasias está entre as três principais causas de óbitos no Brasil (BRAY *et al.*, 2018; INCA, 2019). Segundo dados recém-publicados pelo INCA, são esperados 704 mil casos novos de câncer no Brasil para cada ano do triênio 2023-2025, com destaque para as Regiões Sul e Sudeste, que concentram cerca de 70% da incidência (INCA, 2023).

Dessa forma, considerando a importância dos radiofármacos para a saúde pública brasileira e mundial, este trabalho objetivou identificar os especialistas mais relevantes em território brasileiro atuando na pesquisa de radiofármacos de tecnécio-99m para o diagnóstico de cânceres, a fim de que pudesse ser dado uma melhor dimensão dessa aplicação nas áreas associativas da medicina nuclear e oncologia.

2 Metodologia

A pesquisa foi baseada em cientometria e bibliometria de dados compartilhados abertos em ciência. Utilizou-se *software* gratuito ScriptLattes em ambiente Linux para extração dos dados abertos na Plataforma Lattes brasileira (<http://scriptlattes.sourceforge.net/>). Entre os filtros disponíveis na Plataforma Lattes para extração dos dados dos pesquisadores cadastrados, foram selecionados os seguintes:

- a) nível (doutorado);
- b) nacionalidade (brasileira e estrangeira);
- c) bolsistas de produtividade CNPq (todas as categorias);
- d) presença nos diretórios de Grupos de Pesquisa.

No sentido de identificar, preliminarmente, o número de currículos disponíveis em cada estratégia de pesquisa, entre fevereiro e abril de 2021, foram realizadas buscas no campo “avançado” da Plataforma Lattes utilizando os descritores tanto em inglês quanto em português, resultando na *string* de busca 1 (Tabela 1). Também aplicou-se filtros associados à categoria “Bolsistas de Produtividade do CNPq” e “Presença no Diretório de Grupos de pesquisa”. A identificação e extração foi realizada em março de 2021.

Uma segunda análise foi realizada utilizando os mesmos pares de palavras e filtros contidos na *string* de busca 1, porém, usando ‘entre parênteses’, conectando-os pelo operador “OR”, resultando *string* de busca 2 (Tabela 1). Nesse caso, ligou-se os dois termos da pesquisa e também foram reunidos todos os documentos que incluam pelo menos um deles (BRITO; QUONIAM; MENA-CHALCO, 2016).

Tabela 1 – Descritores selecionados e *string* de busca utilizada

DESCRITORES EM PORTUGUÊS	DESCRITORES EM INGLÊS
99mTc	99mTc
câncer	cancer
tecnécio	technetium
radiofármaco	radiopharmaceutical

String de busca 1

(99mTc AND câncer), (99mTc AND cancer), (Tecnécio AND câncer), (Technetium AND cancer), (Tecnécio AND radiofármaco), (Technetium AND radiopharmaceutical), (99mTC AND radiofármaco), (99mTc AND radiopharmaceutical)

String de busca 2

“(99mTC AND câncer) OR (99mTc AND cancer) OR (Tecnécio AND câncer) OR (Technetium AND cancer) OR (Tecnécio AND radiofármaco) OR (Technetium AND radiopharmaceutical) OR (99mTC AND radiofármaco) OR (99mTc AND radiopharmaceutical)”

Fonte: Elaborada pelos autores deste artigo

Foram categorizados pelo ID Lattes de cada especialista um código de 16 dígitos de identificação dos currículos disponibilizados na plataforma (MENA-CHALCO *et al.*, 2014). Para ser possível utilizar o Scriptlattes, foi necessário criar uma lista desses currículos, elencando um rol com o número de ID de cada pesquisador na Plataforma Lattes, durante a extração, foi observado que vários pesquisadores estão associados a mais de uma área e de uma grande área, então foi utilizado um algoritmo de deduplicação, existente no *software* ScriptLattes (MENA-CHALCO; JUNIOR, 2009). Os dados foram agrupados, organizados, processados e, posteriormente, disponibilizados em ambiente na *web* em formato HTML.

3 Resultados e Discussão

Os resultados da busca utilizando as palavras tecnécio e câncer, nas diversas formas que são apresentadas na literatura, inclusive escrita na língua inglesa, são apresentadas na Tabela 2. Conforme pode ser observado, ao aplicar todos os filtros utilizando os termos de busca com o operador “AND”, foram encontrados um máximo de 64 currículos. Observa-se, ainda, que ao aplicar o filtro “Bolsistas de Produtividade do CNPq”, a quantidade de currículos diminui drasticamente, ao ponto de que em alguns casos, para 1/8 daquela encontrada inicialmente. Acrescentando o filtro “Presença no Diretório de Grupos de pesquisa”, a quantidade de currículos permanece praticamente constante em todos os termos pesquisados.

Tabela 2 – Número de pesquisadores identificados na base de dados Lattes contendo os termos de busca listados no método

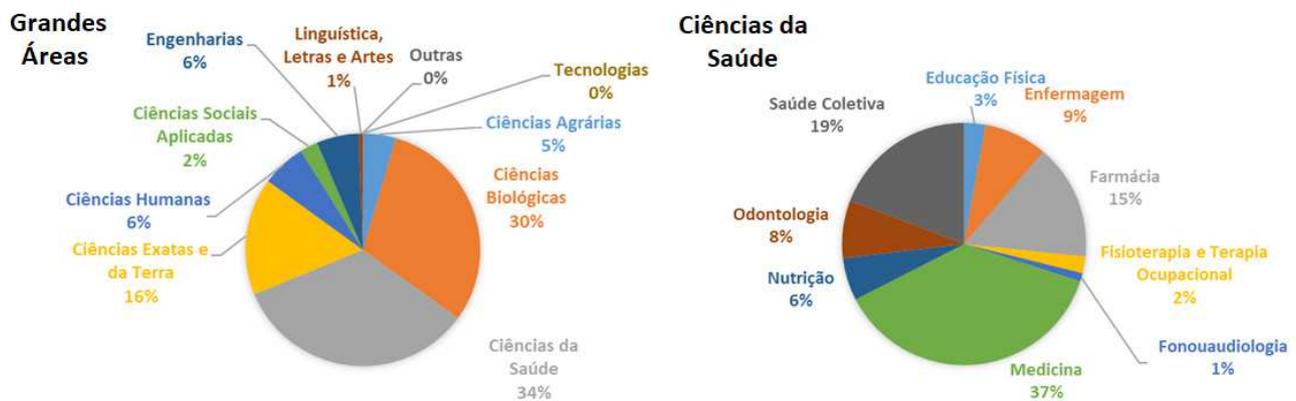
Critérios selecionados para busca na base de dados Lattes	NÚMERO DE CURRÍCULOS IDENTIFICADOS NA PLATAFORMA LATTES UTILIZANDO OS TERMOS DE BUSCA E OPERADORES LISTADOS NO MÉTODO							
	(^{99m} Tc AND câncer)	(^{99m} Tc AND cancer)	(Tecnécio AND câncer)	(Technetium AND cancer)	(Tecnécio AND radiofármaco)	(Technetium AND radiopharmaceutical)	(^{99m} Tc AND radiofármaco)	(^{99m} Tc AND radiopharmaceutical)
Todos (doutores e demais pesquisadores, brasileiros e estrangeiros)	681	681	371	326	165	142	189	174
Doutores brasileiros e estrangeiros	482	482	265	274	113	103	124	129
Doutores brasileiros	469	469	261	264	111	98	122	125
Doutores brasileiros e estrangeiros com qualquer nível de bolsa de produtividade	66	66	40	52	16	15	17	15
Doutores brasileiros com qualquer nível de bolsa de produtividade	66	66	39	52	16	15	17	15
Doutores brasileiros e estrangeiros com qualquer nível de bolsa de produtividade e presença no diretório de GP	64	64	39	52	16	15	17	15
Doutores brasileiros com qualquer nível de bolsa de produtividade e presença no diretório de GP	64	64	38	52	16	15	17	15

Fonte: Elaborada pelos autores deste artigo com base na Plataforma Lattes (2021)

Uma vez que foram encontrados poucos currículos com a primeira estratégia de busca, utilizou-se a segunda estratégia utilizando os mesmos pares de palavras contidos na busca preliminar, porém, usando ‘entre parênteses’, conectando-os pelo operador “OR”, obtendo-se uma maior quantidade de currículos. Assim, adicionou-se mais um filtro: “Atuação Profissional”, no sentido de analisar a quantidade de currículos presentes em cada grande área, bem como nas subáreas presentes (Figura 1).

As três grandes áreas que apresentaram cerca de 80% de currículos do estudo em questão foram: Ciências da Saúde, Ciências Exatas e da Terra e Ciências Biológicas. Como o diagnóstico e terapia do câncer estão principalmente vinculadas à área de Saúde, esta foi desmembrada para suas subáreas, e os resultados são apresentados na Figura 1.

Figura 1 – Percentagem de currículos em cada grande área e na grande área Ciências da Saúde na Plataforma Lattes



Fonte: Elaborada pelos autores deste artigo com base nos currículos encontrados na Plataforma Lattes (2021)

Cabe destacar que, ao somar os currículos de todas as Áreas de uma determinada Grande Área, esta diverge do valor encontrado inicialmente, ou seja, para a Grande Área Ciências da Saúde, foram encontrados inicialmente 1.338 currículos. Como se nota na Tabela 3, ao analisar, separadamente, cada uma de suas subáreas, a soma dos currículos encontrados foi igual a 1.707. Tal comportamento se repetiu para todas as Grandes Áreas. O motivo observado é que o pesquisador pode se cadastrar em mais de uma Área ou Grande área na Plataforma Lattes, resultando, assim, em uma maior quantidade de currículos por área.

Tabela 3 – Quantidade de currículos encontrados em cada Área da Grande Área de Ciências da Saúde

Educação Física	48
Enfermagem	144
Farmácia	263
Fisioterapia e Terapia Ocupacional	38
Fonouaudiologia	19
Medicina	639
Nutrição	97
Odontologia	132
Saúde Coletiva	327
Soma dos currículos de todas as áreas	1707
Total de currículos encontrados inicialmente	1338

Fonte: Elaborada pelos autores deste artigo (2021)

Com esse cenário, foram separados todos os currículos dos pesquisadores das áreas que apresentavam, no mínimo, 12% de currículos em cada Grande Área, a saber:

- I) Ciências da Saúde: Medicina, Farmácia e Saúde Coletiva;
- II) Ciências Biológicas: Bioquímica, Farmacologia e Genética;
- III) Ciências Exatas e da Terra: Química, Física e Ciência da Computação.

As áreas que apresentaram maior quantidade de currículos foram Medicina (Ciências da Saúde), Bioquímica (Ciências Biológicas) e química (Ciências Exatas e da Terra). Foram encontrados 2.954 pesquisadores dentro desses grupos. Durante a extração dos dados dos currículos na Plataforma Lattes, foi observado que vários pesquisadores estão associados a mais de uma área e mais de uma grande área, então, foi utilizado um algoritmo de deduplicação, existente no *software* ScriptLattes (MENA-CHALCO; JUNIOR, 2009). Com essa operação, foram eliminados 801 currículos, permanecendo 2.153 para análise.

Os resultados com informações científicas e tecnológicas das competências elencadas no objetivo na íntegra estão disponíveis no seguinte endereço: <https://pesquisa.ufabc.edu.br/cientometria/tecnecio>.

A página principal gerada pelo ScripLattes no formato HTML pode ser vista na Figura 2. O cabeçalho da página mostra os *links* nos quais se pode ver, por exemplo, a produção bibliográfica, a produção técnica, as orientações, os projetos, entre outras informações dos membros seniores analisados pelo *software*. Pode-se observar pelos resultados obtidos uma vasta publicação na área, com 226.714 artigos publicados em periódicos, e um total de produção bibliográfica de 690.768.

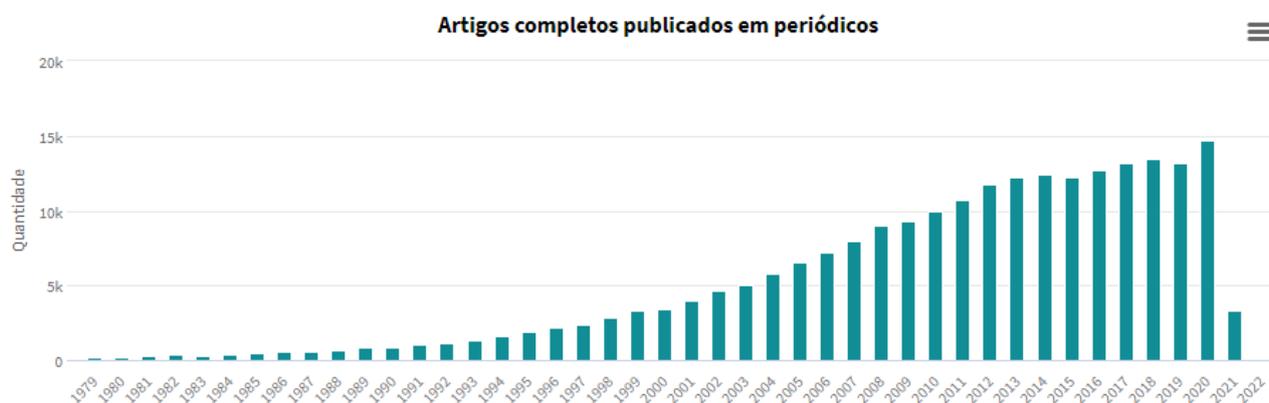
Figura 2 – Menu Principal gerado pelo *software* ScipLattes contendo as informações dos pesquisadores analisados



Fonte: Captura da tela de elaboração própria por meio da ferramenta ScipLattes (2021)

Clicando em qualquer um dos itens desse menu, é possível observar a evolução da produção científica ao longo dos anos (Figura 3). Nesse gráfico, é exibida a quantidade de artigos completos publicados em periódicos por esses pesquisadores desde 1979, e pode ser observado um aumento dessas publicações ano após ano, tendo uma quantidade expressiva nos últimos anos. Os resultados obtidos mostram uma vasta publicação na área, com 226.714 artigos publicados em periódicos e uma produção bibliográfica (artigos, livros, capítulos de livros, dentre outros) de 690.768 itens. Além disso, é possível acessar nessa página a produção por ano a fim de ter a lista de publicações ordenadas alfabeticamente pelo título do artigo publicado.

Figura 3 – Artigos completos publicados em periódicos pelos pesquisadores analisados com o *software* ScipLattes



Fonte: Extraída pelos autores deste artigo por meio da ferramenta ScipLattes

Apesar de a quantidade de artigos publicados ter aumentado ao longo dos anos, o número de novos projetos implementados no Brasil diminuiu principalmente nos últimos anos. Em 2020, foi registrada a menor quantidade de projetos implementados pelos pesquisadores investigados pelo ScipLattes, somando um total de 1.035 projetos. Tal fato pode ser justificado pela queda

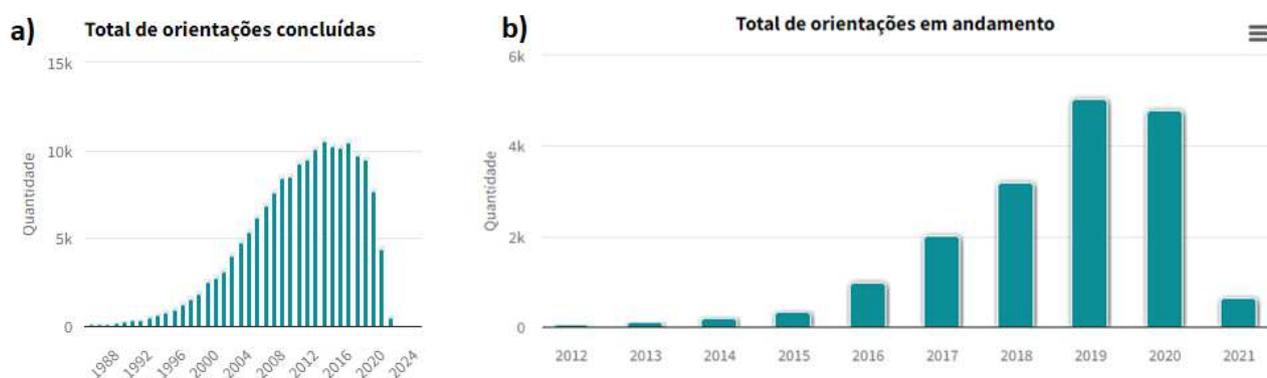
de investimentos em educação no Brasil pelo governo federal, que, entre os anos de 2014 e 2018, diminuiu em torno de 56%, e no governo atual sofreu cortes ainda mais expressivos (MAZIEIRO, 2019; SENADO NOTÍCIAS, 2020).

No que concerne à produção técnica dos especialistas analisados, observa-se os aumentos e as retrações ao longo dos anos, somando um total de 72.757 itens, entre produtos tecnológicos (1.791), processo ou técnicas (1.716), trabalhos técnicos (35.036) e demais tipos de produção técnica (34.214). As primeiras produções técnicas datam de 1983, com um total de 59 produções.

Em relação aos eventos organizados pelos especialistas analisados no presente estudo, observou-se uma grande quantidade de eventos realizados ao longo dos anos, somando um total de 21.206 entre os anos de 1976 e 2021. A partir de 1988 ocorre um aumento constante na quantidade de eventos organizados, com destaque para o ano de 2014, no qual foram registrados 693 eventos. A partir de 2015, ocorreram aumentos e retrações na quantidade de eventos, e no ano de 2020 foi registrada a menor quantidade destes, somando um total de 258 registros. Tal fato pode estar associado à pandemia de Covid-19, que foi relatada pela primeira vez em dezembro de 2019 em Wuhan, na província chinesa de Hubei (AGBEHADJI *et al.*, 2020; ESAKANDARI *et al.*, 2020; STRABELLI; UIP, 2020).

No botão “Orientações” do menu principal, são dispostas as orientações e as supervisões em andamento e concluídas pelos pesquisadores analisados, incluídas orientações de trabalho de conclusão de curso de graduação e especialização (TCC), iniciação científica, mestrado, doutorado, pós-doutorado e orientações de outra natureza. No momento em que foram realizadas as buscas na Plataforma Lattes, foram encontradas um total de 17.561 supervisões/orientações em andamento e 173.004 concluídas. Observa-se certo aumento na quantidade de orientações concluídas ao longo dos anos, apresentando um máximo no ano de 2013 (10.577) e um mínimo no ano de 2020 (4.399). Entre as orientações em andamento, observa-se uma maior quantidade de orientações de alunos de doutorado (7.683), enquanto nas orientações concluídas, observar-se uma maior quantidade de orientações de alunos de iniciação científica (50.676). As primeiras orientações concluídas com a temática em questão aparecem no ano de 1982, somando um total de 110 orientações.

Figura 4 – a) total de orientações concluídas; b) total de orientações em andamento pelos pesquisadores analisados pelo *software* ScriptLattes



Fonte: Extraída pelos autores deste artigo por meio da ferramenta ScriptLattes

O mapa de geolocalização dos especialistas, integrado ao Google Maps, pode ser identificado na Figura 5. Nota-se que os símbolos em verde identificam onde cada pesquisador sênior

atua, ou seja, de acordo com endereço profissional informado em seu currículo (MAGALHÃES *et al.*, 2020). Ao clicar nesse símbolo, é possível ver algumas informações do pesquisador, como o seu nome, a universidade ou instituto em que trabalha e o endereço para acessar o Currículo Lattes. Como pode ser observado no mapa, a maioria dos pesquisadores está concentrada na Região Sudeste do país, especialmente no Estado de São Paulo, que é o estado brasileiro com maior número absoluto de universidades públicas, presentes em pelo menos 25 cidades (FIA BUSINESS SCHOLL, 2019; JORNAL DA USP, 2019). O setor de medicina nuclear do país conta com 432 serviços de medicina nuclear distribuídos por todo o território brasileiro, contando atualmente com 38 radiofármacos. No Brasil, a principal unidade produtora da Comissão Nacional de Energia Nuclear (CNEN) é o Instituto de Pesquisas Energéticas e Nucleares (IPEN), localizado no Campus da USP (São Paulo – SP), que, além de produzir os 38 diferentes radiofármacos contidos no portfólio da CNEN, é ainda o único produtor de geradores de tecnécio-99m no país (CNEN, 2022).

Figura 5 – Mapa de localização de habilidades essenciais em tecnécio no Brasil

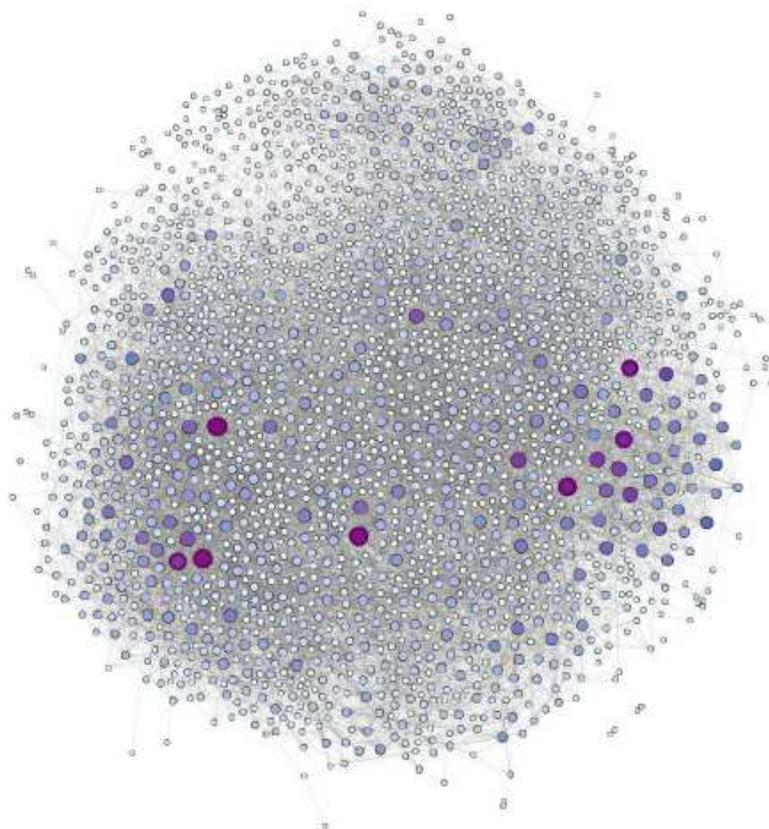


Fonte: Extraída pelos autores deste artigo por meio da ferramenta ScriptLattes (2021)

A colaboração acadêmica entre os pesquisadores selecionados foi identificada com base nas publicações realizadas em coautoria entre eles. Há muito tempo os relacionamentos entre as pessoas têm sido modelados pelas redes, tornando possível identificar, por exemplo, quais são os indivíduos mais importantes para que uma rede permaneça conectada, a distância que separa cada um dos seus colaboradores, quais são os colaboradores da rede que estão conec-

tados ao maior número de outros colaboradores (CAVALCANTE, 2009). Assim, pode-se inferir que a sociedade contemporânea encontra-se imensamente dependente das inúmeras redes de relacionamento atualmente existentes. A ferramenta Screenshot oferece várias possibilidades de visualização gráfica dessas redes, por meio de outras ferramentas computacionais como Treecloud®, Gephi®, Cytoscape® e VOSviewer®, já acopladas ao Screenshot (FERRAZ; QUONIAM; ALVARES, 2014). Na Figura 6 é apresentada a rede de colaboração em que os vértices (2.153) representam os pesquisadores, e as arestas (20.742) representam as ligações entre os vértices. Tanto os diâmetros dos vértices quanto a sua cor estão associados ao número de conexões de cada pesquisador. Um pesquisador bem conectado terá um vértice de tamanho maior e a cor roxa escura, enquanto um pesquisador menos conectado terá um tamanho de vértice pequeno e uma cor roxa clara. Já a centralidade de grau (*centrality degree*) corresponde ao número de arestas incidentes ou ao número de vértices adjacentes a ele (GIORDANO; BRUNING; BORDIN, 2015). Assim, a centralidade de grau aponta quais são os autores que mais colaboraram, publicando conjuntamente com outros autores, levando em consideração o número de coautores que colaboraram com um determinado autor, juntamente com o número de publicações que estes fizeram em parceria (BORDIN; GONÇALVES; TODESCO, 2014). O vértice mais conectado tem a centralidade de grau igual a 130, isto é, ele colabora de forma direta com 130 pesquisadores identificados no contexto desta pesquisa. É importante destacar que essa rede de pesquisadores é coesa, dado que existe uma grande colaboração entre eles. A maior componente conexa da rede está composta de 2.116 pesquisadores (aproximadamente 98%) com 20.739 conexões.

Figura 6 – Rede de coautoria entre os 2.153 pesquisadores relacionados com tecnício no Brasil

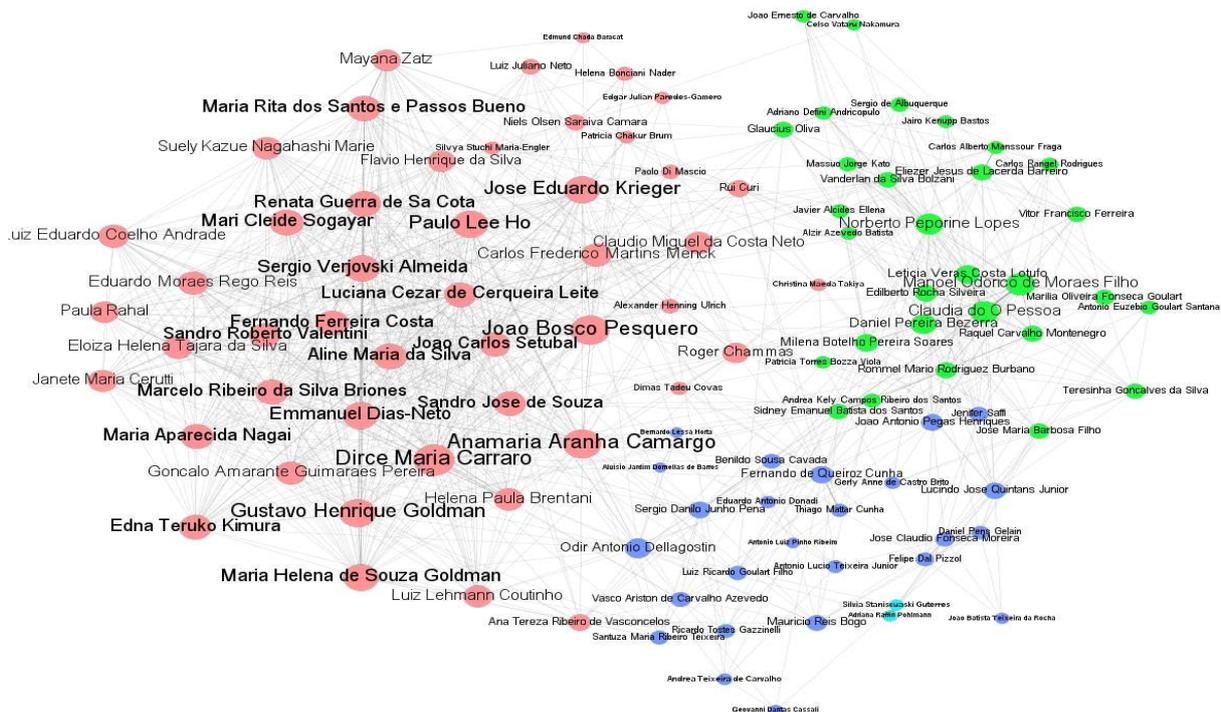


Fonte: Elaborada pelos autores deste artigo (2021)

Da rede de colaborações, foi identificado que existem apenas 105 pesquisadores com centralidade de grau maior ou igual a 50, ou seja, 105 pessoas colaboram com no mínimo outros 50 pesquisadores). Esse grupo seletivo de pesquisadores está representado na Figura 7. O tamanho do vértice representa o número de conexões de cada pesquisador. Já a cor representa um dos quatro grupos identificados pela sua forma de colaboração. Para essa finalidade, foi utilizado o algoritmo de Modularidade para determinar as comunidades baseadas na sua topologia ou configuração da rede. Em redes complexas, a modularidade é uma das possíveis medidas para a detecção de comunidades. Dessa maneira, um conjunto de vértices é classificado como uma comunidade por sua modularidade se a fração de ligações entre eles é maior do que o esperado. A modularidade é uma medida da rede como um todo, dividindo a rede em grupos, de acordo com a força das conexões entre os vários vértices. Nos mais conectados entre si do que aos demais, são incluídos em um mesmo grupo. Nesse tipo de gráfico, uma rede complexa com elevado grau de modularidade aponta uma estrutura de comunidade forte, ou seja, os vértices da comunidade têm uma conexão densa entre si e uma ligação esparsa com as outras comunidades (VINCENZO, 2008). O maior grupo está composto de 48 pessoas e representado em cor vermelha. O segundo grupo em tamanho está representado na cor verde e é composto de 30 pessoas. O terceiro grupo está representado na cor azul e é composto de 25 pessoas. Finalmente, o menor grupo está composto de somente duas pessoas e representado na cor ciano.

Evidencia-se que os pesquisadores com maiores tamanhos em cada agrupamento são líderes de pesquisa nas suas áreas e desenvolvem elevado número de pesquisa colaborativa. Todavia, por meio da análise nominal dos 105 pesquisadores evidenciados, não foram encontrados pesquisadores que trabalham na área de medicina nuclear ou da utilização de radioisótopos, indicando ser uma área que precisa ter seu desenvolvimento científico estimulado.

Figura 6 – Rede de coautoria entre os 105 pesquisadores mais colaborativos



Fonte: Elaborada pelos autores deste artigo (2021)

A ferramenta computacional ScriptLattes atualmente tenta minimizar os problemas encontrados tanto nesse como em outros estudos já reportados na literatura (MAGALHÃES *et al.*, 2020, 2014; MENA-CHALCO; JUNIOR, 2009). Conforme apontado anteriormente, um dos problemas encontrados na Plataforma Lattes é a diferença entre as atualizações dos currículos, uma vez que alguns autores atualizam seus currículos mensalmente e outros anualmente. Dessa maneira, um currículo recém-atualizado pode conter dados desatualizados, dificultando a extração precisa deles. Para minimizar esses problemas, recomenda-se um maior empenho dos pesquisadores e das instituições em padronizar a indexação de seus dados e das informações na plataforma. Esse fato facilitará o processo de identificação mais precisa e, conseqüentemente, a extração e recuperação dos dados mais fidedignos à pesquisa. Dessa forma, novas metodologias deverão ser aprimoradas e desenvolvidas para que a identificação, a extração e a posterior mineração desses dados sejam mais assertivas.

4 Considerações Finais

Devido ao crescimento exponencial no volume de dados adicionados diariamente à *web*, torna-se indispensável na era do *Big Data* a extração, a identificação e o processamento desses dados. Novas metodologias devem ser desenvolvidas para que a mineração desses dados seja mais assertiva. Dessa maneira, o trabalho apresentou um estudo de caso utilizando o ScriptLattes, uma ferramenta computacional de código aberto para extração e visualização de dados científicos e produções dos currículos cadastrado na Plataforma Lattes.

A partir das informações obtidas, foi possível obter importantes indicadores sobre alguns aspectos da ciência no Brasil na área do presente estudo. Foi possível fazer uma análise micro (pesquisador individual e grupos de pesquisadores) e macro (redes de colaboração ou representatividade da área), retratando o que é desenvolvido e publicado no Brasil nos setores de PD&I.

O ScriptLattes mostrou-se eficiente na extração e processamento dos dados dos currículos dos pesquisadores cadastrados na Plataforma Lattes. Todavia, ao introduzir o operador booleano OR para conectar as buscas das palavras entre parênteses, organizadas pelo operador AND, o sistema exacerbou a palavra câncer e suas derivações em detrimento da palavra técnico e suas derivações, evidenciando os pesquisadores que desenvolvem trabalho com câncer e que são importante “*hub*” de conexões. É possível que pesquisadores na área de radiofarmácia e medicina nuclear estejam conectados aos 105 pesquisadores, mas sua rede de conexões é pequena e não pode ser observada na figura para demonstrar sua amplitude.

A Plataforma Lattes permite, por meio dos operadores booleanos, que se obtenham informações para buscas simples. Buscas relacionais mais complexas ainda não são possíveis, uma vez que o motor de busca disponível da plataforma não permite extração e recuperação das informações de forma gerencial, que era o almejado neste estudo, por exemplo, a identificação dos termos de busca numa frase (parágrafo), *abstract* ou título da produção científica ou tecnológica dos especialistas.

Outra limitação observada nesta pesquisa é em relação à indexação de termos na Plataforma Lattes realizada pelos especialistas. Em razão de não apresentar um padrão de citação ou até filiação a GP que mensure e divulgue o escopo de sua pesquisa, alguns currículos essenciais,

provavelmente, não foram identificados. Dessa forma, outros estudos com diferentes metodologias devem ser realizados, com o propósito de abarcar os especialistas que não apareceram nesse grupo.

A real contribuição deste trabalho recaiu na possibilidade de se obter na Plataforma Lattes informações essenciais dos especialistas em oncológicos contendo o radioisótopo ^{99m}Tc no Brasil. Ainda mais se tratando de saúde pública, a gestão do conhecimento torna-se um fator imprescindível para avanços na área, possibilitando que as informações obtidas sejam utilizadas pelos tomadores de decisão de maneira eficaz e em tempo ágil.

5 Perspectivas Futuras

Considerando que os currículos indexados na Plataforma Lattes possuem uma grande quantidade de informações essenciais, este se traduz em um excelente repositório da produção científica brasileira. Como os dados existentes são atualizados constantemente e disponibilizados pelos pesquisadores na plataforma, a base apresenta um rico material para a análise, avaliação e construção de redes, não só de pesquisa, mas como subsídios para a tomada de decisão científica e tecnológica. Portanto, novos trabalhos deveriam ser realizados, a fim de aprofundar a metodologia aqui utilizada e replicar em novas áreas da ciência.

Sugere-se ampliar o escopo de análise para os outros radioisótopos utilizados pelo Sistema Único de Saúde (SUS) no radiodiagnóstico dos vários tipos de cânceres, por exemplo, iodo-123, iodo-131, cromo-51, flúor-18, samário-153, índio-111, lutécio-177, entre outros. No que tange à Gestão em Saúde, os resultados apresentados neste trabalho possibilitarão aos órgãos públicos regulamentadores uma análise apurada sobre a evolução das pesquisas relacionadas ao tema no Brasil, especialmente em saúde pública. Ademais, o mapa de geolocalização dos pesquisadores, obtido com o uso do ScriptLattes, poderá nortear os órgãos fomentadores de pesquisa para a criação de editais específicos baseados na demanda do estado ou de região do país, direcionando, assim, de maneira mais efetiva os recursos financeiros.

Referências

- AGBEHADJI, I. E. *et al.* Review of Big Data Analytics, Artificial Intelligence and Nature-Inspired Computing Models towards Accurate Detection of COVID-19 Pandemic Cases and Contact Tracing. **International Journal of Environmental Research and Public Health**, [s.l.], v. 17, n. 15, p. 5.330, jan. 2020.
- BORDIN, A. S.; GONÇALVES, A. L.; TODESCO, J. L. Análise da colaboração científica departamental através de redes de coautoria. **Perspectivas em Ciência da Informação**, [s.l.], v. 19, p. 37-52, jun. 2014.
- BRAY, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinicians**, [s.l.], v. 68, n. 6, p. 394-424, 2018.
- BRITO, A. G. C. de; QUONIAM, L.; MENA-CHALCO, J. P. Exploração da Plataforma Lattes por assunto: proposta de metodologia. **Transinformação**, [s.l.], v. 28, n. 1, p. 77-86, abr. 2016.

- CAVALCANTE, G. V. Ciência das Redes: Aspectos Epistemológicos. **Revista Ibero-Americana de Ciência da Informação**, [s.l.], v. 2, n. 1, 2009.
- CNEN – COMISSÃO NACIONAL DE ENERGIA NUCLEAR. **RMB e a Produção de Radiofármacos**. 2022. Disponível em: <http://antigo.cnen.gov.br/radiofarmacos>. Acesso em: 27 abr. 2023.
- DIAS, J. A.; DUARTE, P. Big Data Opportunities in Healthcare. How Can Medical Affairs Contribute? **Ver. Port. Farmacoter**, [s.l.], v. 7, p. 230-236, 2015.
- ESAKANDARI, H. *et al.* A comprehensive review of COVID-19 characteristics. **Biological Procedures Online**, [s.l.], v. 22, n. 1, p. 19, dez. 2020.
- FERRAZ, R. R. N. *et al.* Aspectos históricos da criação dos grupos de pesquisa em dengue no Brasil com a utilização da ferramenta computacional ScriptGP. **Ciência & Saúde Coletiva**, v. 23, n. 3, p. 837-848, mar. 2018.
- FERRAZ, R. R. N.; QUONIAM, L.; ALVARES, L. M. A. de R. Avaliação de redes multidisciplinares com a ferramenta scriptlattes: os casos da nanotecnologia, da dengue e de um programa de pós-graduação Stricto Sensu em Administração. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, [s.l.], v. 19, n. 40, p. 67, 2 ago. 2014.
- FIA BUSINESS SCHOLL. **Universidades Públicas: o que são, importância e lista de instituições**, 30 maio 2019. Disponível em: <https://fia.com.br/blog/universidades-publicas/>. Acesso em: 28 set. 2022.
- GALVÃO, A. B.; VALENTIM, R. A. de M. Desafios para os Avanços da Análise de Big Data na Saúde. Anais Estendidos do Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS). In: XIX SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE. SBC, 11 jun. 2019. **Anais [...]**. [S.l.], 2019. Disponível em: https://sol.sbc.org.br/index.php/sbcas_estendido/article/view/6301. Acesso em: 9 jun. 2021.
- GANDOMI, A.; HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. **International Journal of Information Management**, [s.l.], v. 35, n. 2, p. 137-144, abr. 2015.
- GÄRTNER, B.; HIEBL, M. R. W. Issues with Big Data. In: GÄRTNER, B.; HIEBL, M. R. W. **The Routledge Companion to Accounting Information Systems**. [S.l.]: Routledge, 2017.
- GIORDANO, D. M.; BRUNING, E.; BORDIN, A. S. Uso do scriptLattes e Gephi na Análise da Colaboração Científica. **Anais do Computer on the Beach**, p. 239-248, 2015.
- INCA – INSTITUTO NACIONAL DO CÂNCER. **Estimativa 2020**: Apresentação. 2019. Disponível em: <https://www.inca.gov.br/estimativa>. Acesso em: 20 jul. 2021.
- INCA – INSTITUTO NACIONAL DO CÂNCER. **Estimativa 2023**: incidência de câncer no Brasil. 2023. Disponível em: <https://www.inca.gov.br/publicacoes/livros/estimativa-2023-incidencia-de-cancer-no-brasil>. Acesso em: 27 abr. 2023.
- JORNAL DA USP. Universidades públicas paulistas e o Estado de São Paulo. **Jornal da USP**, 16 maio 2019. Disponível em: <https://jornal.usp.br/artigos/universidades-publicas-paulistas-e-o-estado-de-sao-paulo/>. Acesso em: 28 set. 2022.
- MAGALHÃES, J. *et al.* A Management Tool to Aid in the Tropical Outbreak of the 21st Century: Senior Scientists and Their Knowledge of the Triple Threat Dengue, Zika and Chikungunya. **Problems of Management in the 21st Century**, [s.l.], v. 15, n. 1, p. 40-55, 15 jun. 2020.

- MAGALHÃES, J. L. *et al.* Extração e tratamento de dados na base lattes para identificação de core competencies em dengue. **Informação & Informação**, [s.l.], v. 19, n. 3, p. 30, 30 ago. 2014.
- MARTINO, B. D. *et al.* Big data (lost) in the cloud. **International Journal of Big Data Intelligence**, [s.l.], v. 1, n. 1/2, p. 3, 2014.
- MATHER, S. J. Innovation in radiopharmacy: progress and constraints? **European Journal of Nuclear Medicine**, [s.l.], v. 28, n. 4, p. 405-407, 1º abr. 2001.
- MAZIEIRO, G. **Em 4 anos, Brasil reduz investimento em educação em 56%; cortes continuam**. 2019. Disponível em: <https://educacao.uol.com.br/noticias/2019/05/02/em-4-anos-brasil-reduz-investimento-em-educacao-em-56.htm>. Acesso em: 22 ago. 2022.
- MENA-CHALCO, J. P. *et al.* Brazilian bibliometric coauthorship networks. **Journal of the Association for Information Science and Technology**, [s.l.], v. 65, n. 7, p. 1.424-1.445, 2014.
- MENA-CHALCO, J. P.; JUNIOR, R. M. C. scriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, [s.l.], p. 10, 2009.
- RAJA, R.; MUKHERJEE, I.; SARKAR, B. K. A Systematic Review of Healthcare Big Data. **Scientific Programming**, [s.l.], v. 2020, p. 1-15, 13 jul. 2020.
- SENADO NOTÍCIAS. **Senadores criticam corte de R\$ 4,2 bi do orçamento da Educação para 2021**. 2020. Disponível em: <https://www12.senado.leg.br/noticias/materias/2020/08/12/senadores-criticam-corte-de-r-4-2-bi-do-orcamento-da-educacao-para-2021>. Acesso em: 22 ago. 2022.
- STRABELLI, T. M. V.; UIP, D. E. COVID-19 e o Coração. **Arquivos Brasileiros de Cardiologia**, [s.l.], v. 114, p. 598-600, 30 mar. 2020.
- VINCENZO, N. **Modularity for community detection**: history, perspectives and open issues. [S.l.: s.n.], 2008. p. 67.
- WEB, N. N. de D. **Revista Rios – UNIRIOS – Centro Universitário em Paulo Afonso/BA**. 2021. Disponível em: <https://www.unirios.edu.br/revistarios>. Acesso em: 9 jun. 2021.
- WEBSTER, W. D. *et al.* The production of radionuclides for nuclear medicine from a compact, low-energy accelerator system. **Nuclear Medicine and Biology**, [s.l.], v. 41, p. e7-e15, 1º maio 2014.
- WHITE, S. E. A review of big data in health care: challenges and opportunities. **Open Access Bioinformatics**, [s.l.], v. 6, p. 13-18, 31 out. 2014.

Sobre os Autores

Henrique Koch Chaves

E-mail: henrique.chaves@fiocruz.br

ORCID: <https://orcid.org/0000-0003-3035-6799>

Pós-doutor pela Fiocruz em 2021.

Endereço profissional: Instituto de Tecnologia em Fármacos, Farmanguinhos, Fiocruz, Rua Sizenando Nabuco, n. 100, Manguinhos, RJ. CEP: 21041-000.

Fábio Luiz Navarro Marques

E-mail: fabio.marques@fm.usp.br

ORCID: <https://orcid.org/0000-0001-7383-926X>

Doutor em Ciências pela Faculdade de Medicina da USP em 2007.

Endereço profissional: Laboratório de Medicina Nuclear (LIM43), Departamento de Radiologia e Oncologia, Faculdade de Medicina FMUSP, Universidade de São Paulo, Rua Doutor Ovídio Pires de Campos, n. 872, Cerqueira César, São Paulo, SP. CEP: 05403-911.

Jesús Pascual Mena Chalco

E-mail: jesus.mena@ufabc.edu.br

ORCID: <https://orcid.org/0000-0001-7509-5532>

Pós-doutor em Ciência da Computação pela Universidade de São Paulo em 2011.

Endereço profissional: Universidade Federal do ABC, Centro de Matemática, Computação e Cognição. Avenida dos Estados, n. 5.001, Bangu, Santo André, SP. CEP: 09210-580.

Jorge Lima de Magalhães

E-mail: jorge.magalhaes@fiocruz.br

ORCID: <https://orcid.org/0000-0003-2219-5446>

Pós-doutor pela Aix Marseille Université, France (Laboratoire IRSIC EA4262 – Institut de Recherche en Sciences de l'Information et de la Communication) em Inteligência Competitiva para saúde pública em 2013.

Endereço profissional: Instituto de Tecnologia em Fármacos, Farmanguinhos, Fiocruz, Rua Sizenando Nabuco, n. 100, Manguinhos, RJ. CEP: 21041-000.