

Universidade Federal de São Carlos
Departamento de Ciência Da Informação
Bacharelado em Biblioteconomia e Ciência Da Informação

Ingrid Rodrigues

**Indicadores da produção intelectual da UFSCar a partir do Repositório Institucional:
uso da ferramenta Looker Studio**

São Carlos
2023

Ingrid Rodrigues

**Indicadores da produção intelectual da UFSCar a partir do Repositório Institucional:
uso da ferramenta Looker Studio**

Trabalho de Conclusão de Curso apresentado ao Curso de Biblioteconomia e Ciência da Informação para obtenção do título de Bacharel em Biblioteconomia e Ciência da Informação.

Orientação: Prof. Dr. Roniberto Morato do Amaral.

São Carlos

2023

Agradecimentos

Agradeço, primeiro, ao meu companheiro e melhor amigo, Everton Russo, que me conheceu logo no início dessa jornada. Ele é a pessoa que mais me apoia e incentiva, e esteve ao meu lado todos os dias, e estará por muitos mais.

Aos meus pais, que apesar de todas as barreiras, do jeito deles, sempre me apoiaram nos meus sonhos e na minha autonomia, e ao meu irmão, que, no fundo do coração, sabe que me ama e não saberia viver sem mim.

Aos amigos que conheci durante essa jornada. Vocês foram fundamentais para que eu conseguisse seguir em frente. Em especial à Jessica Ramos, Raniel Andrade e Felipe Nachabe, que continuam firmes e fortes após longos 7 anos, e aos novos agregados, Celso Palma e Fillipo Pugliese.

Ao professor e meu orientador, Dr. Roniberto Morato do Amaral, pelos ensinamentos, paciência e puxões de orelha. Obrigado por mesmo depois de muitos sumiços da minha parte, me receber e aconselhar.

À Renata, minha psicóloga, que sempre lidou com meus surtos, ansiedades e procrastinações.

Muito obrigado a todos!

Resumo

Os repositórios institucionais desempenham um papel significativo para a organização, armazenamento, disseminação e preservação da produção intelectual das IES, promovendo o acesso aberto além de prover dados que podem ser empregados na geração de indicadores relativos à produção científica e ao desempenho das instituições. No contexto brasileiro, enquanto muitas pesquisas têm se dedicado à criação de repositórios, a atenção voltada para a exploração desses repositórios como ferramentas geradoras de indicadores ainda é limitada. Este estudo se concentra precisamente nesse aspecto, ao conceber indicadores sobre a produção intelectual da UFSCar a partir de seu Repositório Institucional (RI UFSCar) por meio da análise bibliométrica de 17.474 registros bibliográficos de itens depositados nas coleções de Teses e Dissertações, TCC, Dados de Pesquisa e Artigos até junho de 2023. Foi utilizada para a visualização de indicadores da ferramenta Looker Studio. A análise revelou uma deficiência quanto à completude dos objetos informacionais depositados no RI UFSCar, identificando possíveis desafios associados ao tema. Como conclusão, evidenciou-se que o procedimento utilizado apresenta-se como uma solução para obtenção e visualização de indicadores em repositórios e configura-se como um elemento capaz de contribuir para a melhoria das iniciativas dos RI, em especial o acompanhamento da sua completude e para a ampliação da visibilidade da instituição.

Palavras-chave: Repositório institucional. Indicadores bibliométricos. Comunicação científica.

Abstract

Institutional repositories play a significant role in the organization, storage, dissemination and preservation of the intellectual production of HEIs, promoting open access in addition to providing data that can be used to generate indicators relating to scientific production and the performance of institutions. In the Brazilian context, while much research has been dedicated to the creation of repositories, attention focused on exploring these repositories as indicator-generating tools is still limited. This study focuses precisely on this aspect, by conceiving indicators on the intellectual production of UFSCar from its Institutional Repository (RI UFSCar) through the bibliometric analysis of 17,474 bibliographic records of items deposited in the collections of Theses and Dissertations, TCC, Data from Research and Articles until June 2023. It was used to visualize indicators from the Looker Studio tool. The analysis revealed a deficiency regarding the completeness of the informational objects deposited in the RI UFSCar, identifying possible challenges associated with the topic. In conclusion, it was evident that the procedure used presents itself as a solution for obtaining and viewing indicators in repositories and is an element capable of contributing to the improvement of IR initiatives, in particular the monitoring of their completeness and to increase the institution's visibility.

Keywords: Institutional Repository. Bibliometric indicators. Scientific communication.

Sumário

1 INTRODUÇÃO.....	4
2 COMUNICAÇÃO CIENTÍFICA.....	7
2.1 Editores e periódicos eletrônicos.....	7
2.2 Acesso Aberto.....	9
2.3 Repositórios Institucionais.....	10
2.3.1 Repositórios no Brasil.....	12
2.3.2 Repositório Institucional da Universidade Federal de São Carlos (RI UFSCar)..	13
2.4 Indicadores Bibliométricos.....	17
3 MÉTODO E DESENVOLVIMENTO.....	20
3.1 Abordagem da pesquisa.....	20
3.2 Desenvolvimento.....	20
4 RESULTADOS.....	23
4.1 Indicadores de Teses e Dissertações (TeD).....	26
4.2 Indicadores de Trabalhos de Conclusão de Curso (TCC) dos cursos de graduação.....	30
4.3 Indicadores de Dados de Pesquisa (Dataset).....	34
4.4 Indicadores de Artigos de Periódicos.....	36
5 CONCLUSÕES.....	39
REFERÊNCIAS.....	41

1 INTRODUÇÃO

A comunicação científica sempre esteve atrelada aos publicadores de periódicos científicos. Os periódicos são parte da indústria da informação, um dos setores mais fortes da economia e, atualmente, essas publicações são monopolizadas por gigantes grupos de publicadores. Esse monopólio iniciado nos anos 70 e 90 aumentou significativamente o custo das assinaturas, obrigando as instituições de ensino e pesquisa a reavaliarem suas aquisições, resultando em uma redução no volume de publicações que elas possuem (MARCONDES e SAYÃO, 2009; LAINSON, 2010).

O avanço no uso da internet para compartilhamento de dados alterou o modelo de compra e disseminação da informação, uma vez que com a assinatura de um periódico online por uma instituição, seus membros teriam acesso fácil e rápido a todo o conteúdo disponibilizado, sem a necessidade de ter vários exemplares físicos para consulta. Mas, mesmo com esse avanço tecnológico, as publicações ainda eram restritas ao público pagante (LAWRECNE, 2001).

A cobrança pelo acesso à produção científica é prejudicial para o avanço da ciência, restringindo a disseminação do conhecimento. Nesse sentido, a disponibilidade online e gratuita contribui para maximizar o impacto das pesquisas, reduz a redundância e impulsiona o avanço da ciência de forma mais rápida (CAFÉ et al, 2003; LAWRENCE, 2001).

Em consonância com esse pensamento, na década de 90, surgiu o movimento de acesso aberto, estabelecido pela ideia fundamental de que o conhecimento científico deve ser um bem de domínio público. Esse movimento propões duas formas de divulgação da ciência, conhecidas como Via Verde e Via Dourada, que envolvem a participação dos publicadores, dos autores e outras partes interessadas, e nesse contexto, os RI são um dos meios para acesso às publicações científicas. (MARCONDES e SAYÃO, 2009; CAFÉ et al, 2003; FURNIVAL E HUBBARD, 2011).

Os RI são bibliotecas digitais abertas e interoperáveis, de instituições ou países, que visam guardar, preservar a longo prazo e, principalmente, garantir livre acesso à produção intelectual de seus pesquisadores, professores, estudantes, compreendendo assim uma ampla variedade de tipologias de objetos informacionais, como, por exemplo:

artigos, teses dissertações, relatórios, banco de dados e outros. (MARCONDES e SAYÃO, 2009)

No Brasil o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) tem disponibilizado a ferramenta computacional DSPACE para as universidades brasileiras hospedarem as suas iniciativas de repositórios. Porém, este movimento é recente, o que pode prejudicar a completude das coleções institucionais armazenadas pelas universidades, uma vez que a iniciativa prioriza o texto completo dos objetos informacionais (IBICT, 2023).

Nos últimos anos, os RI têm desempenhado um papel crucial no armazenamento e disseminação da produção intelectual das instituições, contribuindo para o aumento da visibilidade e impacto das publicações institucionais. No Brasil o IBICT disponibiliza a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD). É um sistema de informação que reúne e disponibiliza de forma gratuita e online as teses e dissertações defendidas nos programas de pós-graduação *stricto sensu* aprovados pela CAPES, de diversas instituições de ensino e pesquisa, dando visibilidade à produção científica nacional (IBICT, 2023). Além da BDTD, há o Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto (Oasisbr), uma iniciativa que reúne a produção científica e os dados de pesquisa em acesso aberto, publicados em diversos locais e o Current Research Information System (BrCris), que disponibiliza os dados das publicações em acesso aberto.

Além de fontes de informações científicas, tecnológicas e culturais, os RI disponibilizam registros bibliográficos que podem ser utilizados na elaboração de indicadores, visando o acompanhamento e a avaliação dos resultados das práticas científicas no contexto das Instituições de Ciência e Tecnologia (ICT) (CAFÉ et al, 2003; MARCONDES e SAYÃO ,2009; REIS; SPINÓLA; AMARAL, 2017)

Os indicadores podem ser obtidos através das Bibliometria, uma técnica de análise de informações, envolvendo a contagem da comunicação escrita. Nas práticas de avaliação da atividade científica a Bibliometria é utilizada na elaboração de indicadores bibliométricos, por exemplo, sobre a produção científica e tecnológica, utilizando registros bibliográficos sobre artigos científicos, documentos de patentes entre outros objetos informacionais, recuperados em bases de dados e analisados com abordagem quantitativa, visando a apoiar o processo de tomada de decisão, envolvendo estudos de

prospecção tecnológica, inteligência competitiva, avaliação da produção científica (SPINAK, 1998; FARIA, 2001; MELLO, 2013; TARAPANOFF, 1995; OKUBO, 1997; FARIA, 2001).

Visando contribuir para a sensibilização sobre a importância dos RI para a atuação e reconhecimento das universidades públicas federais, o objetivo desta pesquisa foi elaborar um conjunto de indicadores sobre a produção intelectual da Universidade Federal de São Carlos, utilizando os registros bibliográficos de seu Repositório Institucional, identificando a distribuição das coleções e objetos informacionais, por ano, área do conhecimento, programas de pós-graduação, entre outras referências legítimas a comunidade da UFSCar. Para o tratamento e visualização dos dados foi utilizada a ferramenta Looker Studio.

A contribuição desta pesquisa reside na importância da comunidade acadêmica compreender e aprimorar o uso dos RI, e espera-se que ao acessar os indicadores bibliométricos sobre o conteúdo distribuído nas diversas áreas de conhecimento da universidade, a comunidade se sensibilize e conscientize sobre a importância do RI para o seu projeto institucional e implemente práticas que incentivem a participação dos pesquisadores na disponibilização de seus trabalhos.

Do ponto de vista da Ciência da Informação, esta pesquisa pode contribuir para a consolidação da área em estudos métricos. Embora existam investigações na literatura sobre a implantação e a utilização de repositórios institucionais, poucos estudos se concentram especificamente na utilização de indicadores bibliométricos para avaliar e aprimorar a completude desses repositórios.

2 COMUNICAÇÃO CIENTÍFICA

2.1 Editores e periódicos eletrônicos

Ao longo da história a comunicação dos resultados da ciência esteve centralizada nos periódicos científicos especializados. Em 1660, foi criada a Royal Society em Londres, Inglaterra, conhecida como o primeiro periódico do mundo, que publicava tratados e livros de assuntos relacionados, inicialmente, as áreas de filosofia, matemática, física e biologia (LAINSON, 2010).

Esse primeiro movimento levaria à criação de muitos outros periódicos científicos, que seriam o ator principal de um dos setores mais fortes economicamente, a indústria da informação, que teve significativo crescimento na década de 1960. Esta indústria abrange um amplo conjunto de serviços de intermediação entre produtores e consumidores de informações científicas. É constituída por cientistas e acadêmicos, que se interessam e precisam acessar as publicações atualizadas com os últimos resultados de pesquisas e por editores, que buscam lucrar com a venda de acessos aos artigos científicos. Já as bibliotecas especializadas e acadêmicas ficam responsáveis por adquirir e disponibilizar esses materiais à sua comunidade (MARCONDES e SAYÃO, 2009).

A partir de 1970 observou-se um desequilíbrio no setor de publicadores, quando as movimentações econômicas de compras de empresas por outras, levou à monopolização das publicações por gigantes grupos, como Elsevier e Emerald, antes espalhados entre centenas deles. Como resultado, ocorreu um crescente custo das assinaturas (MARCONDES e SAYÃO, 2009). Entre 1975 e 1995, os aumentos das assinaturas de periódicos cresceram entre 200 e 300%, aumento maior que a inflação (DEWATRIPONT, et al. *apud* MARCONDES e SAYÃO, 2009). Essa crise levou a necessidade de as instituições avaliarem quais assinaturas manter ou excluir, ocasionando também a necessidade de criarem mecanismos de avaliação para esses periódicos.

O crescente e agora sólido uso da web para o compartilhamento de informações, incluindo documentos completos, direcionou os editores para o ambiente *online*. Textos científicos que antes precisavam ser comprados no papel, passaram a ser adquiridos

eletronicamente. Com uma licença de uso de um periódico *online*, por exemplo, os usuários de uma instituição que possua essa licença, têm acesso a todo o acervo, podendo fazer quantos *downloads* quiserem dos materiais de acordo com as condições contratuais.

O uso progressivo dos periódicos *online* agilizou as trocas de informações e o processo de atualização dos resultados das pesquisas, uma vez que os artigos, assim que publicados, já ficavam disponíveis para acesso, recuperação e leitura, e não precisam mais ser enviados em papel para os consumidores. Mesmo com essas evoluções, a disseminação do conhecimento continua restrita a consumidores pagantes, desfavorecendo instituições com menos recursos financeiros e países do terceiro mundo (LAWRENCE, 2001).

Para Café et al (2003), a publicação em periódicos pagos restringe a difusão do conhecimento, uma vez que apenas um número limitado de especialistas é capaz de arcar com os custos de acesso a esses periódicos. Esse meio também não promove mecanismos de diálogos entre esses especialistas, prejudicando a colaboração e o avanço das pesquisas publicadas.

Lawrence (2001) concorda que o acesso pago pelos resultados de pesquisa é contraproducente e complementa dizendo que os autores dos artigos que são lançados nesses periódicos, publicam apenas pelo impacto da pesquisa. Em seu estudo sobre artigos de conferências de Ciência da Computação e disciplinas relacionadas, que correlaciona a quantidade de citações de um artigo com a probabilidade de ele estar disponível gratuitamente na internet, pôde concluir que os artigos disponíveis online possuem mais citações. Alguns fatores como a rápida disponibilidade, a facilidade de acesso e de conectar-se diretamente com os pesquisadores ou grupos de pesquisa contribuem para essa relação. A disponibilidade online e gratuita maximiza o impacto das pesquisas, minimiza a redundância e acelera o progresso da ciência. Os autores não recebem *royalties* pelas publicações, mas buscam relevância, procuram ser citados e participar na construção de outras pesquisas.

Historicamente, os autores não se beneficiaram da necessidade de pagamento para o acesso às pesquisas, e essa é uma barreira que afeta significativamente a trajetória deles, cujas carreiras dependem, muitas vezes, da visibilidade e impacto de suas publicações.

2.2 Acesso Aberto

Em contraponto ao mercado tradicional de publicações científicas surge o acesso aberto ou acesso livre, baseado na premissa fundamental de que o conhecimento científico deve ser um bem público. O termo Acesso Aberto foi consagrado em três declarações: no *Budapest OA Initiative* (2002), na *Bethesda Statement on Open Access Publishing* (2003) e na Declaração de Berlin sobre Acesso Aberto ao Conhecimento nas Ciências e Humanidades (2003) (FURNIVAL E HUBBARD, 2011).

As políticas de livre acesso começaram a ser formuladas por todo o mundo a partir dos anos 2000, e junto delas, surgiram importantes iniciativas para incentivar o acesso aberto, como os RI.

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, poder ser utilizados amplamente, não serem apropriados de forma privada. Assim, cada instituição científica ou acadêmica, e sua correspondente comunidade, deve manter em seu repositório institucional de livre acesso cópias da produção científica de sua comunidade (MARCONDES e SAYÃO, 2009, p. 10).

O livre acesso à produção científica propõe duas formas ou vias de publicação dos trabalhos em livre acesso, as chamadas “via verde” e “via dourada”. A via dourada é uma orientação para os periódicos científicos a publicarem de acordo com a concepção do livre acesso, o que implica na disponibilização gratuita de seus artigos e pesquisas para o público, geralmente por meio de licenças abertas que permitem seu uso, leitura e *download* sem restrições (MARCONDES e SAYÃO, 2009).

A via verde libera o depósito dos artigos nos RI. O depósito da produção científica pelos próprios pesquisadores, também chamado de autodepósito ou autoarquivamento, varia conforme as políticas das instituições ou países, que podem incluir apenas uma sugestão de depósito ou uma obrigatoriedade de fazê-lo. O autoarquivamento permite que o autor envie seu trabalho diretamente para publicação, sem intermédio de terceiros, como ocorre nos periódicos (MARCONDES e SAYÃO, 2009; CAFÉ *et al*, 2003).

A combinação de via verde e dourada ampliam significativamente o alcance e impacto das pesquisas e desempenham um papel fundamental na democratização do

conhecimento. Diferente de apenas disponibilizar o texto na *web*, Costa e Leite (2009) afirmam que as inclusões nos repositórios facilitam a detecção de plágio, que é uma das preocupações dos autores quanto ao lançamento dos resultados de pesquisa diretamente na *web*.

2.3 Repositórios Institucionais

Os RI são considerados bibliotecas digitais, de instituições ou países, que visam guardar, preservar a longo prazo e, principalmente, garantir livre acesso, via internet, à produção científica. São bases de dados na *web* em que uma instituição de pesquisa deposita sua produção acadêmica (MARCONDES e SAYÃO, 2009). Diferente de apenas um catálogo *online* de uma biblioteca, os RI recuperam apenas itens digitais, não há nenhum acervo físico (TARTAROTTI, 2020).

O termo Repositório Institucional, segundo Costa e Leite (2009, p. 163), é utilizado para representar “um novo serviço bibliotecário cuja ênfase é constituir, gerenciar e, sobretudo, disseminar amplamente coleções digitais de informação científica, de modo que a comunicação, acesso e uso de resultados de pesquisa fossem expandidos”. Café, *et al* (2003) complementa que os RI contribuem com a transparência e acessibilidade da instituição, uma vez que disponibilizam sua produção científica a todo o público.

É importante ressaltar que a avaliação por pares, por exemplo, ocorre no processo de seleção e publicação dos artigos em periódicos científicos e não é diretamente aplicada aos RI. Esse processo de avaliação continua sendo uma prática essencial no contexto da publicação científica em periódicos, enquanto os RI desempenham um papel importante na preservação, disseminação e acesso aberto aos trabalhos científicos institucionais, complementando o processo de revisão por pares (LIMA, 2009).

Os RI podem armazenar trabalhos de conclusão de curso (graduação, mestrado, doutorado), artigos de periódicos, artigos de eventos, entre outros objetos informacionais, que sejam relevantes para a instituição, de toda a natureza, como “texto, imagens, vídeo, áudio, apresentações, programas de computador, *datasets* etc” (MARCONDES; SAYÃO, 2009, p. 24). O conteúdo é heterogêneo no tipo e na multidisciplinaridade a depender do

objetivo do repositório. Os principais tipos de documentos são os registros intelectuais de estudantes e pesquisadores, sejam pesquisas ou materiais didáticos, e podem conter qualquer informação de interesse da instituição, como atividades e eventos produzidos nas instituições (CAFÉ *et al*, 2003; WEITZEL, 2018)

A gestão do RI se dá com a utilização de *softwares* que auxiliam na inclusão, armazenamento e recuperação da informação. Esses *softwares* desempenham um papel fundamental na gestão eficiente e acessível de repositórios, permitindo que instituições organizem e disponibilizem seus materiais de forma ampla e livre. Dentre alguns desses softwares, os mais utilizados são *DSPACE*, *EPRINTS*, *GREENSTONE*, *FEDORA* e *NOU-RAU*. Um número considerável dessas ferramentas foi desenvolvido em universidades e centros de pesquisa, cobrindo necessidades específicas delas. Embora cada uma mantenha sua especificidade, esses *softwares* são distribuídos gratuitamente, e são de código aberto, possibilitando customizações (MARCONDES e SAYÃO, 2009).

Os RI desempenham papel relevante no fortalecimento das universidades e instituições de pesquisa ao fornecerem uma visão abrangente da sua produção científica e tecnológica, que aumenta a visibilidade, a reputação e a influência da instituição, enquanto promove o compartilhamento e disseminação dessa produção.

Informação é um insumo essencial para o desenvolvimento da ciência e, conseqüentemente, para o desenvolvimento social e econômico de um país. O livre acesso à informação científica é um dos seus instrumentos, os repositórios institucionais, são atualmente um dos mais importantes desafios políticos e profissionais com que se deparam hoje os profissionais de informação no Brasil (MARCONDES e SAYÃO, 2009, p. 19).

O uso dos RI também pode introduzir novos mecanismos de avaliação da ciência, como indicadores de número de *downloads*, citações, índices de apoio às pesquisas e outros, e esses mecanismos podem servir como estímulos para o uso do repositório, tanto para fonte de informação, como para o autodepósito de materiais (MARCONDES e SAYÃO, 2009).

2.3.1 Repositórios no Brasil

O Acesso Aberto, afirma Guédon (2010), têm colaborado positivamente para o aumento do poder científico dos países em desenvolvimento, que agora podem ter acesso gratuito aos resultados de pesquisa do mundo todo. Mesmo com o cenário positivo, o autor afirma que ainda há um longo caminho para percorrer, para superar a estrutura dos países dominantes, e destaca a potencialidade dos Repositórios Institucionais para aumentar o impacto e a visibilidade a produção científica desses países.

Ainda nos países emergentes, o potencial de crescimento destes pode ser observado no BRICS (Brasil, Rússia, Índia, China e África do Sul), que representam 22% da economia mundial e 24% das publicações científicas da base Scopus (SCHÖPFEL, 2017).

O Brasil é reconhecido como pioneiro na introdução do Acesso Aberto no mundo devido às suas estratégias bem-sucedidas. Weitzel (2018) destaca ações do Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (Bireme) e do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) como as principais nessa estratégia. O IBICT, ao longo dos últimos 15 anos, tem promovido eventos, lançado manifestos e organizado diversos diretórios e provedores de serviços para promover a adoção do acesso aberto e, em 2009, lançou um convênio através da Financiadora de Estudos e Projetos (FINEP) para financiar iniciativas de Acesso Aberto, tanto pela via verde, quanto pela via dourada (WEITZEL, 2018).

Apesar desse pioneirismo e os avanços notados, dados extraídos em abril de 2017 no *Directory of Open Access Journals (DOAJ)* e do *Directory of Open Access Repositories (OpenDOAR)* mostram que o Brasil, comparado a outros países, possui mais dados provenientes da via dourada, evidenciando a necessidade de investimento nos repositórios em funcionamento. (BRUMATTI, 2016; WEITZEL, 2018). Nesse contexto, Costa e Leite (2017) enfatizam a relevância da implementação de mais políticas e programas nacionais que promovam o acesso aberto e incentivem o debate público sobre o assunto, a fim de fortalecer o acesso aberto verde.

Um estudo feito por Weitzel (2018), analisou os repositórios registrados em 5 principais fontes sobre o assunto, e recuperou um total de 130 repositórios, dos quais 29 não puderam ser analisados, por estarem defasados ou não funcionarem mais (seja por *link* quebrado ou por descontinuação). Dos 101 restantes, observou que 92% são repositórios institucionais e, destes, 2% são de acesso restrito, e 12% contém apenas teses e dissertações. O estudo também atenta para a falta de uniformidade nas informações trazidas pelos repositórios: indicadores como o número total de itens, o número total de itens por tipo de documento ou por áreas do conhecimento conseguem ser apenas parcialmente obtidos.

Outro levantamento, realizado por Reis, Spinóla e Amaral (2017) em suas avaliações sobre os indicadores bibliométricos presentes nos repositórios brasileiros, que analisou 81 repositórios registrados até abril de 2016 no OpenDOAR (Diretório de Repositórios de Acesso Aberto) constatou que em 91% era possível obter indicadores de duas dimensões, que relacionavam dois ou mais indicadores de produção. Mas, os indicadores de citação, como índice H e Fator de Impacto, foram encontrados em apenas 3,5% deles. Outro aspecto analisado foi a apresentação desses indicadores, e apenas 9,9% possuem gráficos estáticos e 6,2% gráficos dinâmicos.

A partir das observações apresentadas, fica evidente a necessidade de aprimorar a operabilidade, gestão e a divulgação dos RI. É imprescindível investir em melhorias contínuas, para promover uma cultura acadêmica mais colaborativa, transparente e inclusiva, que contribua efetivamente para o avanço da ciência e o desenvolvimento do país.

2.3.2 Repositório Institucional da Universidade Federal de São Carlos (RI UFSCar)

O Repositório Institucional da UFSCar (RI-UFSCar) é um sistema de informação que visa armazenar, preservar, organizar e disseminar amplamente a produção intelectual dos diversos setores e segmentos da comunidade da UFSCar, provendo o acesso aberto à informação produzida na instituição e registrada como científica, tecnológica, didática, artístico-cultural e técnico-administrativa (Repositório Institucional UFSCar, 2016).

O RI UFSCar, conforme conhecido hoje, foi criado em 2016, após sua aprovação no Conselho Universitário da UFSCar, que em 04 de março de 2016 lançou uma resolução que dispôs sobre a criação, as políticas de implantação e a governança do repositório. O repositório objetiva organizar, armazenar, preservar e disseminar toda a produção intelectual e técnico-administrativa de toda a comunidade da universidade, e assim contribuir com a preservação da memória da instituição, ampliar sua visibilidade nacional e internacional, sua transparência, e compromisso com o acesso aberto (UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2016).

Entende-se como produção intelectual toda a produção científica, tecnológica, didática, artístico-cultural e técnico-administrativa produzida, tais como: teses e dissertações, artigos científicos, livros, capítulos de livros, relatórios técnicos e administrativos, documentos técnico-administrativos, portarias, publicações em anais de eventos, trabalhos de conclusão de cursos, conjuntos de dados primários ou brutos (*datasets*), *softwares*, patentes, audiovisuais e outros tipos de produção intelectual, previamente classificados pela UFSCar como sendo de acesso aberto ao público (UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2016).

O depósito de materiais no RI pode ocorrer por meio do autodepósito ou mediado por uma instância da universidade, e deve estar acompanhado do termo de consentimento do autor, concedendo à universidade o direito, não exclusivo, de incluir o documento em acesso aberto na internet. Atualmente, apenas a inclusão dos Trabalhos de Conclusão de Curso (TCC) é obrigatório, e o depósito é de responsabilidade do docente orientador. Já para o depósito de Teses e Dissertações (TeD) dos Programas de Pós-Graduação, há apenas as orientações para realizar o autodepósito, e é de responsabilidade dos discentes dos programas fazê-lo (UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2022 e UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2020).

A gestão do RI-UFSCar está sob responsabilidade do Sistema Integrado de Bibliotecas da Universidade Federal de São Carlos (SIBi), e todo o material recebido é inspecionado pelos responsáveis antes de ser lançado ao público. O RI segue a legislação nacional vigente quanto aos direitos autorais, e todo material é depositado sob a licença *Creative Commons*, sendo recomendado o uso da CC BY-NC-ND, que permite o *download* e compartilhamento, desde que atribuído os créditos aos autores, e também protege o material de possíveis modificações em seu conteúdo (UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2022).

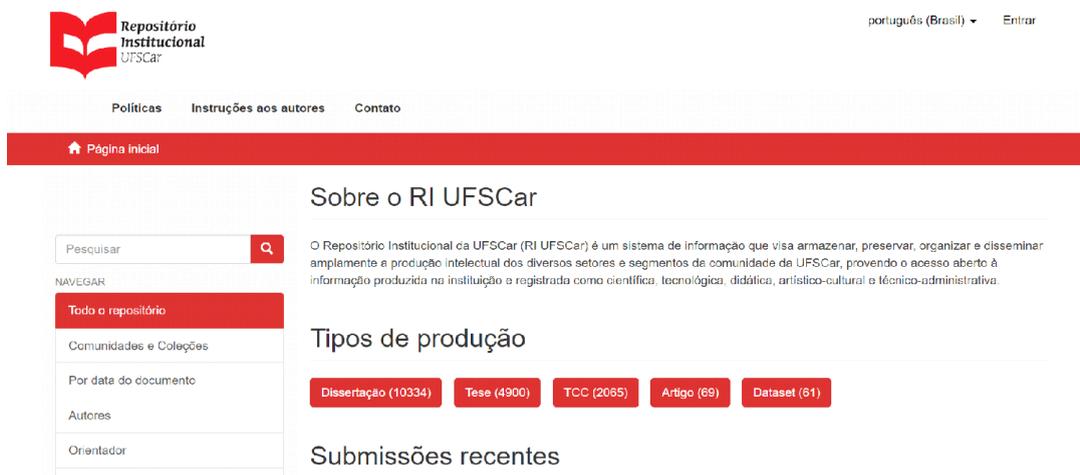
Embora o RI tenha sido criado em 2016, ele possui itens que foram publicados desde os anos 2000. A maioria desses itens pertence à coleção de Teses e Dissertações, cujas publicações já integravam a Biblioteca Digital de Teses e Dissertações (BDTD), criada oficialmente em 2004 (AMARAL *et al*, 2006).

Um problema apontado por Amaral *et al* (2006) relacionado à BDTD foi, em partes, solucionado com a criação do RI: a falta de espaço físico para armazenamento dos trabalhos, uma vez que os depósitos realizados no RI são apenas de itens em formatos eletrônicos (como documentos em PDF).

O RI contém apenas dissertações, teses, TCCs, artigos e dados de pesquisa, deixando de cumprir, em parte, com seu propósito, pois não inclui, por exemplo, livros, dados de patentes ou documentos técnico-administrativos, caracterizados como parte componente da produção intelectual da UFSCar.

O acesso ao RI (Figura 01) é aberto para todos que possuam acesso à internet e está disponível em português, inglês e espanhol. Em sua página inicial, há uma pequena apresentação sobre o que é o repositório, e é possível realizar buscas e navegar pelas coleções.

Figura 1 - Captura de tela da página inicial do RI-UFSCar.



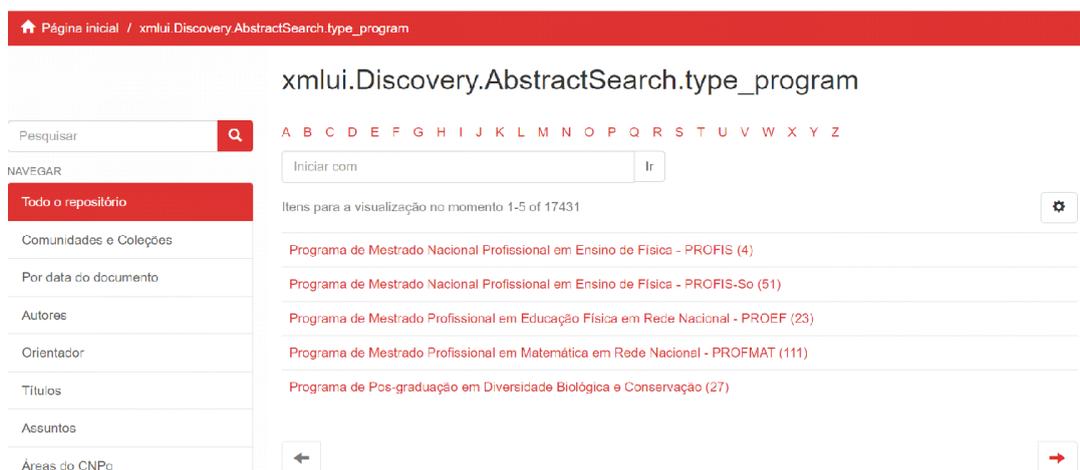
Fonte: Elaborado pela autora.

A navegação pode ser realizada por tipos de documentos (Dissertação, Tese, TCC, Artigo ou *Dataset*), pelas comunidades e coleções, (autodepósitos, cursos de graduação ou pós-graduação), autor, assunto, data de publicação e áreas do Conselho Nacional de

Desenvolvimento Científico e Tecnológico (CNPq) e PPG. Também é possível realizar buscas textuais que abrangem todo o repositório, ou apenas a navegação escolhida, e há suporte para buscas avançadas utilizando operadores booleanos, mas essa função não está especificada nas orientações do RI.

Dentro das navegações disponibilizadas, é possível visualizar o número total de itens disponíveis para consulta, e a quantidade de itens em cada resultado encontrado. No filtro por “PPG”, por exemplo (figura 2), há a quantidade de itens de todos os programas e a quantidade de itens de cada programa, mas não há outra forma de visualização desses dados, e nem a possibilidade de extração para o tratamento e desenvolvimento de indicadores.

Figura 2 - Captura de tela da navegação por programas de pós-graduação no RI-UFSCar.



Fonte: Elaborado pela autora.

Para realizar o autodepósito, a pessoa interessada deve acessar o site do RI, logar (apenas quem possui vínculo ativo com a UFSCar consegue realizar esse passo) e iniciar o processo de depósito. Durante o processo, é necessário selecionar a coleção à que pertence o item a ser depositado, incluir todas as informações necessárias (como autor, título, título em inglês, e outras), incluir uma cópia eletrônica do item, selecionar o tipo de licença de publicação, se há algum impedimento para a publicação no RI e aceitar a publicação. Após, as informações serão revisadas por um responsável e o material será incluído no RI ou devolvido para revisão.

2.4 Indicadores Bibliométricos

A Bibliometria é um termo usado para designar uma gama de métodos utilizados para medir a produção científica e tecnológica através da coleta e análise de dados da literatura científica e de patentes (OKUBO, 1997). Também é definida por Terapanoff (1995) como a análise dos elementos numéricos relacionados à criação, disseminação e utilização da informação registrada, utilizando abordagens matemáticas, com o objetivo de auxiliar na tomada de decisões. Ela investiga as propriedades de criação e utilização de documentos, realiza a quantificação da produção de documentos e aplica métodos matemáticos e estatísticos para análise.

A análise bibliométrica utiliza inúmeros parâmetros, como literatura científica (artigos, etc.), coautoria, patentes, citações, cocitações e copalavras. Esses parâmetros são medidas indiretas da comunidade científica, sua estrutura e sua produção (OKUBO, 1997, p. 20, tradução da autora).

Como um método para elaboração e análise de indicadores, a Biblioteconomia teve sua origem no século XX como uma abordagem adotada por bibliotecários e cientistas para acompanhar o acelerado progresso da ciência, quando reconheceram que era necessário um novo meio de acompanhar esse avanço devido ao crescente volume de conhecimento científico produzido (Faria 2001).

Há outras áreas que se ocupam da medição da informação, como a Cientometria e Infometria. Faria (2001) descreve a Cientometria como um método que aplica as mesmas técnicas da bibliometria, mas objetiva estudar apenas documentos científicos, a fim de estudar a ciência e seus processos de comunicação, e a Infometria como estudo quantitativo da informação em qualquer formato e grupos social, não se restringindo apenas aos textos e referências registrados pela ciência. Ainda, para o autor, as demarcações entre bibliometria, Cientometria e Infometria não estão claramente estabelecidas, o que leva a uma certa confusão terminológica nessa área.

Mello (2013) concorda que há entre os pesquisadores algumas confusões quanto às terminologias utilizadas nesse campo, e cita a Cientometria, a Bibliometria, Infometria, a Tecnometria e a Webometria como métodos quantitativos capazes de gerar indicadores

sobre a ciência e tecnologia. Acrescenta ainda que os indicadores obtidos são de grande valor e capazes de “representar o estado desta ciência, prever seu futuro e fazer comparações” (MELLO, 2013).

A organização e frequência dos dados em um texto ou conjuntos de referência bibliográficas acompanha certos padrões, e alguns pesquisadores, observando esse fenômeno, desenvolveram algumas leis bibliométricas (GALDINO, GARCIA e AMARAL, 2022). A Lei de Bradford, que descreve o comportamento repetitivo e ocorrências uma área de estudo, e analisa assuntos e tendências, a Lei de Lotka, que analisa a contribuição dos autores para o progresso da ciência e a Lei de Zipf, conhecida como lei quantitativa fundamental da atividade humana, subdividida em outras duas leis e usa expressões matemáticas para analisar a frequência em que as palavras aparecem em um texto (QUOANIN *et al*, 2001).

Através da bibliometria, conjuntos de centenas ou milhares de registros bibliográficos de artigos científicos, patentes, notícias e outros documentos podem ser analisados para dar origem a novas informações bastante sintéticas e de alto valor agregado, chamados indicadores, que não dizem respeito a um documento em particular, mas ao conjunto dos documentos analisados. Os indicadores bibliométricos podem ser usados para estudos de prospecção tecnológica, inteligência competitiva e análise da produção científica (GALDINO, GARCIA e AMARAL, 2022, p. 5).

Os indicadores representam uma abordagem indireta para avaliar fenômenos intangíveis, como ciência e tecnologia, como um sistema que gera informação, conhecimento e inovação, dependendo de insumos para operar e produzir resultados, é possível construir indicadores por meio da mensuração dos insumos empregados e dos resultados alcançados (SPINAK, 1998).

Faria (2001) classifica os indicadores como de uma ou duas dimensões. Os de uma dimensão, são divididos entre os indicadores de atividade e indicadores de impacto, e os de duas dimensões são indicadores de ligação. Okubo (1997) agrupou os indicadores em apenas dois grupos, sendo o primeiro designado como indicadores de atividade ou de produção, e o segundo, indicadores de relação. Em seu trabalho, também listou quais são os indicadores utilizados.

Os indicadores de atividades ou de produção (de um indivíduo, uma instituição ou um país) podem ser: número de publicações, número de citações, número de co-autores

e número de patentes. O número de publicações compreende a contagem de publicações produzidas (que podem ser livros, periódicos, jornais, resenhas, relatórios, artigos etc.) e fornece a quantidade de trabalhos produzidos, e pode ser usado para monitorar tendências de um determinado campo ou disciplina ao longo do tempo. O número de citações é a quantidade de citações de uma pesquisa em outros trabalhos e pode ser considerado uma forma de medir o impacto da produção científica. O número de co-autores quantifica as publicações produzidas por dois ou mais autores, usado para medir o crescimento ou declínio de pesquisas colaborativas. Já o número de patentes é a contagem de patentes produzidas e pode ser utilizado para medir a inovação e a capacidade tecnológica (OKUBO, 1997)

Os indicadores de relação (entre indivíduos, instituições ou países), podem ser co-publicação, índice de afinidade, co-citação, e coocorrência de citações. A co-publicação mede as interações e relações científicas entre redes, equipes, instituições e países e possibilita delinear relações e identificar os principais parceiros de pesquisa. O índice de afinidade avalia o intercâmbio científico entre dois atores em um determinado período e a análise das variações do índice de afinidade ao longo do tempo pode indicar mudança nas relações científicas entre esses atores. A co-citação mensura a quantidade de vezes que duas publicações são citadas simultaneamente em uma mesma publicação e pode ser usada para representar a influência e impacto dos autores. Por fim, a co-ocorrência de palavras mede quantas vezes duas determinadas palavras são utilizadas juntas em artigos ou patentes de uma determinada área de pesquisa e é utilizada na criação de mapas de redes de palavras que representam os principais temas de um campo de estudo, e suas relações (OKUBO, 1997).

As publicações científicas têm sido a principal fonte de dados utilizada para gerar indicadores que possibilitam a análise dos resultados e da qualidade da produção científica. O aumento na disponibilidade de bases de dados de periódicos eletrônicos também aumentou o uso desses dados na construção de indicadores, devido ao fácil acesso e extração, e é possível extraí-los de bases como *Web of Science* (WoS), Scopus, Scientific Electronic Library Online (SciELO) e outras (LIMA, VELHO e FARIA, 2012, REIS, 2017).

O avanço dos estudos métricos e da informática proporcionou o surgimento de novas ferramentas/*softwares* que auxiliam na automatização dos processos dentro da

análise bibliométrica, como o VantagePoint, Gephi e planilhas eletrônicas (REIS, 2017). Com as novas tecnologias, as etapas de análise também se transformaram. Para Faria (2001) as análises atuais contam com 5 etapas: recuperação dos dados; preparação dos dados; tratamento bibliométrico; tratamento estatístico, e representação gráfica dos resultados.

3 MÉTODO E DESENVOLVIMENTO

3.1 Abordagem da pesquisa

Esta pesquisa pode ser classificada como quantitativa em relação a sua abordagem, e de natureza exploratória. É quantitativa, pois quantifica os dados de toda a produção intelectual depositada no Repositório Institucional da UFSCar, e é de natureza exploratória, porque investiga novos cenários e possibilidades relacionados ao RI e seus indicadores.

O RI da UFSCar foi escolhido pela facilidade de acesso aos dados e aos profissionais envolvidos na sua gestão, e por integrar a instituição que comporta o curso de Bibliometeonomia e Ciência da Informação, além de ser umas das melhores universidades do país e estar entre as 1000 melhores universidade do mundo, segundo a edição de 2023 da lista do CWUR (*Center for World University Rankings*) (CWUR, 2023).

3.2 Desenvolvimento

A análise dos dados dos itens incluídos no RI UFSCar contou com as etapas descritas por Quoniam (2001), que são: preparação, tratamento bibliométrico e pós-tratamento. Foram analisados os dados do RI de publicações realizadas até junho de

2023, com um total de 17.474 itens. O repositório fornece informações em formato de metadados utilizando o padrão *Dublin Core*, como pode ser visto na Figura 3.

A coleta dos dados em massa não pôde ser realizada diretamente no site do RI UFSCar. Apesar do uso de todos os protocolos de acesso aberto, a interface do RI não compreende a extração dos metadados. Para a extração, foi solicitado à Secretaria de Informática (SIn) da UFSCar, responsável pelos sistemas da UFSCar, que extraísse os dados para a análise. Este serviço é disponibilizado à comunidade UFSCar via Portal de Serviços, e foi solicitado por pesquisadores do Núcleo de Informação Tecnológica em Materiais da UFSCar. Como todos os dados estão disponíveis no repositório e são de livre acesso, e apenas a extração em massa não pôde ser realizada, não há problemas em relação às licenças e segurança da informação. Os dados foram disponibilizados em formato CSV, que pode ser aberto em planilhas e/ou importados em programas que auxiliem na análise dos dados.

Os dados exportados então estruturados em formato *Dublin Core*, e contam com 171 campos de informações. Alguns campos não são de preenchimento obrigatório para determinados itens (como, por exemplo, os artigos de periódicos, que não possuem orientador, então o campo "*dc.contributor.advisor1*" não será preenchido), e por isso vieram sem dados, mas são informações complementares e que não prejudicam a análise. Também há alguns dados faltando ou com formatação fora do padrão, possivelmente por quebra dos dados durante a extração, que dificulta a análise e necessita de ação humana para serem ajustados.

O tratamento dos dados foi realizado com uso de planilhas do *Google Sheets* e do *Looker Studio* (Antigo *Google Data Studio*), uma ferramenta gratuita da Google utilizada para criação de relatórios, gráficos e painéis informativos personalizáveis. O *Looker Studio* utilizada bancos de dados em planilha ou outras ferramentas do Google como fontes de informação.

Para utilizar o *Looker Studio* foi necessário realizar o *upload* do arquivo *csv* disponibilizado pela SIn em uma planilha do *Google Sheets* e realizar a integração dessa planilha com o *Looker Studio*. Todos os dados trabalhados e analisados na ferramenta para esse trabalho tiveram essa planilha como única fonte de dados. Após essa integração, é realizado o tratamento dos dados, já unificado à visualização dos mesmo, pois a ferramenta possibilita que as duas etapas sejam realizadas em conjunto.

Durante o tratamento, foi necessário identificar quais campos *Dublin Core* seriam utilizados, e quais as relações entre os campos. Para a elaboração de indicadores de publicação por autor, por exemplo, foi necessário utilizar o campo “*dc.contributor.author*”, e para indicadores de produção por área do conhecimento, distribuídos por ano de publicação, foi necessário utilizar os campos “*dc.date.issued*” e “*dc.subject.cnpq*” (Figura 03).

Figura 3 - Formato de metadados no padrão *Dublin Core* dos itens incluídos no RI-UFSCar

Incipiência da visualização de indicadores bibliométricos e altmétricos nos Repositórios Institucionais brasileiros

dc.contributor.author	Reis, José Eduardo dos	
dc.contributor.author	Spinola, Adriana Tahereh Pereira	
dc.contributor.author	Amaral, Roniberto Morato do	
dc.coverage.spatial	Porto Alegre	por
dc.date.accessioned	2022-09-29T13:54:05Z	
dc.date.available	2022-09-29T13:54:05Z	
dc.date.issued	2017-01-21	
dc.identifier	http://dx.doi.org/10.19132/1808-5245230.213-234	por
dc.identifier.citation	REIS, José Eduardo dos; SPINOLA, Adriana Tahereh Pereira; AMARAL, Roniberto Morato do. Incipiência da visualização de indicadores bibliométricos e altmétricos nos Repositórios Institucionais brasileiros. Em <i>Questão</i> , Porto Alegre, v. 23, n. Edição Especial 5 EBBC, p. 213–234, 2017. Disponível em: https://repositorio.ufscar.br/handle/ufscar/16734 .	*
dc.identifier.issn	1808-5245	por
dc.identifier.uri	https://repositorio.ufscar.br/handle/ufscar/16734	
dc.description.abstract	Os Repositórios Institucionais apresentam significativo potencial como fontes de informação para a elaboração dos estudos métricos, visando à compreensão das dinâmicas da atividade científica institucional. Contudo, sinalizam deficiências no tocante à externalização visual de	por

Fonte: Elaborado pela autora.

O uso do *Looker Studio* facilita as etapas de tratamento e visualização dos dados, uma vez que ambos podem ser realizados no mesmo local e, ao mesmo tempo. Análises mais robustas, com um número ainda maior de dados para trabalhar também podem ser feitas na ferramenta, mas requerem um conhecimento mais profundo e a integração outras fontes de dados.

4 RESULTADOS

Foram considerados na análise todos os itens do RI UFSCar depositados até junho de 2023, compreendendo 17.473 itens, distribuídos entre Teses, Dissertações, Dados de Pesquisa (*Dataset*), Trabalhos de Conclusão de Curso de Graduação (TCC), e Artigos.

Analisado a distribuição por tipo de trabalho acadêmico, nota-se que a maior concentração é de Teses e Dissertações, que, juntas, representam 87% do total de publicações, como é possível observar no Quadro 1. Essa representatividade alta acontece por conta dos trabalhos que já estavam na BDTD e foram migrados para o repositório após sua criação (AMARAL *et al*, 2006).

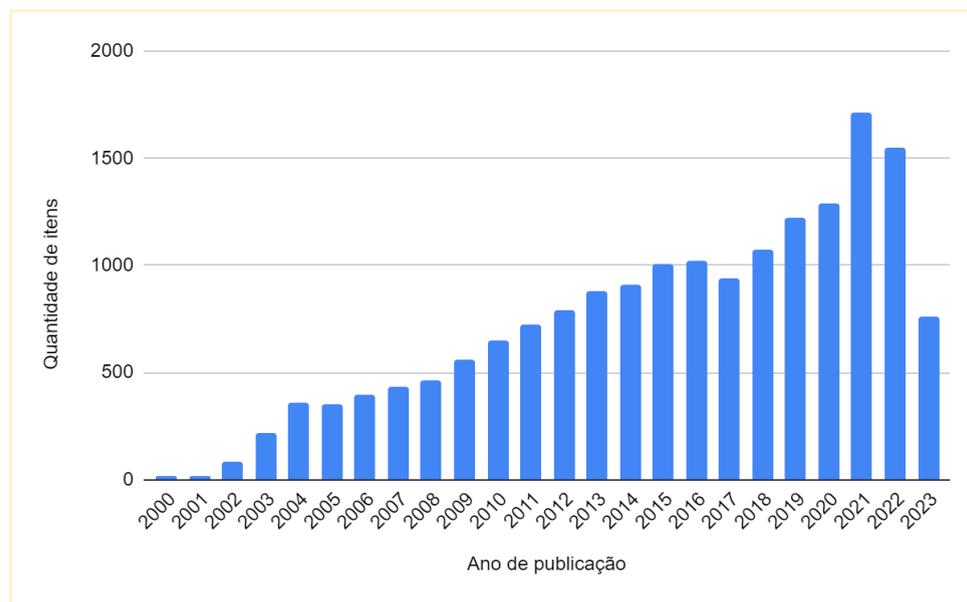
Quadro 1 - Itens do RI UFSCar distribuídos por tipo de publicação

Tipo	Itens	Participação
Dissertação	10356	59,27%
Tese	4911	28,11%
TCC	2073	11,86%
Artigo	69	0,39%
Dataset	44	0,25%
Não definido	20	0,11%
Total	17473	100,00%

Fonte: Elaborado pela autora.

Quanto à distribuição dos trabalhos por ano de publicação (Figura 4), nota-se um crescimento linear, com exceção dos anos de 2017 e 2021, que tiveram uma queda e um significativo aumento, respectivamente. O crescimento visto em 2021 pode ser uma consequência das atualizações das políticas de autodepósito da instituição, realizadas a partir de 2020, que culminaram na criação das coleções *Dataset*, TCC e Artigos.

Figura 4 - Itens do RI UFSCar distribuídos por ano de publicação



Fonte: Elaborado pela autora.

Na distribuição por Área de pesquisa (Quadro 2), definidas pelo CNPQ, nota-se concentração das publicações nas grandes áreas do conhecimento de Ciências Humanas (a de maior destaque), Engenharias, Ciências Exatas e da Terra e Ciências Biológicas. As 4 áreas, juntas, representam 76% das publicações.

Quadro 2 - Itens do RI UFSCar distribuídos por grandes áreas do conhecimento

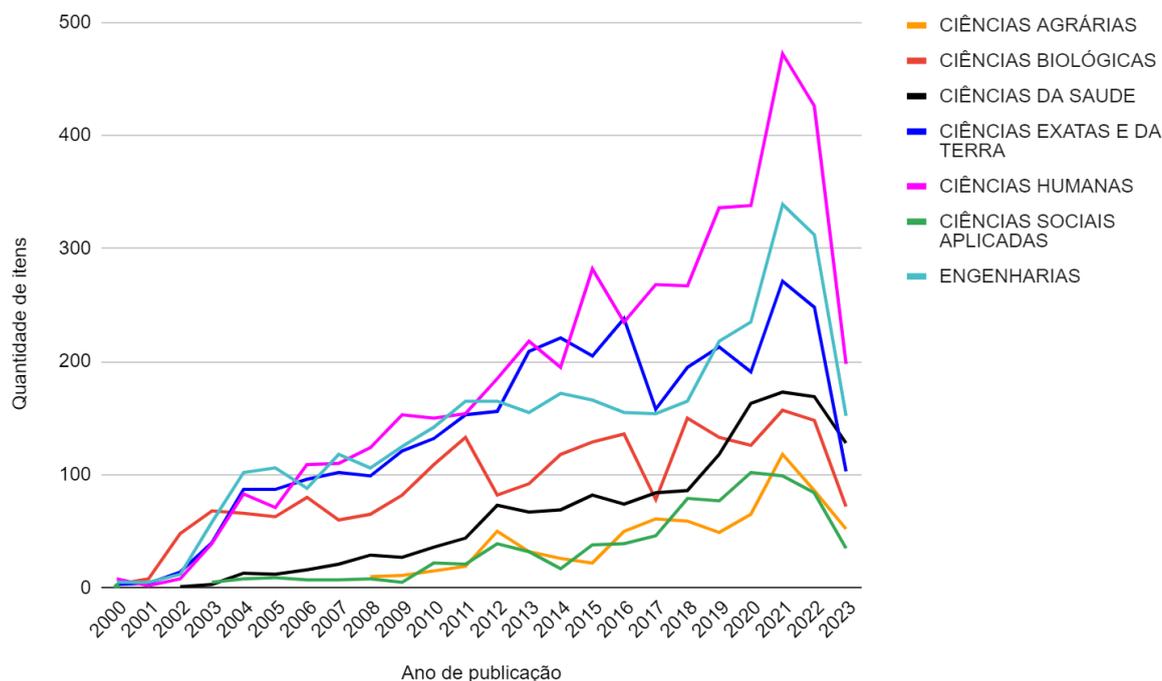
Grandes áreas do conhecimento	Itens	Participação no Total
CIÊNCIAS HUMANAS	4434	25,38%
ENGENHARIAS	3425	19,60%
CIÊNCIAS EXATAS E DA TERRA	3350	19,17%
CIÊNCIAS BIOLÓGICAS	2216	12,68%
CIÊNCIAS DA SAÚDE	1491	8,53%
CIÊNCIAS SOCIAIS APLICADAS	784	4,49%
LINGUÍSTICA, LETRAS E ARTES	742	4,25%
CIÊNCIAS AGRÁRIAS	726	4,15%
OUTROS	283	1,62%
Não definido	22	0,13%

Fonte: Elaborado pela autora.

Na distribuição dessas áreas por ano de publicação (Figura 5), observa-se que mais publicações de Ciências Biológicas foram realizadas nos anos da 2001 a 2003, após, há um aumento das Engenharias, Ciências Exatas e da Terra e Ciências Humanas, que seguem com grande volume até 2012, quando Engenharias têm uma queda na quantidade de publicações, e volta a crescer em 2018.

Ciências Exatas e da Terra e Ciências Humanas seguem com maiores números, até 2016, quando temos um grande aumento das publicações de Ciências Humanas, e uma queda das Ciências Exatas e da Terra. Nota-se que todas as áreas seguiram a mesma curva de crescimento, que aumentou de 2019 para 2020, e teve pico em 2021. Com esse indicador é possível avaliar o crescimento de cada área do conhecimento dentro da UFSCar.

Figura 5 - Itens do RI UFSCar distribuídos por área do conhecimento e por ano



Fonte: Elaborado pela autora.

4.1 Indicadores de Teses e Dissertações (TeD)

O RI UFSCar possui 15.267 Trabalhos de Conclusão de Curso de mestrado e doutorado, que, juntos, formam 87% das publicações depositadas. Desses 87%, 67,8% são dissertações e 32% são teses (Quadro 3). A partir do total, já é possível notar que essa coleção está incompleta. Uma busca simples por “UFSCar” no catálogo de Teses e Dissertações da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) é possível recuperar 19.691 trabalhos, um déficit de 4.424 itens no RI UFSCar.

A falta de completude desses dados, principalmente no período anterior à criação da RI UFSCar pode-se dar pelo fato dos autores optarem pelo depósito de itens físicos na bibliotecas, prejudicando a migração desse material para o ambiente digital. Outro ponto importante é a portaria publicada pela CAPES em 2006, que tornou obrigatória a disponibilização das TeD pelos Programas de Pós-Graduação (PPG) (CAPES, 2006).

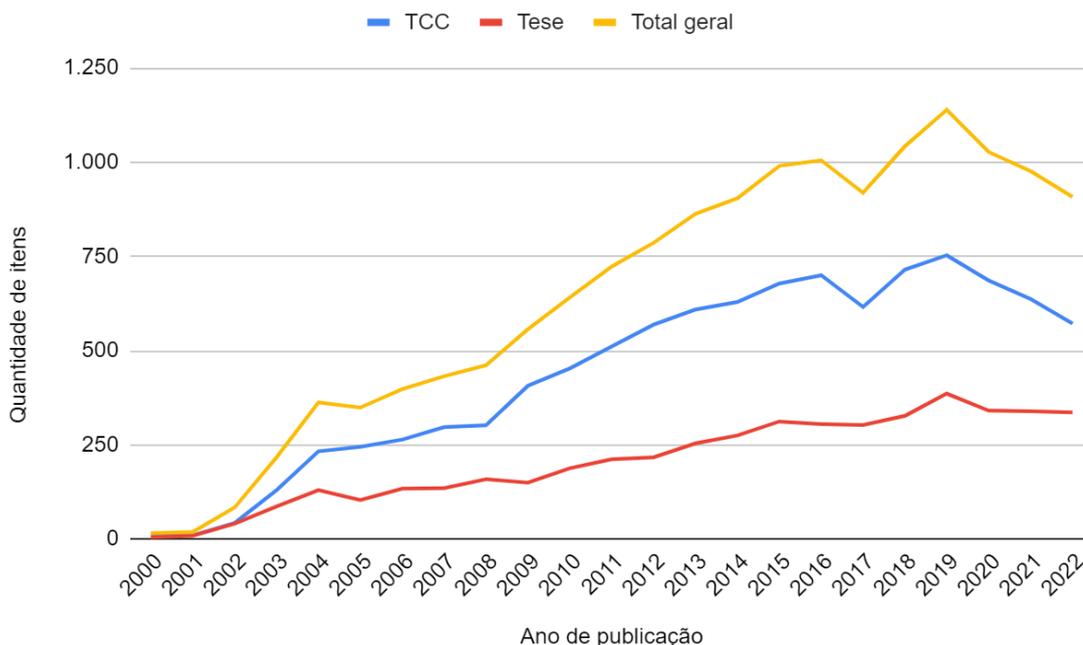
Quadro 3 - Distribuição de itens do RI UFSCar na coleção de Teses e Dissertações

Tipo	Itens	Participação
Dissertação	10356	67,83%
Tese	4911	32,17%

Fonte: Elaborado pela autora.

Analisando a distribuição por anos (Figura 6), vemos que desde 2002 o número de dissertações publicadas sempre foi maior que o de teses. Também é possível notar a mesma queda de 2017 que o gráfico geral de todas as publicações (Figura 5). Diferente do total, nas TeD, observa-se um pico em 2019 e uma queda entre 2020 e 2021. A queda de 2020/2021 possivelmente se deu pela pandemia de Covid-19, as atividades acadêmicas da instituição.

Figura 5 - Teses e Dissertações do RI UFSCar distribuídas por ano de publicação



Fonte: Elaborado pela autora.

Nas grandes áreas do conhecimento (Quadro 4) notas-se que as TeD possuem a mesma distribuição de todo o RI UFSCar, com as principais áreas também sendo Engenharias, Ciências Exatas e da Terra, Ciências Humanas e Ciências Biológicas, que somam 78% das publicações.

Quadro 4 - Teses e Dissertações do RI UFSCar distribuídas pelas grandes áreas do conhecimento

Grandes áreas do conhecimento	Tese	Dissertação	Total	Participação na coleção TeD
CIÊNCIAS HUMANAS	1346	2604	3950	25,89%
CIÊNCIAS EXATAS E DA TERRA	1042	2043	3085	20,22%
ENGENHARIAS	909	2005	2914	19,10%
CIÊNCIAS BIOLÓGICAS	882	1139	2021	13,25%
CIÊNCIAS DA SAÚDE	383	844	1227	8,04%
LINGUÍSTICA, LETRAS E ARTES	156	489	645	4,23%
CIÊNCIAS SOCIAIS APLICADAS	84	542	626	4,10%
CIÊNCIAS AGRÁRIAS	42	486	528	3,46%
OUTROS	63	195	258	1,69%

Fonte: Elaborado pela autora.

Outros importantes indicadores de produção para as TeD são a quantidade de publicações por (PPG) e por orientador. Nos dados analisador, foram encontrados 66 PPG relacionados às publicações de TeD. No Quadro 5 é possível observar os 15 programas com maior número de publicações. Nenhum PPG possui mais de 10% de representatividade em relação ao total de itens do RI UFSCar (todas as coleções).

Quadro 5 - Ranking dos 15 PPG da UFSCar com mais publicações depositadas no RI

PPG	Tese	Dissertação	Total	Participação na coleção TeD
PPG em Educação - PPGE	525	695	1220	7,99%
PPG em Química - PPGQ	546	466	1012	6,63%
PPG em Ecologia e Recursos Naturais - PPGERN	526	362	888	5,82%
PPG em Engenharia de Produção - PPGEp	327	463	790	5,18%
PPG em Educação Especial - PPGEEs	250	427	677	4,43%
PPG em Ciência e Engenharia de Materiais - PPGCEM	248	424	672	4,40%
PPG em Ciência da Computação - PPGCC	91	543	634	4,15%
PPG em Engenharia Química - PPGEQ	274	357	631	4,13%
PPG em Fisioterapia - PPGFt	277	291	568	3,72%
PPG em Linguística - PPGL	143	295	438	2,87%
PPG em Engenharia Urbana - PPGEU	66	347	413	2,71%
Programa Interinstitucional de Pós-Graduação em Ciências Fisiológicas - PIPGCF	177	229	406	2,66%
PPG em Genética Evolutiva e Biologia Molecular - PPGGEv	145	219	364	2,38%
PPG em Sociologia - PPGS	146	192	338	2,21%

Fonte: Elaborado pela autora.

Os dados apresentados, quando completos, indicam quantos mestres e doutores se formam na instituição, em qual área de pesquisa, e qual PPG mais forma pessoas,

uma vez que a apresentação do trabalho é elemento obrigatório para a finalização do curso.

Nos dados por orientador, foram encontrados 1175 orientadores. No quadro 6 pode-se observar os 11 orientadores que mais se destacam por quantidade de trabalhos orientados.

Quadro 6 - Ranking dos 11 orientadores com mais publicações de TeD

Orientador	Tese	Dissertação	Total	Participação na Coleção TeD
Galetti Júnior, Pedro Manoel	26	37	63	0,41%
Mendes, Enicéia Gonçalves	29	28	57	0,37%
Silva, Elson Longo da	33	20	53	0,35%
Souza, Deisy das Graças de	16	35	51	0,33%
Miotello, Valdemir	15	34	49	0,32%
Louzada Neto, Francisco	18	31	49	0,32%
Rose, Julio Cesar Coelho de	28	21	49	0,32%
Aguiar, Mônica Lopes	18	29	47	0,31%
Dias, Angélica Maria Penteado Martins	22	24	46	0,30%
Santos, José Eduardo dos	35	11	46	0,30%

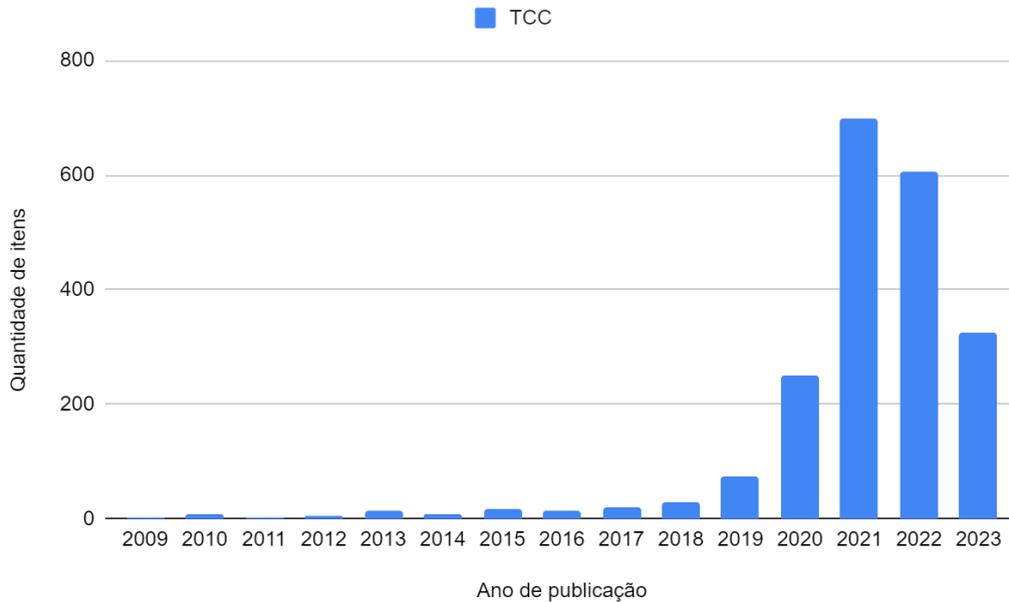
Fonte: Elaborado pela autora.

4.2 Indicadores de Trabalhos de Conclusão de Curso (TCC) dos cursos de graduação

Os TCC da graduação representam 11% das publicações do RI, mas seus depósitos começaram a ter maior volume apenas em 2020 (Figura 6), dada a obrigatoriedade do depósito que começou a ser aplicada no mesmo ano. Apenas em

2021, foram publicados 688 trabalhos. Assim como na coleção de TeD, os indicadores nos mostram quantas pessoas se formam na UFSCar e em qual curso.

Figura 6 - TCC do RI UFSCar distribuídos por ano



Fonte: Elaborado pela autora

A distribuição por área de pesquisa difere um pouco das TeD. Ainda, prevalecem as áreas de Ciências Humanas, Engenharias, e Ciências Exatas e da Terra, mas a quarta posição fica com Ciências da Saúde (Quadro 7). Diferente do que também foi mostrado por outros indicadores, nos TCC a área com mais publicações é Engenharias

Quadro 7 - TCC do RI UFSCar distribuídos por grandes áreas do conhecimento

Grandes áreas do conhecimento	TCC	Participação na Coleção TCC
ENGENHARIAS	508	24,49%
CIÊNCIAS HUMANAS	469	22,61%
CIÊNCIAS EXATAS E DA TERRA	250	12,05%
CIÊNCIAS DA SAÚDE	209	10,08%
CIÊNCIAS AGRÁRIAS	196	9,45%
CIÊNCIAS BIOLÓGICAS	189	9,11%
CIÊNCIAS SOCIAIS APLICADAS	132	6,36%
LINGUÍSTICA, LETRAS E ARTES	95	4,58%
OUTROS	25	1,21%

Fonte: Elaborado pela autora.

Quanto à distribuição por curso de graduação, o curso de maior destaque é Engenharia Química, com 8,2% do total de publicações. No quadro 8 é possível observar os 15 cursos com maior contribuição. A partir desses, quando completos, é possível saber quais cursos mais formam pessoas. É importante destacar a ausência do curso de graduação Bacharelado em Biblioteconomia e Ciência da Informação na lista dos cursos com mais depósitos, que pode estar relacionado a recente obrigatoriedade de depósitos.

Quadro 8 - Ranking dos 15 cursos de graduação da UFSCar com mais depósitos de TCC no RI

Curso	TCC	Participação na Coleção TCC
Engenharia Química - EQ	170	8,20%
Engenharia Agrônômica - EAgr-Ar	118	5,69%
Medicina - Med	89	4,29%
Psicologia - Psi	89	4,29%
Engenharia de Materiais - EMa	78	3,76%
Engenharia Civil - ECiv	77	3,71%
Pedagogia - PedL-So	73	3,52%
Terapia Ocupacional - TO	65	3,14%
Engenharia Mecânica - EMec	63	3,04%
Biotecnologia - Biotec	56	2,70%
Engenharia Ambiental - EAm-LS	53	2,56%
Química - Q	47	2,27%
Engenharia Elétrica - EE	44	2,12%
Engenharia de Computação - EC	44	2,12%
Ciências Biológicas - CB-So	44	2,12%

Fonte: Elaborado pela autora.

Dos orientadores, foi possível recuperar a relação 573 professores. O Quadro 9 traz a relação dos 10 orientadores com mais publicações relacionadas.

Quadro 9 - Ranking dos 10 orientadores com mais publicações de TCC.

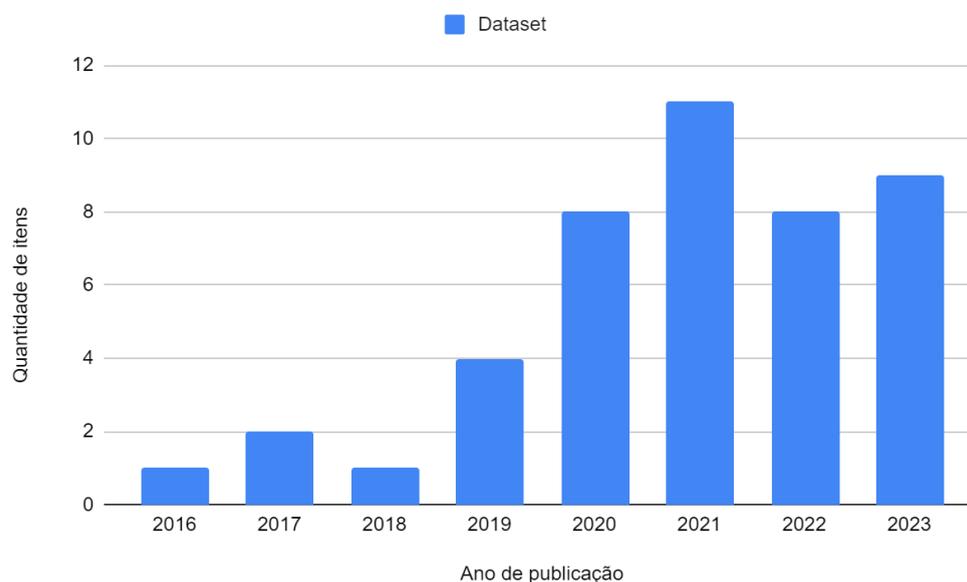
Orientador	TCC	Participação na Coleção TCC
Soares, Marcio Roberto	19	0,92%
D'Affonseca, Sabrina Mazo	17	0,82%
Lombardi, Lucia Maria Salgado dos Santos	16	0,77%
Bernardo, André	16	0,77%
Shiki, Sidney Bruce	16	0,77%
Fontes, Andréa Regina Martins	15	0,72%
Hirosue, Fernando Hideki	15	0,72%
Cabello, Janaina	12	0,58%
da Silva, Antônio Fernando Gouvêa	12	0,58%

Fonte: Elaborado pela autora.

4.3 Indicadores de Dados de Pesquisa (Dataset)

Os dados de pesquisa são os trabalhos menos depositados no RI-UFSCar. São 41 itens, que representam apenas 0,25% do RI. Na Figura 7 é possível observar que houve um aumento na quantidade de publicações a partir de 2019. Na área de pesquisa, a de maior destaque é ciências da saúde (43%), seguida de ciência sextas e da terra (25%) e ciências biológicas (13%) (quadro 10).

Figura 7 - Dataset do RI UFSCar distribuídos por ano de publicação



Fonte: Elaborado pela autora.

Quadro 10 - Dataset do RI UFSCar distribuídos por grandes áreas do conhecimento

Grandes áreas do conhecimento	Dataset	Participação na Coleção Dataset
CIÊNCIAS DA SAÚDE	19	43,18%
CIÊNCIAS EXATAS E DA TERRA	11	25,00%
CIÊNCIAS BIOLÓGICAS	6	13,64%
CIÊNCIAS SOCIAIS APLICADAS	4	9,09%
ENGENHARIAS	2	4,55%
LINGUÍSTICA, LETRAS E ARTES	1	2,27%
CIÊNCIAS AGRÁRIAS	1	2,27%

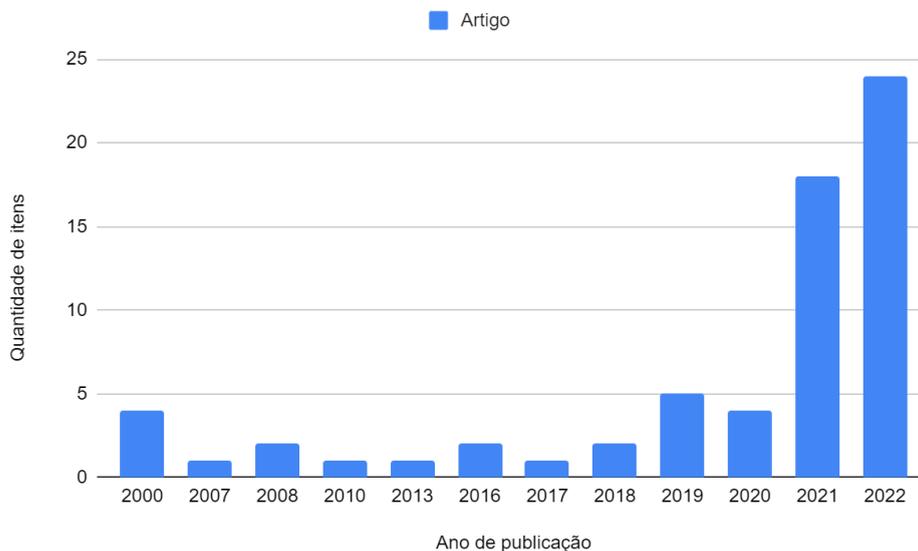
Fonte: Elaborado pela autora.

4.4 Indicadores de Artigos de Periódicos

Olhando para os artigos de periódicos, observa-se um número muito pequena de textos depositados no RI, apenas 69, e que começaram a ter maior volume de depósitos apenas em 2021, um número muito pequeno e que faz pensar quanto material deixa de ser depositado no RI.

Maciel (2018) realizou uma análise bibliométrica, a partir de dados da Plataforma *Lattes*, de artigos em que ao menos um dos autores são docentes dos PPG ativos da UFSCar e recuperou 16.275 artigos de periódicos publicados entre 2000 e 2017. Esses dados apontam para o tamanho da incompletude do RI. Esses indicadores, quando completos, são muito importantes para avaliarmos o avanço ou retrocesso da produção científica da instituição.

Figura 8 - Artigos do RI UFSCar distribuídos por ano de publicação.



Fonte: Elaborado pela autora.

Quanto à área de pesquisa, também se observa uma diferença em relação aos outros tipos de publicação, com a maioria dos artigos (30%) pertencendo às ciências sociais aplicadas. No quadro 11 também é possível notar alguns artigos classificados como não definidos, o que ocorre, provavelmente, pela quebra dos dados durante a extração.

Quadro 11 - Artigos do RI UFSCar distribuídos por grandes áreas do conhecimento.

Grandes áreas do conhecimento	Artigo	Participação na Coleção Artigos
CIÊNCIAS SOCIAIS APLICADAS	21	30,43%
CIÊNCIAS DA SAÚDE	19	27,54%
CIÊNCIAS HUMANAS	14	20,29%
Não definido	9	13,04%
CIÊNCIAS EXATAS E DA TERRA	4	5,80%
ENGENHARIAS	1	1,45%
LINGUÍSTICA, LETRAS E ARTES	1	1,45%

Fonte: Elaborado pela autora.

A quantidade de publicações por autor é um indicador de produção muito importante para a universidade e, principalmente, para os próprios pesquisadores. O Quadro 12 traz a relação dos 10 autores (de um total de 159) com mais publicações de artigos no RI UFSCar.

Quadro 12 - Ranking dos 10 autores com mais artigos depositados no RI UFSCar.

Autores	Artigos	Participação na Coleção Artigos
Silva-Nunes, Mônica da	10	3,88%
Furnival, Ariadne Chloe Mary	8	3,10%
Driusso, Patricia	7	2,71%
Ramalho, Alanderson Alves	7	2,71%
Carvalho, Joelson Gonçalves de	6	2,33%
Roitberg, Guilherme Prado	6	2,33%
Amaral, Roniberto Morato do	5	1,94%
Candido, Silvio Eduardo	5	1,94%
Lima, Paulo Gomes	5	1,94%
Silva, Jordana Barbosa da	5	1,94%

Fonte: Elaborado pela autora.

5 CONCLUSÕES

A pesquisa visou conceber indicadores sobre o RI UFSCar, por meio da análise bibliométrica dos registros bibliográficos de itens depositados pela comunidade acadêmica, contribuindo para a completude de suas coleções, através do uso de indicadores sobre a produção intelectual da instituição nele depositada. A investigação realizada possibilitou a identificação dos principais indicadores sobre o RI UFSCar, por área do conhecimento, por tipologia de coleções (itens) e unidades organizacionais da UFSCar.

Através da literatura, foi possível notar a relevância dos RI para as universidades, uma vez que são importantes fontes de dados e informações científicas e tecnológicas, que além de serem usadas para transparência e visibilidade dos resultados sobre a atuação da comunidade acadêmica, podem ser utilizados para gerar indicadores sobre a produção intelectual e o avanço da ciência, prever tendências e auxiliar nos processos de tomada de decisão.

Em relação aos RI no Brasil, observou-se que há inúmeros repositórios no país, e muitas pesquisam abrangendo a criação e a importância de toda instituição possuir o seu, mas pouco se investiga sobre o pós-implementação em relação ao uso, problemas que podem ocorrer na gestão, e como incentivar o autodepósito, principalmente de itens como artigos de periódicos. Apesar dos progressos, o país enfrenta desafios na implementação do acesso aberto pela via verde, e os levantamentos realizados apontam para a necessidade de investimento e fortalecimento dos RI já existentes.

O RI UFSCar apresenta tendência de crescimento, e é possível notar o aumento dos autodepósitos nos últimos anos, mas ainda há um *deficit* significativo de itens, observável quando se compara a quantidade trabalhos no RI UFSCar com outros catálogos que possuem publicações de discentes e docente da UFSCar.

É importante ressaltar que foi possível notar o desenvolvimento das áreas de pesquisa, dos cursos de graduação e pós-graduação e ter uma visão completa dos indicadores, mas essa visão não representa o total da produção científica da instituição.

A falta de grandes incentivos para o autodepósito, aliada ao processo ainda trabalhoso necessário para realizá-lo e a falta de incentivo ao uso dos RI como fonte de

informação, contribuem para esse *deficit*. A navegabilidade do RI UFSCar também contribui para o pouco uso, pois a busca em sua base de dados é um pouco limitada, o que pode levar os usuários a optarem por pesquisar em outras bases, com melhor desempenho. E em relação aos indicadores, há uma escassez desses elementos e a visualização dos mesmos é confusa.

Os dados obtidos apontam para uma necessidade de ampliar a divulgação dos RI, tanto na comunidade da UFSCar quando em âmbito nacional, e criar mais incentivos ao autodepósito, que precisam também estar aliados a uma melhoria da apresentação e recuperações dos dados e da *interface* do RI UFSCar. É necessário refletir sobre as políticas de Ciência e Tecnologia no Brasil e a relação das práticas da comunidade acadêmica e o acesso aberto.

REFERÊNCIAS

AMARAL, R. M et al. Criação de Indicadores sobre a Biblioteca Digital de Teses e Dissertações da BCo/UFSCar através da análise bibliométrica Automatizada. 2006. São Carlos: Universidade Federal de São Carlos, 2006.

BRUMATTI, J. D. O acesso aberto verde no Brasil: um estudo descritivo da produção científica depositada em repositório institucional. 2016. 119 f. Dissertação (Mestrado Profissional em Biblioteconomia)—Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2016.

CAFÉ, L. et al. Repositórios institucionais: nova estratégia para publicação científica na Rede. In: ENCONTRO NACIONAL DE INFORMAÇÃO EM CIÊNCIAS DA COMUNICAÇÃO, 13., 2003, Belo Horizonte. Anais... Belo Horizonte: INTERCOM, 2003.

CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Portaria nº 034, de 30 de maio de 2006. Institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos. Brasília, 2006. Disponível em: <http://cad.capes.gov.br/ato-administrativo-detalhar?idAtoAdmElastic=822>. Acesso em: 10 out. 2023.

COSTA, S. M. de S.; LEITE, F. C. L. Insumos conceituais e práticos para iniciativas de repositórios institucionais de acesso aberto à informação científica em bibliotecas de pesquisa. In: SAYÃO, L. F. (Org.). Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p.163-202

Center for World University Rankings (CWUR). Global 2000 list by the Center for World University Rankings. 2023. Disponível em: cwur.org/2023.php. Acesso em: 03 ago. 2023.

FARIA, L. I. L. Prospecção tecnológica em materiais: aumento da eficiência do tratamento bibliométrico. São Carlos: UFSCar, 2001. 187 f. Tese (Doutorado em Ciência e Engenharia de Materiais) - Centro de Ciências Exatas e Tecnologia, Universidade Federal de São Carlos, São Carlos, 2001.

FURNIVAL, A. C. M.; HUBBARD, B. Acesso aberto às publicações científicas: vantagens, políticas e advocacy. InCID: Revista de Ciência da Informação e Documentação, v. 2 n. 2, n. 2, p. 160-177, 2011.

GALDINO, R.; GARCIA, L. G.; AMARAL, R. M. Contribuições da bibliometria ao tratamento de dados institucionais não sistematizados de produção científica: o caso do instituto federal de educação, ciência e tecnologia de são paulo (ifsp). Brazilian Journal of Information Science, v. 16, 2022.

GERHARDT, T. E.; SILVEIRA, D. T. Métodos de pesquisa. Porto Alegre: Editora da UFRGS, 2009

GUÉDON, Jean-Claude. Acesso aberto e divisão entre ciência predominante e ciência periférica. In: FERREIRA, S. M. S. P.; TARGINO, M. G. Acessibilidade e visibilidade de revistas científicas eletrônicas. São Paulo: SENAC, 2010. p. 21-77.

IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia). Sobre a BDTD. 2023. Disponível em: <https://bdttd.ibict.br/vufind/>. Acesso em: 14 ago. 2023.

LAINSON, Ralph. Os 350 anos da Royal Society of London. Rev Pan-Amaz Saude, Ananindeua, v. 1, n. 3, p. 1-2, set. 2010.

LAWRENCE, S. Free online availability substantially increases a paper's impact. Nature, 31 May 2001.

LIMA, M. F. Consequência do movimento pelo livre acesso - open access - e o direito à informação científica. In: SAYÃO, L. et al. Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 9-21.

LIMA, R. A.; VELHO, L. M. L. S.; FARIA, L. I. L. Bibliometria e “avaliação” da atividade científica: um estudo sobre o índice h. Perspectivas em Ciência da Informação, Belo Horizonte, v. 17, n. 3, p. 3-17, set. 2012.

MACIEL, Raquel Santos. A Plataforma Lattes como recurso estratégico para a gestão dos Programas de Pós-Graduação: uma análise baseada na produção de artigos científicos. 2018. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos, São Carlos, 2018.

MARCONDES, C. H.; SAYÃO, L. À guisa de introdução: repositórios institucionais e livre acesso. In: SAYÃO, L. et al. Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 9-21.

MELLO, L. C. Análise da produção científica brasileira sobre o conhecimento tradicional. 2013. 134 f. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) – Universidade Federal de São Carlos, São Carlos, 2013.

OKUBO, Y. Bibliometric indicators and analysis of research systems: methods and examples. Paris: ODCE/GD, 1997.3

QUONIAM, L., et al. Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil. Ciência da Informação, Brasília, v.30, n.2, p.24, maio/ago. 2001.

REIS, J. E. Incipiência da disponibilidade de indicadores bibliométricos e altmétricos nos repositórios institucionais brasileiros. 2017. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) – Universidade Federal de São Carlos, São Carlos, 2017.

REIS, J. E.; SPINOLA, A. T. P.; AMARAL, R. M. Incipiência da visualização de indicadores bibliométricos e altmétricos nos repositórios institucionais brasileiros. Em Questão, v. 23, p. 213-234, 2017.

Repositório Institucional UFSCar. 2016 Disponível em: <https://repositorio.ufscar.br/>. Acesso em: 18 jul. 2023.

SCHÖPFEL, Joachim. Open Access to Scientific Information in Emerging Countries. D-Lib Magazine, v. 23, n. ¾, Mar./Apr. 2017.

SPINAK, E. Indicadores cientificos. Ciência da Informação, Brasília, v.27, n.2, p.141-148, maio/ago, 1996.

TARTAROTTI, R. C. D. Representação temática em repositórios institucionais de bibliotecas universitárias: a percepção de gestores e bibliotecários catalogadores-indexadores da USP, UNESP e UNICAMP . RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, SP, v. 18, n. 00, p. e020030, 2020. DOI: 10.20396/rdbci.v18i00.8660803.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. Conselho Universitário. Resolução ConSUni nº 835, de 04 de março de 2016. Dispõe sobre a criação, a política de implantação e a governança do Repositório Institucional da UFSCar. São Carlos: Conselho Universitário, 2016. Disponível em: https://sei.ufscar.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?R29RRw7JFgWgzbsZUdVJoMvGfN9_v70TIHOFRncT_1-J7f6Nu4_ZpfZkuVpGbKQotHOKOckAyTtdO7y4BOHnUho2_nLzq45ULHDGwjmm5MF3E2J_-TUEULUjcWdaBxW. Acesso em: 08 jul. 2023.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. Conselho de Graduação. Resolução nº 322, de 27 de abril de 2020. Dispõe sobre a obrigatoriedade e a responsabilidade de depósito dos Trabalhos de Conclusão de Curso no repositório Institucional da Universidade Federal de São Carlos. São Carlos: Conselho de Graduação, 2020. Disponível em: <https://repositorio.ufscar.br/static/Resolucao-CoG-no-322.pdf>. Acesso em: 15 jul. 2023.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. Sistema Integrado de Bibliotecas. Resolução nº 2, de 16 de dezembro de 2022. Dispõe sobre a Política de Autodepósito das Teses e Dissertações dos Programas de Pós-Graduação Stricto Sensu no Repositório Institucional da UFSCar. São Carlos: Sistema Integrado de Bibliotecas, 2022. Disponível em: https://repositorio.ufscar.br/static/SEI_FUFSCar-0907054-Resolucao-2.pdf. Acesso em: 10 jul. 2023.

WEITZEL, S. R. O mapeamento dos repositórios institucionais brasileiros: perfil e desafios. Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação, v. 24, n. 54, p. 105-123, 2018.