

Universidade Federal de São Carlos
Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade

**Avaliação da Plataforma Lattes como fonte de coleta de
metadados para povoamento automatizado em
repositórios institucionais**

Denilson de Oliveira Sarvo

São Carlos – SP
2018

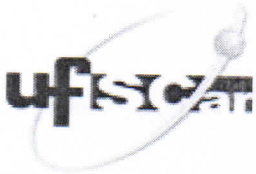
DENILSON DE OLIVEIRA SARVO

Avaliação da Plataforma Lattes como fonte de coleta de metadados para povoamento automatizado em repositórios institucionais

Dissertação apresentada ao Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, do Centro de Educação e Ciências Humanas, da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência, Tecnologia e Sociedade.

Orientador: Prof. Dr. Roniberto Morato do Amaral

São Carlos – SP
2018



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Denilson de Oliveira Sarvo, realizada em 27/02/2018:

Prof. Dr. Roniberto Morato do Amaral
UFSCar

Prof. Dr. Douglas Henrique Milanez
UFSCar

Profa. Dra. Aline Grasielle Cardoso de Brito
CEUCLAR

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Aline Grasielle Cardoso de Brito e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ao) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof. Dr. Roniberto Morato do Amaral

AGRADECIMENTOS

A minha família, por sempre me acompanhar e apoiar nessa jornada.

Ao meu orientador, Prof. Dr. Roniberto Morato do Amaral, pelos conhecimentos compartilhados ao longo desses anos, por toda a confiança e oportunidade dada que refletem em meu desenvolvimento intelectual e profissional.

A toda equipe do NIT-Materiais, Leandro Innocentini Lopes de Faria, Douglas Henrique Milanez, Vera Aparecida Lui Guimarães, Mesailde Souza de Oliveira Matias, Celise Vila dos Santos, Bráulio Salumão de Oliveira, Lucas Faccioni Chanchetti, Adriana Spinola, Francisco Rocha Pirolla, José Eduardo dos Reis, Marcela Bassoli, Nathalia Mendes Gerotti Franco, Raquel Santos Maciel e Thais da Silva Justino por toda aprendizagem, troca de experiências e conhecimento. Agradeço a todos pela oportunidade de trabalhar em um ambiente tão desafiador e colaborativo.

A equipe do SIBi-UFSCar e BCo-UFSCar, em especial a Dr^a Eliane Colepicolo, por toda a oportunidade, apoio e ensinamento.

Ao Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, por todo o conhecimento e enriquecimento intelectual após esse desafio. Agradeço a Prof^ª. Dr^ª. Ariadne Chloe Mary Furnival por sempre compartilhar seus conhecimentos e ser uma fonte inspiradora para o trabalho com o acesso aberto à informação. Ao Paulo Augusto Lazaretti pelos esclarecimentos e atenção durante todo o decorrer do mestrado.

Agradeço a Terezinha Gagliardi que acompanhou o início de toda essa jornada, obrigado pelo apoio, confiança e amizade em todos esses anos.

RESUMO

Os repositórios institucionais (RI) estão ligados com a estratégia da Via Verde adotada pelo Movimento de Acesso Aberto que visa criar arquivos digitais para o armazenamento e disponibilização de informação científica e tecnológica. Os RI trazem benefícios como a eliminação de barreiras de acesso à informação, aumento de visibilidade, preservação, além de se configurar como uma fonte estratégica para a confecção de indicadores sobre produção científica institucional. A implementação de RI está relacionada com aspectos tecnológicos, humanos, políticos e de gestão, tendo como um dos principais pontos críticos a garantia de seu povoamento, que compreende a coleta, inserção, validação e disponibilização de publicações em texto completo em suas coleções. Ações de povoamento têm se mostrado ineficazes diante das dificuldades para a recuperação da produção científica institucional, indexadas em parte em bases de dados de informação científica e tecnológica. Diante da hipótese de que a Plataforma Lattes pode minimizar a deficiência na recuperação da produção científica de instituições de ensino superior, a pesquisa teve como objetivo avaliar a capacidade de recuperação de metadados a partir da Plataforma Lattes, comparando-a com as bases de dados *Web of Science*, Scopus e SciELO. Para a comparação foi realizado um estudo de caso, tendo como unidade-caso a Universidade Federal de São Carlos (UFSCar), sendo utilizada a ferramenta SyncLattes para obtenção de registros que compõem a amostra analisada e aplicada a bibliometria como técnica de análise. Os resultados alcançados compreendem: 1) a recuperação da produção científica da UFSCar registrada na Plataforma Lattes; 2) a identificação da parcela de registros da Plataforma Lattes também indexadas nas bases de dados *Web of Science*, Scopus e SciELO; 3) a identificação das publicações da UFSCar em fontes de acesso aberto; e 4) a atribuição do fator de embargo nos registros recuperados. Conclui-se que a Plataforma Lattes é uma fonte de coleta de metadados estratégica para o povoamento de RI, uma vez que garantiu exclusivamente a cobertura de uma parcela significativa da produção científica da unidade-caso estudada.

Palavras-chave: Acesso Aberto. Repositórios Institucionais. Recuperação da informação. Plataforma Lattes.

ABSTRACT

Institutional Repositories (IR) are related to the Green Road strategy adopted by the Open Access Movement, which aims to create digital archives for storing and providing access to scientific and technological information. IR bring benefits such as the elimination of barriers to information access, enhancement of visibility and preservation, and also appear as a strategic source for the development of indicators of institutional scientific production. Their implementation is associated with technological, human, political and managerial aspects, and a crucial factor for this is ensuring that they are populated with data. This process comprehends collection, insertion, validation and provision of full text publications in their collections. Actions intended to populate IR have shown themselves ineffective in face of the difficulties to retrieve institutional scientific production, partly indexed on scientific and technological information databases. Facing the hypothesis that the Lattes Platform might minimize deficiency in the retrieval of scientific production of higher education institutions, this study aimed to assess the capacity to retrieve metadata from the Lattes Platform, comparing it to Web of Science, Scopus and SciELO databases. For the sake of comparison, a case study was conducted, our unit case being the Federal University of São Carlos (UFSCar). We have utilized the SyncLattes tool for obtaining the registers that constitute the analyzed sample, and have applied bibliometry as our method of analysis. The results obtained comprise: 1) the retrieval of the scientific production from UFSCar registered in the Lattes Platform; 2) identification of the share of registers from the Lattes Platform also indexed on the Web of Science, Scopus and SciELO databases; 3) identification of UFSCar publications in open access sources; and 4) the assignment of the embargo factor on the registers retrieved. It has been concluded that the Lattes Platform is a metadata collection source that is strategic for populating IR, since it alone has guaranteed the coverage of a significant share of the scientific production in the unit-case analyzed.

Key-words: Open Access. Institutional Repositories. Information retrieval. Lattes Platform.

LISTA DE FIGURAS

Figura 1 - Número de registros depositados em RI nacionais.....	29
Figura 2 - Interface de busca do Currículo Lattes (Busca simples).....	35
Figura 3 - Interface de busca do Currículo Lattes (Busca avançada).....	35
Figura 4 - Interface de busca do Diretório de Instituições	36
Figura 5 - Interface de busca do Diretório de Grupos de Pesquisa	37
Figura 6 - Resultado de busca do Diretório de Grupos de Pesquisa a nível de pesquisador....	37
Figura 7 - Resultado de busca do Diretório de Grupos de Pesquisa a nível de grupo.....	38
Figura 8 - Processo de execução do ScriptLattes	40
Figura 9 - Interface de busca do Somos (UFMG)	41
Figura 10 - Resultados de busca do Somos (UFMG).....	41
Figura 11 - Interface de busca da plataforma Stela Experta.....	42
Figura 12 - Interface de busca do WeR-USP	43
Figura 13 - Processos de coleta, tratamento e sincronização do SyncLattes.....	52
Figura 14 - Processo de coleta e tratamento de dados realizado por Bassoli (2017)	53
Figura 15 - Processo de comparação entre Plataforma Lattes, WOS, Scopus e SciELO.....	54
Figura 16 - Interface gráfica do <i>software</i> VantagePoint.....	55
Figura 17 - Relação de metadados de periódicos indexados na WOS	57
Figura 18 - Fluxo de comparação de registros entre a Plataforma Lattes e a WOS.....	58
Figura 19 - Número e distribuição geográfica de periódicos indexados na Scopus e WOS	59
Figura 20 - Relação de metadados de periódicos indexados na Scopus.....	59
Figura 21 - Fluxo de comparação de registros entre a Plataforma Lattes e a Scopus	60
Figura 22 - Relação de metadados de periódicos indexados na SciELO	61
Figura 23 - Fluxo de comparação de registros entre a Plataforma Lattes e a SciELO	62
Figura 24 - Relações de ISSN obtidas a partir das tabelas de ISSN-L.....	64
Figura 25 - Processo de construção do tesauro de ISSN-L	65
Figura 26 - Relação de metadados de periódicos indexados no DOAJ.....	67
Figura 27 - Relação de metadados de periódicos indexados no SHERPA/ROMEO	67
Figura 28 – Produção científica da UFSCar recuperada na Plataforma Lattes (1971-2017) ...	69
Figura 29 – Taxa acumulada de criação de cursos de pós-graduação na UFSCar	70
Figura 30 – Produção científica da UFSCar informada na Plataforma Lattes e indexada na	

WOS, Scopus e SciELO	71
Figura 31 – Produção científica da UFSCar informada na Plataforma Lattes e indexada na WOS, Scopus e SciELO (1971-2017)	72
Figura 32 – Produção científica da UFSCar indexada na WOS e Scopus	73
Figura 33 – Produção científica da UFSCar indexada na WOS, Scopus e SciELO.....	74
Figura 34 – Produção científica da UFSCar em fontes de acesso aberto	76
Figura 35 – Taxa de publicação da UFSCar em acesso aberto por departamentos.....	77
Figura 36 – Taxa de publicação da UFSCar em acesso aberto por fontes de informação	78
Figura 37 – Fator de embargo da produção científica da UFSCar (SHERPA/ROMEIO).....	79
Figura 38 – Produção científica da UFSCar informada somente na Plataforma Lattes e indexada.....	81
Figura 39 – Produção científica da UFSCar informada somente na Plataforma Lattes	81
Figura 40 – Produção científica da UFSCar informada somente na Plataforma Lattes por centros acadêmicos	83
Figura 41 – Produção científica da UFSCar informada somente na Plataforma Lattes por áreas do conhecimento (CAPES).....	84

LISTA DE TABELAS

Tabela 1 - Número de ISSN e registros da WOS, Scopus e SciELO após concatenação com ISSN-L.....	66
---	----

LISTA DE QUADROS

Quadro 1 - Sistema de classificação de fator de embargo do SHERPA/ROMEO	24
Quadro 2 - Requisitos para implementação de repositórios institucionais.....	25
Quadro 3 - Aplicações envolvendo a Plataforma Lattes como fonte de informação	44

LISTA DE ABREVIATURAS E SIGLAS

A&HCI - Arts & Humanities Citation

BDICTs - Bases de Dados de Informação Científica e Tecnológicas

BDTD - Biblioteca Digital Brasileira de Teses e Dissertações

BIREME/OPAS/OMS - Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde

BOAI - Budapeste Open Access Initiative

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CBISSN - Centro Brasileiro do ISSN

CCA - Centro de Ciências Agrárias

CCBS - Centro de Ciências Biológicas e da Saúde

CCET - Centro de Ciências Exatas e de Tecnologia

CCGT - Centro de Ciências em Gestão e Tecnologia

CCHB - Centro de Ciências Humanas e Biológicas

CCN - Centro de Ciências da Natureza

CCTS - Centro de Ciências e Tecnologias para a Sustentabilidade

CECH - Centro de Educação e Ciências Humanas

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

CSIC - Conselho Superior de Investigação Científica

CT&I - Ciência, Tecnologia e Inovação

CTIT - Coordenadoria de Transferência e Inovação Tecnológica

DOAJ - Directory of Open Access Journals

DOI - Digital Object Identifier

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo

FI - Fator de Impacto

HTML - HyperText Markup Language

IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia

IES - Instituições de Ensino Superior

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

ISSN - International Standard Serial Number

ISSN-L - International Standard Serial Number Linking

JCR - Journal Citation Reports

JSONB - JavaScript Object Notation

MAA - Movimento de Acesso Aberto

MEC - Ministério da Educação

NIH - National Institute of Health

NIT/Materiais - Núcleo de Informação Tecnológica em Materiais

OAI - Open Archives Initiative

OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting

OAIS - Open Archival Information System

OpenDOAR - Directory of Open Access Repositories

PAAP - Programa de Apoio à Aquisição de Periódicos Eletrônicos

RI - Repositórios Institucionais

RIS - Research Information Systems

ROAR - Registry of Open Access Repositories

RUF - Ranking Universitário Folha

SciELO - Scientific Electronic Library Online

SCI-Expanded - Science Citation Index Expanded

SSCI - Social Sciences Citation Index

STI - Superintendência de Tecnologia da Informação

UFMG - Universidade Federal de Minas Gerais

UFSCar - Universidade Federal de São Carlos

UNESP - Universidade Estadual Paulista "Júlio de Mesquita Filho"

UNICAMP - Universidade Estadual de Campinas

USP - Universidade de São Paulo

WOS - Web of Science

XML - eXtensible Markup Language

SUMÁRIO

1 INTRODUÇÃO.....	14
2 REVISÃO DE LITERATURA	18
2.1 Repositórios institucionais.....	18
2.2 Movimento de Acesso Aberto	20
2.3 Fator de embargo	23
2.4 Implementação de repositórios institucionais	25
2.5 Sistemáticas automatizadas de povoamento.....	30
2.6 Bases de dados de informação científica e tecnológica.....	31
2.7 A Plataforma Lattes	33
2.7.1 Recuperação da informação a partir da Plataforma Lattes	34
2.7.2 Ferramentas para extração de dados da Plataforma Lattes.....	39
2.7.3 A Plataforma Lattes como fonte de informação.....	44
3 MÉTODO E DESENVOLVIMENTO	47
3.1 Delineamento da pesquisa	47
3.2 Unidade-Caso: Universidade Federal de São Carlos.....	49
3.3 Etapas da pesquisa	50
3.3.1 Coleta de dados.....	50
3.3.2 Organização de dados.....	54
3.3.3 Análise dos dados	55
3.3.4 Criação de tesauro <i>Web of Science</i>	56
3.3.5 Criação de tesauro Scopus.....	58
3.3.6 Criação de tesauro SciELO.....	60
3.3.7 Criação de tesauro ISSN-L.....	62

3.3.8 Criação de tesouro de acesso aberto	66
3.3.9 Criação de tesouro de fator de embargo	67
3.3.10 Criação de tesouro de vínculo institucional	68
4 RESULTADOS	69
4.1 Produção científica da UFSCar informada na Plataforma Lattes	69
4.2 Produção científica da UFSCar indexada na WOS, Scopus e SciELO	70
4.3 Produção científica da UFSCar em acesso aberto e fator de embargo	76
4.4 Produção científica da UFSCar recuperada somente na Plataforma Lattes	80
5 CONSIDERAÇÕES FINAIS	85
REFERÊNCIAS	87

1 INTRODUÇÃO

Os repositórios institucionais (RI) podem ser compreendidos como um tipo de biblioteca digital que comporta a produção científica de uma instituição, em seus mais diversos formatos, garantido a reunião e organização dos registros produzidos. Os RI trazem entre seus benefícios a promoção do acesso aberto à informação, a preservação da memória e o aumento da visibilidade da produção e dos pesquisadores da instituição. Para o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), órgão que atua diretamente na promoção do acesso aberto no Brasil, os RI podem ser consideradas como bases de dados voltadas para a reunião da produção científica institucional (INSTITUTO..., 2012a).

Apesar de sua origem ser datada anteriormente a criação do Movimento de Acesso Aberto (MAA), os RI estão diretamente ligados a promoção da Via Verde, estratégia que promove a criação de repositórios digitais em acesso aberto visando o depósito de publicações e a consequente eliminação de barreiras para o acesso à informação (LEITE, 2009; AUTRAN; BORGES, 2014). O MMA busca criar e promover alternativas diante do modelo tradicional de comunicação científica, calcado no comércio de assinaturas de periódicos e bases de dados, que representam em muitos casos barreiras para o acesso à informação (GARCIA; SUNYE, 2003; MORENO; LEITE; ARELLANO, 2006; KURAMOTO, 2006; SAYÃO; MARCONDES, 2009; AUTRAN; BORGES, 2014).

A presente pesquisa foi desenvolvida diante do contexto em que instituições de ensino superior (IES) passam cada vez mais a receber demandas para a implementação de seus RI, visando a promoção do acesso à informação, o aumento da visibilidade de sua produção científica, a preservação de sua memória institucional e o atendimento das obrigações instituídas por órgãos governamentais e agências de fomentos (LEITE, 2009; KURAMOTO 2009; LEITE et al., 2012; FURNIVAL; HUBBARD, 2013). Entre os desafios para a implementação e gestão de RI, que compreende fatores políticos, de recursos humanos, tecnológicos e de gestão, destaca-se a dificuldade das IES para realizarem o povoamento de suas coleções, ato que compreende ações como: 1) identificar, recuperar, inserir e avaliar metadados; 2) verificar fatores de embargo; e 3) disponibilizar em acesso aberto o texto completo de publicações institucionais, conforme suas permissões (LEITE et al., 2012). De modo geral, os RI adotam como estratégia para povoamento a abordagem centralizada, que consiste na identificação, recuperação, avaliação e inserção de registros e textos completos das publicações institucionais por uma equipe própria; ou o autodepósito, que compreende a realização de depósito das publicações pelos próprios autores (CARVALHO; CARVALHO,

2012). Ambas as estratégias, porém, tem se mostrado ineficazes para garantir o povoamento das coleções dos RI nacionais. A estratégia centralizada apresenta limitações quanto a capacidade de formação de equipes capazes de recuperar e disponibilizar toda a produção científica institucional de maneira manual. Quanto ao autodepósito, são necessários grandes esforços para a conscientização da comunidade acadêmica para que depositem suas publicações, sendo em muitos casos necessário a adoção de políticas que instituem a sua obrigatoriedade (ASSIS MEDEIROS, FERREIRA, 2014).

Diante das limitações das estratégias de povoamento (centralizado ou autodepósito), algumas IES têm buscado soluções automatizadas para o povoamento de seus RI, desenvolvendo sistemáticas voltadas para a extração de metadados da produção científica institucional a partir de bases de dados de informação científica e tecnológicas (BDICTs). Como exemplo, pode-se citar os casos da Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), Universidade de São Paulo (USP) e Universidade Estadual de Campinas (UNICAMP), que utilizaram a extração automatizada de metadados oriundos de BDICTs para o povoamento de seus RI (ASSUMPCÃO et al., 2014; VIDOTTI et al., 2015; 2016; UNIVERSIDADE..., 2017a, 2017b). Como elementos em comum dessas iniciativas está o uso das BDICTs como fonte para a identificação e recuperação da produção científica institucional. Nos casos citados, as universidades utilizaram como fonte para a coleta de metadados as bases de dados *Web of Science* (WOS), Scopus e *Scientific Electronic Library Online* (SciELO).

O povoamento automatizado de repositórios institucionais tem se mostrado mais eficientes do que a estratégia de povoamento realizado pelas equipes de RI ou por meio do autodepósito. Porém, as delimitações quanto ao conteúdo das BDICTs limitam a recuperação de metadados, o que reflete diretamente na representatividade do povoamento quanto a totalidade da produção científica da IES. Uma vez que essas delimitações restringem a recuperação da totalidade da produção científica e as práticas de povoamento automatizado de RI, a pesquisa foi desenvolvida a partir da seguinte questão: quais as alternativas para a obtenção de metadados voltados ao povoamento automatizado de repositórios institucionais?

A partir dessa questão a pesquisa trabalhou com a hipótese de que, devido as suas características, a Plataforma Lattes poderia configurar-se como uma fonte estratégica para a obtenção de metadados voltados para o povoamento automatizado de RI, uma vez que não apresenta restrições quanto a inserção de registros em seu conteúdo. A Plataforma Lattes é uma iniciativa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) que reúne em um único sistema informações sobre a carreira e produção intelectual de

pesquisadores que atuam no Brasil. Entre suas finalidades, apresenta-se como uma fonte de referência para a avaliação e monitoramento das atividades de pesquisa do país, além de servir como uma importante fonte para a obtenção de metadados da produção científica nacional (BRASIL, 2017a).

Justifica-se a realização da pesquisa uma vez que, a partir da revisão de literatura, foram encontradas iniciativas ainda incipientes quanto a avaliação ou uso da Plataforma Lattes como fonte para a coleta de metadados. Deste modo, não foi possível obter informações precisas quanto a capacidade de recuperação da produção científica de IES a partir da Plataforma. Nota-se também a carência de estudos comparativos entre a Plataforma Lattes e outros mecanismos de recuperação da produção científica de IES, como as BDICTs. Um dos desafios para a realização de estudos comparativos desse caráter está relacionado as dificuldades para extração de dados a partir da Plataforma Lattes. Desafio esse minimizado com o desenvolvimento de ferramentas como ScriptLattes, que possibilita a obtenção de dados da plataforma para a geração de relatórios (MENA-CHALCO; CESAR JUNIOR, 2009) e o SyncLattes, sistemática automatizada de extração de metadados voltados para o povoamento de RI (MATIAS, 2015; MATIAS et al., 2016; MATIAS; AMARAL; MATIAS, 2017).

A pesquisa teve como **objetivo geral** avaliar a capacidade de recuperação de metadados a partir da Plataforma Lattes, comparando-a com as bases de dados *Web of Science*, Scopus e SciELO para um conjunto de publicações científicas da unidade-caso: a Universidade Federal de São Carlos (UFSCar). O intuito da comparação foi a de verificar a capacidade de recuperação de registros da produção científica da universidade a partir da Plataforma Lattes em relação a quantidade de registros disponíveis a partir das bases de dados analisadas. Este objetivo geral pode ser desdobrado nos seguintes **objetivos específicos**:

- a) Identificar o estado da arte sobre as práticas de povoamento em RI nacionais; as iniciativas de povoamento automatizado de RI; e os métodos e fontes de informação utilizados para a recuperação de metadados da produção científica de IES;
- b) Desenvolver um procedimento experimental para comparar a capacidade de recuperação da produção científica institucional a partir da Plataforma Lattes em relação as BDICTs; e

- c) Caracterizar a produção científica institucional recuperada a partir da Plataforma Lattes quanto ao tipo de acesso (aberto ou não), ao embargo e a cobertura por áreas de conhecimento, visando identificar a potencialidade dos metadados recuperados para o povoamento automatizado de RI.

Essa dissertação está dividida em cinco seções: 1) Introdução, que apresenta o contexto em que a pesquisa foi elaborada; 2) Revisão de Literatura, na qual são apresentadas a contextualização dos RI: seus conceitos, ligação com o Movimento de Acesso Aberto, práticas de implementação e povoamento; os desafios quanto a recuperação de informação científica; e o uso da Plataforma Lattes como fonte para obtenção de metadados da produção científica de IES; 3) Método e Desenvolvimento, que descreve o processo metodológico empregado para o desenvolvimento do estudo de caso na Universidade Federal de São Carlos (UFSCar) e a análise bibliométrica realizada com o intuito avaliar a capacidade da Plataforma Lattes quanto a recuperação da produção científica da unidade-caso em comparação as BDICTs analisadas; 4) Resultados, apresentando os indicadores obtidos visando a comparação da Plataforma Lattes com as bases de dados *Web of Science*, Scopus e SciELO; e 5) Considerações Finais, com os apontamentos sobre a análise realizada a partir da comparação entre as fontes de coleta de metadados avaliadas visando o povoamento de RI.

2 REVISÃO DE LITERATURA

2.1 Repositórios institucionais

Costa e Leite (2015, p. 17) definem RI como “[...] sistemas de informação compostos por coleções digitais, que são desenvolvidas a partir de serviços de gestão relacionados com a coleta, organização, disseminação e preservação da produção acadêmica dos membros de uma instituição”. Para Leite et al. (2012) um RI pode ser compreendido como um tipo de biblioteca digital que possui em seu conteúdo somente publicações vinculadas a instituição. Para o autor um repositório institucional é um “[...] serviço de informação científica - em ambiente digital e interoperável - dedicado ao gerenciamento da produção científica e/ou acadêmica de uma instituição [...] contempla a reunião, armazenamento, organização, preservação, recuperação e, sobretudo, a ampla disseminação da informação científica produzida na instituição” (LEITE et al., 2012, p. 7).

Implementar um repositório institucional não significa somente criar uma biblioteca digital que apresente como conteúdo a produção científica da instituição. Um RI possui atributos e características que vão além do conceito de uma biblioteca digital, assim como apontado por Leite (2009, p. 21) “[...] todo repositório institucional de acesso aberto pode ser considerado um tipo de biblioteca digital, mas nem toda biblioteca digital pode ser considerada um repositório institucional”. Para que o ambiente digital seja considerado um RI este deve ser instituído oficialmente, ser aberto, interoperável e disponibilizar em seu conteúdo texto completo da produção científica institucional em formato digital (CROW, 2002 apud LEITE, 2009). Costa e Leite (2015) identificam como características básicas de um RI: armazenar exclusivamente a produção científica dos membros da instituição, aceitando como tipo de material em suas coleções documentos como artigos científicos (*preprints* e *posprints*), publicações em anais de eventos, teses e dissertações, livros, dados brutos de pesquisa, material audiovisual e recursos educacionais. Para os autores o RI deve disponibilizar o texto completo das publicações nele depositadas, garantindo sua preservação, e adotar o protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) com a finalidade de ser interoperável.

Entre os atributos que caracterizam os RI, destaca-se o critério do acesso aberto, que consiste em disponibilizar seu conteúdo para “[...] leitura, *download*, cópia, distribuição, impressão, busca ou criação de links para os textos completos” (LEITE, 2009, p. 15). O acesso aberto apresenta-se como uma alternativa diante das barreiras de acesso à informação,

como as instituídas pelo modelo tradicional de comunicação científica, na qual as publicações são disponibilizadas em periódicos científicos e bases de dados que na sua maioria são propriedades de editoras privadas e apresentam elevados custos para aquisição.

Historicamente a criação de RI está ligada ao contexto da denominada “crise dos periódicos”. Iniciada em meados de 1986 e presente no contexto das IES até os dias atuais, a crise foi causada pelo aumento substancial dos preços para aquisição de periódicos científicos, consequência da fusão de editoras científicas, que passaram a atuar em um mercado menos competitivo. O aumento de valores para a aquisição trouxe dificuldades para que IES garantissem a cobertura de publicações de seu interesse, forçando a adoção de um maior rigor de seleção e cortes de assinaturas, resultando em limitações quanto o acesso à informação científica (AUTRAN; BORGES, 2014). Diante das dificuldades de acesso à informação algumas iniciativas visando eliminar tais barreiras foram iniciadas. Entre as ações voltadas para a ampliação do acesso à informação destaca-se a criação de ambientes digitais em acesso aberto em meados dos anos 90, viabilizados a partir da evolução das tecnologias de informação que deram o respaldo tecnológico para seu desenvolvimento e implementação (GOMES; ROSA, 2010). A partir da possibilidade de criação desses ambientes “[...] era possível publicar na internet, a um custo mínimo, com um alcance mundial e com uma rapidez entre a submissão do artigo e sua publicação consideravelmente maior [deste modo] a lógica das cobranças de assinaturas, começa a ser confrontada por vários setores da comunidade acadêmica” (SAYÃO; MARCONDES, 2009, p. 14).

O ArXiv pode ser considerado como o primeiro exemplo de arquivo digital de acesso aberto com conteúdo científico, sendo incorporado posteriormente como modelo para a definição de propostas da *Open Archives Initiative* (OAI). O ArXiv foi criado em 1991 no EUA pelo físico Paul Ginsparg com a finalidade de permitir aos autores depositarem e trocarem informações sobre versões prévias de artigos a serem publicados. Entre as características que o diferencia destaca-se a prática de autodepósito adotada desde os seus primórdios, que estabelece aos próprios autores a responsabilidade de informar os metadados e depositar o texto completo da publicação no ambiente (SUNYE et al., 2009; AUTRAN; BORGES, 2014). O ArXiv continua em funcionamento até os dias atuais, sendo mantido pela biblioteca da Universidade de Cornell, tendo cobertura nas áreas de física, matemática, ciência da computação, biologia, finanças e estatística (SUNYE et al., 2009; AUTRAN; BORGES, 2014).

2.2 Movimento de Acesso Aberto

Os encontros e debates entre grupos de interesse deram início a formulação de acordos e manifestos voltados para a promoção de arquivos digitais em acesso aberto, o que resultaria na constituição do denominado Movimento de Acesso Aberto (MAA). Um dos principais marcos do movimento foi a Convenção de Santa Fé, realizada em 1999, tendo como resultado a criação da OAI. O movimento foi criado com o objetivo de promover alterações nos padrões de comunicação científica, utilizando para isso a aplicação de tecnologias, o estabelecimento de diretrizes para gestão e a construção de políticas necessárias para a estruturação de uma rede de publicações científicas em acesso aberto (KURAMOTO, 2006; AUTRAN; BORGES, 2014). Entre os resultados da criação da OAI destacam-se a definição do padrão de descrição de metadados *Dublin Core* e o OAI-PMH, protocolo que permite a comunicação entre provedores de dados e serviços, viabilizando a interoperabilidade dos arquivos digitais. A adoção do OAI-PMH permite a realização do processo denominado *harvesting*, que consiste na coleta, indexação e exibição de metadados do arquivos digital por outros sistemas de informação, potencializando deste modo a recuperação das publicações nele depositado (KURAMO, 2006). Uma vez que os metadados depositados em arquivos digitais de acesso aberto são coletados por iniciativas nacionais e internacionais, o processo de *harvesting* representa para os autores uma maior visibilidade de seus resultados de pesquisa e o consequente aumento de impacto de suas publicações (GARCIA; SUNYE, 2003; MORENO; LEITE; ARELLANO, 2006; KURAMOTO, 2006; SAYÃO; MARCONDES, 2009; AUTRAN; BORGES, 2014). Batista (2009, p. 72) aponta a importância da adoção desse padrão de comunicação no contexto dos arquivos digitais de acesso aberto:

[...] se as plataformas de implementação e os dados presentes nos repositórios forem interoperáveis, a possibilidade de pesquisa simultânea entre repositórios é facilitada, permitindo maximizar o potencial dos recursos documentais arquivados individualmente em cada repositório, na medida em que se torna possível a pesquisa em simultâneo.

Outro marco importante para o MAA foi o *Budapest Open Access Initiative* (BOAI), evento realizado em 2002, que estabeleceu como recomendações a adoção de duas estratégias visando a criação de canais alternativos para publicação e disponibilização de conteúdo em acesso aberto (LEITE, 2009):

- a) **A Via Dourada (*Golden Road*):** que busca promover a criação de periódicos científicos de acesso aberto, utilizando o mesmo modelo adotado pelos periódicos científicos tradicionais, porém sem instaurar barreiras de acesso ao seu conteúdo; e
- b) **A Via Verde (*Green Road*):** que compreende a criação de RI, visando promover ambientes para o armazenamento e disponibilização de produções científicas em acesso aberto.

Cabe ressaltar que os RI não buscam substituir os canais tradicionais de comunicação científica, mas sim representam um mecanismo para disponibilizar conteúdo em acesso aberto. Deste modo, não há informações publicadas em RI, mas sim publicadas em outras fontes como em periódicos científicos, sendo posteriormente depositadas em ambientes digitais de acesso aberto como os RI (LEITE, 2009; AUTRAN; BORGES, 2014).

No Brasil, foi lançado em 2005 pelo IBICT o "Manifesto Brasileiro de Apoio ao Acesso Livre ao Conhecimento Científico", que reúne recomendações a serem adotadas pela comunidade científica nacional em prol ao movimento de acesso à informação. O Manifesto ressalta e defende o acesso livre à informação científica, destacando a sua importância para o contínuo desenvolvimento científico (KURAMOTO, 2006). Nesse sentido Sayão e Marcondes (2009) apontam a importância do acesso à informação científica, provenientes de resultados de pesquisas, para a retroalimentação do processo de fazer ciência. Os autores defendem a lógica de que “[...] os resultados da atividade científica, na forma das diferentes publicações, resultados estes muitas vezes obtidos à custa de pesados investimentos públicos, devem necessariamente também ser públicos [podendo] ser utilizados amplamente [e não] apropriados de forma privada” (SAYÃO; MARCONDES, 2009, p. 10). Quando lançado, o "Manifesto Brasileiro de Apoio ao Acesso Livre ao Conhecimento Científico" defendia a criação de uma política nacional de acesso livre à informação científica, recomendando que as instituições e agências de fomento exigissem o depósito em repositórios de acesso livre dos resultados de pesquisas que receberam financiamento público (KURAMOTO, 2009). Apesar das recomendações apresentadas em 2005, o Brasil ainda não possui uma política nacional específica voltada para a disponibilização de informação científica.

Nesse sentido há um projeto de Lei datado de 2011 sendo tramitado, que visa instaurar a obrigatoriedade do registro e disseminação da produção científica de IES e unidades de pesquisa nacionais. O projeto de Lei nº 287/2011:

Obriga as instituições de educação superior de caráter público, bem com as unidades de pesquisa a construir repositórios institucionais de acesso livre, nos quais deverão ser depositados o inteiro teor da produção técnico-científica conclusiva dos estudantes aprovados em cursos de mestrado, doutorado, pós-doutorado ou similar, assim como da produção técnico-científica, resultado de pesquisa científicas realizadas por professores, pesquisadores e colaboradores, apoiados com recursos públicos para acesso livre na rede mundial de computadores; entende-se por produção técnico-científica monografias, teses, dissertações e artigos publicados em revistas, nacionais e internacionais, com revisão por pares (BRASIL, 2017b).

Um exemplo de política nacional voltada para a garantia de acesso aberto ao conteúdo científico é a Portaria nº 013 da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), de 15 de fevereiro de 2006. O documento institui a obrigatoriedade aos programas de pós-graduação reconhecidos pela CAPES de depositar as teses e dissertações em ambientes digitais de acesso aberto, sendo este um dos critérios considerados no processo de avaliação dos programas (BRASIL, 2006). Cabe ressaltar que esta portaria reflete diretamente no grande volume de teses e dissertações disponíveis nos repositórios institucionais nacionais.

Apesar de não ser uma política voltada especificamente para informação científica, no ano de 2011 foi instituída no Brasil a Lei nº 12.527, de 18 de novembro de 2011, conhecida como a Lei de Acesso à Informação. Franco (2015) aponta uma possível ligação desta lei com os RI, uma vez que prevê em seu conteúdo a disponibilização pública de “[...] relatórios técnicos, de pesquisa, publicações governamentais, traduções avulsas, *preprints*, dissertações, teses e literatura originada de encontros científicos, como os anais de congressos”. Nessa perspectiva a Lei configura-se como uma ferramenta em potencial que pode contribuir para o povoamento de RI.

Diante da ausência de uma política nacional que institui a disponibilização da produção científica em arquivos digitais em acesso aberto, Assis Medeiros e Ferreira (2014) apontam a possibilidade da IES criar uma política própria, decretando a nível institucional a obrigatoriedade de depósito da produção científica no repositório da instituição.

Em um âmbito mais amplo destaca-se também as iniciativas de órgãos públicos e agências de fomento de instituir mandatos que determinam a obrigatoriedade de depósito dos resultados de pesquisa por eles financiados em arquivos digitais de acesso aberto (FURNIVAL; HUBBARD, 2013). A base de dados SHERPA/JULIET indexa informações sobre mandatos de agências de fomento a nível mundial, estando atualmente cadastradas 161 políticas. Dessas, 68% exigem a publicação ou depósito dos resultados de pesquisas

financiadas em fontes de acesso aberto; 12% incentivam a publicação e depósito nesse tipo de fonte; e 20% não apresentam orientações sobre a disponibilização dos resultados das pesquisas financiadas. Entre os países que possuem maior concentração de políticas no SHERPA/JULIET estão o Reino Unido, Canadá e Estados Unidos, não sendo identificadas políticas de agências de fomento nacionais cadastradas (SHERPA/JULIET, 2017). Os mandados de depósito compulsório têm se apresentado como um fator determinante na decisão e investimento sobre a implementação de RI, assim como, configuram-se como uma importante ferramenta para garantir o povoamento de suas coleções. Kuramoto (2009) cita como exemplo o mandato adotado pelo *National Institute of Health* (NIH), o qual determina a obrigatoriedade de depósito dos resultados de pesquisa financiados pelo instituto no PubMed Central, ambiente digital de texto completo em acesso aberto da biblioteca nacional de medicina dos Estados Unidos. Cabe ressaltar que ao serem instituídos, os mandatos de depósito compulsório consideraram e estão em concordância com as políticas de direitos autorais e fatores de embargo impostas pelas fontes de publicação, que irão determinar as permissões que o autor possui quanto à disponibilização do conteúdo publicado em outras fontes, como no caso do depósito em RI.

2.3 Fator de embargo

Além dos desafios de recuperação da produção científica institucional, somente a coleta de registros não é suficiente para o povoamento dos RI. Após a identificação, busca e recuperação do registro, torna-se necessário verificar quais as permissões concedidas pelas fontes de publicação sobre a disponibilização do texto completo em arquivos digitais de acesso aberto. Essas permissões são concedidas a partir do fator de embargo da publicação, que determina quais as permissões concedidas quanto ao tempo necessário de espera e a versão da publicação que pode ser disponibilizada em arquivos digitais de acesso aberto. De modo geral, são liberadas para depósito o *preprint* (versão da publicação submetida ao periódico), o *posprint* (versão do texto após as considerações da revisão dos pares) ou a versão final (texto final publicado ou o próprio PDF diagramado pela fonte de publicação) (FURNIVAL; GUIRRA, 2017).

A necessidade de consultar as regras de embargo das publicações quanto a disponibilização de conteúdo em outras fontes, como os RI, ocorre devido a concessão de direitos patrimoniais concedido pelo autor à editora científica responsável pela fonte de publicação. O direito patrimonial trata-se da parte dos direitos de autor que permite

“a possibilidade de ceder ou licenciar a obra, definitiva ou temporariamente, podendo explorá-la economicamente como desejar” (JUNGMANN, BONETTI, 2010, p. 27). Leite (2009, p. 25) ressalta que no processo de publicação científica os autores concedem à editora o “[...] direito de utilizar, fruir e dispor da obra (direito autoral patrimonial)” e com isso torna-se necessária a autorização por parte das editoras para a realização de depósitos em RI, além de outras atividades que envolvam o artigo publicado. Furnival e Guirra (2017) indicam a necessidade de consulta da política editorial do periódico antes da realização do depósito em RI. Os autores ressaltam que, apesar das limitações estabelecidas pelas editoras, muitos periódicos concedem autorização para o depósito de alguma das versões da publicação após um período de tempo em que foi publicada.

Para a consulta do fator de embargo cabe ao autor ou a equipe do RI verificar a política editorial da fonte em que o conteúdo foi publicado. Como alternativa o SHERPA/ROMEO reúne política editoriais de periódicos científicos a nível mundial, oferecendo como resumo uma classificação dos periódicos segundo o nível de concessões que a fonte de publicação fornece. A partir de um sistema de cores disponibilizados é possível identificar qual a versão do artigo pode ser depositada (Quadro 1).

Quadro 1 - Sistema de classificação de fator de embargo do SHERPA/ROMEO

Classificação	Permissões
Verde	Pode arquivar a versão <i>preprint</i> e <i>postprint</i> ou versão/PDF do editor
Azul	Pode arquivar a versão <i>postprint</i> (i.e., o rascunho final após o <i>peer-review</i>) ou versão/PDF do editor
Amarelo	Pode arquivar a versão <i>preprint</i> (i.e., antes do <i>peer-review</i>)
Branco	O arquivo não é suportado formalmente

Fonte: SHERPA/ROMEO (2017).

Furnival e Guirra (2017) destacam que mesmo para os periódicos indicados como de acesso aberto, como no caso do título indexados no *Directory of Open Access Journals* (DOAJ), é necessária a consulta sobre o fator de embargo, pois estes podem permitir o livre acesso ao conteúdo publicado, porém apresentar restrições quanto ao depósito do conteúdo em outras fontes de informação como os RI.

2.4 Implementação de repositórios institucionais

A implementação de RI se dá por meio de um conjunto de iniciativas envolvendo aspectos políticos, de gestão, tecnológicos e de recursos humanos, sendo o desenvolvimento e alinhamento desses diversos aspectos um fator determinante para o seu sucesso. Diversos autores apresentam recomendações para a implementação de RI, que podem ser distribuídas em três fases principais: o planejamento, a implementação e a manutenção. O Quadro 2 apresenta uma compilação de fases e recomendações elencadas a partir da revisão de literatura sobre requisitos e ações necessárias na implementação de um RI.

Quadro 2 - Requisitos para implementação de repositórios institucionais

1 DEFINIR MISSÃO E OBJETIVOS
Devem ser estabelecidos com base na demanda da instituição. São associados com a busca pela maior visibilidade da produção científica, o controle da produção institucional e com a inserção da instituição no contexto da gestão de conteúdos digitais. Ao definir os objetivos, torna-se necessária a tomada de decisão entre duas abordagens que irão delimitar o conteúdo do repositório: a rígida, que permite somente o depósito de publicações avaliadas pelos pares, ou a flexível, que, além do conteúdo aprovado pelos pares, incorpora outros materiais não avaliados formalmente pela comunidade científica (LEITE, 2009).
2 DEFINIR EQUIPE TÉCNICA
As equipes de repositórios são caracterizadas por serem enxutas, interdisciplinares e com alta capacitação, sendo geralmente constituídas por bibliotecários, analistas de sistemas e profissionais de comunicação. Leite (2009) aponta como principais componentes da equipe: os gestores, responsáveis pela definição de políticas e gestão de recursos, e os administradores, responsáveis por gerenciar os aspectos tecnológicos do repositório. Shintaku e Meirelles (2010) recomendam a instauração de um grupo gestor para as tomadas de decisões, composto pela equipe técnica, usuários e os <i>stakeholders</i> - indivíduos que possuem influência sobre o repositório, como reitores, gestores e docentes.
3 ESCOLHA DO SOFTWARE
Diante da grande oferta de <i>softwares</i> disponíveis para a implementação de repositórios, torna-se necessário o estabelecimento de critérios para sua seleção. Sayão e Marcondes (2009) recomendam a formação de um comitê avaliador, com a representação de todos os atores envolvidos, certificando-se, deste modo, que a escolha irá considerar diferentes perspectivas e demandas. Para os autores, são critérios pertinentes à seleção de <i>software</i> : a infraestrutura, capacidade de organização do conteúdo, qualidade da interface com o usuário e os recursos de interoperabilidade.
3.1 Infraestrutura
Sayão e Marcondes (2009) apontam como principais características do <i>software</i> a ser analisado: a escalabilidade, extensibilidade, facilidade de implantação, plataforma computacional necessária, satisfação de usuários, suporte do sistema, base de conhecimento das comunidades envolvidas, estabilidade da organização de desenvolvimento, perspectivas de inovações, limites do sistema, documentação disponível/cursos/publicações. Além dos aspectos referentes ao <i>hardware</i> e <i>software</i> , é necessário o mapeamento das competências da equipe, o que possibilitará o planejamento sobre a contratação de serviços externos e capacitação interna.

3.2 Organização do conteúdo
Deve-se considerar a capacidade e flexibilidade do sistema quanto à arquitetura de informação a ser definida: os tipos de objetos comportados, os fluxos de trabalhos permitidos e a possibilidade de organizar o conteúdo segundo comunidades e coleções.
3.3 Interface - Usuários
Item relacionado com a qualidade da interface do sistema e a possibilidade de customização; os idiomas disponíveis; a capacidade de recuperar informação, identificando quais os tipos de buscas permitidas; a possibilidade de navegar por índices; se existem recursos que permitem a disseminação seletiva da informação; quais os filtros apresentados para aplicar nos resultados de uma pesquisa e o formato em que os resultados são apresentados.
3.4 Interoperabilidade
O <i>software</i> deve estar em conformidade com os protocolos de interoperabilidade, como OAI-PMH e Z39.50, o que possibilitará a troca de informações com outros sistemas de informação.
4 PLANEJAR CUSTOS
Os custos para implementação de um repositório estão associados com a aquisição de <i>hardware</i> e <i>software</i> ; contratação de recursos humanos, principalmente equipe de TI para a instalação e customização, e a necessidade de treinamento e capacitação. Para a manutenção, são apresentados custos para as ações de <i>marketing</i> , manutenção do sistema, segurança da informação, preservação digital, e o desenvolvimento de funcionalidades e serviços (LEITE, 2009; TOMAÉL; SILVA, 2007).
5 ANÁLISE CONTEXTUAL
Para definir a política de gerenciamento, é necessário mapear todos os atores envolvidos no repositório - tomadores de decisões, equipe técnica e usuários em potencial. A partir do mapeamento, é possível identificar qual a demanda por parte dos atores e como essas proporcionam ameaças e oportunidades ao projeto (LEITE, 2009).
6 POLÍTICA DE FUNCIONAMENTO
No momento do planejamento, é necessário definir qual o tipo de conteúdo aceito; quem pode depositá-lo; como são os fluxos de trabalho; as responsabilidades de cada ator envolvido; os serviços oferecidos e se há custos para uso. A política de funcionamento é concebida na forma de um documento, contendo todas as regras do repositório, o que deve refletir a política de informação da instituição. Sua publicação irá formalizar o repositório diante da comunidade acadêmica (ROSA, 2011).
7 POLÍTICA DE CONTEÚDO
A política de conteúdo determina quais são os tipos de documentos aceitos para depósito, refletindo a decisão sobre a abordagem a ser adotada (rígida ou flexível). Devido às questões de direitos autorais, é imprescindível o acompanhamento do setor jurídico da instituição para sua definição (VARGAS, 2009).
8 POLÍTICA DE DEPÓSITO
A política de depósito define questões relacionadas à organização do conteúdo: o estabelecimento de comunidades e coleções, os esquemas de metadados adotados para cada material, os fluxos de submissão de conteúdo e a resolução dos conflitos de direitos autorais.

8.1 Comunidades e coleções
As universidades utilizam as relações hierárquicas entre faculdades, institutos, departamentos e centros de pesquisas para definir as comunidades, subcomunidades e coleções de seus repositórios, sendo outros critérios possíveis para a organização as relações entre áreas do conhecimento ou as comunidades de interesse (LEITE et al., 2012).
8.2 Definir metadados
O Dublin Core é o padrão de metadados comumente adotado pelos repositórios, sendo a base para a interoperabilidade entre sistemas desse caráter (LEITE, 2009; CATARINO; SOUZA, 2012). Deve-se estabelecer uma estrutura de metadados para cada tipo de material aceito, identificando no próprio documento sua descrição, identificação, explicação e localização, o que facilitará sua recuperação em sistemas de informação. Recomenda-se, como parâmetro para estabelecer os esquemas de metadados, o uso das diretrizes <i>DRIVER guidelines 2.0</i> e <i>OpenAIRE guidelines 2.0</i> (LEITE et al., 2012).
8.3 Fluxo de submissão
Existem três etapas básicas nos modelos de submissão: 1) catalogação, etapa obrigatória na qual são realizadas pelo autor a submissão dos metadados, do documento e a concessão de direitos para depósito; 2) avaliação, que consiste na verificação da compatibilidade do conteúdo depositado com a política do repositório; e 3) revisão, na qual serão avaliados os metadados dos documentos aprovados para depósito. Esta etapa é crucial para garantir uma boa recuperação dos documentos (SHINTAKU; MEIRELLES, 2010).
8.4 Direitos autorais
Na etapa de submissão, é necessário verificar quais as permissões para a realização do depósito, pois, apesar de ser comum a prática de concessão de direitos patrimoniais, no ato da publicação, dos autores aos editores, em muitos casos é concedida permissão para o autoarquivamento de uma cópia do trabalho em servidores institucionais ou pessoais (UNIVERSIDADE..., 2013). Recomenda-se incluir na etapa de submissão uma confirmação por parte do autor autorizando a realização do depósito e se comprometendo com sua legalidade, para verificar as permissões de depósito concedidas pelas fontes de publicações científicas. Recomenda-se o uso de diretórios de políticas editoriais como o Diadorim ou SHERPA/ROMEO.
9 POLÍTICA DE PRESERVAÇÃO
Nova, Galindo e Ribeiro (2011) apontam como papel do repositório preservar a memória da instituição, garantindo a preservação, em longo prazo, do conteúdo nele armazenado. Para os autores, cabe à política de preservação determinar quais as diretrizes em relação aos aparatos tecnológicos envolvendo <i>hardware</i> e <i>software</i> sob a perspectiva da memória, visando garantir o acesso em longo prazo, a confiabilidade dos dados e a autenticidade de conteúdo. O <i>Open Archival Information System</i> (OAIS) é um modelo conceitual que descreve as ações necessárias para a preservação, em longo prazo, em um repositório institucional. A adoção dos padrões apresentados pelo modelo garante a disponibilidade, durabilidade e confiabilidade dos dados, além da manutenção, o compartilhamento e a distribuição do material preservado (ARELLANO, 2008; CASTRO et al., 2009; SAYÃO; MARCONDES, 2009).

10 ESTRATÉGIAS DE <i>MARKETING</i>
<p>O planejamento de <i>marketing</i> deve contemplar ações em todo o ciclo de desenvolvimento do projeto: na concepção do repositório, ao buscar incentivos da instituição e atores para o seu desenvolvimento; na etapa de implementação, ao definir as estratégias para seu projeto piloto; em sua fase de manutenção, quando deverá divulgar o repositório buscando novos adeptos e incentivando os usuários que o utilizam; além de se comprometer, num aspecto mais amplo, em promover a filosofia de acesso aberto em toda a instituição (LEITE, 2009; RODRIGUES, 2010).</p>
10.1 Definir produtos e serviços
<p>Na fase de planejamento são definidos quais os serviços oferecidos pelo repositório, os quais devem buscar agregar valor ao projeto, incentivando a adesão de novos usuários e se caracterizando como uma recompensa para os que o utilizam. Entre os serviços oferecidos por repositórios estão os treinamentos e suporte aos usuários; serviço de identificadores persistentes; importação de dados por lote; digitalização de documentos; orientação sobre direitos autorais, e depósito mediado (LEITE, 2009). Diante da possibilidade de inúmeros serviços, a escolha por ofertá-los deve ser realizada com cautela, sendo crucial avaliar as condições e capacidades da equipe para que a oferta destes não se torne pontos fracos e/ou ameaças ao sucesso do repositório.</p>
10.2 Indicadores de desempenho
<p>O repositório institucional é uma importante fonte de indicadores sobre a produção científica, e possibilita o monitoramento da relevância científica, social e econômica das atividades de pesquisa da instituição (CROW, 2002 apud MELIS, 2013). Leite et al. (2012) destacam o uso do <i>Ranking Web of Repositories</i> para a geração de indicadores que servirão como subsídios na avaliação do repositório. Visando o aumento de visibilidade e conseqüente melhoria das posições nos <i>rankings</i> institucionais, Melis (2013) aponta como possível estratégia o cadastro do repositório em diretórios internacionais como o <i>Registry of Open Access Repositories (ROAR)</i> e <i>Directory of Open Access Repositories (OpenDOAR)</i>. Com a finalidade de avaliação, é por meio dos indicadores de desempenho que a equipe irá obter o <i>feedback</i> das estratégias e decisões tomadas, o que possibilitará a realização de diagnóstico indicando quais as melhorias necessárias para o sucesso do projeto.</p>

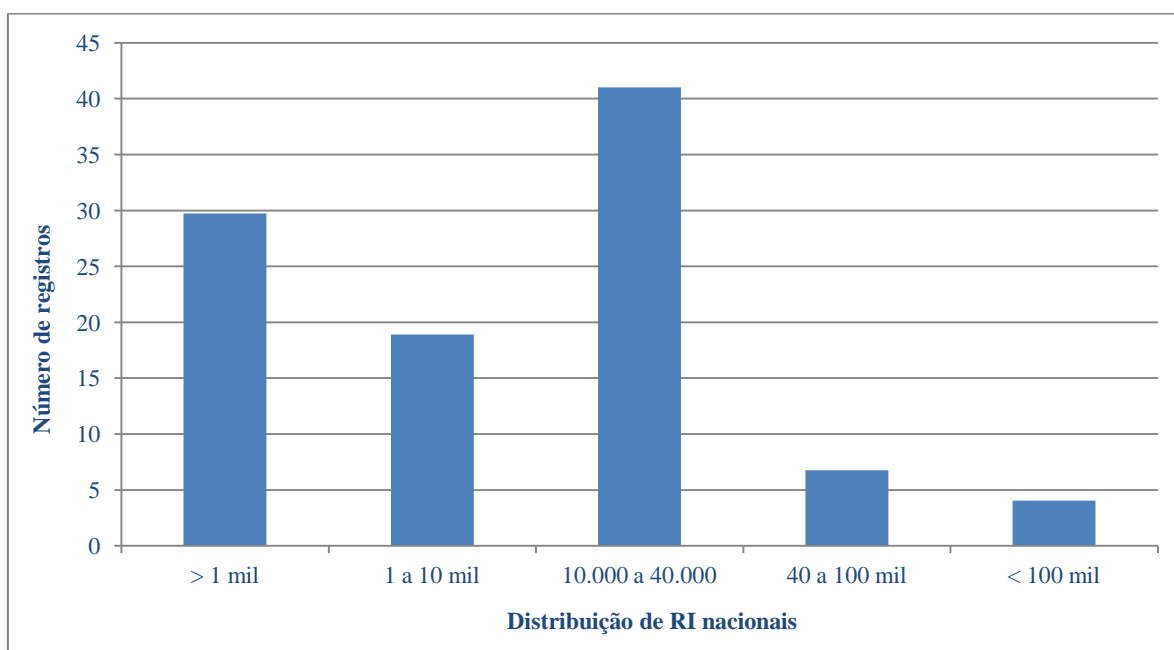
Fonte: Sarvo e Amaral (2014).

Leite et al. (2012) ressaltam que a implementação de RI está muito além das questões tecnológicas, apontando não ser suficiente apenas a instalação de *software* e disponibilização de infraestrutura que o comporte. Para os autores a maior preocupação na constituição de um repositório deve estar associada aos fatores humanos e a criação de mecanismos voltados para a garantia de seu funcionamento e povoamento. Entre os pontos críticos citados, o povoamento consiste no conjunto de ações realizadas para a recuperação, tratamento e disponibilização dos metadados e texto completo nas coleções do RI. De modo geral, o povoamento é realizado de maneira centralizada, no qual uma equipe fica responsável pela identificação, coleta e inserção de registros no RI ou por meio do autodepósito, que consiste no arquivamento das publicações por parte dos autores (CARVALHO; CARVALHO, 2012). O método de povoamento a ser utilizado pelo RI deverá ser estabelecido pela política de

funcionamento do repositório. Weitzel (2006) ressalta a importância da IES prever um crescimento descentralizado com base no autodepósito, cabendo à equipe promover mecanismos que facilitem e incentivem a participação da comunidade acadêmica no processo de povoamento.

Segundo o *Directory of Open Access Repositories* (OpenDOAR), diretório oficial de repositórios acadêmicos de acesso aberto, existem atualmente 92 repositórios brasileiros cadastrados, sendo 74 classificados como institucional (DIRECTORY..., 2017a). Nota-se que apesar das estratégias de povoamento, o volume de itens nas coleções dos RI nacionais ainda é baixo, dos 74 RI cadastros: 22 possuem menos de mil registros; 14 apresentam até 10 mil registros; 30 até 40 mil registros; 05 até 100 mil registros; e 03 apresentam mais de 100 mil registros em suas coleções (Figura 1).

Figura 1 - Número de registros depositados em RI nacionais



Fonte: OpenDOAR (DIRECTORY..., 2017a).

Em relação ao tipo de material depositados nos RI nacionais, nota-se uma maior pluralidade quanto a cobertura de tipos de publicações nos repositórios mais povoados. No caso dos RI com menor volume de itens é notada a predominância de teses e dissertações em suas coleções, o que pode estar associada a facilidade de obtenção desses itens devido à dois fatores:

1) A criação da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), em 2002, construída com a finalidade de coletar metadados e reunir em um portal informações de teses e dissertações produzidas por programas de pós-graduação brasileiros; e

2) A determinação pela CAPES da obrigatoriedade de disponibilização em acesso aberto dos resultados de pesquisas dos programas e pós-graduação por ela reconhecidos, estabelecida pela Portaria nº 013 de 15 de fevereiro de 2006.

Quanto aos outros tipos de materiais, nota-se uma limitação quanto ao povoamento de artigos de periódicos, dos 74 RI nacionais cadastrados no OpenDOAR 43 (58%) não apresentam artigos em suas coleções.

Entre as dificuldades encontradas pelas instituições e equipes para garantir o povoamento de RI estão o convencimento dos autores sobre a importância da disponibilização de suas publicações nesse tipo de ambiente e a falta de esclarecimento sobre o funcionamento do modelo de acesso aberto. Furnival e Hubbard (2013) apontam entre os motivos para a não aderência por parte dos autores às práticas de acesso aberto:

1) A associação pelos pesquisadores de que as publicações em canais de acesso aberto possuem baixa qualidade, considerando que a revisão pelos pares ocorre somente nos periódicos que adotam o modelo tradicional de comunicação científica;

2) A busca por publicar em fontes de informação que forneçam métricas científicas, como o caso do Fator de Impacto atribuído para publicações indexadas na *Web of Science*;

3) O receio de infringirem questões de direitos concedidos aos canais de comunicação em que publicaram; e

4) A falta de tempo para a realização do autodepósito.

2.5 Sistemáticas automatizadas de povoamento

Diante do baixo número de itens apresentados nos RI nacionais, algumas IES têm buscado o desenvolvimento de sistemáticas automatizadas visando garantir um povoamento mais eficiente de seus RI. Como exemplo, pode-se citar a iniciativa adotada pela Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) que desenvolveu uma sistemática automatizada para o povoamento de seu RI a partir da conversão de registros disponíveis nas

bases de dados *Web of Science*, *Scopus* e *SciELO*. A produção científica da UNESP foi recuperada a partir das três bases de dados e incorporadas nas coleções de seu RI. Posteriormente, a carga de metadados foi complementada utilizando a Plataforma Lattes e outras fontes de metadados como o catálogo de suas bibliotecas e a BDTD (VIDOTTI et al., 2015; 2016; ASSUMPÇÃO et al., 2014). Além da experiência de povoamento do RI a partir de BDICTs, Assumpção et al. (2014) e Vidotti et al. (2015; 2016) apontam os resultados significativos obtidos pela UNESP quando utilizada a Plataforma Lattes como fonte para coleta de metadados visando complementar o povoamento do seu RI. A IES obteve resultado significativo quanto a recuperação de registros de sua produção científica quando utilizada a Plataforma Lattes, tendo destaque a obtenção de metadados para as áreas de humanidades.

O uso de sistemáticas automatizadas de coleta de metadados voltadas para o povoamento de RI também foi relatado pela Universidade de São Paulo (USP) e Universidade Estadual de Campinas (UNICAMP), que utilizaram metadados oriundos das bases de dados WOS e SciELO para o início do povoamento de seus repositórios (UNIVERSIDADE..., 2017a, 2017b). A partir dos relatos de experiência fica evidente a vantagem das sistemáticas automatizadas quando comparadas com as estratégias de povoamento centralizado, realizado de maneira manual, ou o autodepósito. O uso dessas sistemáticas pode configurar-se como uma solução diante ao baixo número de itens apresentados nas coleções dos RI nacionais.

Nota-se como ponto crítico de tais sistemáticas a necessidade de identificação de fontes de informação adequadas e as dificuldades encontradas para a recuperação da informação. Parte dessa dificuldade ocorre devido aos esforços necessários para mensurar qual a totalidade da produção científica da IES, uma vez que geralmente não existem controles da produção científica institucional e seus registros não se encontram reunidos em um único sistema de informação. Parte devido ao conteúdo das bases de dados utilizadas pelas sistemáticas automatizadas apresentarem restrições quanto ao tipo de publicações, período de cobertura e áreas do conhecimento disponíveis em seu conteúdo. Julga-se deste modo que, para o desenvolvimento de sistemáticas automatizadas visando o povoamento de RI, torna-se crucial a compreensão da estrutura e conteúdo das fontes de informações selecionadas para a identificação, recuperação e coleta de metadados da produção científica institucional.

2.6 Bases de dados de informação científica e tecnológica

Entende-se como bases de dados a coleção de publicações armazenadas em um sistema de informação, geralmente disponíveis *online*, que possuem como finalidade

armazenar, organizar e disponibilizar aos usuários o conteúdo por ela armazenado, oferecendo sistemáticas estruturadas para a busca e recuperação da informação (COLEPICOLO, 2015, p. 15). Cunha e Cavalcanti definem as bases de dados bibliográficas como sendo:

Bases de dados que contém registros automatizados, relativos a documentos e itens bibliográficos. Pode, ou não, conter resumos e é formada por “uma série de registros bibliográficos ligados entre si, onde cada um em geral apresenta uma combinação dos seguintes componentes: número do documento; título; autor; referência da fonte; resumos; textos integral; termos ou expressões de indexação; citações ou quantidade de referências; instituição de origem do documento, ou endereço do autor, ou ambos, língua do documento-fonte; informação de uso interno, como números de classificação ou localização (CUNHA; CAVALCANTI, 2008, p. 44).

De modo geral, as características de uma base de dados podem variar em relação ao: *tipo*, apenas referencial ou disponibilizando o texto completo; *conteúdo*, armazenando diferentes fontes de informação como artigos de periódicos, teses e dissertações, livros, patentes, entre outros; *cobertura*, com enfoque em áreas do conhecimento específicas ou multidisciplinar; *acesso*, podendo ser de acesso privado ou livre; e diferenças quanto ao *período de cobertura* referente a data de publicação dos itens que disponibilizam. Colepicolo (2015, p. 15) aponta que as bases de dados de informação científico-tecnológica (BDICTs)

[...] são portais internacionais mantidos por grandes editoras científicas que indexam uma série de publicações de cunho científico e tecnológico em todas as áreas do conhecimento [...] a maioria destas bases utiliza como critérios para aceite de publicações a revisão pelos pares, periodicidade regular, heterogeneidade de corpo editorial como forma de garantir a qualidade e confiabilidade destas. Estas bases geralmente apresentam indicadores bibliométricos, tais como fator de impacto, índice de citações, entre outros. Ainda que haja controvérsias quanto à validade destes indicadores, atualmente são os que mais sustentam a qualidade e confiabilidade das publicações.

Esse tipo de sistema de informação tem como vantagem permitir a realização de pesquisas complexas, viabilizada pela capacidade de combinação entre campos e termos. Em contrapartida, a precisão dos resultados de pesquisas nas bases de dados é comprometida devido a ausência de mecanismos para o controle e padronização na descrição da informação. Como exemplo, podemos citar a falta de controle de autoridade, a não disponibilização de tesouros ou vocabulários controlados e o desconhecimento por parte dos usuários dos recursos de pesquisa disponíveis, como fatores que implicam na precisão de recuperação de informação em bases de dados (ARAÚJO, 2012). Matias (2015) aponta as dificuldades

encontradas pelas IES na identificação e recuperação de suas publicações, destacando que as práticas de povoamento automatizado têm recorrido a BDCITs como *Web of Science*, Scopus e SciELO para identificar, recuperar e coletar metadados de sua produção científica.

Além da disponibilização de informação, as BDICTs configuram-se como importantes fontes para a extração de indicadores. Gingras (2016) aponta que a partir da informatização da pesquisa bibliográfica, com a criação de sistemas que constituíram as primeiras BDICTs, torna-se possível potencializar a aplicação da bibliometria e a geração de indicadores voltados para o acompanhamento e avaliação das atividades científicas e tecnológicas. Esses indicadores passam a configurar-se ao longo do tempo como base para o desenvolvimento da cientometria, área voltada aos estudos sobre a própria ciência envolvendo, por exemplo, a avaliação de instituições, de pesquisadores e a formulação de políticas públicas em Ciência e Tecnologia.

As BDICTs apresentam, de modo geral, elevados custos de assinatura para acesso ao seu conteúdo. No Brasil a CAPES disponibiliza um Portal de Periódicos visando minimizar as barreiras de acesso à informação ocasionadas devido a esses elevados custos. O Portal está disponível para IES credenciadas e reúne diversas bases de dados e periódicos científicos assinadas pela CAPES. O Portal de Periódicos foi criado em 2000, vinculado ao Programa de Apoio à Aquisição de Periódicos Eletrônicos (PAAP), com o intuito de subsidiar o acesso à informação científica e tecnológica dos programas de pós-graduação brasileiros. Sua criação foi motivada visando agilizar o processo de aquisição de periódicos, minimizar custos, sendo sua constituição viabilizada em grande parte pelo advento da internet no país. Por meio do Portal de Periódicos a CAPES busca promover o acesso à informação, eliminando restrições quanto a limitações físicas de acesso à acervos e buscando complementar as limitações do desenvolvimento de coleções, consequência das restrições quanto aos recursos financeiros das IES. No ano de 2017 o Portal de Periódicos CAPES disponibilizou 533 bases de dados distribuídas entre diversas áreas do conhecimento (ALMEIDA; GUIMARÃES; ALVES, 2011; BRASIL, 2015).

2.7 A Plataforma Lattes

A Plataforma Lattes é uma iniciativa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) que reúne informações em um único sistema sobre a rede científica nacional, a nível de pesquisadores por meio do Currículo Lattes; de grupos de pesquisa, por meio do Diretório de Grupos de Pesquisa; e a nível de instituições por meio de

seu Diretório de Instituições. Disponível desde 1999, a Plataforma Lattes trata-se da evolução de iniciativas do CNPQ datadas desde o início da década de 80 que visavam estabelecer um currículo comum a toda comunidade acadêmica do país. A Plataforma possui como objetivo a padronização de informações sobre pesquisadores brasileiros sob um único formato a fim de fornecer dados utilizados como subsídios para as avaliações por parte do CNPq. No Currículo Lattes são reunidas informações sobre a carreira e produção intelectual de pesquisadores que atuam no país como sua formação, atuação profissional, registros de produções intelectuais, participações em eventos, organizações de eventos e participações em bancas examinadoras (BRASIL, 2017a).

Por ser “[...] uma base de dados pública, tanto no que se refere ao ingresso quanto à recuperação das informações através da internet” (BRASIL, 2017a), a Plataforma Lattes vem sendo utilizada para outras finalidades além da de diagnóstico para avaliação de pesquisadores e instituições, como por exemplo, a de servir como uma importante fonte para a obtenção de dados sobre a produção científica nacional voltados para a confecção de indicadores científicos.

2.7.1 Recuperação da informação a partir da Plataforma Lattes

A Plataforma Lattes disponibiliza em sua interface¹ acesso direto a busca no Currículo Lattes, Diretório de Instituições e Diretório dos Grupos de Pesquisa. Para a busca por Currículos Lattes são oferecidas como opções de pesquisa a busca simples ou avançada. A busca simples permite a realização de pesquisas por nome do pesquisador ou assunto, utilizando palavras-chave da produção científica. A interface apresenta como opções de filtro a titulação, nacionalidade (nacional ou estrangeiro) ou a delimitação do país de origem do pesquisador (Figura 2).

¹ <http://lattes.cnpq.br>

Figura 2 - Interface de busca do Currículo Lattes (Busca simples)

Buscar Currículo Lattes (Busca Simples)

Buscar por:
 Seleccione o modo de busca Nome Assunto(Título ou palavra chave da produção)

Nas bases

 Doutores Demais pesquisadores (Mestres, Graduados, Estudantes, Técnicos, etc.)

Nacionalidade:

 Brasileira Estrangeira

Pais de nacionalidade:

Todos

Tipo de filtro

Filtros

Preferências

Fonte: Brasil (2017c).

A busca avançada apresenta como opção a construção de expressões de busca por assuntos, a partir de campos pré-estabelecidos: todas essas palavras, esta frase exata, qualquer uma dessas palavras, nenhuma dessas palavras ou expressão booleana (Figura 3).

Figura 3 - Interface de busca do Currículo Lattes (Busca avançada)

Busca Avançada (por Assunto)

Construa uma consulta com:

todas essas palavras :

esta frase exata:

qualquer uma dessas palavras:

e nenhuma dessas palavras:

esta expressão booleana:

Nas bases

 Doutores Demais pesquisadores (Mestres, Graduados, Estudantes, Técnicos, etc.)

Nacionalidade:

 Brasileira Estrangeira

Fonte: Brasil (2017c).

O Diretório de Instituições permite a recuperação de informações sobre as instituições cadastradas no CNPq. Apresenta como opções de pesquisa o nome da instituição, sigla ou CNPJ, recuperando as informações sobre o nome da instituição, sigla, CNPJ, razão social, endereço, dados para contato, natureza, informações sobre sua fundação, missão, responsável e hierarquia institucional (Figura 4).

Figura 4 - Interface de busca do Diretório de Instituições

CADI
Cadastro de Informações Institucionais (CADI)

DI Diretório de Instituições
Sistema-Cliente

Você está em: Plataforma Lattes > Diretório de Instituições > Cadastro de Informações Institucionais

Cadastro de Informações Institucionais

Preencha o campo abaixo, escolha o tipo de busca e clique em Pesquisar.

ufscar Pelo nome e/ou sigla CNPJ

Número de instituições encontradas: 62 **Legenda:** Instituições | Subunidades

Sigla Inst.	Nome da Instituição ou Subunidade - Sigla	UF/País	Situação*	Titular
FAI-UFSCAR	Centro de Caracterização e Desenvolvimento de Materiais - CCDM/FAI-UFSCAR	SP/Brasil	Ativa	Nelson Guedes de Alcântara
FAI-UFSCAR	Fundação de Apoio Institucional ao Desenvolvimento Científico e Tecnológico - FAI-UFSCAR	SP/Brasil	Ativa	Nelson Guedes de Alcântara
FAI-UFSCAR	Incubadora Regional de Cooperativas Populares - FAI-UFSCAR	SP/Brasil	Ativa	Nelson Guedes de Alcântara

Fonte: Brasil (2017c).

O portal do Diretório de Grupos de Pesquisa disponibiliza em sua interface as opções de consulta de nomes de grupos, linhas de pesquisa, pesquisadores, estudantes, técnicos e colaboradores estrangeiros, que podem ser aplicados nos campos: nome do grupo, nome da linha de pesquisa, palavra-chave da linha de pesquisa, objetivo da linha de pesquisa, nome do líder e nome do colaborador estrangeiro (Figura 5). Como resultados são recuperadas informações sobre os pesquisadores que constituem os grupos de pesquisa: nome, titulação e grupos de pesquisa a qual pertence (Figura 6); e a nível de grupo de pesquisa: seus líderes, áreas de atuação, instituições vinculadas, linhas de pesquisa e ano de formação (Figura 7).

Figura 5 - Interface de busca do Diretório de Grupos de Pesquisa

Consultar - Base corrente

Termo de Busca Todas as palavras ▼

* Consultar por

Aplicar a busca nos campos

- Nome do grupo
- Nome da linha de pesquisa
- Palavra-chave da linha de pesquisa
- Repercussões do grupo
- Nome do líder
- Nome do pesquisador
- Nome do estudante
- Nome do técnico
- Nome do colaborador estrangeiro
- Nome da Instituição Parceira

Situação Certificado Não-atualizado

Fonte: Brasil (2017c).

Figura 6 - Resultado de busca do Diretório de Grupos de Pesquisa a nível de pesquisador

Pesquisador: Roniberto Morato do Amaral

Formação acadêmica: Doutorado

Grupo de pesquisa: Lab4u - UFSCAR

Grupo de pesquisa: Núcleo de Informação em Ciência, Tecnologia, Inovação e Sociedade - UFSCAR

Grupo de pesquisa: Núcleo de Informação Tecnológica em Materiais - NIT / Materiais - UFSCAR

Grupo de pesquisa: PERITO - Pesquisa em Inteligência Tecnológica e Organizacional - UFSCAR

Total de registros: 1

Fonte: Brasil (2017c).

Figura 7 - Resultado de busca do Diretório de Grupos de Pesquisa a nível de grupo

Grupo de pesquisa

Lab4u

Endereço para acessar este espelho: dgp.cnpq.br/dgp/espelhogrupo/0031120984394992

Identificação

Situação do grupo:	Em preenchimento
Ano de formação:	2009
Data da Situação:	12/12/2017 15:30
Data do último envio:	10/03/2016 14:57
Líder(es) do grupo:	Luc Quoniam  Leandro Innocentini Lopes de Faria 
Área predominante:	Ciências Sociais Aplicadas; Ciência da Informação
Instituição do grupo:	Universidade Federal de São Carlos - UFSCAR
Unidade:	Departamento de Ciências da Informação

Fonte: Brasil (2017c).

Apesar do acesso público ao conteúdo e as ferramentas de busca disponíveis em sua interface, a Plataforma Lattes apresenta algumas limitações quanto a capacidade de recuperação de informação. Brito, Quoniam e Mena-Chalco (2016) apontam limitações identificadas no uso do sistema de busca da Plataforma Lattes, identificadas a partir de um estudo de caso sobre a identificação e recuperação do tema "nanotecnologia" em registros de publicações informados na Plataforma. Segundo os autores a interface não possibilita o desenvolvimento e aplicação de estratégias de maneira satisfatória para a recuperação da informação registradas nos Currículos Lattes, apontando como pontos críticos:

- 1) O não reconhecimento de operadores booleanos e de truncagem, essenciais para elaboração de estratégias de busca, mesmo os indicados pela própria Plataforma como compatíveis com o sistema;

- 2) As limitações quanto ao número máximo de registros que podem ser apresentados nas páginas de resultados; e
- 3) A impossibilidade de escolha de campos específicos para a execução da estratégia de busca, gerando o aumento da revocação e a baixa precisão nos resultados obtidos.

2.7.2 Ferramentas para extração de dados da Plataforma Lattes

Diante das limitações de recuperação da informação por meio da interface de busca, algumas alternativas voltadas para a extração de dados da Plataforma Lattes são apresentadas. Oficialmente o CNPq oferece duas formas para que as IES possam obter os dados sobre os seus pesquisadores que estão cadastrados na Plataforma Lattes (BRASIL, 2017c):

- 1) O **espelhamento**, permite a duplicação da parcela da Plataforma Lattes que contem dados sobre os pesquisadores da IES, obtida mediante autorização do CNPq; e
- 2) A **extração de cv e grupos da instituição**, permite o acesso aos os dados dos Currículos Lattes dos pesquisadores da IES no formato *eXtensible Markup Language* (XML) por meio do estabelecimento de um protocolo de cooperação técnica.

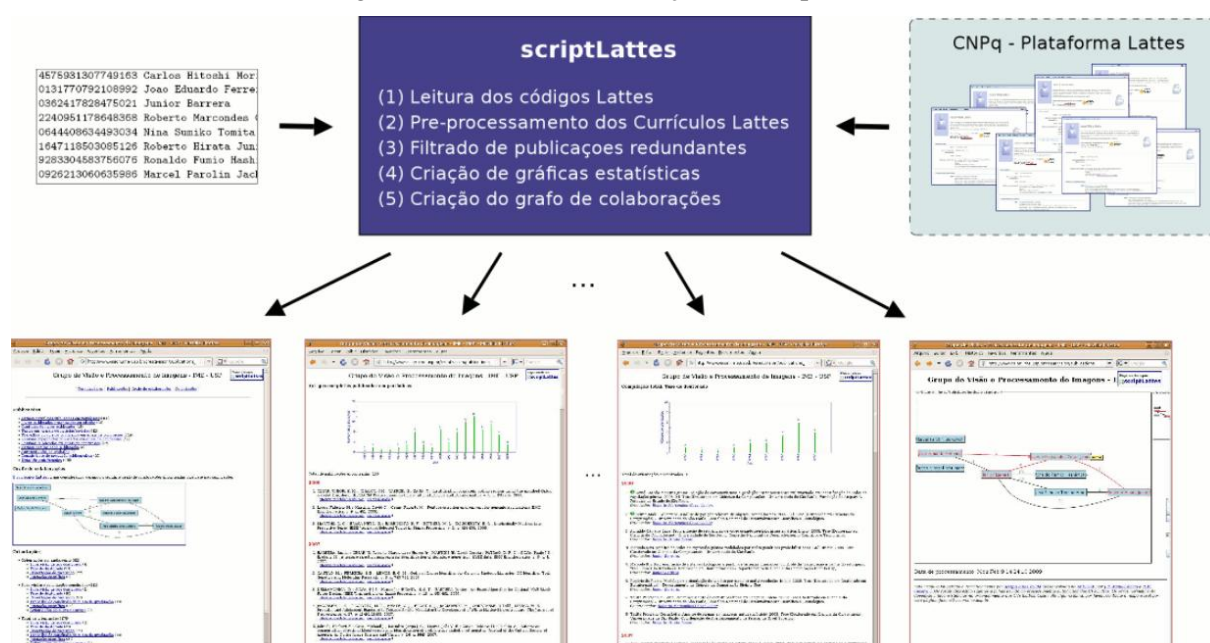
Além dos canais oficiais fornecidos pelo CNPq outras iniciativas voltadas para a extração de dados da Plataforma Lattes como o ScriptLattes, Somos, Plataforma Stela Experta, WeR-USP e SyncLattes são identificadas.

O ScripLattes pode ser considerado como uma iniciativa pioneira no país. A ferramenta foi desenvolvida por Jesús Pascual Mena-Chalco e Roberto Marcondes Cesar Junior com a finalidade de extrair e compilar automaticamente metadados de registros cadastrados na Plataforma Lattes. Segundo os autores o ScriptLattes trata-se de

[...] um sistema de código aberto para criar relatórios acadêmicos de grupos com base nos currículos da base de dados Lattes. O sistema scriptLattes é composto pelos seguintes módulos: (a) seleção de dados, (b) pré-processamento de dados, (c) tratamento de redundância, (d) geração de grafos de colaboração entre os membros do grupo, (e) geração de mapas de pesquisa com base em informações geográficas, e (f) criação automática de relatórios de bibliografia, técnica e produção artística e supervisão acadêmica (MENA-CHALCO; CESAR JUNIOR, 2009, p. 31, tradução nossa).

A ferramenta de extração, desenvolvida em Python e disponível² de maneira livre na internet, teve sua primeira versão disponibilizada em 2005 e se encontra em sua versão 8. O processo de execução do ScriptLattes consiste na instalação do *software* e seus requisitos; criação de lista de IDs Lattes, que irá determinar em quais Currículos Lattes serão extraídos os dados; preenchimento do arquivo de configuração, o qual irá determinar qual será o arquivo de entrada e saída, o período de tempo considerado para a realização da coleta, entre outras funcionalidades; e execução do programa, por meio de linha de comando. Em seu processo de coleta o ScriptLattes realiza a leitura dos IDs Lattes informados; coleta e processa os dados dos currículos; realiza tratamento de redundâncias dos dados coletados; e gera relatórios em formato *HyperText Markup Language* (HTML), com as listas de publicações e redes de colaboração identificadas (Figura 8). Além do formato HTML também é possível obter um arquivo com os registros recuperados em formato da *Research Information Systems* (RIS) (SCRIPTLATTES, 2015).

Figura 8 - Processo de execução do ScriptLattes



Fonte: ScriptLattes (2015).

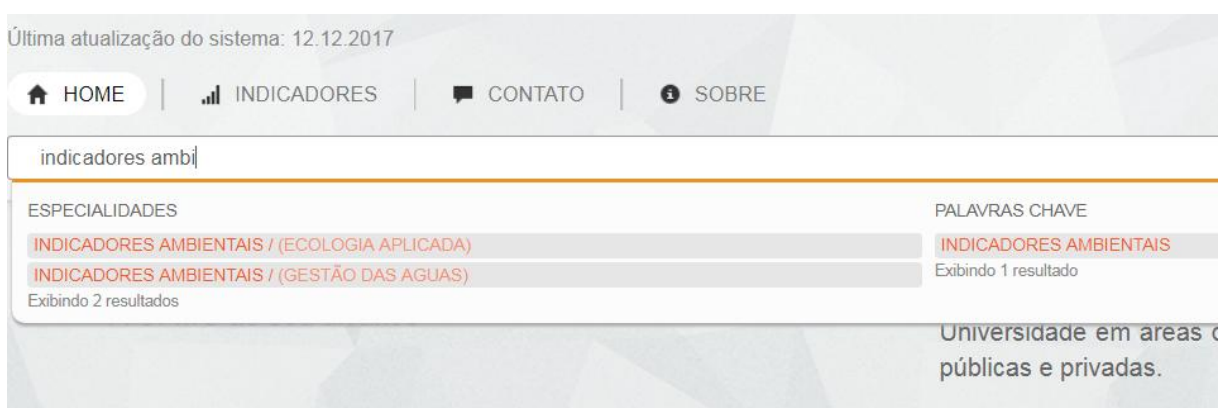
O Somos³ é uma ferramenta de visualização de dados desenvolvida pela Coordenadoria de Transferência e Inovação Tecnológica (CTIT) da Universidade Federal de

² <http://scriptlattes.sourceforge.net/>

³ <http://www.somos.ufmg.br>

Minas Gerais (UFMG). A ferramenta utiliza dados da Plataforma Lattes para o desenvolvimento de indicadores sobre a produção científica de IES, tendo como principal finalidade facilitar o mapeamento de competências e potencializar a formação de redes de colaboração. Por meio do Somos “[...] é possível identificar os pesquisadores, suas especialidades e produção científica, além de informações sobre unidades, departamentos, ativos de propriedade intelectual, infraestrutura instalada nos laboratórios, dentre outras informações” (UNIVERSIDADE..., 2017c). A ferramenta oferece em sua interface de busca (Figura 9) a opção de recuperação por nome de especialistas ou palavras-chave. Como resultados são apresentadas as listas de docentes e as redes de palavras-chave que se relacionam com o tema pesquisado (Figura 10). Além da interface de busca, o Somos disponibiliza uma seção “indicadores” que apresenta um conjunto de métricas sobre o corpo docente, unidades e produção acadêmica como a publicação de livros, artigos, patentes, entre outros itens.

Figura 9 - Interface de busca do Somos (UFMG)



Fonte: UFMG (UNIVERSIDADE..., 2017c).

Figura 10 - Resultados de busca do Somos (UFMG)



Fonte: UFMG (UNIVERSIDADE..., 2017c).

A Plataforma Stela Experta⁴ é uma ferramenta que utiliza dados da Plataforma Lattes para realizar a gestão da produção científica, auxiliando no mapeamento de competências e dando subsídios para a instituição diante aos critérios de avaliação aplicados por órgãos como o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e CAPES. Desenvolvido pelo Grupo Stela, a ferramenta gera relatórios e indicadores com a produção científica, orientações e participações em projetos de pesquisadores, os relacionando com métricas como Qualis e Fator de Impacto (STELATEK, 2017) (Figura 11).

Figura 11 - Interface de busca da plataforma Stela Experta

The screenshot shows the Stela Experta search interface. At the top, there is a navigation bar with 'Página inicial', 'Análise', and 'Administração'. Below this, there are tabs for 'Pessoa', 'Produção', and 'Projeto'. The main search area is titled 'Quais são os itens de produção da instituição?'. It shows a search result for 'TEIXEIRA, A. O.; GRACIANO, S. A.; SILVA, J. C. S.; PEREIRA, S. R. M.; Coelho M. J. APLICABILIDADE DA TIPOLOGIA DE CUIDADOS EM PACIENTES IDOSOS., 2012.' and lists other results like 'Miranda, Fernando,P.; Almeida, Fabrício, F.; Crepaldi, Ana, M.; CRUZ, A. P. M. C.; De Souza, Wânia, C.; Cheniaux E; Nardi, A. E.; Landeira-Fernandez, J. Avaliação da percepção de dor aguda em pacientes com transtorno de pânico., 2012.' and 'Fioravanti-Bastos, Ana Carolina Monnerat; Cheniaux E; Landeira-Fernandez, J. Development and validation of a short-form version of the Brazilian State-Trait validation of a short-form version of the Brazilian State-Trait., 2012.' The interface also includes a search bar, filters for 'Tipo' (Produção bibliográfica, Produção técnica, etc.), 'Área', 'Ano de publicação', and 'Com registro ou patente'. The top right corner has a 'Sair' button.

Fonte: StelaTek (2017).

O WeR-USP⁵ trata-se de um sistema desenvolvido pela Superintendência de Tecnologia da Informação (STI) da Universidade de São Paulo (USP) voltado para a gestão da produção científica. O WeR-USP “[...] está disponível a todos os pesquisadores da Universidade, funciona como um visualizador dos indicadores de produção bibliográfica (com base nas informações da Plataforma Lattes), citações (com base nas informações do Google

⁴ <http://site.stelaexperta.com.br>

⁵ <https://uspdigital.usp.br/datausp/publico/indicador/indicadores.jsp>

Scholar, Scopus e *Web of Science*) e número de teses e dissertações na USP” (FERREIRA, 2015). Disponível para docentes e gestores da USP, o WeR-USP apresenta indicadores com base na estrutura organizacional da IES (por unidades de ensino, pesquisa e departamentos) e a nível individual, apresentando indicadores da produção científica dos pesquisadores. Por meio de sua interface de busca é possível obter indicadores da instituição, número de docentes, produção bibliográfica por ano, tipo de material, número de citações recebidas nas bases de dados Google Scholar, Scopus e WOS, número de mestrados e doutorados por ano, distribuição de docentes por área de conhecimento e número de colaborações internacionais (Figura 12).

Figura 12 - Interface de busca do WeR-USP

Indicator in 2018	Total
Number of Teaching Staff	5739
Number of Teaching Staff with Curriculum Lattes	5704 (99.39%)
Number of Teaching Staff with Google Scholar ID	4256 (74.15%)
Number of Teaching Staff with ORCID	1003 (17.47%)
Number of Teaching Staff with Scopus ID	4129 (71.94%)
Number of Teaching Staff with Web of Science (ResearcherID)	3256 (56.73%)
Number of Units	52

Source: Marte System (USP), Lattes Platform, Google Scholar, Scopus and Web of Science.

Fonte: USP (UNIVERSIDADE..., 2017a).

O SyncLattes⁶ é uma ferramenta de extração de metadados a partir da Plataforma Lattes. Desenvolvido por Matias (2015), possui como funcionalidade a identificação de registros, extração, tratamento quanto a duplicidade, aplicação de controle de autoridade e sincronização com o *software* DSpace. Além da sincronização, o SyncLattes oferece possibilidades de obtenção de dados em outros formatos de exportação, como no formato RIS (MATIAS, 2015; MATIAS et al., 2016; MATIAS; AMARAL; MATIAS, 2017). O funcionamento da ferramenta, utilizada para a coleta de dados nessa pesquisa, está descrita no item 3.3.1 "Coleta de Dados" dessa dissertação.

Cabe ressaltar que apesar das várias iniciativas voltadas para a extração de dados a partir da Plataforma Lattes, somente o ScriptLattes e o SyncLattes estão disponíveis para acesso público e gratuito.

2.7.3 A Plataforma Lattes como fonte de informação

O Quadro 3 apresenta algumas aplicações identificadas na literatura que utilizaram a Plataforma Lattes como fonte de informação. O levantamento demonstra a pluralidade de tipos de aplicações a partir da Plataforma, o que ressalta seu caráter estratégico para os estudos e o acompanhamento do desenvolvimento da ciência nacional.

Quadro 3 - Aplicações envolvendo a Plataforma Lattes como fonte de informação

Ano	Autoria	Título	Fonte	Aplicação
2017	HADDAD, E. A. MENA-CHALCO, J. P. SIDONE, O.	Produção científica e redes de colaboração dos docentes vinculados aos programas de pós-graduação em Economia no Brasil	Estudos Econômicos (São Paulo)	Utiliza dados coletados da Plataforma Lattes, obtidos por meio do ScriptLattes, para avaliar o desempenho científico de pesquisadores brasileiros, confeccionando uma série de indicadores bibliométricos
2017	LETA, J. COSTA, E. H. S. MENA-CHALCO, J. P.	Artigos em Periódicos de Acesso Aberto: um Estudo com Pesquisadores Bolsistas de Produtividade do CNPq	Revista Eletrônica de Comunicação, Informação & Inovação em Saúde	A partir de registros obtidos na Plataforma Lattes, por meio do ScriptLattes, realiza uma análise de correlação entre o índice de produtividade do CNPQ com a prática de publicação em periódicos de acesso aberto

⁶ <https://github.com/nitmateriais/synclattes>

2016	AMARAL, R. M. BRITO, A. G. C. ROCHA, K. G. S. QUONIAM, L. M. FARIA, L. I. L.	Panorama da inteligência competitiva no Brasil: os pesquisadores e a produção científica na Plataforma Lattes	Perspectivas em Ciência da Informação	Com o uso do ScriptLattes utiliza a Plataforma Lattes como fonte de informação para coleta de dados, identificando e analisando o perfil de pesquisadores de Inteligência Competitiva no Brasil segundo sua formação, atuação profissional e produção científica
2016	MOTTA, L. J. FERRAZ, R. R. N. QUONIAM, L. MENA-CHALCO, J. P.	Funded projects on dengue fever in Brazil: use of the Scriptlattes Tool to support the planning of public health research	Revista Espacios	Utiliza dados da Plataforma Lattes, por meio do ScriptLattes, para mapear e analisar projetos financiados sobre dengue no Brasil, utiliza o ScriptLattes como ferramenta para extração de dados
2016	SIDONE, O. J. G. HADDAD, E. A. MENA-CHALCO, J. P.	A ciência nas regiões brasileiras: evolução da produção e das redes de colaboração científica	Transinformação	A partir de dados da Plataforma Lattes, obtidos por meio do ScriptLattes, realiza uma análise da evolução da produção e colaboração científica brasileira a partir do ponto de vista geográfico
2016	VIDOTTI, S. A. B. G. BASTOS, F. M. GRISOTO, A. P. ARAKAKI, F. A. FERREIRA, J. B.	Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT	Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB)	Apresenta o processo de conversão de registros obtidos a partir da Plataforma Lattes, por meio da coleta automática de dados e uso de folha de estilo XSLT para a importação de registros no <i>software</i> DSpace
2015	VIDOTTI, S. A. B. G. ASSUMPÇÃO, F. S. FERREIRA, J. B. GRISOTO, A. P. SILVA, R. E. RODRIGUES, V. S. MAY, O. L. BASTOS, F. M.	Coleta de dados a partir dos currículos da Plataforma Lattes: procedimentos utilizados no Repositório Institucional da UNESP	Ponto de Acesso	Relata a experiência da UNESP com o povoamento de seu RI a partir de dados obtidos na Plataforma Lattes, obtidos em formato XML, convertidos e importados para o <i>software</i> DSpace
2014	MENA-CHALCO, J. P. DIGIAMPIETRI, L. A. LOPES, F. M. CESAR JUNIOR, R. M.	Brazilian bibliometric coauthorship networks	Journal of the Association for Information Science and Technology	Obtém a partir da Plataforma Lattes dados de 1.131.912 Currículos Lattes de pesquisadores associados as oito áreas de conhecimento e realiza uma análise de rede identificando a dinâmica de coautoria nas áreas de: ciências agrícolas; ciências biológicas; ciências exatas e da terra; humanidades; ciências sociais aplicadas; ciências da saúde; engenharia; e linguística, letras e artes.

2013	MENA-CHALCO, J. DIGIAMPIETRI, L. OLIVEIRA, L.	Perfil de produção acadêmica dos programas brasileiros de pós-graduação em Ciência da Computação nos triênios 2004-2006 e 2007-2009	Revista em Questão	Por meio do ScriptLattes, utiliza a Plataforma Lattes para obter a relação de publicações e orientações de docentes associados aos programas de pós-graduação brasileiro em Ciência da Computação, traçando um perfil da produção acadêmica da área.
------	---	---	--------------------	--

Fonte: Elaborado pelo autor.

3 MÉTODO E DESENVOLVIMENTO

Para o desenvolvimento da pesquisa foi desenvolvido um estudo de caso com o objetivo de avaliar a capacidade da Plataforma Lattes quanto a recuperação de metadados da produção científica de IES. Para isso a Universidade Federal de São Carlos foi selecionada como unidade-caso, sendo recuperado metadados de sua produção científica a partir da Plataforma Lattes e comparados com os disponíveis nas bases de dados *Web of Science*, Scopus e SciELO.

Justifica-se a escolha da unidade-caso devido aos seus esforços institucionais visando a implementação do seu RI. Além de seu perfil, a UFSCar foi selecionada diante do intuito da pesquisa de dar continuidade ao processo de avaliação da Plataforma Lattes, realizado a partir dos trabalhos de Matias (2015) e Bassoli (2017). Os autores utilizaram dados da produção científica da UFSCar, informados na Plataforma Lattes, para o desenvolvimento de uma sistemática voltada para a criação de bases de dados referencial visando o povoamento de RI e para avaliar a potencialidade da Plataforma como fonte de coleta de dados voltados para a confecção de indicadores. A presente pesquisa foi desenvolvida no âmbito do Núcleo de Informação Tecnológica em Materiais (NIT-Materiais), grupo de pesquisa da UFSCar que desenvolve entre suas atividades estudos métricos e possui a Plataforma Lattes como enfoque de estudos, avaliando-a e utilizando-a como fonte de informação (NÚCLEO..., 2011).

3.1 Delineamento da pesquisa

A pesquisa caracteriza-se como um estudo de caso exploratório, com abordagem quantitativa, utilizando a bibliometria como técnica de análise. Os estudos de caso focam na delimitação e análise de um ou poucos objetos, possibilitando o aprofundamento de conhecimentos sobre suas características. Gil (2002, 2008) aponta o uso desse tipo de pesquisa para os casos em que o objeto de estudo não possui delimitações claras com o contexto em que está inserido. Para o autor o estudo de caso “[...] é encarado como o delineamento mais adequado para a investigação de um fenômeno contemporâneo dentro do seu contexto real, onde os limites entre o fenômeno e o contexto não são claramente percebidos” (GIL, 2008, p. 54).

As pesquisas de caráter exploratório buscam obter maior familiaridade com o objeto de estudo, o tornando mais explícito e possibilitando a formulação de hipóteses e ideias. Os estudos exploratórios possuem características flexíveis quanto ao seu planejamento, o que

permite a consideração e exploração dos mais variados aspectos sobre o objeto estudado (GIL, 2002; APPOLINÁRIO, 2006). Appolinário (2006) aponta a classificação de pesquisas como quantitativa para aquelas que possuem como objeto de estudo fatos mensuráveis, passíveis de serem investigados cientificamente por meio da realização da coleta de dados e análise estatística, que resultam no conhecimento sobre suas características ou causas.

A bibliometria trata-se de uma ciência voltada para a medição da comunicação escrita, possibilitando a identificação de padrões e análises a partir da contagem de frequências. Auxilia no mapeamento da produção, disseminação e uso da informação, utilizando como fonte para análise o próprio conteúdo registrado ou metadados como nomes de autores, títulos e dados de publicação (FARIA, 2001). Entre os desafios para o desenvolvimento de estudos bibliométricos estão as barreiras de acesso à informação, a dispersão e descentralização de conteúdo em diversas bases de dados e a necessidade de limpeza e padronização de dados buscando garantir resultados confiáveis. Entre as características comuns dos estudos bibliométricos é apontada a necessidade de uma equipe multidisciplinar para o desenvolvimento de análises quantitativas e qualitativas a partir de diferentes técnicas e perspectivas; os requisitos técnicos como uso de *softwares* e linguagens de programação para extração de dados; a capacidade para armazenamento; o acesso às bases que configuram-se como fontes confiáveis de dados, o que geralmente apresenta altos custos; além da necessidade de tratamento de dados, com a limpeza e padronização, a fim de garantir a integridade e confiabilidade dos indicadores e análises obtidas (COSTAS, 2017).

A área da bibliometria é consolidada a partir da necessidade crescente de análise da informação, que apresenta aumento constante em sua produção e distribuição, sendo o resultado de sua aplicação a confecção de indicadores. No âmbito científico, Mugnani, Fujino e Yumiko (2017) apontam a “importância dos indicadores bibliométricos e cientométricos para subsidiar políticas de pesquisa e avaliação de Ciência, Tecnologia e Inovação (CT&I)”. Nesse contexto, a confecção de indicadores de CT&I está associada com a exploração de fontes de informação e aplicação de métodos de análise que permitem a obtenção de indicadores voltados para a avaliação da pesquisa científica e fornecimento de subsídios para o desenvolvimento de políticas institucionais e governamentais.

3.2 Unidade-Caso: Universidade Federal de São Carlos

A UFSCar é constituída por quatro campus: São Carlos (fundado em 1968), Araras (1991), Sorocaba (2005) e Lagoa do Sino (2011). A universidade oferece 64 cursos de graduação e possui 52 programas de pós-graduação reconhecidos pela CAPES, distribuídos entre 48 departamentos que estão divididos em uma estrutura organizacional formada por oito centros acadêmicos (UNIVERSIDADE..., 2016):

- a) Centro de Ciências Biológicas e da Saúde (CCBS), Centro de Ciências Exatas e de Tecnologia (CCET) e Centro de Educação e Ciências Humanas (CECH) localizados no campus São Carlos;
- b) Centro de Ciências e Tecnologias para a Sustentabilidade (CCTS), Centro de Ciências em Gestão e Tecnologia (CCGT) e Centro de Ciências Humanas e Biológicas (CCHB) localizados no campus Sorocaba;
- c) Centro de Ciências Agrárias (CCA) localizado no campus Araras; e
- d) Centro de Ciências da Natureza (CCN) localizado no campus Lagoa do Sino.

Em relação a atividade de pesquisa a universidade teve crescimento na taxa de produção científica na última década, o que pode demonstrar o foco e a intensa atividade de pesquisa desenvolvida pela instituição. No ano de 2016, a UFSCar apresentou melhoras nas posições em *rankings* universitários que apontam o desempenho acadêmico e de pesquisa de instituições de ensino superior. Segundo seu relatório anual de 2016:

No cenário mundial, a UFSCar ocupa a 753ª posição entre mais de 12.000 universidades avaliadas pelo *Webometrics Ranking*, mostrando melhoria significativa em relação à posição do ano anterior. A UFSCar manteve-se posicionada na faixa da 600ª à 700ª posição segundo o ranking QS [*World University Ranking*]. No âmbito da América Latina, ambos os rankings QS e *Webometrics* apontam melhoria de posição da UFSCar em relação ao ano anterior, posicionando-a entre as 30 melhores universidades da região. Entre as universidades brasileiras, a UFSCar melhorou seu posicionamento segundo os 3 rankings consultados [*Ranking* Universitário Folha, *Webometrics Ranking of World Universities* e *QS World University Rankings*]. Destaca-se o bom posicionamento da UFSCar nos indicadores específicos para a avaliação das atividades de Pesquisa presentes tanto no RUF [*Ranking* Universitário Folha] como no *Webometrics*, em que a universidade ocupa a 8ª e a 14ª posição (UNIVERSIDADE..., 2016, p. 141).

A constituição desses *rankings* utiliza como critérios não somente taxas de produção científica, o *Webometrics Ranking* desenvolvido pelo Conselho Superior de Investigação Científica (CSIC) da Espanha aplica métricas de webometria para avaliar a geração de conteúdos e atividades desenvolvidas por instituições de ensino superior. O *ranking* quantifica a disponibilização de conteúdo de acesso aberto distribuídos pelas IES na internet, considerando mais de 12.000 instituições no mundo (RANKING..., 2017). O *QS World University Ranking* elaborado pela empresa Quacquarelli Symonds se propõe a avaliar as melhores universidades do mundo, o *ranking* é elaborado a partir de métricas baseadas na reputação das instituições entre especialistas, alunos, qualidade de ensino, produção científica e internacionalização (QUACQUARELLI SYMONDS, 2017). O *Ranking* Universitário Folha (RUF) apresenta o *ranking* de universidades brasileiras construído a partir de cinco indicadores: pesquisa, internacionalização, inovação, ensino e mercado. Para a sua elaboração são coletados dados a partir de documentos de patentes, publicações em periódicos científicos, avaliações do Ministério da Educação (MEC) e dados de opiniões oriundas de pesquisas desenvolvidas pelo instituto Datafolha (FOLHA..., 2016).

3.3 Etapas da pesquisa

Appolinário (2006) aponta como fases para a realização de pesquisas quantitativas: a coleta de dados, por meio do uso de instrumentos ou técnicas específicas; a organização dos dados, sendo realizada a tabulação em planilhas ou *softwares* específicos; a análise de dados, identificando suas variáveis; e a elaboração do relatório final da pesquisa.

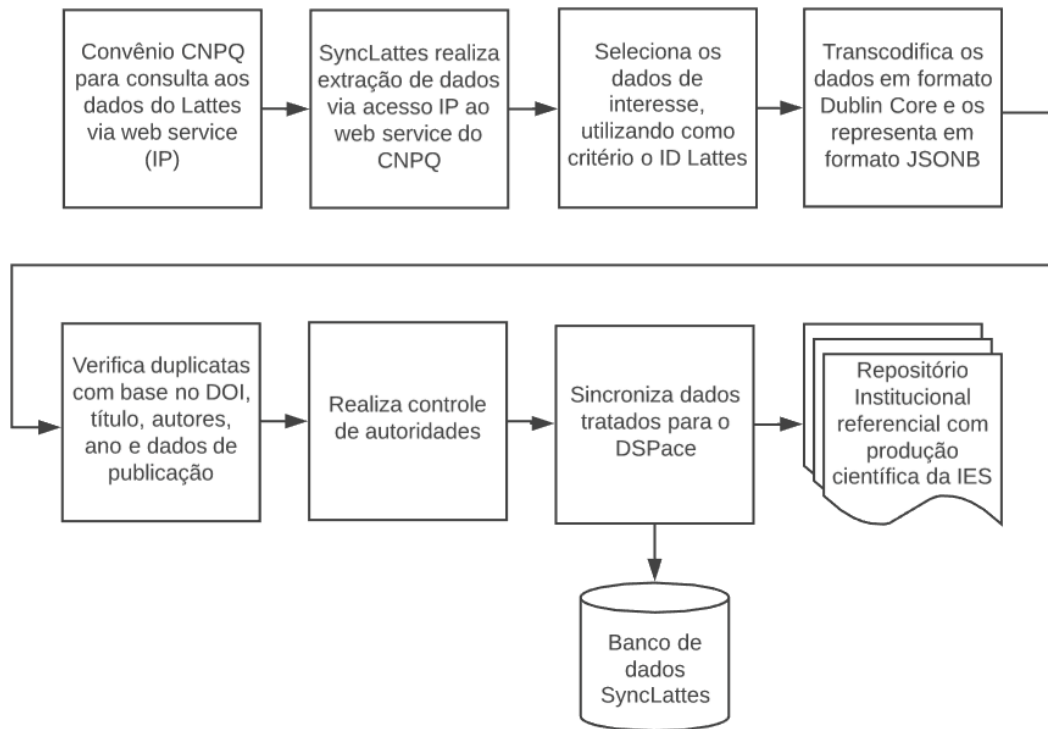
3.3.1 Coleta de dados

Com o objetivo de dar continuidade a avaliação e diagnóstico da Plataforma Lattes como fonte de metadados, a pesquisa foi desenvolvida com base no método de coleta de dados utilizado por Bassoli (2007). Segundo o método apresentado, foi utilizada a ferramenta SyncLattes para a obtenção de metadados da produção científica da UFSCar a partir da Plataforma Lattes. O SyncLattes foi desenvolvido por Matias (2015) com a finalidade de obtenção de metadados voltados para a construção de uma base de dados referencial destinada ao povoamento de RI (MATIAS, 2015; MATIAS et al., 2016; MATIAS; AMARAL; MATIAS, 2017). Trata-se de uma sistemática de carga automatizada de metadado, podendo ser definida como:

[...] uma metodologia de consulta ao *web service* da Plataforma Lattes via protocolo SOAP, utilizando as linguagens Java e Python; Desenvolvimento de um proxy para compartilhar o acesso ao *web services* da plataforma Lattes [...]; Desenvolvimento e integração com o DSpace de um sistema de autoridade de nomes baseado na Plataforma Lattes; Desenvolvimento de scripts para extração, tratamento e sincronização de metadados da Plataforma Lattes com o DSpace (MATIAS, 2015, p. 55).

Para o desenvolvimento do SyncLattes Matias (2015) obteve registros de publicações de docentes da UFSCar por meio de um convênio disponibilizado pelo CNPq que permite às IES terem acesso direto aos dados dos Currículos Lattes dos pesquisadores vinculados a instituição. A partir da possibilidade de acesso ao *web service* da Plataforma Lattes, um conjunto de *scripts* em linguagem Python foi desenvolvido possibilitando a extração de dados da Plataforma Lattes. A ferramenta utiliza como critério para a extração de dados os IDs Lattes dos pesquisadores, selecionando os dados referentes aos currículos indicados, os transcodificando para o padrão Dublin Core qualificado, e os representando no formato *JavaScript Object Notation* (JSONB). Após a extração e transcodificação, o SyncLattes realiza a verificação de duplicidades de registros utilizando como critério a comparação entre o *Digital Object Identifier* (DOI) ou a similaridade entre títulos, autores e ano das publicações. Os casos de duplicidade ocorrem quando uma mesma publicação foi informada em mais de um currículo, sendo essa funcionalidade essencial para garantir a integridade da extração de dados. Com a remoção dos registros duplicados, o SyncLattes aplica um controle de autoridade e sincroniza os dados com o *software* DSpace. Além da sincronização, a ferramenta dispõe de uma base de dados que permite a consulta direta aos registros extraídos, possibilitando a sua aplicação para outras finalidades, como a elaboração de diagnósticos e indicadores sobre a produção científica da IES. A Figura 13 apresenta a síntese do processo realizado pelo SyncLattes para a extração, tratamento e sincronização de dados da Plataforma Lattes com o DSpace.

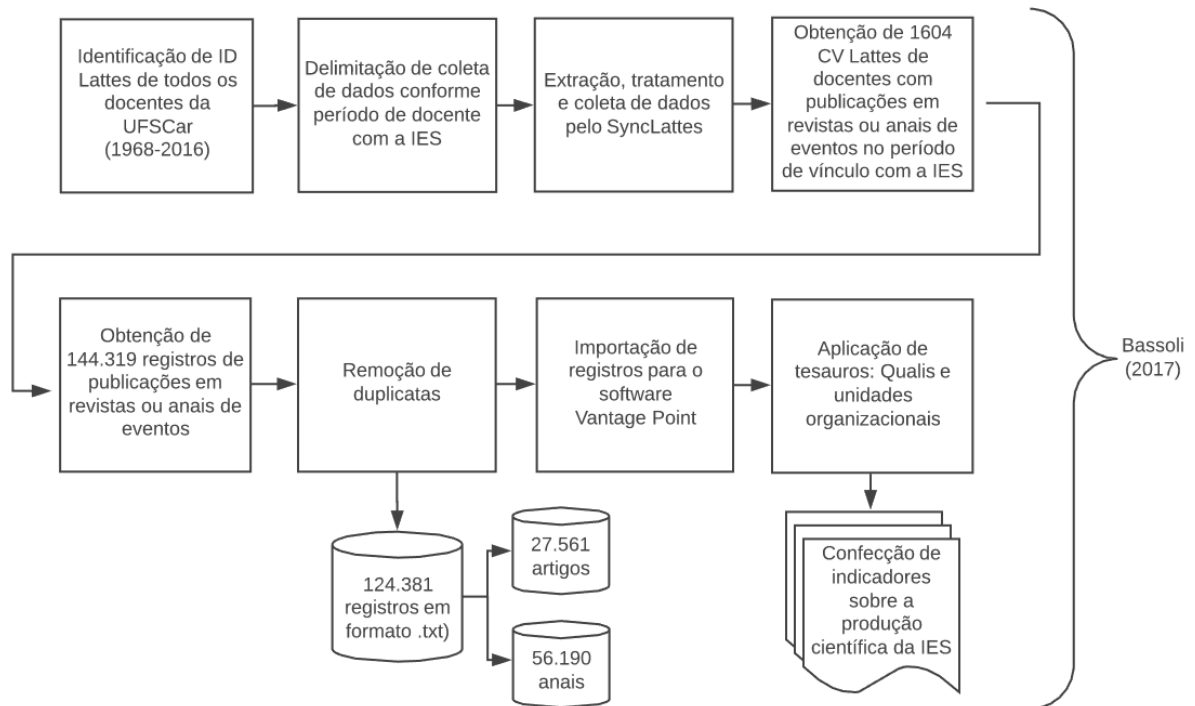
Figura 13 - Processos de coleta, tratamento e sincronização do SyncLattes



Fonte: Elaborado pelo autor.

A Figura 14 apresenta a síntese dos processos realizados por Bassoli (2017), utilizado como referência para o desenvolvimento da presente pesquisa. Para a avaliação da Plataforma Lattes como fonte de coleta de dados para a confecção de indicadores a autora utilizou uma relação de 1.604 docentes vinculados a UFSCar. Para a extração dos metadados foi utilizada a ferramenta SyncLattes sendo obtido 144.319 registros da produção científica da IES. Um tratamento quanto a duplicidade foi realizado, resultando em uma amostra de 124.381 registros (27.561 artigos e 56.190 publicações em eventos). Os dados foram importados para o VantagePoint e aplicados tesouros voltados a identificação de atributos como Qualis e unidades organizacionais da UFSCar.

Figura 14 - Processo de coleta e tratamento de dados realizado por Bassoli (2017)

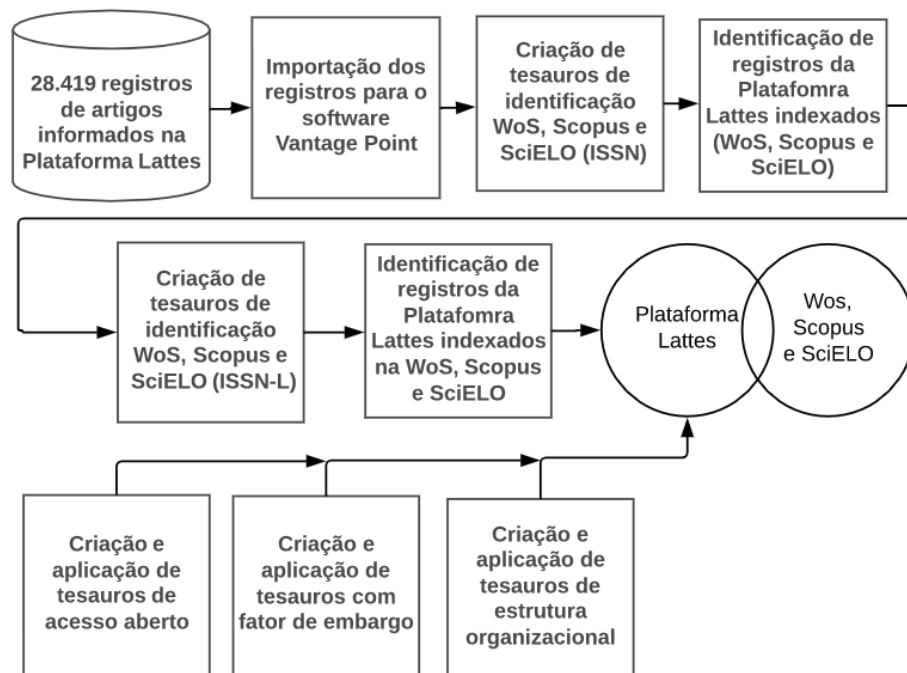


Fonte: Elaborado pelo autor.

Com base no método apresentado por Bassoli (2017) uma coleta de dados foi realizada em outubro de 2017 para a composição da amostra utilizada na pesquisa. Foram recuperados registros de publicações em periódicos científicos e anais de eventos, desses foram selecionados para a composição da amostra o tipo de publicação "artigos de periódicos", totalizando 28.419 registros. A escolha da amostragem ocorreu devido a predominância ou exclusividade desse tipo de publicação no conteúdo das BDICTs. Os registros obtidos em formato de texto puro (TXT) estavam estruturados nos seguintes campos: idioma; título; autor; ID Lattes; fonte; ano; ISSN; referência bibliográfica; DOI; tipo de publicação; palavras-chave; e classificação.

Os dados foram importados para o *software* VantagePoint, sendo aplicados recursos denominados “tesauros” para adicionar atributos aos registros, como por exemplo, identificar se estavam ou não indexados e qual seu vínculo em relação aos departamentos e programas de pós-graduação da UFSCar. A Figura 15 apresenta a síntese dos processos realizados na presente pesquisa para a comparação entre a Plataforma Lattes e as BDICTs analisadas.

Figura 15 - Processo de comparação entre Plataforma Lates, WOS, Scopus e SciELO



Fonte: Elaborado pelo autor.

3.3.2 Organização de dados

Para a organização dos dados coletados foi realizada a importação dos registros obtidos na Plataforma Lattes para o VantagePoint, *software* voltado para o apoio de atividades de mineração de dados e de texto, que auxilia na limpeza e análise de grandes volumes de informação. Como resultado da importação foi obtido um conjunto de listas e matrizes correspondentes a estrutura do arquivo TXT obtido pelo SyncLattes (Figura 16).

Figura 16 - Interface gráfica do *software* VantagePoint

The screenshot shows the VantagePoint software interface. At the top, there is a menu bar with options: File, Edit, View, Sheets, Fields, Groups, Tools, Scripts, Window, Help. Below the menu bar is a toolbar with various icons. The main window displays a table with the following data:

Field	Number of items	Derived	Data Type	Meta Tags
~Raw Record	474			
22 Áreas de Conhecimento (a partir de Áreas de Pesquisa)	37		Category	
22 Áreas de Conhecimento (a partir de Categorias Web of Science)	25		Category	
Ano da publicação	1		Year	
Áreas de Pesquisa	90		Category	
Autores	2270			
Categorias Web of Science	142		Category	
Código UT (Código único do registro)	474			
Estados Brasileiros	63			
Estados Brasileiros (Microregiões) - cidade	273			
Fonte	328			
Instituições	550			
Instituições (e País)	560			
Instituições Brasileiras	247			
Número de citações recebidas (até a data da coleta)	4		Number	
Número de referências citadas	96		Number	
País	119			
Palavras chave automáticas (ID)	2576			
Palavras-chave do Autor (DE)	1741			
Periódicos das Referências citadas	285			
Referências citadas	19836			
Resumos	461			
Tipo de documento	8			
Título	473			

Fonte: Elaborado pelo autor.

3.3.3 Análise dos dados

Com a importação para o VantagePoint, foram realizados tratamentos nos registros da produção científica e incluídos atributos que possibilitaram verificar sua disponibilidade nas bases de dados WOS, Scopus e SciELO. A análise comparativa foi realizada a partir da aplicação do recurso denominado “tesauros”. Esse recurso permite a criação de dicionários de sinônimos possibilitando a união de variações de um mesmo termo para padronização de listas, por exemplo, um nome de um país que apresenta diversas variações como United States, USA e EUA, pode ser unificado por meio da aplicação de um tesouro. O intuito da criação e aplicação de tesauros nos registros foi o de identificar quais as publicações informadas na Plataforma Lattes também estavam indexadas nas bases de dados.

O *International Standard Serial Number* (ISSN) foi utilizado como identificador comum para o processo de comparação de registros entre as fontes de metadados. O ISSN é um código internacional constituído por oito dígitos que é atribuído aos títulos de publicações seriadas com a finalidade de individualizá-los, facilitando sua identificação, recuperação e controle bibliográfico (INTERNATIONAL..., 2017). A atribuição do ISSN é coordenada pelo *ISSN International Centre*, organização criada em 1976 e sediada em Paris. No Brasil, o

Centro Brasileiro do ISSN (CBISSN) vinculado ao Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), é o órgão responsável pela atribuição do identificador para as publicações seriadas do país (INSTITUTO..., 2012b; INTERNATIONAL..., 2017). A nomenclatura do ISSN é constituída por dois grupos de quatro dígitos separados entre si por um hífen e antecidos pela sigla ISSN (ISSN XXXX-XXXX). Um número ISSN é atribuído a cada título de periódico considerando o suporte em que está vinculado, deste modo, um periódico terá um ISSN distinto para a sua versão impressa, *online*, em CD-ROM, etc.

3.3.4 Criação de tesauro *Web of Science*

A *Web of Science* (WOS) é de uma base de dados distribuída pela empresa Clarivate Analytics que indexa em seu conteúdo cerca de 12.000 títulos de caráter multidisciplinar. Sua coleção principal é formada por três índices: *Science Citation Index Expanded (SCI-Expanded)*, *Social Sciences Citation Index (SSCI)* e *Arts & Humanities Citation (A&HCI)*. A WOS é uma das bases de dados mais utilizadas em estudos cientométricos a nível mundial, apresentando uma vasta gama de metadados de seus registros que permite inúmeras possibilidades de cruzamento de dados e análise de indicadores (COSTAS, 2017). Além de oferecer a pesquisa por registros bibliográficos, a base de dados é conhecida por disponibilizar ferramentas para diagnóstico e índices de avaliação sobre o conteúdo indexado em sua coleção. Entre as ferramentas para diagnósticos que disponibiliza, tem destaque o *Journal Citation Reports (JCR)* que apresenta o Fator de Impacto (FI) de periódicos indexados. Essa métrica calcula a taxa de citação de uma publicação a partir do número de artigos publicados e citações recebidas ao longo de um determinado período de tempo, demonstrando o nível de visibilidade do periódico para sua área de conhecimento (CLARIVATE ANALYTICS, 2017). O FI é utilizado como base para geração de diversos indicadores, por exemplo, utilizado no Brasil como base de cálculo do Qualis de diversas áreas do conhecimento, métrica utilizada pela CAPES para a estratificação de periódicos e avaliação da produção científica dos programas de pós-graduação nacionais (BRASIL, 2009).

Para a construção do tesauro de identificação da WOS foram obtidos os dados disponíveis no *Master Journal List*, base de dados disponibilizada pela Clarivate Analytics, que permite a busca por informações sobre os periódicos indexados em suas coleções. Foram obtidos por meio do *Master Journal List* os documentos com a relação de periódicos indexados nas coleções que compõem a *Web of Science: Science Citation Index Expanded (SCI-Expanded)*, *Social Sciences Citation Index (SSCI)* e *Arts & Humanities Citation*

(A&HCI). Os documentos, em formato *Portable Document Format* (PDF), apresentavam os metadados: título, editor, ISSN, e-ISSN, país e idioma. A partir dos documentos, foram coletados os metadados dos títulos de periódicos indexados até 2016 na base de dados, sendo realizado um tratamento quanto à duplicidade. Ao ser realizada a extração de dados foram identificadas algumas inconsistências, como a atribuição de um mesmo número de ISSN para mais de um título de periódico. Devido a essas inconsistências uma relação oficial de periódicos indexados na WOS foi solicitada para a editora Clarivate Analytics, sendo obtida uma nova relação em formato *eXtensible Style Language* (XLS) com os metadados dos periódicos indexados na *Web of Science* até 2016. A relação apresentava os seguintes campos: título do periódico, código do periódico, editor, URL, ISSN, e-ISSN, país, região, idioma e categoria WOS (Figura 17).

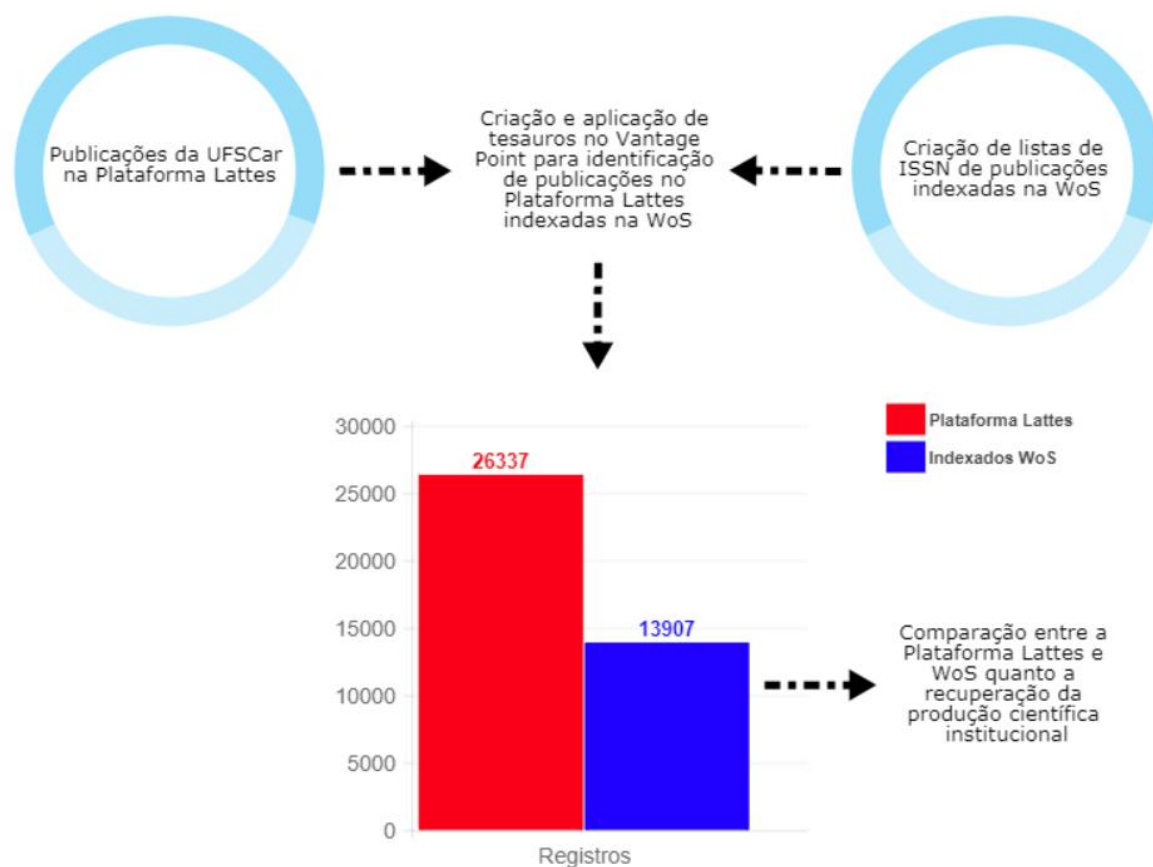
Figura 17 - Relação de metadados de periódicos indexados na WOS

B	C	D	E	F	G	H	I	J	K	L
JOURNAL TITLE	PUB CODE	PUBLISHER	Edition	JOURNAL URL	ISSN	E-ISSN	COUNTRY	REGION	LANGUAGE	WOS CATEGOF
2D Materials	NW493	IOP PUBLISHING LTD	WoS 3	http://iopscience.iop.org/2053-1583/	2053-1583	2053-1583	ENGLAND	EU	English	Materials Science, Multidisciplinary
3 Biotech	ZX651	SPRINGER HEIDELBERG	WoS 3	http://www.springerlink.com/content/2190-572X/	2190-5738	2190-572X	GERMANY	EU	English	Biotechnology & Applied Microbiology
3C Empresa	ADL66	AREA INNOVACION & DESARROLLO	ESCI	http://www.3ciencias.com/revistas/categoria/empresa/	2254-3376	2254-3376	SPAIN	EU	Spanish	Business
3D Printing and Additive Manufacturing	PZ785	MARY ANN LIEBERT, INC	WoS 3	http://www.liebertpub.com/3Dp	2329-7662	2329-7670	UNITED STATES	NA	English	Engineering, Manufacturing

Fonte: Clarivate Analytics (2017).

No total foram obtidos dados de 17.904 periódicos indexados na WOS, contabilizando 31.190 números de ISSN. Cabe ressaltar que o número de ISSN é maior que o número de títulos de periódicos devido aos títulos apresentarem em alguns casos mais de um ISSN, referentes aos seus diferentes suportes. Os dados dos periódicos indexados foram importados para o VantagePoint e criado um tesauro com a relação de ISSN das publicações. O tesauro de periódicos indexados na WOS foi aplicado aos registros da produção científica da UFSCar, o que possibilitou identificar quais registros informados na Plataforma Lattes também estavam disponíveis na WOS (Figura 18).

Figura 18 - Fluxo de comparação de registros entre a Plataforma Lattes e a WOS

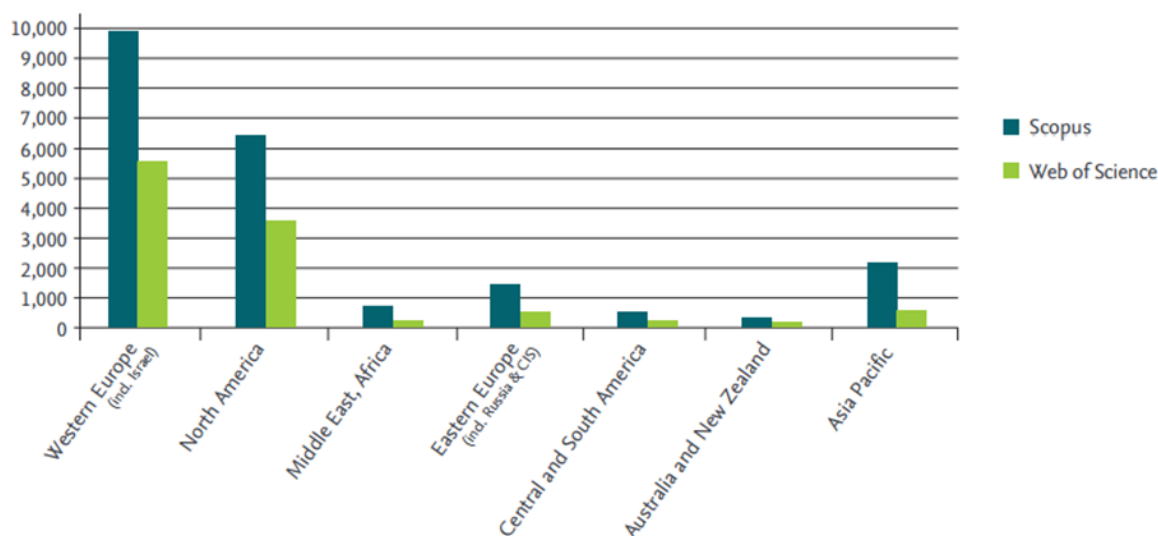


Fonte: Elaborado pelo autor.

3.3.5 Criação de tesauro Scopus

A Scopus é uma base de dados multidisciplinar da editora Elsevier, que assim como a WOS, oferece ferramentas para a extração e análise de seus registros bibliográficos, caracterizando-se como uma importante fonte para a elaboração de indicadores. A base de dados possui como objetivo ter uma grande cobertura da produção científica mundial, diferenciando-se da WOS quanto ao volume de periódicos indexados. Na Figura 19 é apresentado uma análise comparativa entre os números de títulos indexados nas duas bases de dados, WOS e Scopus, demonstrando os perfis distintos: a de possuir uma seleção dos principais periódicos científicos mundiais, no caso da WOS, e o de oferecer uma grande cobertura das áreas de conhecimento, no caso da Scopus (COSTAS, 2017; ELSEVIER, 2016).

Figura 19 - Número e distribuição geográfica de periódicos indexados na Scopus e WOS



Fonte: Elsevier (2016).

Para a construção do tesouro de identificação foram obtidos os metadados das publicações indexadas na Scopus a partir da *Scopus Source List*. O documento, em formato *eXtensible Style Language* (XLS), está disponível para *download* na interface da base de dados e apresenta entre os metadados disponíveis: títulos, ISSN e e-ISSN das publicações indexadas (Figura 20).

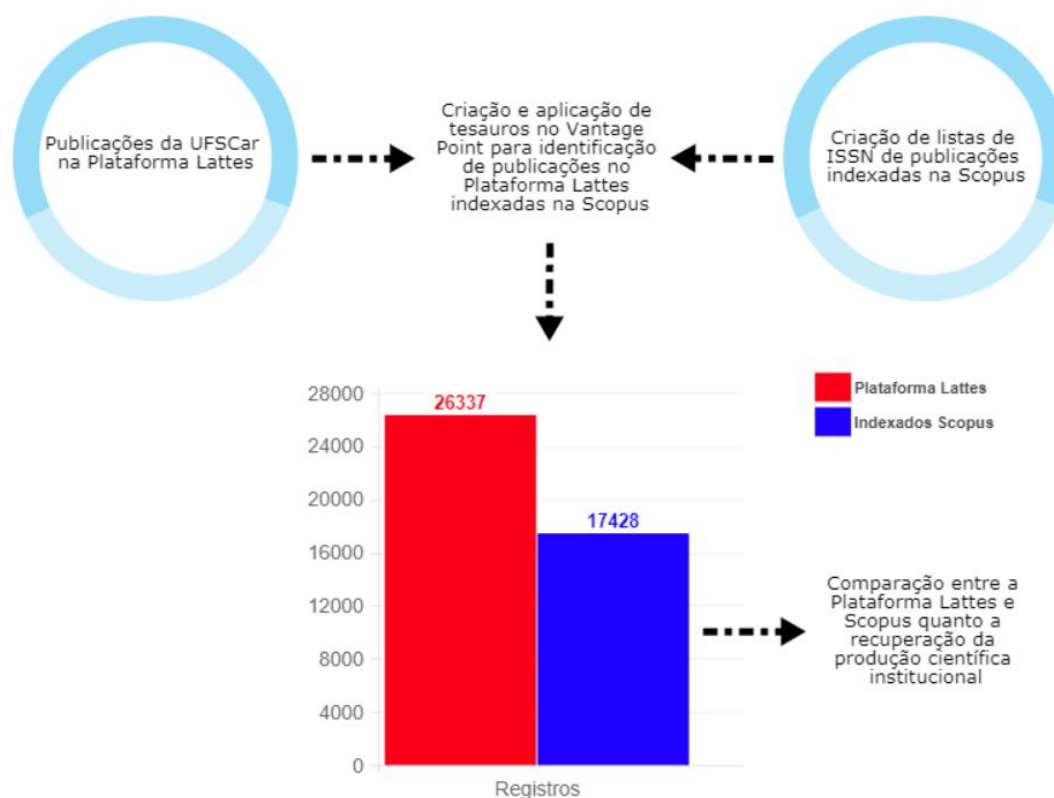
Figura 20 - Relação de metadados de periódicos indexados na Scopus

Sourcerecord id	Source Title (Medline-sourced journals are indicated in Green) Titles indicated in bold red do not meet the Scopus quality criteria anymore and therefore Scopus discontinued the forward capturing	Print-ISSN	E-ISSN	Active or Inactive	Coverage	Article language in source (three-letter ISO language codes)
18500162600	21st Century Music	15343219		Inactive	2002-2011	ENG
21100404576	2D Materials		20531583	Active	2014-ongoing	ENG
21100447128	3 Biotech	2190572X	21905738	Active	2015-ongoing	ENG
21100779062	3D Printing and Additive Manufacturing	23297662	23297670	Active	2014-ongoing	ENG
21100229836	3D Research		20926731	Active	2010-ongoing	ENG
19700200922	3L: Language, Linguistics, Literature	01285157		Active	2008-ongoing	ENG
145295	4OR	16194500	16142411	Active	2003-ongoing	ENG
21100399164	A & A case reports	23257237		Active	2015-ongoing	ENG
16400154734	A + U-Architecture and Urbanism	03899160		Active	2002-ongoing	JPN, ENG
5700161051	A Contrario. Revue interdisciplinaire de sciences sociales	16607880		Active	2009-ongoing	FRE, ENG
19600162043	A.M.A. American Journal of Diseases of Children	00968994		Inactive	1945-1955	
19400157806	A.M.A. archives of dermatology	00965359		Inactive	1955-1959	
19600162081	A.M.A. Archives of Dermatology and Syphilology	00965979		Inactive	1950-1954	
19400157807	A.M.A. archives of industrial health	05673933		Inactive	1954-1960	
19600162082	A.M.A. Archives of Industrial Hygiene and Occupational Medicine	00966703		Inactive	1950-1954	

Fonte: Elsevier (2016).

A partir do *Scopus Source List* foram obtidos os metadados de um total de 36.294 títulos de periódicos e anais de eventos (*proceedings*), que foram importados para o VantagePoint e realizado o mesmo procedimento aplicado para o tesauro da WOS, o que permitiu identificar quais registros recuperados a partir da Plataforma Lattes estavam indexados na Scopus (Figura 21).

Figura 21 - Fluxo de comparação de registros entre a Plataforma Lattes e a Scopus



Fonte: Elaborado pelo autor.

3.3.6 Criação de tesauro SciELO

A SciELO é resultado de um programa da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) em parceria com o Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME/OPAS/OMS). A base de dados foi criada com o intuito de promover maior “[...] visibilidade, acessibilidade, qualidade, credibilidade, uso e impacto nacional e internacional de periódicos de qualidade por meio da sua indexação, publicação e interoperabilidade na web” (SCIENTIFIC..., 2014, p. 4). A SciELO é voltada

para a publicação *online* de periódicos científicos em acesso aberto, tendo como objetivo [...] contribuir para o desenvolvimento da pesquisa científica por meio do aperfeiçoamento e da ampliação dos meios, infraestruturas e capacidades de comunicação e avaliação dos seus resultados veiculados por periódicos de qualidade crescente do Brasil publicados em acesso aberto (SCIENTIFIC..., 2014, p. 4). Iniciada em 1998 a partir da coleção SciELO Brasil, atualmente a Rede SciELO comporta coleções de 14 países: África do Sul, Argentina, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, Espanha, México, Peru, Portugal, Uruguai e Venezuela, além de Paraguai que possui sua coleção em desenvolvimento. A base oferece também a coleção Saúde Pública, coleção temática que reúne revistas científicas sobre saúde pública de países ibero-americanos. Todos os periódicos indexados na Rede SciELO adotam licenças *Creative Commons* que garantem o acesso aberto ao seu conteúdo.

Para a construção do tesouro de identificação foram obtidos os metadados dos periódicos indexados na base de dados a partir da ferramenta SciELO Analytics, que reúne e disponibiliza indicadores sobre as coleções da Rede SciELO. Uma relação de metadados de 1.584 títulos indexados até 10 de maio de 2017 foi obtida a partir da consulta do item “*journal network*”, disponível na seção “*reports*” da ferramenta (Figura 22).

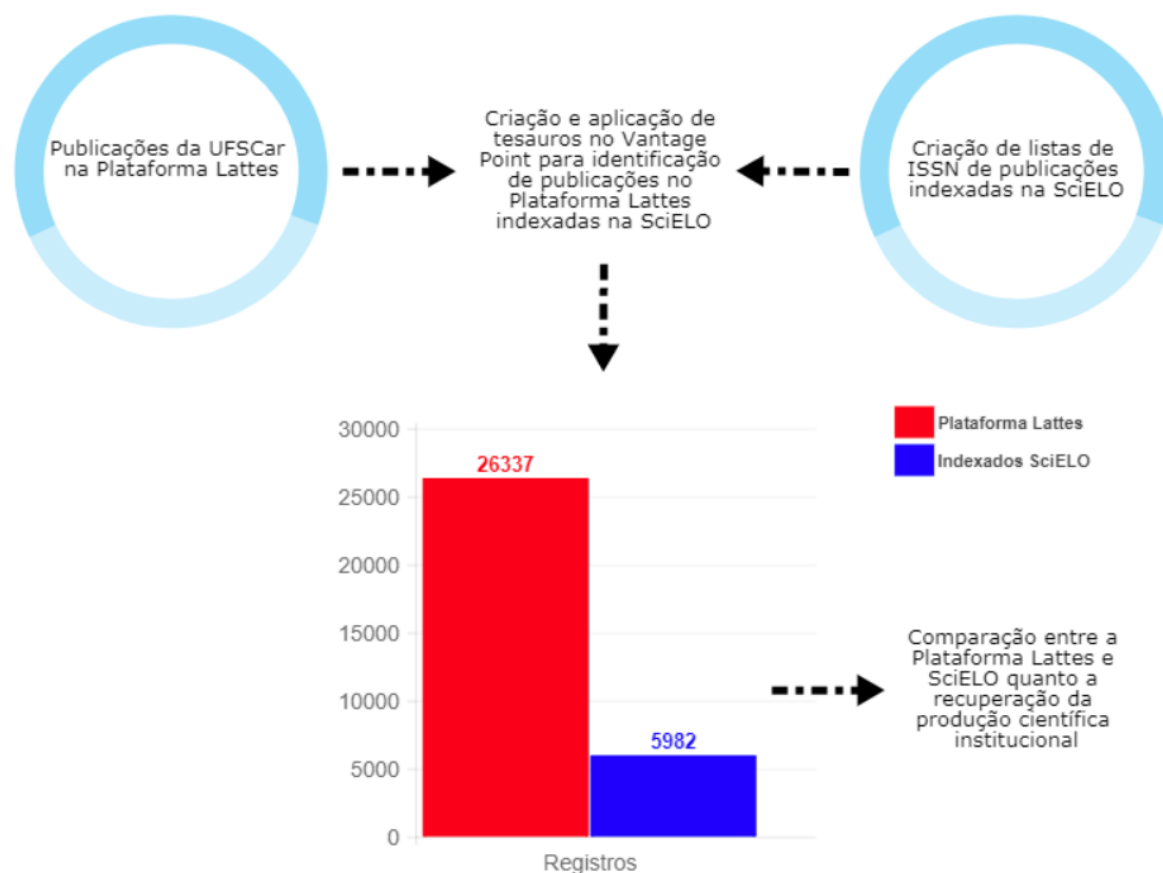
Figura 22 - Relação de metadados de periódicos indexados na SciELO

extraction date	study unit	collection	ISSN SciELO	ISSN's	title at SciELO	title thematic areas
2017-05-10	journal	psi	2175-6104	1775-1851;2175-6104	A Peste : Revista de Psicanálise e Sociedade	Applied Social Sciences
2017-05-10	journal	mex	2448-6132	2007-428X;2448-6132	Abanico veterinario	Agricultural Sciences;Bi
2017-05-10	journal	scl	0102-6720	0102-6720;2317-6326	ABCD. Arquivos Brasileiros de Cirurgia Digestiva	Health Sciences
2017-05-10	journal	esp	1578-908X	1578-908X;2255-1271	Acción Psicológica	Health Sciences
2017-05-10	journal	cub	1024-9435	1024-9435;1561-2880	ACIMED	Applied Social Sciences;
2017-05-10	journal	col	0120-2812	0120-2812	Acta Agronómica	Agricultural Sciences
2017-05-10	journal	scl	0044-5967	0044-5967;1809-4392	Acta Amazonica	Agricultural Sciences;Bi
2017-05-10	journal	chl	1726-569X	1726-569X	Acta bioethica	Applied Social Sciences;
2017-05-10	journal	col	0120-548X	0120-548X	Acta Biológica Colombiana	Biological Sciences
2017-05-10	journal	arg	0325-2957	0325-2957;1851-6114	Acta bioquímica clínica latinoamericana	Biological Sciences;Heal
2017-05-10	journal	scl	0102-3306	0102-3306;1677-941X	Acta Botanica Brasilica	Biological Sciences
2017-05-10	journal	mex	0187-7151	0187-7151;2448-7589	Acta botánica mexicana	Biological Sciences
2017-05-10	journal	ven	0084-5906	0084-5906	Acta Botánica Venezuelica	Biological Sciences

Fonte: SciELO Analytics (SCIENTIFIC..., 2017).

Os metadados dos periódicos indexados na SciELO foram importados para o VantagePoint, sendo realizado o mesmo procedimento aplicado para o tesouro da WOS, o que permitiu identificar quais registros recuperados a partir da Plataforma Lattes estavam indexados na SciELO (Figura 23).

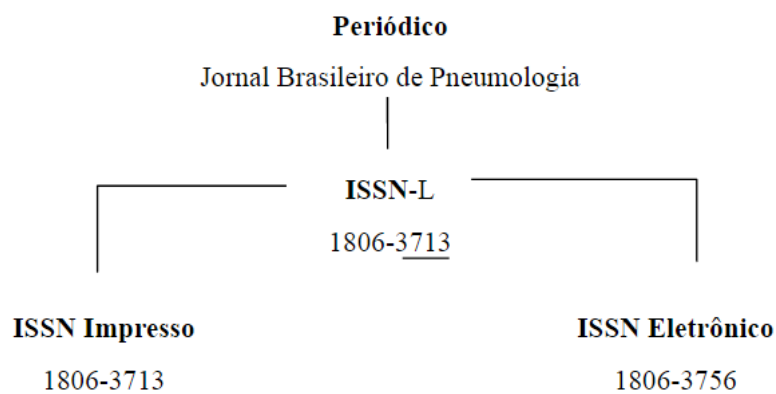
Figura 23 - Fluxo de comparação de registros entre a Plataforma Lattes e a SciELO



Fonte: Elaborado pelo autor.

3.3.7 Criação de tesouro ISSN-L

O *International Standard Serial Number Linking* (ISSN-L) tem como objetivo minimizar as dificuldades de recuperação de informação geradas devido as variações de ISSN dos diferentes suportes de uma mesma publicação. O ISSN-L é um identificador atribuído a uma publicação seriada independente das variações de suporte que possui, o que garante a exatidão na representação e recuperação do título em bases de dados e sistemas de informação. O identificador possui a mesma nomenclatura do número ISSN, sendo formado por 8 dígitos separados entre si por hífen e antecedido pela sigla ISSN-L (ISSN-L XXXX-XXXX). Sua atribuição não gera um novo código de identificação para a publicação, mas sim eleger entre os números de ISSN existentes qual será considerado o identificador de ligação do título. Por exemplo, o *Jornal Brasileiro de Pneumologia* possui o número 1806-3713 para identificar a versão impressa da publicação e também seu ISSN-L, sendo a versão eletrônica identificada pelo número 1806-3756, o que determina a seguinte relação:



A vinculação de diferentes números de ISSN para suportes distintos de um periódico poderia ocasionar problemas na integridade da comparação entre a Plataforma Lattes e as bases de dados, uma vez que um mesmo título poderia estar descrito pelo número de ISSN impresso e em outra fonte pelo número de ISSN eletrônico. Devido a essa possível falta de integridade foi realizada a concatenação dos tesouros criados a partir dos números de ISSN obtidos nas bases de dados com uma relação de ISSN-L.

O trabalho de Gabriel Junior (2016) comprova o impacto do ISSN-L para os resultados na comparação de conteúdo entre fontes de informação. O autor analisou o conteúdo das bases de dados WOS (JCR), Scopus (Scimago) e Qualis quanto as variações na descrição de conteúdo, constatando o uso de diferentes ISSN para um mesmo título, correspondentes aos seus diferentes suportes. Ao aplicar o ISSN-L na comparação de conteúdo entre as fontes “[...] o índice do SJR em relação ao Qualis passa de 30% para 54%, um acréscimo de 24% de compatibilidade dos registros, e na comparação do JCR com o Qualis, o índice aumenta de 37,6% para 43,7%, incremento de 6,1%” (GABRIEL JUNIOR, 2016, p. 6). Além da constatação de variações no conteúdo indexado, o autor também identificou a predominância de ISSN impresso na WOS e Qualis enquanto a Scopus apresentou maior frequência no uso de ISSN eletrônico.

Para a construção do tesouro de ISSN-L foi solicitada a relação entre os números de ISSN e ISSN-L, fornecidas pelo *ISSN International Centre*. Foram obtidos dois arquivos em formato TXT (Figura 24) com a seguinte relação:

- 1) **ISSN-L-to-ISSN**: apresentando uma lista de 1.837.070 números de ISSN-L e todos os respectivos números de ISSN vinculados a ele, correspondentes aos diferentes suportes de uma publicação; e

- 2) **ISSN-to-ISSN-L**: apresentando uma relação de 1.933.426 registros de ISSN e seu respectivo ISSN-L.

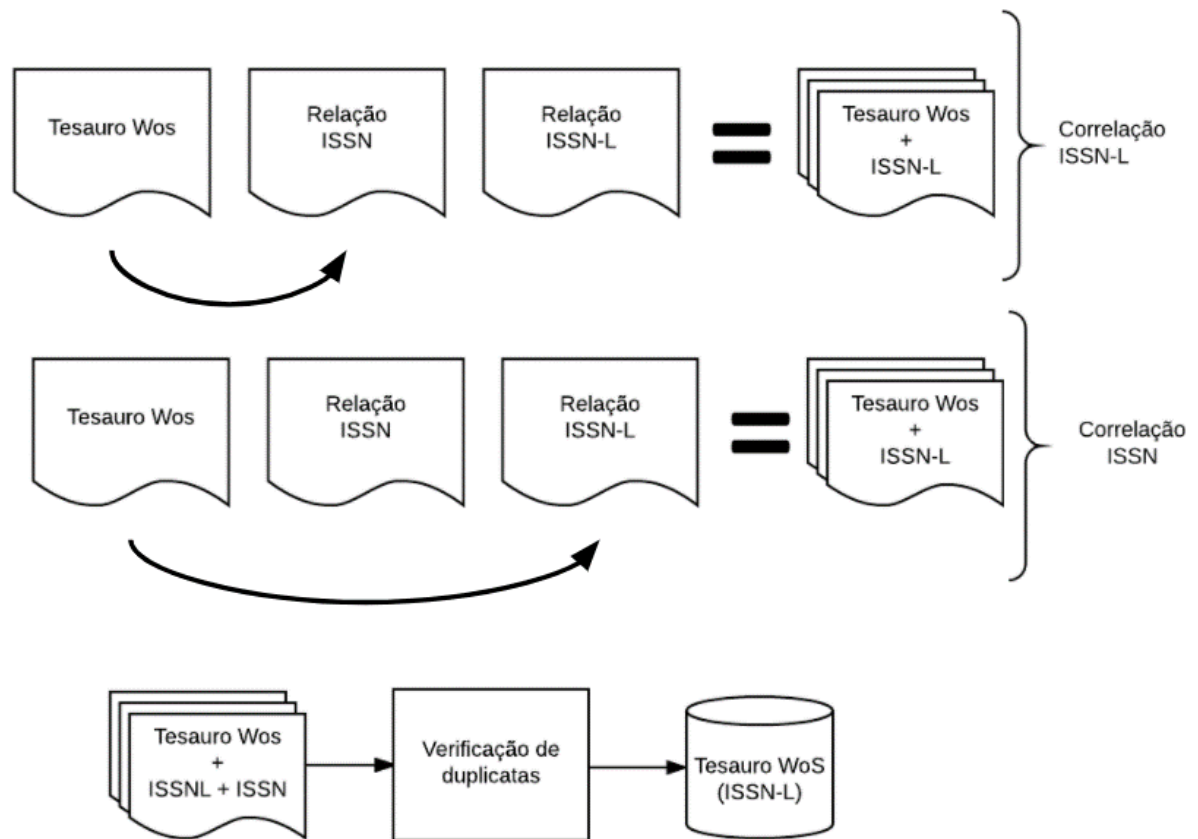
Figura 24 - Relações de ISSN obtidas a partir das tabelas de ISSN-L

ISSN-L	ISSN			
0000-0019	0000-0019	2150-4008		
0000-0027	0000-0027			
0000-1155	0000-0671	0000-1155	0065-759X	0068-0540
0000-1686	0000-1686	0000-1732		
0000-1694	0000-1694			
0001-026X	0001-026X	2326-9014		
0001-0456	0001-0456	1076-5476		
0001-0464	0001-0464			
0001-0472	0001-0472			
ISSN	ISSN-L			
0000-0019	0000-0019			
0000-0027	0000-0027			
0000-0043	0000-0043			
0000-0051	0000-0051			
0000-006X	0000-006X			
0000-0078	0000-0078			
0000-0094	0000-0094			
0000-0108	0000-0108			
0000-0140	0000-0140			

Fonte: *ISSN International Centre* (INTERNATIONAL..., 2017).

Para a concatenação entre os diferentes números de ISSN das publicações, referente aos seus suportes, foi utilizado o arquivo ISSN-to-ISSN-L. Os 1.933.426 registros apresentados no arquivo foram importados para o *software* Microsoft Excel. Posteriormente foram inseridos os ISSN dos periódicos indexados nas bases de dados WOS, Scopus e SciELO, sendo realizadas as seguintes concatenações: união do tesauro com a relação de periódicos indexados nas bases de dados com a coluna ISSN do arquivo; e união do tesauro com a relação de periódicos indexados nas bases de dados com a coluna ISSN-L do arquivo. Foi realizada a remoção de duplicidades, sendo obtido como resultado um tesauro para cada base de dados com todas as diferentes versões de ISSN do título indexado (Figura 25).

Figura 25 - Processo de construção do tesauro de ISSN-L



Fonte: Elaborado pelo autor.

Os tesauros com a relação de periódicos indexados em cada base de dados, atualizados com as diferentes versões de ISSN dos títulos, foram aplicados novamente aos registros da produção científica da UFSCar que foram importada no VantagePoint. Após a concatenação houve acréscimo de 6.165 números de ISSN nos tesauros. Quanto ao impacto na comparação entre as fontes, houve um aumento de 373 registros em comum entre a Plataforma Lattes e a WOS; 105 registros para a Scopus; e 178 registros para a SciELO (Tabela 1.)

Tabela 1 - Número de ISSN e registros da WOS, Scopus e SciELO após concatenação com ISSN-L

	WOS	Scopus	SciELO	Total
Número de ISSN a partir das bases de dados	31.190	45.199	2.200	56.771
Número de ISSN após a concatenação (ISSN-L)	32.754	55.559	2.576	62.936
Aumento no número de ISSN após concatenação (ISSN-L)	1.564	10.360	376	6.165
Número de registros a partir das bases de dados	13.617	16.947	5.982	18.901
Número de registros após a concatenação (ISSN-L)	13.907	17.428	6.072	19.038
Aumento no número de registros após concatenação (ISSN-L)	290	481	90	137

Fonte: Elaborado pelo autor.

3.3.8 Criação de tesauro de acesso aberto

Considerando a finalidade de realizar o povoamento de repositórios institucionais a partir dos metadados coletados na Plataforma Lattes, foi aplicado um tesauro para identificar quais os registros recuperados estavam em fontes de acesso aberto. Para a identificação foi criado um tesauro a partir dos metadados de periódicos disponíveis no DOAJ. O diretório indexa título que são de acesso aberto, tendo como objetivo aumentar a visibilidade dos periódicos de acesso aberto para a comunidade científica e, conseqüentemente, aumentar o impacto dessas publicações (DIRECTORY..., 2017b). Para a construção do tesauro foram obtidos os metadados de 9.467 periódicos (Figura 26), que foram importados para o VantagePoint e realizado o mesmo procedimento de construção e aplicação de tesouros utilizados para a identificação dos registros indexados nas três bases de dados. A aplicação do tesauro permitiu identificar quais os registros recuperados a partir da Plataforma Lattes estavam em fontes de acesso aberto.

Figura 26 - Relação de metadados de periódicos indexados no DOAJ

Journal title	Journal ISSN (print version)	Journal EISSN (online version)	Publisher
Revista de Microbiologia	0001-3714	1678-9881	Sociedade Brasileira de Microbiologia
Anais da Academia Brasileira de Ciências	0001-3765	1678-2690	Academia Brasileira de Ciências
ACME	0001-494X	2282-0035	Università degli Studi di Milano
Acta Dermato-Venereologica	0001-5555	1651-2057	Society for Publication of Acta Dermato-Venereologica
Acta Mycologica	0001-625X	2353-074X	Polish Botanical Society
Acta Societatis Botanicorum Poloniae	0001-6977	2083-9480	Polish Botanical Society
Acta Stomatologica Croatica	0001-7019	1846-0410	University of Zagreb. School of Dental Medicine
Acta Veterinaria Brno	0001-7213	1801-7576	"University of Veterinary and Pharmaceutical Sciences
Africa Spectrum	0002-0397	1868-6869	German Institute of Global and Area Studies (GIGA)
Proceedings of NAS RA : Mechanics	0002-3051		National Academy of Sciences of the Republic of Armenia
Alfa: Revista de Lingüística	0002-5216	1981-5794	Universidade Estadual Paulista Júlio de Mesquita Filho
Análise Social	0003-2573	2182-2999	Instituto de Ciências Sociais da Universidade de Lisboa
Angle Orthodontist	0003-3219		E.H Angle Education and Research Foundation

Fonte: DOAJ (DIRECTORY..., 2017b).

Os 13.497 números de ISSN identificados a partir do DOAJ foram concatenados com os dados informados no arquivo de ISSN-L, sendo obtido como resultado final um tesouro com 14.258 números de ISSN.

3.3.9 Criação de tesouro de fator de embargo

Um tesouro de fator de embargo foi criado com a finalidade de identificar quais os níveis de embargo e a potencial carga de registros que poderia ser depositada e disponibilizada no RI da unidade-caso. Para a construção do tesouro foram coletados dados de periódicos classificados na SHERPA/ROMEO, plataforma que disponibiliza informações sobre as políticas editoriais de periódicos científicos no que tange às permissões concedidas para a disponibilização de conteúdo em acesso aberto. As informações sobre os periódicos foram obtidas a partir da seção “revistas” do SHERPA/ROMEO, que apresenta os metadados de título, ISSN, e-ISSN, classificação, editor e notas (Figura 27).

Figura 27 - Relação de metadados de periódicos indexados no SHERPA/ROMEO

A - B - C - D - E - F - G - H - I - J - K - L - M - N - O - P - Q - R - S - T - U - V - W - X - Y - Z - Outro -					
Título da revista	ISSN	ESSN	Cores RoMEO	Editor	Notas
A Current Bibliography on African Affairs	0011-3255	-	Verde	SAGE Publications (UK and US)	Outras partes
A Economia em Revista - AERE	1413-6090	2236-2029	Azul	Universidade Estadual de Maringá	-
A Life in the Day	1366-6282	-	Verde	Emerald	-
A obra nasce	1645-8729	-	Azul	Edições Universidade Fernando Pessoa	-
AACE Journal	1551-3696	-	Verde	Association for the Advancement of Computing in Education	-
AACN Advanced Critical Care	1559-7768	1559-7776	Amarelo	American Association of Critical Care Nurses	Outras partes
AADE in Practice	2325-1603	-	Verde	SAGE Publications (UK and US)	Outras partes

Fonte: SHERPA/ROMEO (2017).

Foram coletados 43.875 números de ISSN a partir do SHERPA/ROMEIO, importados para o VantagePoint e concatenados com os dados apresentados no arquivo de ISSN-L, como resultado foi obtido um tesouro com 46.387 números de ISSN. A aplicação do tesouro de embargo permitiu classificar os registros da produção científica da UFSCar informados na Plataforma Lattes, segundo o sistema de cores apresentados pelo SHERPA/ROMEIO, demonstrando desse modo a potencialidade dos registros para o povoamento do seu RI.

3.3.10 Criação de tesouro de vínculo institucional

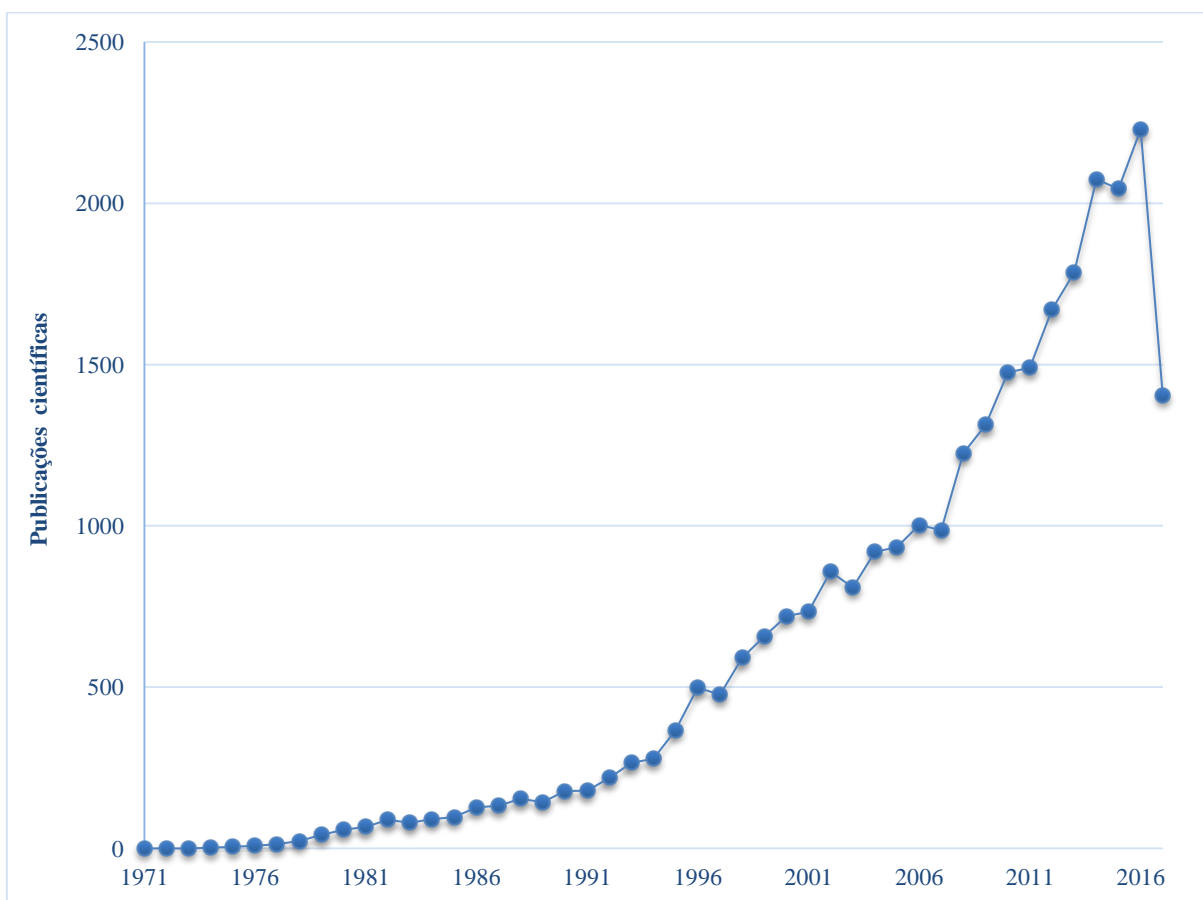
Após a análise comparativa quanto a disponibilidade de registros da produção científica da UFSCar na Plataforma Lattes e nas bases de dados foram construídos tesouros segundo a estrutura organizacional da universidade, a fim de identificar a relação dos registros e a cobertura das fontes conforme as áreas de conhecimento pesquisadas na IES. A construção dos tesouros foi realizada a partir da lista de docentes (IDs Lattes) utilizada para a coleta de dados a partir do SyncLattes. A partir da identificação do vínculo institucional foi possível criar tesouros para a identificação dos 4 campus, 8 centros acadêmicos, 48 departamentos e 54 programas de pós-graduação. A aplicação do tesouro de vínculo institucional permitiu o aprofundamento da análise quanto as características da produção científica da UFSCar recuperada a partir da Plataforma Lattes.

4 RESULTADOS

4.1 Produção científica da UFSCar informada na Plataforma Lattes

A amostra referente a produção científica da UFSCar recuperada a partir da Plataforma Lattes, consistiu em 28.419 publicações. A Figura 28 apresenta a distribuição da amostra segundo o ano de publicação dos registros.

Figura 28 – Produção científica da UFSCar recuperada na Plataforma Lattes (1971-2017)

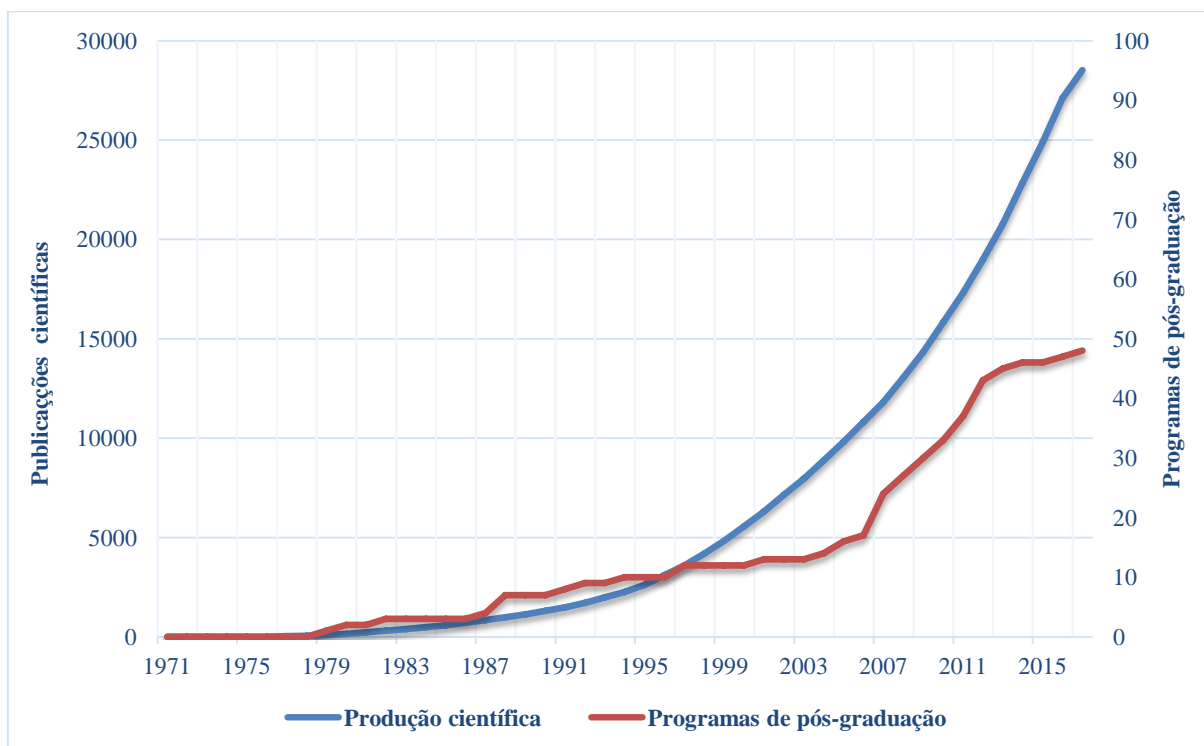


Fonte: Elaborada pelo autor.

Nota-se um crescimento acentuado do número de registros a partir da década de 1990. Esse crescimento pode estar associado com a expansão de campus da universidade, assim como pelo aumento do número de programas de pós-graduação. Atualmente a UFSCar conta com 54 programas de pós-graduação, que oferecem 85 cursos distribuídos entre mestrado (44), doutorado (31) e mestrado profissional (10), sendo apresentada uma maior concentração de criação de cursos a partir da década de 2000 (UNIVERSIDADE..., 2014; BRASIL, 2016). A Figura 29 apresenta a relação entre o crescimento acumulado no número de produção

científica informada na Plataforma Lattes e o número de programas de pós-graduação existentes na universidade.

Figura 29 – Taxa acumulada de criação de cursos de pós-graduação na UFSCar



Fonte: Elaborado pelo autor.

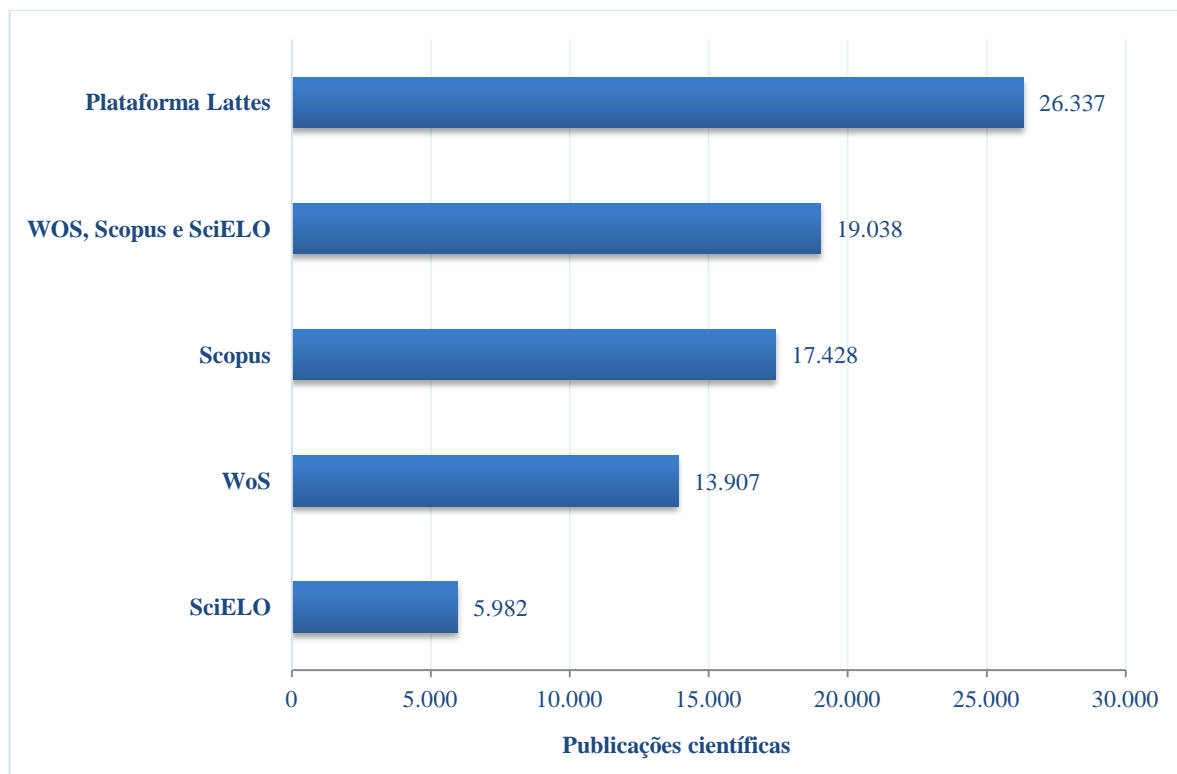
O fato da Plataforma Lattes ter sido lançada no final da década de 90 também pode influenciar a concentração do número de publicação a partir desse período, uma vez que o registro de publicações que o antecedem são frutos do preenchimento retroativo realizado pelo pesquisador.

4.2 Produção científica da UFSCar indexada na WOS, Scopus e SciELO

Da amostragem de dados (28.419 registros), 26.337 apresentavam número de ISSN, identificador comum utilizados para a comparação entre a Plataforma Lattes e as bases de dados WOS, Scopus e SciELO. O número de registros com ISSN, e passíveis de comparação, corresponde a 93% dos registros de artigos científicos recuperados a partir da Plataforma Lattes. A partir da parcela de registros com ISSN foram aplicados os tesouros de identificação das bases de WOS, Scopus e SciELO, permitindo identificar que dos 26.337 registros 19.038 (72%) encontram-se indexados em pelo menos uma das três bases de dados analisadas. Quanto a concentração dos registros indexados, a Scopus apresentou cobertura para 66% dos

registros informados na Plataforma Lattes; a WOS 53% e a SciELO 23%, proporção essa esperada devido ao número de itens indexados em cada base de dados (Figura 30).

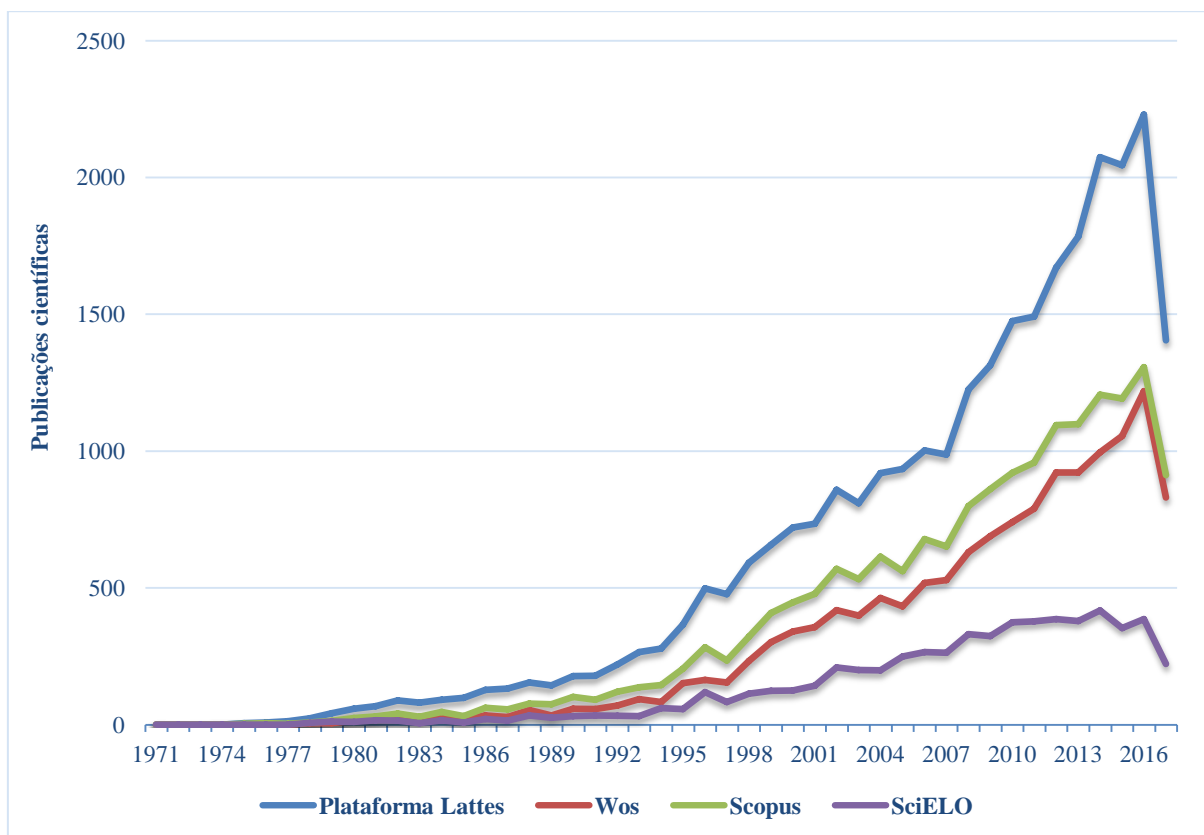
Figura 30 – Produção científica da UFSCar informada na Plataforma Lattes e indexada na WOS, Scopus e SciELO



Fonte: Elaborada pelo autor.

O número de artigos indexados nas bases de dados acompanha a taxa de crescimento da produção científica da UFSCar informada na Plataforma Lattes. Nota-se um menor crescimento no número de registros na base de dados SciELO, o que pode ser decorrente da delimitação estabelecida pela SciELO quanto ao crescimento de suas coleções (Figura 31).

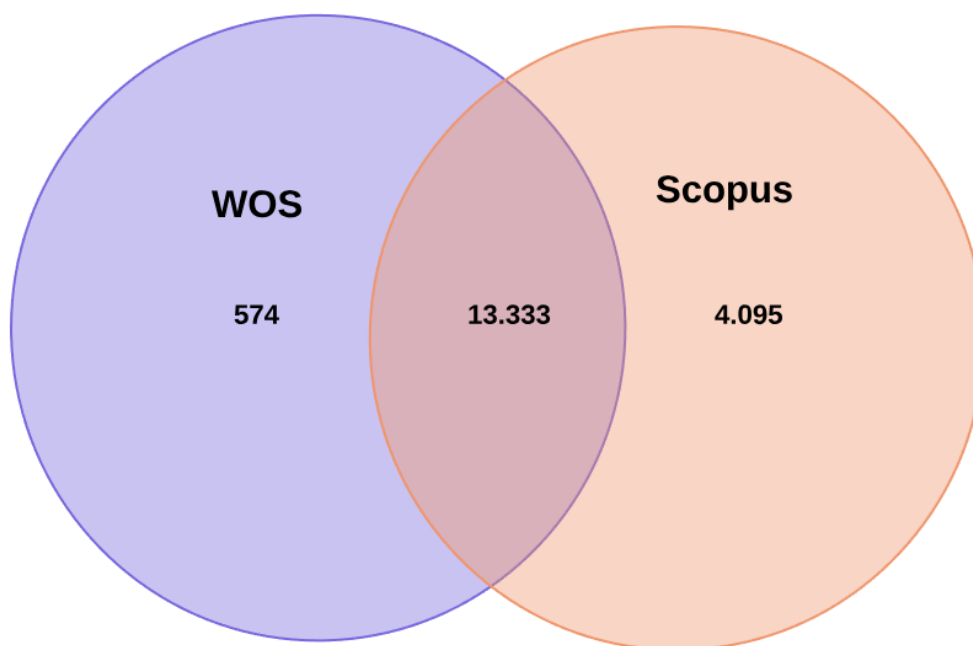
Figura 31 – Produção científica da UFSCar informada na Plataforma Lattes e indexada na WOS, Scopus e SciELO (1971-2017)



Fonte: Elaborada pelo autor.

Considerando que as experiências de povoamento automatizado relatadas a partir da literatura realizam a coleta de metadados individualmente para cada base de dados utilizada, foi realizada a distribuição da parcela da produção científica da UFSCar informada na Plataforma Lattes e indexada conforme cada fonte de informação. Essa distribuição teve como objetivo identificar a capacidade de recuperação da produção científica da UFSCar em cada fonte analisada. Como primeira etapa da análise foi realizada a distribuição dos registros informados na Plataforma Lattes conforme a indexação nas bases de dados WOS e Scopus, uma vez que são consideradas duas das principais fontes para a identificação, recuperação e extração de registros da produção científica de IES. A Figura 32 apresenta a distribuição de registros da produção científica da UFSCar passíveis de serem recuperados a partir da WOS e Scopus.

Figura 32 – Produção científica da UFSCar indexada na WOS e Scopus

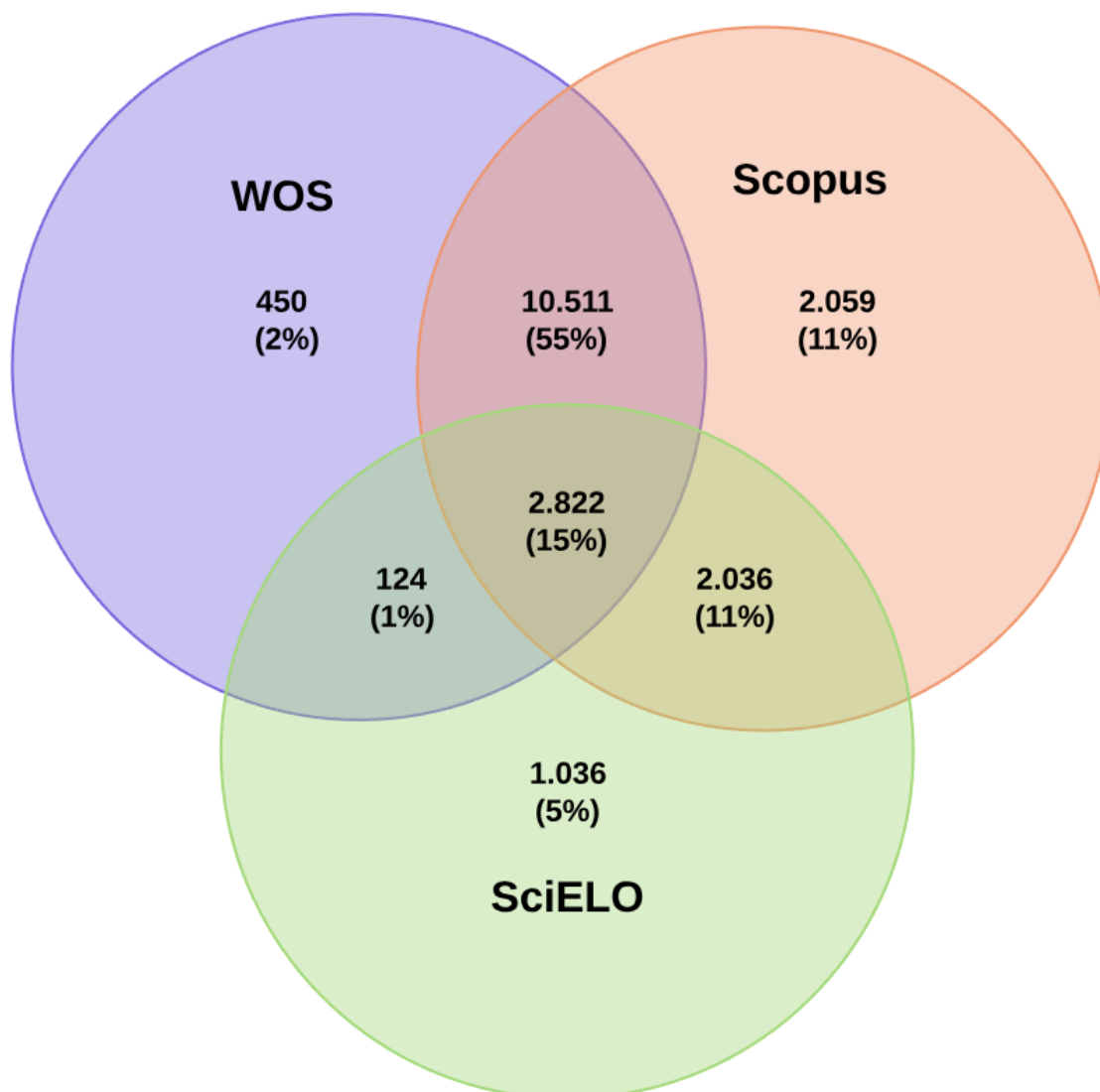


Fonte: Elaborada pelo autor.

A partir das duas bases de dados é possível recuperar 18.002 registros da produção científica da UFSCar indexada em bases de dados, 13.907 a partir da WOS e 17.428 a partir da Scopus. Cabe ressaltar que as duas fontes apresentam grande parcela de registros em comum indexados em seu conteúdo, sendo 13.333 registros passíveis de serem recuperados pelas duas bases de dados. Nota-se que a combinação entre as bases de dados traz impacto significativo quanto a obtenção de metadados da produção científica da IES.

A partir da Figura 33 é possível observar que a combinação entre Scopus e WOS possibilita a recuperação da grande maioria dos registros da UFSCar indexados. Dos 19.038 registros informados na Plataforma Lattes e também disponíveis nas bases de dados analisadas, 95% são passíveis de serem recuperados a partir da combinação entre a WOS e Scopus.

Figura 33 – Produção científica da UFSCar indexada na WOS, Scopus e SciELO



Fonte: Elaborada pelo autor.

Se analisadas de maneira isolada é possível identificar que a WOS é a base de dados com menor número de registros exclusivos e que, conseqüentemente, traz um menor impacto quanto a recuperação da produção científica da IES caso não utilizada como fonte de coleta de metadados. Essa baixa taxa de registros exclusivos identificada é reflexo do tamanho da coleção da Scopus, que possui um maior número de periódicos indexados e garante a cobertura de grande parcela de títulos indexados na WOS. Constam na coleção da Scopus um total de 36.294 títulos de periódicos indexados, 23.402 títulos ativos e 12.892 inativos (que não estão atualmente indexados na base de dados, mas possuem conteúdo disponível), além

dos títulos de periódicos a Scopus indexa também 32.780 títulos de *proceedings*. Ao realizar uma análise comparativa entre o conteúdo indexado na WOS e Scopus, a partir da comparação entre os títulos de periódicos indexados, dos 17.911 títulos apresentados na WOS uma parcela de 13.333 (74%) também está indexada na Scopus.

Se considerada a proporção entre o número de periódicos indexados e a disponibilidade de registros exclusivos há destaque para a SciELO, que apresenta 1.036 registros da produção científica da UFSCar passíveis de serem recuperados exclusivamente quando comparada com a cobertura da Scopus e WOS. Constam na coleção da SciELO 1.584 periódicos indexados, desses 195 títulos (10%) também estão indexados na WOS e 323 títulos (20%) na Scopus. Essa exclusividade de registros demonstra que, apesar do baixo número de itens indexados em relação a WOS e Scopus, a SciELO configura-se como uma importante fonte de coleta de metadados voltada para o povoamento de RI.

A exclusividade de registros recuperáveis a partir da SciELO pode estar ligada ao perfil dos periódicos indexados na base de dados, Packer (2009 apud PACKER, 2014, p. 23) aponta que “a abrangência dos periódicos latino-americanos é mais genérica, no que tange às disciplinas ou áreas temáticas, quando comparada ao universo dos periódicos internacionais [...] no que tange ao público-alvo, a maioria dos periódicos latino-americanos está centrada em comunidades de pesquisa nacionais e regionais”. Considerando a área de biomedicina Stegemann (2013) aponta que os periódicos oriundos da América Latina indexados na SciELO apresentam características particulares em relação aos periódicos internacionais, apontando que “[...] existe uma diferença substancial entre os clássicos e prestigiosos periódicos biomédicos internacionais firmemente estabelecidos”, principalmente no que tange seus aspectos gerenciais e financeiros” e que mesmo diante dessas características “[...] seus editores realizam grandes esforços para chegar o mais próximo possível de reconhecidos critérios internacionais de qualidade e conteúdo científico, incluindo o processo de arbitragem, estrutura e conteúdo, visibilidade e presença em bibliotecas, índices e bases de dados, “Fator de Impacto”, etc.”.

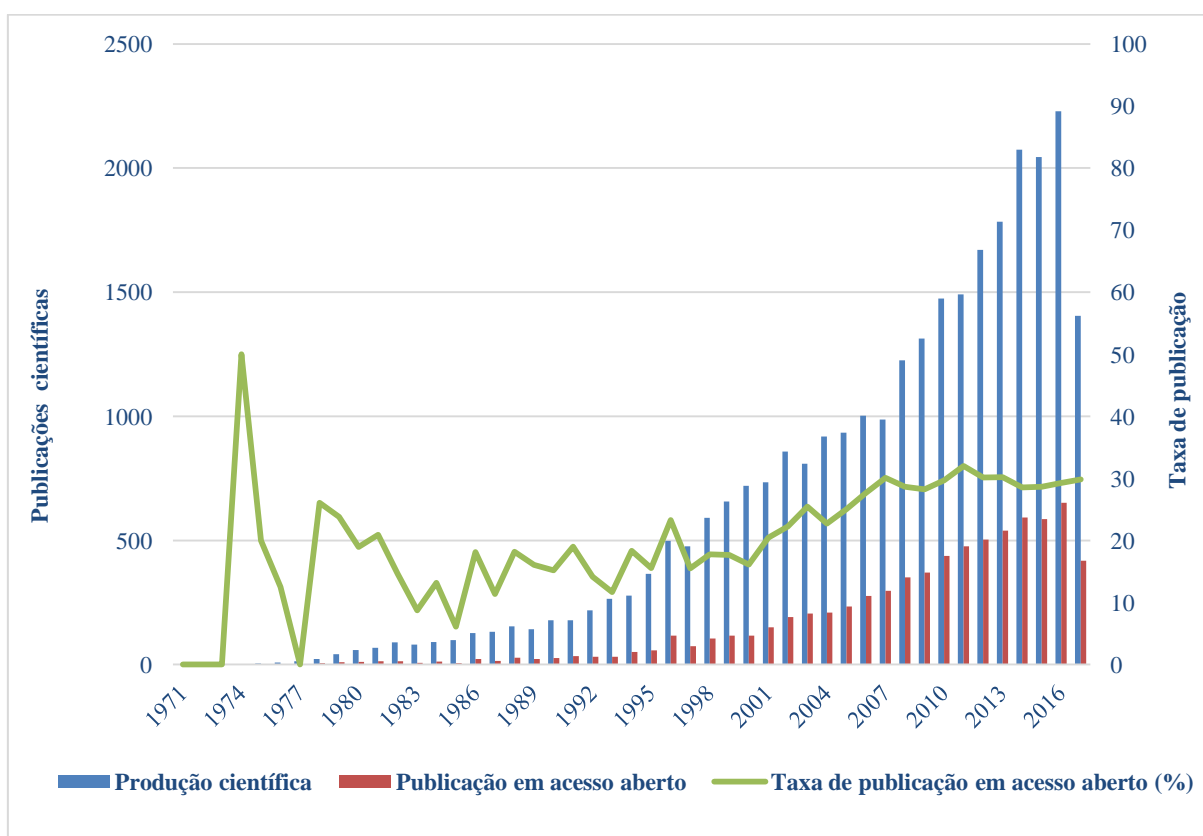
Além da cobertura, cabe ressaltar a importância de avaliar as opções de extração dos registros da produção científica recuperados, uma vez que as bases de dados podem apresentar barreiras para a obtenção dos metadados. Por exemplo, no relato de experiência da UNESP de povoamento de seu RI a partir de registros recuperados nas bases de dados WOS, Scopus e SciELO, é relatada a necessidade de aquisição dos registros identificados na Scopus, uma vez que a base de dados não dispunha do formato de arquivos necessário para a sistemática automatizada de povoamento utilizada pela IES (ASSUMPÇÃO et al., 2014).

4.3 Produção científica da UFSCar em acesso aberto e fator de embargo

Os tesouros de acesso aberto (DOAJ) e fator de embargo (SHERPA/ROMEO) foram aplicados na produção científica da UFSCar recuperada a partir da Plataforma Lattes com o objetivo de identificar a potencialidade dos registros obtidos quanto ao povoamento do RI.

A aplicação do tesouro de acesso aberto identificou uma parcela de 7.421 artigos da UFSCar publicados em fontes de acesso aberto, o que corresponde a 26% de toda a produção científica da IES identificada na Plataforma Lattes. Em relação a taxa de publicações por ano, nota-se um crescimento, porém não acentuado, no número de publicações em fontes de acesso aberto ao longo do período (Figura 34).

Figura 34 – Produção científica da UFSCar em fontes de acesso aberto

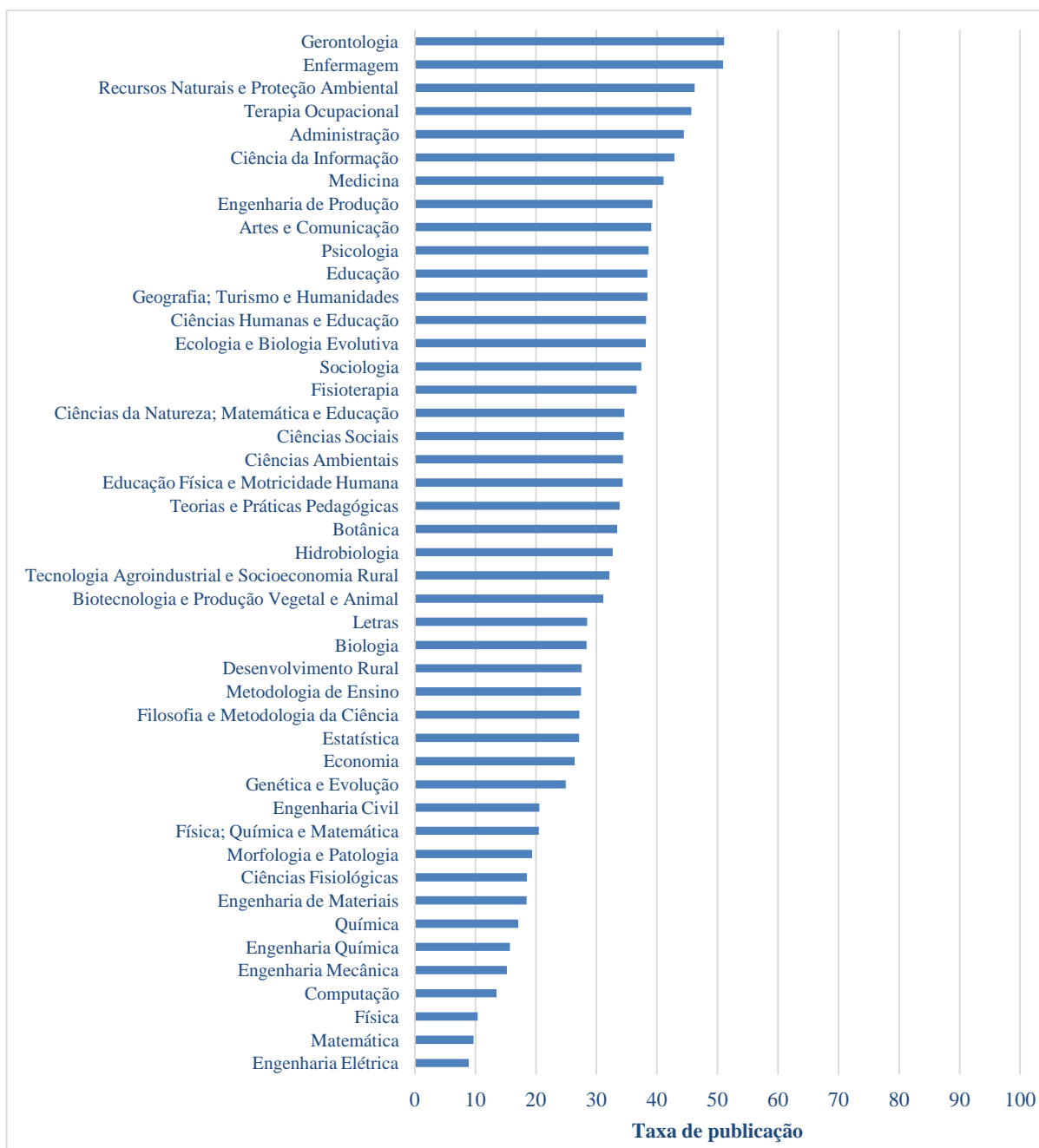


Fonte: Elaborado pelo autor.

A fim de identificar as práticas de publicação em acesso aberto foi realizada uma correlação a partir da estrutura organizacional da UFSCar entre a taxa de publicação em acesso aberto por departamentos (Figura 35). Nota-se que não há predominância da prática de publicação nesse tipo de fonte por áreas de conhecimento, uma vez que é demonstrada

heterogeneidade em relação as áreas de conhecimento relacionadas aos departamentos que apresentam maior taxa de publicações.

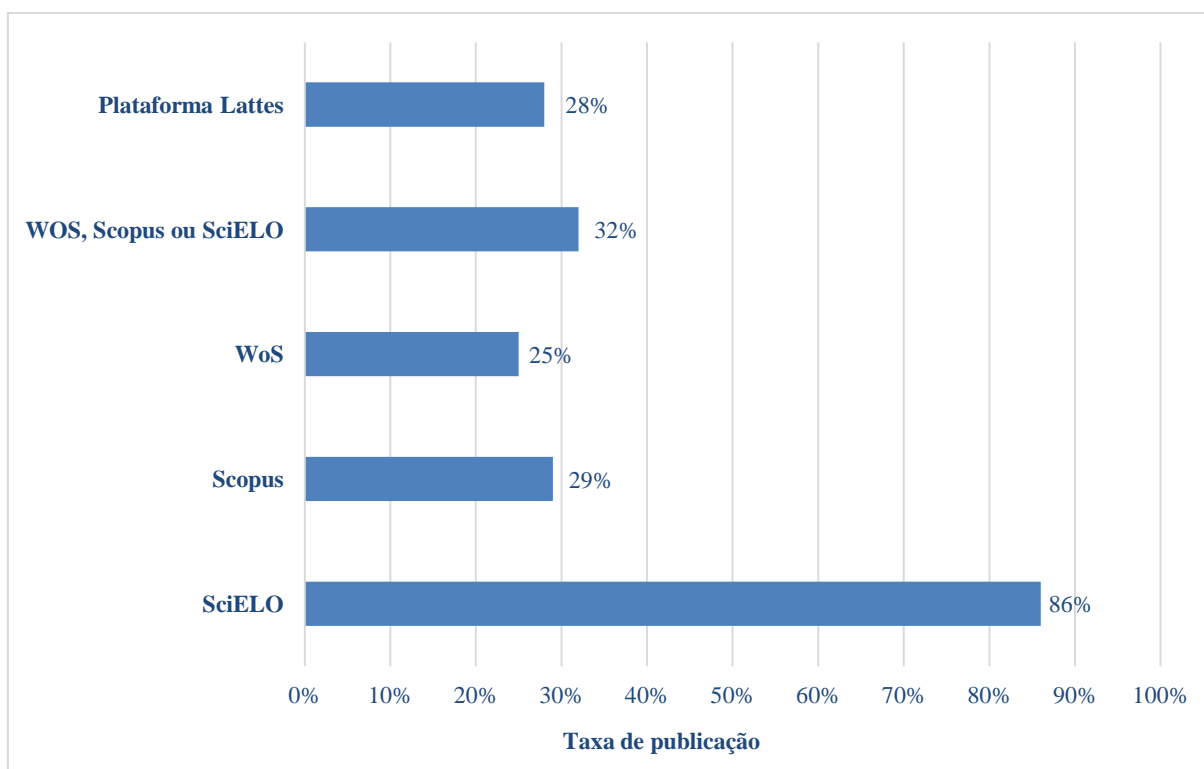
Figura 35 – Taxa de publicação da UFSCar em acesso aberto por departamentos



Fonte: Elaborado pelo autor.

Seguindo a análise foi identificada qual a parcela de publicação em acesso aberto segundo cada fonte de informação analisada. A Figura 36 apresenta a porcentagem de registros da UFSCar em acesso aberto disponíveis na Plataforma Lattes, WOS, Scopus e SciELO.

Figura 36 – Taxa de publicação da UFSCar em acesso aberto por fontes de informação

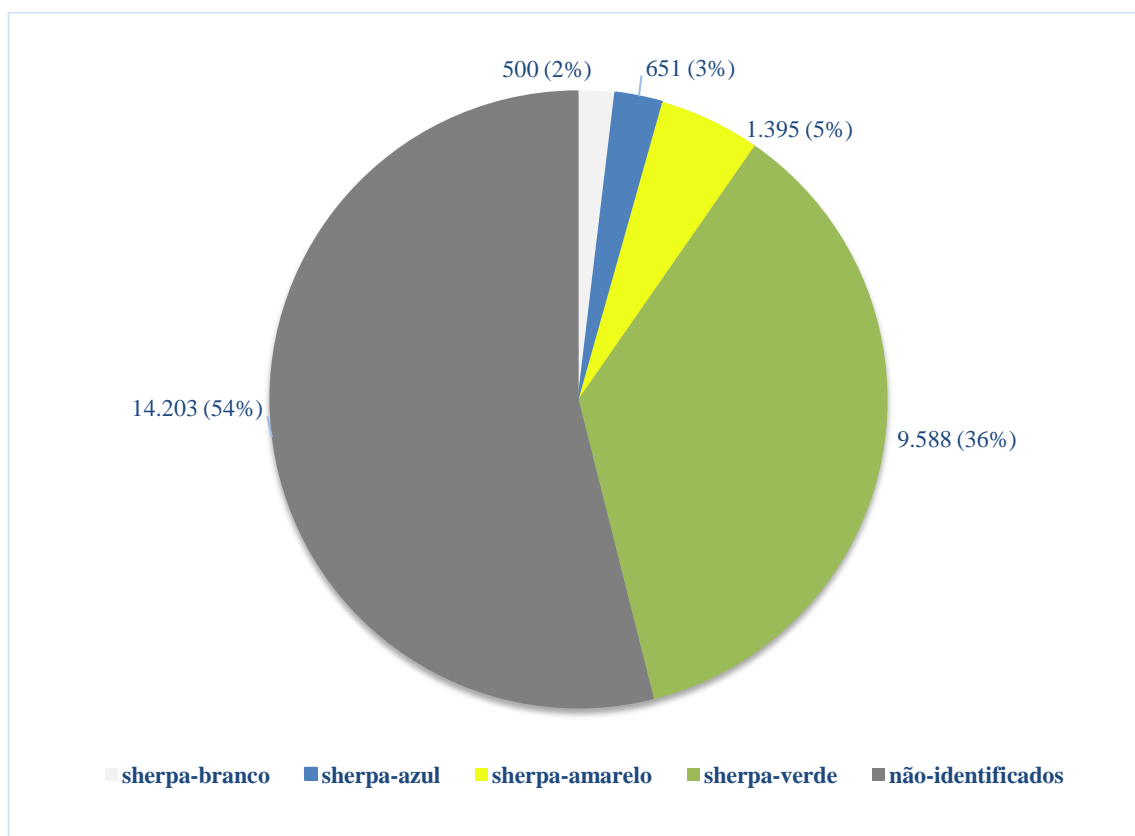


Fonte: Elaborado pelo autor.

Cabe ressaltar que no caso da SciELO a taxa de 86% dos registros da UFSCar indexados na base de dados corresponde a relação de periódicos também indexados na DOAJ, fonte de referência utilizada para a criação do tesauro, uma vez que a base de dados adota a licença *Creative Commons* que garante o acesso aberto para 100% do seu conteúdo.

Além da identificação das fontes de acesso aberto foi aplicado nos registros da produção científica da UFSCar o tesauro de fator de embargo. Para o povoamento de RI a consulta do fator de embargo das publicações é necessária até mesmo para os títulos em acesso aberto. O embargo irá determinar quais as permissões quanto ao tempo e versão da publicação que poderá ser incorporada ao RI. A Figura 37 apresenta a distribuição de fatores de embargo da produção científica da UFSCar segundo a classificação estabelecida pelo SHERPA/ROMEO.

Figura 37 – Fator de embargo da produção científica da UFSCar (SHERPA/ROMEO)



Fonte: Elaborado pelo autor.

Dos registros identificados pelo tesauro de fator de embargo:

- 1) **Sherpa azul:** 3% dos registros estão classificados como azul, o que significa que podem ser depositadas na sua versão *posprint*, que foi avaliada pelos pares e aprovada para a publicação. Em alguns casos, é liberada a versão final do PDF que foi diagramado para a publicação na revista.
- 2) **Sherpa amarelo:** 5% dos registros estão classificados como amarelo, sendo liberada para depósito a sua versão *preprint*, versão do texto enviada no momento da submissão sem que tenha recebido a avaliação pelos pares.
- 3) **Sherpa verde:** uma parcela de 36% está classificada como verde, sendo permitido nesse caso o depósito da versão *preprint* e *posprint* da publicação.
- 4) **Sherpa branco:** 2% estão classificados no como branco”, o que significa que os periódicos estão cadastrados no SHERPA/ROMEO, porém não apresentam em suas

políticas editoriais regras claras quanto a disponibilização de seu conteúdo em repositórios de acesso aberto.

Além da taxa de 2% dos periódicos classificados como branco, 54% dos periódicos identificados na produção científica da UFSCar não foram passíveis de identificação a partir do tesouro do SHERPA/ROMEO. Deste modo, estima-se a necessidade de consulta quanto ao embargo diretamente na política editorial da fonte de publicação de 14.703 registros da produção científica da UFSCar, o que demonstra os esforços necessários para a equipe de RI ou para os autores, no caso de autodepósito, para disponibilizarem a produção científica institucional.

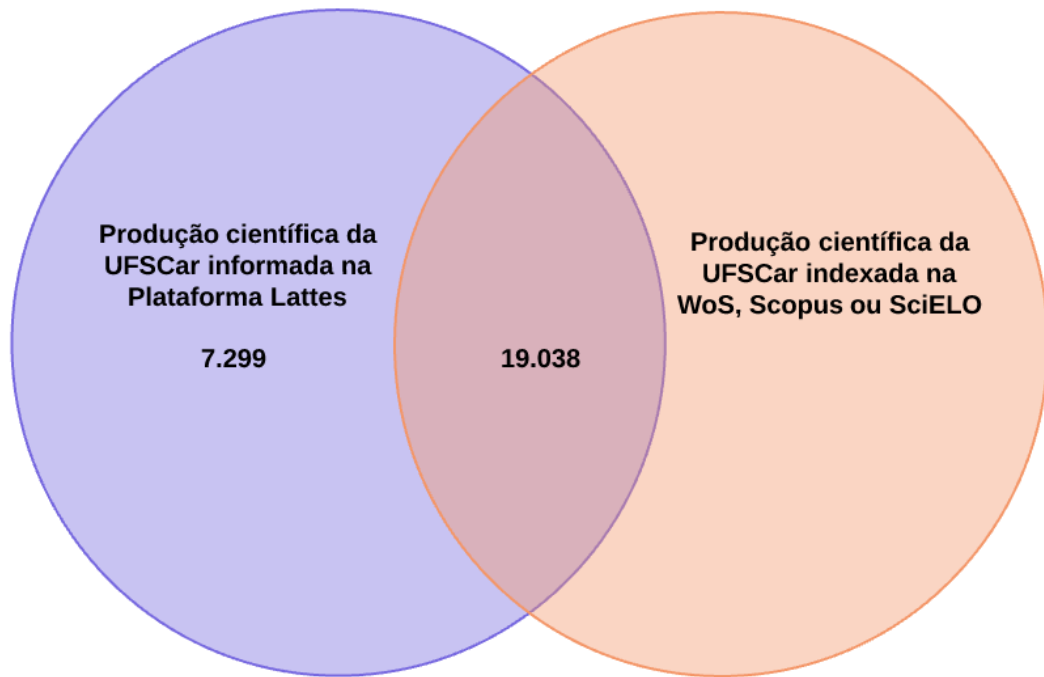
Como potenciais registros que poderiam ser inseridos no RI da IES, estima-se um número de 11.634 publicações, que estão distribuídas entre as classificações azul, amarelo e verde sendo prevista a possibilidade de depósito de ao menos uma de suas versões (*preprint* ou *posprint*).

4.4 Produção científica da UFSCar recuperada somente na Plataforma Lattes

A partir da comparação entre as fontes de informação foi identificado uma parcela de 7.299 registros da produção científica da UFSCar recuperáveis exclusivamente pela Plataforma Lattes, não estando indexados na WOS, Scopus ou SciELO. A Figura 38 apresenta a distribuição dos registros indexados e informados somente na Plataforma.

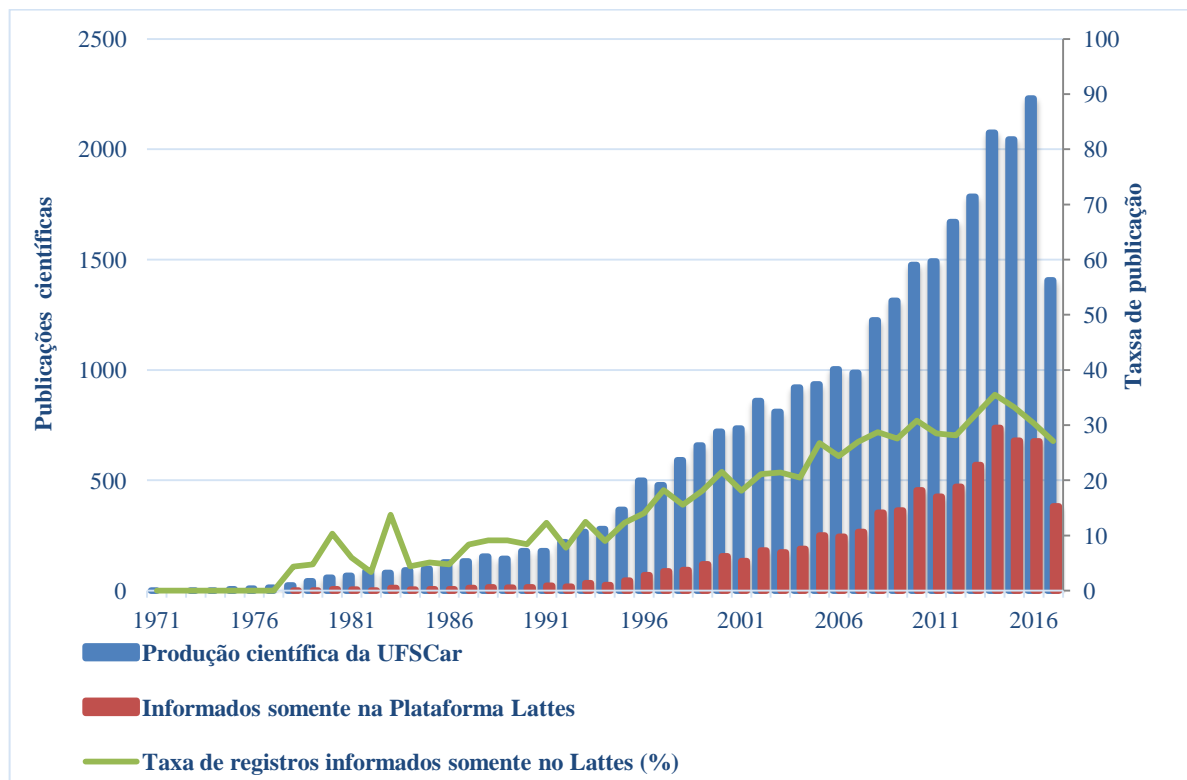
Buscando identificar as características dos artigos informados somente na Plataforma Lattes, os 7.299 registros foram separados em uma nova amostra no VantagePoint. Quando analisados os registros segundo o ano de publicação, é identificada a mesma tendência de crescimento em relação aos registros que também estão disponíveis nas bases de dados. A Figura 39 demonstra a taxa de publicação por ano para o total da produção científica da UFSCar, calculando a taxa de registros que estão informados somente na Plataforma Lattes e não indexados na WOS, Scopus ou SciELO.

Figura 38 – Produção científica da UFSCar informada somente na Plataforma Lattes e indexada



Fonte: Elaborada pelo autor.

Figura 39 – Produção científica da UFSCar informada somente na Plataforma Lattes



Fonte: Elaborada pelo autor.

Considerando que as bases de dados de informação científica possuem maior cobertura em determinadas áreas do conhecimento em detrimento de outras, foi considerada a possibilidade de que os artigos informados exclusivamente na Plataforma Lattes poderiam estar publicados em periódicos indexados em outras fontes de informação e/ou estar relacionados com áreas do conhecimento em que as três bases de dados apresentam baixa cobertura. Além desses fatores, é preciso considerar que nem todos os periódicos científicos estão indexados em bases de dados, o que pode justificar a presença de registros informados exclusivamente na Plataforma Lattes.

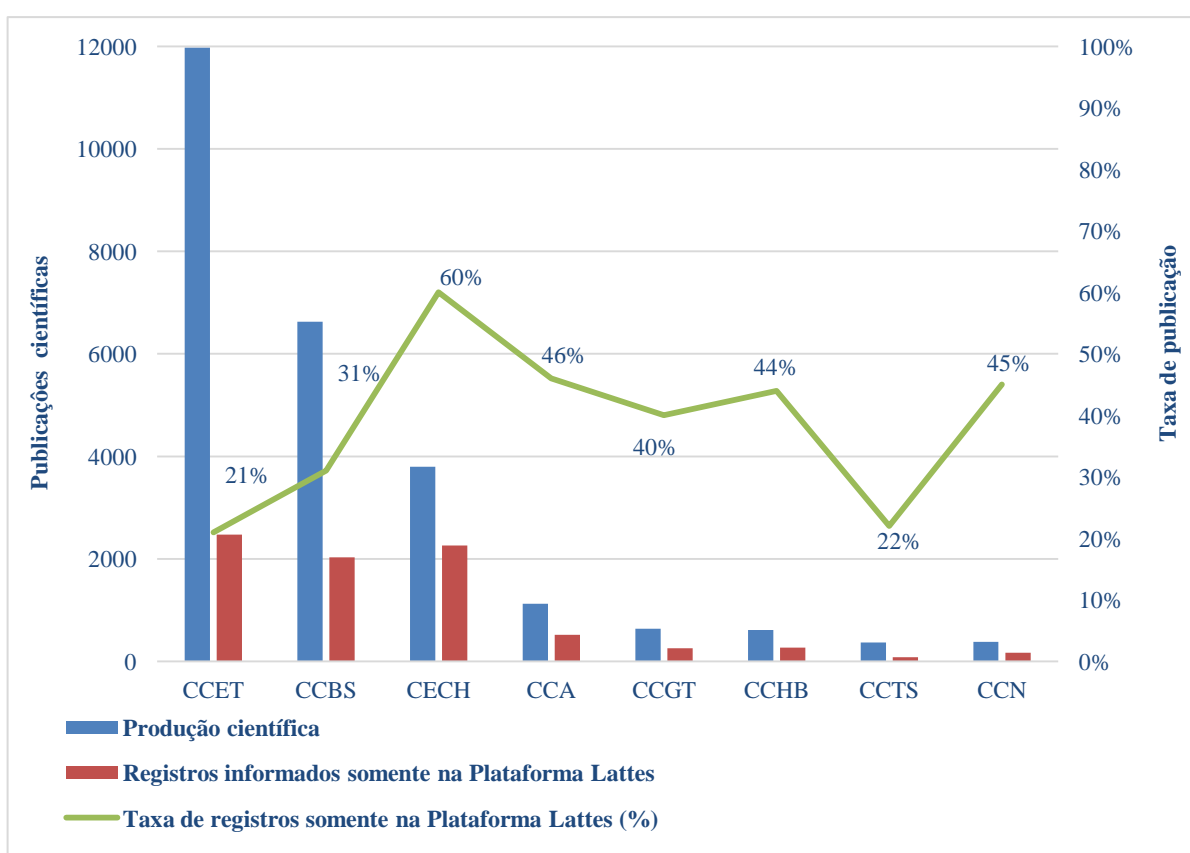
Uma distribuição dos registros da produção científica da UFSCar foi realizada segundo a estrutura organizacional da universidade, a fim de identificar possíveis relações entre a presença de registros informados exclusivamente na Plataforma Lattes com áreas do conhecimento. Para a distribuição foram aplicados tesouros a nível de campus, centros acadêmicos e programas de pós-graduação.

Com a aplicação do tesouro a nível de campus foi identificada a concentração de 80% dos registros informados somente na Plataforma Lattes para São Carlos, 10% para Sorocaba, 7% para Araras e 3% para Lagoa do Sino. A aplicação do tesouro de centros possibilitou a análise da distribuição de suas produções científicas conforme a cobertura nas bases de dados e exclusivamente na Plataforma Lattes (Figura 40).

Quando analisada a taxa de cobertura da produção científica dos centros acadêmicos, conforme as fontes de informação analisadas, há destaque para a cobertura da produção científica do Centro de Educação e Ciências Humanas (CECH), que apresenta um índice elevado de publicação e uma taxa de 60% de registros informados exclusivamente na Plataforma Lattes. Essa cobertura demonstra a contribuição que a Plataforma Lattes pode trazer para a recuperação de registros relacionados à área das ciências humanas, e o consequente impacto na capacidade de cobertura do povoamento do RI. Além do CECH, os centros acadêmicos que apresentam um índice menor de produção científica como o Centro de Ciências Agrárias (CCA), Centro de Ciências em Gestão e Tecnologia (CCGT), Centro de Ciências Humanas e Biológicas (CCHB) e Centro de Ciências da Natureza (CCN), apresentam uma média de 44% de sua produção científica recuperáveis somente a partir da Plataforma Lattes. No caso do Centro de Ciências Exatas e de Tecnologia (CCET) e Centro de Ciências Biológicas e da Saúde (CCBS), apesar de apresentarem menor cobertura de registros informados exclusivamente na Plataforma Lattes, essa taxa passa a ser significativa quando considerada o número de publicações que representam. No caso do CCET estima-se que 2.474 deixariam de ser inseridos no RI caso não fossem coletados os metadados da Plataforma

Lattes quando consideradas as bases WOS, Scopus e SciELO como fonte para coleta. Para o CCBS, estima-se a não inclusão de 2.028 registros no povoamento do RI. As taxas de recuperação demonstram que a Plataforma Lattes impacta diretamente na cobertura da produção científica de todos os centros acadêmicos da UFSCar. Deste modo a Plataforma se apresenta como uma alternativa para a recuperação de registros da produção científica da IES, principalmente para áreas das humanidades, na qual teve destaque de recuperação a partir dos registros obtidos para o CECH.

Figura 40 – Produção científica da UFSCar informada somente na Plataforma Lattes por centros acadêmicos⁷



Fonte: Elaborado pelo autor.

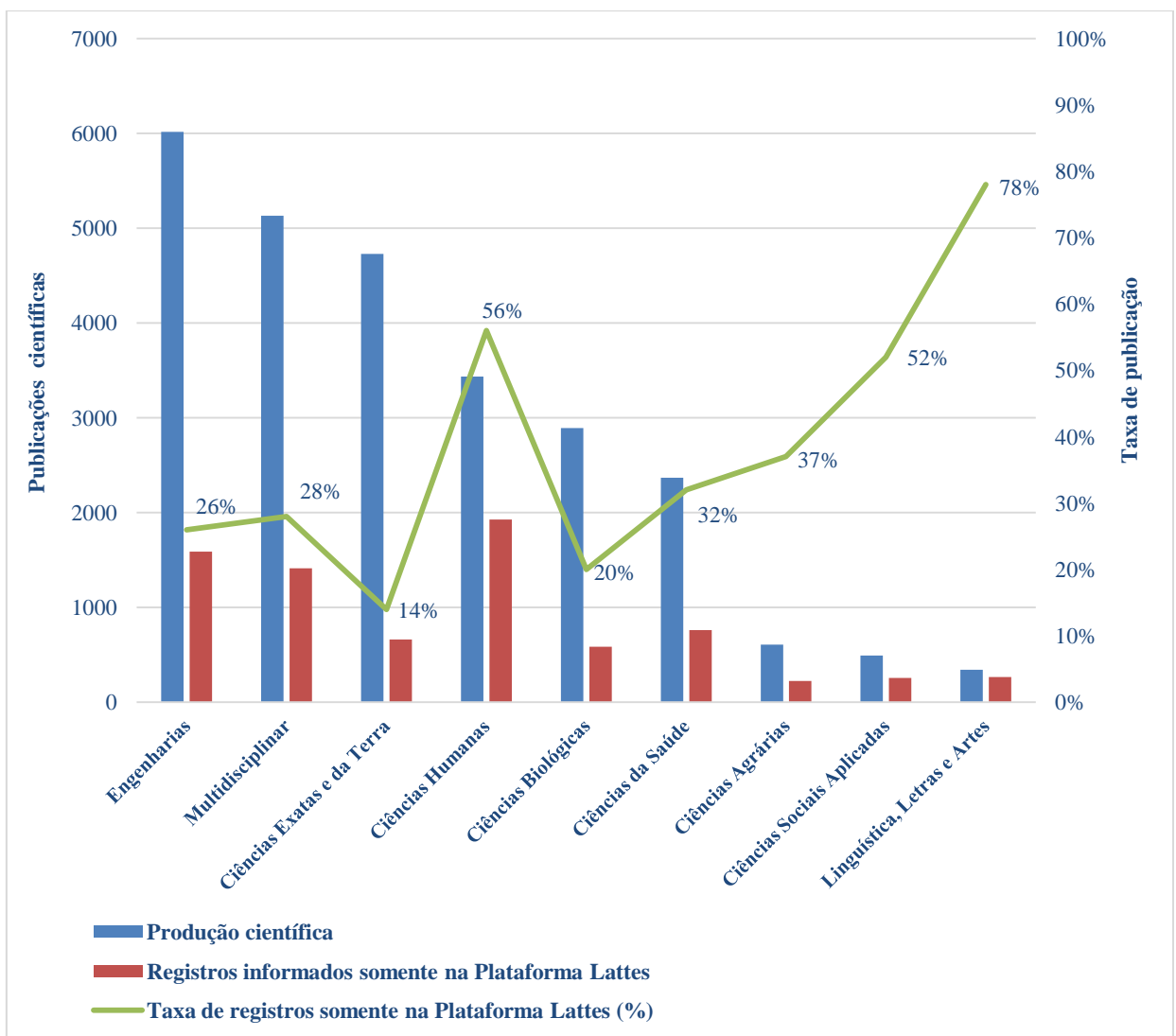
Complementando a análise da correlação entre áreas do conhecimento com a cobertura da produção científica a partir da Plataforma Lattes, uma distribuição foi realizada na parcela de registros da produção científica da UFSCar vinculada aos seus programas de pós-

⁷ Centro de Ciências Exatas e de Tecnologia – CCET; Centro de Ciências Biológicas e da Saúde – CCBS; Centro de Educação e Ciências Humanas – CECH; Centro de Ciências Agrárias – CCA; Centro de Ciências em Gestão e Tecnologia – CCGT; Centro de Ciências Humanas e Biológicas – CCHB; Centro de Ciências e Tecnologias para a Sustentabilidade – CCTS; Centro de Ciências da Natureza - CCN

graduação. Após a distribuição entre produção científica e programas de pós-graduação foi realizada a classificação dos programas segundo a Tabela de Áreas de Conhecimento/Avaliação da CAPES (BRASIL, 2018). A Figura 41 apresenta a distribuição da produção científica da UFSCar segundo a classificação dos programas de pós-graduação a qual está vinculada.

A distribuição ressalta que a obtenção de registros a partir da Plataforma Lattes impacta na cobertura de todas as áreas do conhecimento identificadas, tendo destaque para a recuperação da informação nas áreas relacionadas as "Ciências Humanas", "Ciências Sociais Aplicadas" e "Linguística, Letras e Artes", com um aumento respectivo na taxa de registros recuperados a partir da Plataforma de 56%, 52% e 78%.

Figura 41 – Produção científica da UFSCar informada somente na Plataforma Lattes por áreas do conhecimento (CAPES)



Fonte: Elaborado pelo autor.

5 CONSIDERAÇÕES FINAIS

A partir dos resultados obtidos é possível afirmar que a Plataforma Lattes se configura como uma alternativa de fonte de coleta de metadados voltados para o povoamento automatizado de RI quando considerada a prática de recuperação da produção científica de IES a partir de bases de dados de informação científica e tecnológica (BDICTs).

Os resultados da pesquisa apontam que a recuperação de registros da UFSCar a partir da Plataforma Lattes trouxe resultados significativos quanto a cobertura de sua produção científica. Quando analisada a distribuição dessa produção conforme sua origem, é notado o impacto que a Plataforma Lattes traz para todos os centros acadêmicos e suas respectivas áreas de conhecimento, tendo como destaque o aumento na recuperação de registros para as áreas das humanidades.

A Plataforma Lattes demonstra potencial para ser a principal fonte de coleta de metadados de IES, uma vez que não apresenta delimitações quanto cobertura ou tipo de publicação informados em seu conteúdo. Essa característica possui vantagem quando comparada com a recuperação de registros da produção científica de IES a partir de BDICTs, além disso demonstra-se como essencial para a capacidade de representação da totalidade da produção científica da IES no povoamento de RI. A possibilidade de utilizar a Plataforma Lattes como fonte principal na coleta de metadados da produção científica significa também a redução de esforços para o desenvolvimento de sistemáticas automatizadas de povoamento, uma vez que para cada base de dados selecionada como fonte de coleta torna-se necessária a adequação da sistemática para as possibilidades de recuperação e extração de metadados fornecidas pela base.

Um dos desafios para o trabalho com a Plataforma Lattes é a garantia de seu preenchimento e atualização por parte dos pesquisadores, desse modo é recomendável que as IES instituem mecanismos voltados para a garantia do preenchimento de Currículos Lattes. Nesse sentido, destaca-se o uso do SyncLattes como ferramenta para a extração de metadados, uma vez que dispõe entre suas funcionalidades da remoção de duplicidades e controle de autoridade que pode associar uma publicação a um docente mesmo que este não tenha preenchido o registro em seu currículo, mas que os metadados estejam descritos no currículo de um dos coautores da publicação. Essas funcionalidades eliminam, em parte, a problemática de preenchimento e atualização dos currículos da Plataforma Lattes.

Entre as limitações e a perspectiva de desenvolvimentos futuros destaca-se o fato de que a pesquisa considerou a totalidade da produção científica da UFSCar com aquela registrada na Plataforma Lattes, não sendo realizada a recuperação da produção científica da IES a nível de registros nas bases de dados analisadas (WOS, Scopus e SciELO). Essa recuperação a partir da BDICTs poderia demonstrar a existência de registros indexados, porém não informados na Plataforma.

Cabe ressaltar que a capacidade de recuperação de metadados a partir da Plataforma Lattes contribui para uma das problemáticas envolvidas no povoamento de RI, a da dificuldade para identificar a totalidade da produção científica de IES e recuperar seus metadados. Porém, além dos metadados torna-se necessária para completar o processo de povoamento a recuperação e disponibilização do texto completo das publicações na interface do RI.

Conclui-se que a Plataforma Lattes é uma alternativa para a coleta de metadados da produção científica de IES e pode contribuir para o desenvolvimento de sistemáticas automatizadas de povoamento de RI. Além disso, destaca-se o uso potencial da Plataforma Lattes em diversas ações estratégicas para a atuação da IES, como por exemplo, para a elaboração de indicadores institucionais, maximizando a visibilidade da produção científica e a melhoria na avaliação de *rankings* universitários, no apoio em processos internos como o de progressão de carreira, no mapeamento dos pontos fortes e identificação dos avanços necessários para as atividades de pesquisa e pós-graduação.

REFERÊNCIAS

ALMEIDA, E. C. E.; GUIMARÃES, J. A.; ALVES, I. T. G. Dez anos do Portal de Periódicos da Capes: histórico, evolução e utilização. **Revista Brasileira de Pós-Graduação**, v. 7, n. 13, p. 218-246, 2011. Disponível em: <<http://ojs.rbpg.capes.gov.br/index.php/rbpg/article/view/194>>. Acesso em: 25 jul. 2017.

AMARAL, R. M. et al. Panorama da inteligência competitiva no Brasil: os pesquisadores e a produção científica na plataforma Lattes. **Perspectivas em Ciência da Informação**, v. 21, n. 4, p. 97-120, 2016. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362016000400097&lng=pt&tlng=pt>. Acesso em: 08 dez. 2017.

APPOLINÁRIO, F. **Metodologia da ciência: filosofia e prática da pesquisa**. São Paulo: Cengage Learning, 2006. Disponível em: <link.galegroup.com/apps/doc/CX2137900016/GVRL?u=capes&sid=GVRL&xid=32bc6116>. Acesso em: 28 dez. 2017.

ARAÚJO, V. M. A. P. Sistemas de recuperação da informação: uma discussão a partir de parâmetros enunciativos. **Transinformação**, v. 24, n. 2, p. 77-90, 2012. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/771>>. Acesso em: 24 jul. 2017.

ASSIS MEDEIROS, S.; FERREIRA, P. A. Política pública de acesso aberto à produção científica: um estudo sobre a implementação de repositórios institucionais em instituições de ensino superior. **Perspectivas em Gestão & Conhecimento**, v. 4, n. 2, p. 195-217, 2014. Disponível em: <<http://periodicos.ufpb.br/index.php/pgc/article/view/16852>>. Acesso em: 18 jul. 2017.

ASSUMPÇÃO, F. S. et al. A conversão de registros na implantação de repositórios institucionais: o caso do repositório institucional UNESP. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS (SNBU), 18., 2014. **Anais...** Belo Horizonte: UFMG, 2014.

AUTRAN, M. M. M.; BORGES, M. M. Comunicação da ciência: (r)evolução ou crise? **RECIIS - Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 8, n. 2, p. 122-138, 2014. Disponível em: <<http://hdl.handle.net/10316/26596>>. Acesso em: 17 jul. 2017.

BASSOLI, M. **Avaliação do Currículo Lattes como fonte de informação para construção de indicadores: o caso da UFSCar**. 2017. 132 f. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) - Universidade Federal de São Carlos, São Carlos, 2017.

BATISTA, A. A. A Falar nos entendemos – a interoperabilidade entre repositórios digitais. In: SAYÃO, L. et al. **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 71-90.

BRASIL. Conselho Nacional de Desenvolvimento Científico e Tecnológico. **Sobre a Plataforma Lattes**. Brasília, DF: CNPQ, 2017a. Disponível em: <<http://lattes.cnpq.br/>>. Acesso em: 26 jul. 2017.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Portaria nº 013, de 15 de fevereiro de 2006**. Institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos. Brasília, DF: CAPES, 2006. Disponível em: <http://repositorio.unb.br/documentos/Portaria_N13_CAPES.pdf>. Acesso em: 26 jul. 2017.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Qualis**. Brasília, DF: CAPES, 2009. Disponível em: <<http://www.capes.gov.br/sala-de-imprensa/noticias/2550-capes-aprova-a-nova-classificacao-do-qualis>>. Acesso em: 17 jul. 2017.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Portal de Periódicos CAPES**: perguntas frequentes. Brasília, DF: CAPES, 2015. Disponível em: <http://www-periodicos-capes-gov-br.ez31.periodicos.capes.gov.br/?option=com_pfaq>. Acesso em: 26 jul. 2017.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Plataforma Sucupira**. Cursos Avaliados e Reconhecidos. CAPES, 2016. Disponível em: <<https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/programa/quantitativos/quantitativoPrograma.jsf?areaAvaliacao=0&cdRegiao=3&sgUf=SP&ies=338687>>. Acesso em: 03 dez. 2017.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Portal de Periódicos**: buscar base. 2017c. Disponível em: <<http://www-periodicos-capes-gov-br.ez31.periodicos.capes.gov.br/index.php>>. Acesso em: 26 dez. 2017.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Tabela de Áreas de Conhecimento/Avaliação da CAPES**. Brasília, DF: CAPES, 2018. Disponível em: <<http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/tabela-de-areas-do-conhecimento-avaliacao>>. Acesso em: 18 fev. 2018.

BRASIL. Congresso Nacional. Senado Federal. **Projeto de Lei do Senado nº 387, de 2011**. Brasília-DF, 2017b. Disponível em: <<http://www25.senado.leg.br/web/atividade/materias/-/materia/101006>>. Acesso em: 25 jul. 2017.

BRITO, A. G. C. DE; QUONIAM, L.; MENA-CHALCO, J. P. Exploração da Plataforma Lattes por assunto: proposta de metodologia. **Transinformação**, v. 28, n. 1, p. 77-86, 2016. Disponível em: <<https://seer.sis.puc-campinas.edu.br/seer/index.php/transinfo/article/view/2519>>. Acesso em: 25 jul. 2017.

CARVALHO, P. C; CARVALHO, R. A. Construção de políticas para repositórios institucionais: análise da ferramenta do OpenDOAR. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, v. 26, n. 2, p. 105-138, 2012. Disponível em: <http://www.brapci.inf.br/_repositorio/2014/05/pdf_5f13dd6a85_0024352.pdf>. Acesso em: 13 jul. 2016.

CLARIVATE ANALYTICS. **Principal Coleção do Web of Science**. 2017. Disponível em: <https://images.webofknowledge.com/WOKRS524B8/help/pt_BR/WOS/hp_search.html>. Acesso em: 17 jul. 2017.

COLEPICOLO, E. Análise bibliométrica sobre a produção científica no campo das habilidades sociais. **Biblos: Journal of Librarianship and Information Science**, v. 57, n. 57, p. 12-21, 2015. Disponível em: <<http://biblios.pitt.edu/ojs/index.php/biblios/article/view/191>>. Acesso em: 24 jul. 2017.

COSTA, M. P.; LEITE, F. C. L. Repositórios institucionais de acesso aberto à informação científica: proposta de modelo de avaliação. **RECIIS – Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, v. 9, n. 3, p. 1-20, 2015. Disponível em: <<https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/996>>. Acesso em: 20 set. 2017.

COSTAS, R. Discussões gerais sobre as características mais relevantes de infraestruturas de pesquisa para a cientometria. In: MUGNANI, R.; FUJINO, A.; YUMIKO, K. (Org.). **Bibliometria e cientometria no Brasil: infraestrutura para avaliação da pesquisa científica na era do Big Data**. São Paulo: ECA/USP, 2017. p. 19-42.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Brasília, DF: Briquet de Lemos, 2008.

DIRECTORY OF OPEN ACCESS REPOSITORIES. **Search or Browse for Repositories**. OPENDOAR, 2017a. Disponível em: <<http://www.opendoar.org/find.php>>. Acesso em: 28 dez. 2017.

DIRECTORY OF OPEN ACCESS JOURNALS. **About DOAJ**. DOAJ, 2017b. Disponível em: <<https://doaj.org/about>>. Acesso em: 20 dez. 2017.

ELSEVIER. **Scopus content coverage guide**. Elsevier, 2016. Disponível em: <https://www.elsevier.com/__data/assets/pdf_file/0007/69451/scopus_content_coverage_guide.pdf>. Acesso em: 19 jul. 2017.

FARIA, L. I. L. **Prospecção tecnológica em materiais: aumento da eficiência do tratamento bibliométrico. Aplicação na análise de tratamentos de superfície resistentes ao desgaste**. 2001. 213 f. Tese (Doutorado em Ciência e Engenharia de Materiais) - Universidade Federal de São Carlos, São Carlos, 2001.

FERREIRA, I. **Ferramentas facilitam prestação de contas e avaliação da produção científica na USP**. São Paulo: USP, 2015. Disponível em: <<http://www5.usp.br/101872/ferramentas-facilitam-prestacao-de-contas-e-avaliacao-da-producao-cientifica-na-usp/>>. Acesso em: 27 dez. 2017.

FOLHA DE SÃO PAULO. **RUF: Ranking das Universidades do Brasil**. São Paulo: Folha de São Paulo, 2016. Disponível em: <<http://ruf.folha.uol.com.br/2016/>>. Acesso em: 15 jul. 2017.

FRANCO, I. M. **Repositórios digitais institucionais: um viés pela lei de acesso à informação**. 2015. 140 f. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) - Universidade Federal de São Carlos, São Carlos, 2015.

FURNIVAL, A. C.; GUIRRA, D. A. R. As percepções e práticas de publicação em acesso aberto dos pesquisadores de dois programas de pós-graduação em engenharia. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 15, n. 2, p. 469, 2017. Disponível em: <<http://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646398>>. Acesso em: 26 jul. 2017.

FURNIVAL, A. C.; HUBBARD, B. Acesso Aberto às Publicações Científicas: vantagens, políticas e advocacy. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 7, n. 1, p. 160-177, 2013. Disponível em: <<http://www.revistas.usp.br/incid/article/view/42358>>. Acesso em: 11 jun. 2016.

GABRIEL JUNIOR, R. F. Metodologia para uso do ISSN-L como ligação entre bases de dados qualificadoras. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016, São Paulo. **Anais...** São Paulo: USP, 2016.

GARCIA, P. A. B.; SUNYE, M. S. O protocolo OAI-PMH para interoperabilidade em bibliotecas digitais. In: CONGRESSO DE DADOS E METADADOS DO CONE SUL - CONGED, 1., 2003. **Anais...** Ponta Grossa: UEPG, 2003.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GINGRAS, Y. **Os desafios da avaliação da pesquisa: o bom uso da bibliometria**. Rio de Janeiro: EDUF RJ, 2016.

GOMES, M. J.; ROSA, F. **Repositórios institucionais: democratizando o acesso ao conhecimento**. Salvador: EDUFBA, 2010.

HADDAD, E. A.; MENA-CHALCO, J. P.; SIDONE, O. Produção científica e redes de colaboração dos docentes vinculados aos programas de pós-graduação em Economia no Brasil. **Estudos Econômicos (São Paulo)**, v. 47, n. 4, p. 617-679, 2017. Disponível em: <<http://dx.doi.org/10.1590/0101-416147414ejo>>. Acesso em: 05 dez. 2017.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Repositórios Digitais**. 2012a. Disponível em: <<http://www.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao/repositorios-digitais>>. Acesso em: 28 dez. 2017.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Centro Brasileiro do ISSN**. Brasília, DF: IBICT/MCTI, 2012b.

INTERNATIONAL STANDARD SERIAL NUMBER INTERNACIONAL CENTRE. **Understanding the ISSN**. Paris: ISSN International Centre, 2017. Disponível em: <<http://www.issn.org/understanding-the-issn/what-is-an-issn/>>. Acesso em: 16 jul. 2017.

JUNGMANN, D. M., BONETTI, E. A. **Inovação e propriedade intelectual: guia para o docente**. Brasília, DF: SENAI, 2010. Disponível em: <http://www.inpi.gov.br/sobre/arquivos/guia_docente_iel-senai-e-inpi.pdf>. Acesso em: dez. 2017.

KURAMOTO, H. Informação científica: proposta de um novo modelo para o Brasil. **Ciência da Informação**, v. 35, n. 2, p. 91-102, 2006. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652006000200010&lng=pt&nrm=iso&tlng=pt>. Acesso em: 13 jul. 2016.

KURAMOTO, H. Repositórios institucionais: políticas e mandatos. In: SAYÃO, L. et al. **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA 2009. p. 203-218.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. Brasília: IBICT, 2009.

LEITE, F. C. L. et al. **Boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: IBICT, 2012.

LETA, J.; DOS SANTOS COSTA, E. H.; MENA-CHALCO, J. P. Artigos em periódicos de acesso aberto: um estudo com pesquisadores bolsistas de produtividade do CNPq. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, v. 11, sup., 2017. Disponível em: <<https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1408>>. Acesso em: 01 dez. 2017.

MATIAS, M. S. O. **Base referencial para o povoamento de repositórios institucionais**: coleta automatizada de metadados da Plataforma Lattes. 2015. 86 f. Dissertação (Mestrado em Gestão de Organizações e Sistemas Públicos) - Universidade Federal de São Carlos, São Carlos, 2015.

MATIAS, M. S. O. et al. Base referencial para o povoamento de repositórios institucionais: coleta automatizada de metadados da Plataforma Lattes: In: CONFERÊNCIA LUSO-BRASILEIRA SOBRE ACESSO ABERTO (CONFOA), 6., 2015. **Anais...** Salvador: UFBA, 2016.

MATIAS, M. S. O.; AMARAL, R. M.; MATIAS, P. Proxy customizado para acesso ao web service da Plataforma Lattes. In: WORKSHOP DE TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÃO DAS INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR (WTICIFES), 11., 2017. **Anais...** Recife: 2017.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, v. 15, n. 4, p. 31-39, 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-65002009000400004>. Acesso em: 10 dez. 2017.

MENA-CHALCO, J. P.; DIGIAMPIETRI, L.; OLIVEIRA, L. Perfil de produção acadêmica dos programas brasileiros de pós-graduação em Ciência da Computação nos triênios 2004-2006 e 2007-2009. **Em Questão**, v. 18, n. especial, p. 215-229, 2013. Disponível em: <<http://seer.ufrgs.br/index.php/EmQuestao/article/view/33289>>. Acesso em: 27 dez. 2017.

MENA-CHALCO, J. P. et al. Brazilian bibliometric coauthorship networks. **Journal of the Association for Information Science and Technology**, v. 65, n. 7, p. 1424-1445, 2014. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.23010/full>>. Acesso em: 10 dez. 2017.

MORENO, F. P.; LEITE, F. C. L.; ARELLANO, M. A. M. Acesso livre a publicações e repositórios digitais em ciência da informação no Brasil. **Perspectivas em Ciência da Informação**, v. 11, n. 1, p. 82-94, 2006. Disponível em: <http://www.scielo.br/pdf/pci/v11n1/v11n1_a07.pdf>. Acesso em: 10 dez. 2017.

MOTTA, L. J. et al. Funded projects on dengue fever in brazil: use of scriptlattes tool to support the planning of public health research. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 22, n. 50, p. 114-127, 2017. Disponível em: <<http://www.brapci.inf.br/v/a/26922>>. Acesso em: 23 mar. 2018.

MUGNANI, R.; FUJINO, A.; YUMIKO, K. Apresentação. In: MUGNANI, R.; FUJINO, A.; YUMIKO, K. (Org.). **Bibliometria e cientometria no Brasil: infraestrutura para avaliação da pesquisa científica na era do Big Data**. São Paulo: ECA/USP, 2017. p. 07-13.

NÚCLEO DE INFORMAÇÃO TECNOLÓGICA EM MATERIAIS. **Institucional**. São Carlos: NIT/MATERIAIS, 2011. Disponível em: <<http://pesquisa.nit.ufscar.br/portal/index.php/institucional>>. Acesso em: 13 fev. 2018.

PACKER, A. L. (Org.). **SciELO – 15 Anos de acesso aberto: um estudo analítico sobre acesso aberto e comunicação científica**. Paris: UNESCO, 2014.

QUACQUARELLI SYMONDS. **QS World University Rankings**. Quacquarelli Symonds 2017. Disponível em: <<https://www.topuniversities.com/qs-world-university-rankings/methodology>>. Acesso em: 15 jul. 2017.

RANKING WEB OF UNIVERSITIES. **About Us**. Madrid: Cybermetrics Lab, 2017. Disponível em: <http://www.webometrics.info/en/About_Us>. Acesso em: 15 jul. 2017.

SARVO, D. O.; AMARAL, R. M. Roteiro de implementação de repositórios institucionais em instituições de ensino superior. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS (SNBU), 18., 2014. **Anais...** Belo Horizonte: UFMG, 2014.

SAYÃO, L.; MARCONDES, C. H. Introdução: repositórios institucionais de acesso livre. In: SAYÃO, L. et al. **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA, 2009. p. 9-22.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE. **Crêterios, política e procedimentos para a admissãe e a permanência de periódicos científicos na Coleção SciELO Brasil**. São Paulo: SciELO, 2014. Disponível em: <http://www.scielo.br/avaliacao/20141003NovosCriterios_SciELO_Brasil.pdf>. Acesso em: 19 jul. 2017.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE. **SciELO Analytics**. São Paulo: SciELO, 2017. Disponível em: <<https://analytics.scielo.org>>. Acesso em: 10 mar. 2017.

SCRIPTLATTES. **Uma ferramenta para extração e visualização de conhecimento a partir de Currículos Lattes**. 2015. Disponível em: <<http://scriptlattes.sourceforge.net/>>. Acesso em: 26 jul. 2017.

SHERPA/JULIET. **Some JULIET Statistics**. SHERPA Juliet, 2017. Disponível em: <<http://www.sherpa.ac.uk/juliet/stats.php?la=en&mode=advanced>>. Acesso em: 26 jul. 2017.

SHERPA/ROMEO. **Publisher copyright policies & self-archiving**. SHERPA Romeo, 2017. Disponível em: <<http://www.sherpa.ac.uk/romeo/>>. Acesso em: 10 jul. 2017.

SIDONE, O. J. G.; HADDAD, E. A.; MENA-CHALCO, J. P. A ciência nas regiões brasileiras: evolução da produção e das redes de colaboração científica. **Transinformação**, v. 28, n. 1, p. 15-32, 2016. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862016000100015&lng=en&nrm=iso>. Acesso em: 20 mar. 2017.

STEGEMANN, H. “Small Journals”, índices e e-cancer: uma oportunidade para a América Latina. **SciELO em Perspectiva**, 2013. Disponível em: <<http://blog.scielo.org/blog/2013/12/26/small-journals-indices-e-ecancer-uma-oportunidade-para-a-america-latina/>>. Acesso em: 30 dez. 2017.

STELATEK. **StelaExperta**. Santa Catarina: Instituto Stela, 2017. Disponível em: <<http://site.stelaexperta.com.br/materiais-gratuitos/>>. Acesso em: 27 dez. 2017.

SUNYE, M. et al. A experiência da UFPR na construção de repositórios digitais: a implantação integrada das ferramentas Dspace e Open Journal System. In: SAYÃO, L. et al. **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 107-122.

UNIVERSIDADE DE SÃO PAULO. Sistema integrado de bibliotecas. **Repositório Cruesp**. São Paulo: USP, 2017a. Disponível em: <<https://www.sibi.usp.br/bibliotecas/digitais-sistemicas/portal-producao-cruesp/>>. Acesso em: 26 jul. 2017.

UNIVERSIDADE ESTADUAL DE CAMPINAS. Sistemas de Bibliotecas da UNICAMP. **Repositório da Produção Científica e Intelectual da UNICAMP**. Campinas: UNICAMP, 2017b. Disponível em: <<http://repositorio.unicamp.br/static/historico.jsp>>. Acesso em: 26 jul. 2017.

UNIVERSIDADE FEDERAL DE MINAS GERAIS. **Bem-vindo ao Somos UFMG**. 2017c. Disponível em: <<http://somos.ufmg.br>>. Acesso em: 27 dez. 2017.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. Pró-Reitoria de Pós-Graduação. **Programas de Pós-Graduação Stricto Sensu**. São Carlos: UFSCar, 2014. Disponível em: <<http://www.propg.ufscar.br/propg>>. Acesso em: 03 dez. 2017.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. **Relatório anual de atividades 2016**. São Carlos: UFSCar, 2016.

VIDOTTI, S. A. B. G. et al. Coleta de dados a partir dos currículos da Plataforma Lattes: procedimentos utilizados no Repositório Institucional da UNESP. **Ponto de Acesso**, v. 9, n. 3, p. 117-132, 2015. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/15164>>. Acesso em: 19 maio 2016.

VIDOTTI, S. A. B. G. et al. Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 17., 2016. **Anais...** Salvador: UFBA, 2016.

WEITZEL, S. R. O papel dos repositórios institucionais e temáticos na estrutura da produção científica. **Em Questão**, v. 12, n. 1, p. 51-71, 2006. Disponível em: <<http://seer.ufrgs.br/index.php/EmQuestao/article/view/19/7>>. Acesso em: 10 mar. 2017.