

VISUALIZAÇÃO DE DADOS ABERTOS NO CONTEXTO DA PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO: análise bibliométrica dos estudos defendidos

Visualization of open data in the context of postgraduate in Information Science: bibliometric analysis of the defended studies

Francis Bento Marques⁽¹⁾, Yuri Bento Marques⁽²⁾, Benildes Coura Moreira dos Santos Maculan⁽³⁾ e Renato Rocha Souza⁽⁴⁾

- (1) Programa de Pós-Graduação em Gestão & Organização do Conhecimento - UFMG, Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte - MG, 31270-901, fbmarques@gmail.com.
- (2) Programa de Pós-Graduação em Gestão & Organização do Conhecimento - UFMG, Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte - MG, 31270-901, yuri.marques@ifnmg.edu.br.
- (3) Programa de Pós-Graduação em Gestão & Organização do Conhecimento - UFMG, Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte - MG, 31270-901, benildes@gmail.com.
- (4) Programa de Pós-Graduação em Gestão & Organização do Conhecimento - UFMG, Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte - MG, 31270-901, rsouzaufmg@gmail.com.

Resumo:

A visualização de dados abertos vem se tornando um indicativo de credibilidade por parte das instituições nos diversos segmentos sociais, incluindo as instituições de pesquisa. No âmbito desta temática, este estudo teve como objetivo apresentar o panorama da produção acadêmica na Pós-Graduação brasileira na área da Ciência da Informação. Definiu-se como recorte de análise os estudos defendidos nas teses e dissertações nos Programas de Pós-Graduação dessa área, a partir da análise das seguintes facetas: ano, modalidade do programa, grau acadêmico dos cursos, tipos de produção, programas e instituições ofertantes. Trata-se de um estudo descritivo e quantitativo, cujos dados foram coletados por meio de um levantamento bibliométrico realizado no Catálogo de Teses e Dissertações, no âmbito do Plano de Dados Abertos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Os resultados revelaram um crescimento gradativo entre a produção geral, os programas e os cursos, prevalência das dissertações e uma distribuição heterogênea entre os programas e as instituições, localizadas nas diferentes regiões brasileiras. Concluiu-se que as pesquisas defendidas nos Programas de Pós-Graduação estão visíveis no Plano de Dados Abertos, que a área da Ciência da Informação está em desenvolvimento e que os dados permitem análises transversais que podem beneficiar as políticas educacionais.

Palavras-chave: visualização de dados abertos; programas de Pós-Graduação; Ciência da Informação; plano de dados abertos da Capes.

Abstract:

The visualization of open data has become an indicator of credibility on the part of institutions in different social segments, including research institutions. Within the scope of this theme, this study aimed to present the panorama of academic production in Brazilian Graduate Studies in the area of Information Science. The studies defended in the theses and dissertations in the Postgraduate Programs in this area were defined as an analysis cut, based on the analysis of the following facets: year, type of program, academic degree of the courses, types of production, programs and offering institutions. This is a descriptive and quantitative study, whose data were collected through a bibliometric survey carried out in the Catalog of Theses and Dissertations, within the scope of the Open Data Plan of the Coordination for the Improvement of Higher Education Personnel. The results revealed a gradual growth between the general production, programs and courses, prevalence of dissertations and a heterogeneous distribution between programs and institutions, located in different Brazilian regions. It was concluded that the research defended in the Graduate Programs are visible in the Open Data Plan, that the area of Information Science is in development and that the data allow cross-sectional analyzes that can benefit educational policies.

Keywords: open data visualization; Graduate programs; Information Science; Capes open data plan.

1 Introdução

A visualização de dados constitui uma estratégia de comunicação e de transparência acerca do que é produzido pelas instituições. No âmbito científico, essa tendência fortaleceu-se nos últimos anos, por diversos fatores, como: a necessidade de se

preservar a memória institucional, o grande volume de dados produzidos, a facilidade oportunizada pelas tecnologias e a obrigação em dar transparência às atividades desenvolvidas, sobretudo em instituições do setor público.

Silva (2018) corrobora esse argumento, mencionando que as instituições públicas de pesquisa, ao disponibilizarem os dados produzidos, comunicam à sociedade os resultados de suas ações, com a finalidade de justificar os investimentos e garantir que tais dados sejam reutilizados.

A fim de permitir que os dados produzidos sejam comunicados à sociedade, essas instituições sistematizam esses dados em repositórios, muitas vezes, em ambientes digitais, dando a eles acesso aberto (disponibilizado gratuitamente, sem restrições). Nesses espaços, a sociedade pode acessar os dados e realizar análises e visualizações, que tornam as informações claras e promovem o entendimento sobre “[...] as interconexões e os relacionamentos causa-efeito que estão por trás de cada atividade ou conjunto de políticas públicas” (SILVA, 2018, p. 523).

O acesso aos dados abertos promove a transparência pública, legitima o trabalho das instituições e permite a participação da sociedade, sobretudo com o uso de ferramentas digitais de manipulação e gerenciamento de dados (MACEDO *et al.*, 2020). No contexto científico, a disponibilização dos dados de pesquisa pelas instituições de fomento, por exemplo, além de enaltecer o trabalho da instituição, demonstra o desenvolvimento das diversas atividades realizadas pelas áreas do conhecimento.

Uma das agências de fomento que se preocupa com a disponibilização dos dados de pesquisa é a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Essa agência, no ano de 2017, elaborou o Plano de Dados Abertos (PDA), com o intuito de publicizar as suas ações e estratégias, que nortearão as atividades de implementação e promoção da abertura de dados. A priorização dos conjuntos de dados a serem disponibilizados foi definida considerando a relevância das informações para o cidadão e o grau de maturidade dos conjuntos de dados dentro da instituição (CAPES, 2022).

Na visão de Torino, Trevisan e Vidotti (2019), os dados abertos da CAPES representam uma iniciativa de adequação à Lei de Acesso à Informação, Lei n. 12.527,

de 18 de novembro de 2011, que determinou a obrigatoriedade dos órgãos públicos em dar transparência de suas atividades aos cidadãos. Além disso, a legislação exigiu às instituições a facilidade de acesso, “[...] no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas” (BRASIL, 2011, não paginado).

Assim, o PDA pode ser considerado uma importante estratégia para demonstração das atividades, do desempenho, dentre outros elementos que caracterizam uma área do conhecimento. Mesmo apresentando falhas no que tange ao tratamento dos metadados, conforme resultados do estudo de Torino, Trevisan e Vidotti (2019, p. 45), no PDA, “[...] todos os dados estão estruturados e representados para que possam ser facilmente compreendidos pelos consumidores [...]”.

Nesse sentido, entende-se que os dados disponibilizados pelas agências possibilitam a medição do desempenho das áreas científicas, por meio de diferentes indicadores como: produtividade, autores envolvidos, instituições, temáticas, períodos, colaborações, dentre outras categorias. Esse processo pode ser concretizado por meio dos estudos bibliométricos que, segundo Wolski *et al.* (2021), constituem o conjunto de técnicas estatísticas que mensuram índices de produção e disseminação do conhecimento científico.

As análises dos conjuntos de dados abertos oportunizam novas verificações e práticas. Para Gonçalves *et al.* (2018), o trabalho com esses dados pode ser realizado por diferentes áreas do conhecimento, tendo em vista identificar o conhecimento útil em grandes volumes de dados. Os autores mencionam que conjuntos específicos de dados poderão proporcionar descobertas valiosas, a partir do desdobramento de novos estudos. Assim, nota-se que a abrangência do PDA vislumbra oportunidades de mapeamento das características de cada área do conhecimento, embora a literatura ainda padeça de falta de estudos que quantifiquem os dados referentes a áreas específicas, como a Ciência da Informação, por exemplo.

Nesse contexto, o presente estudo versa sobre a visualização de dados abertos à luz do PDA e com foco na área da Ciência da Informação. O estudo constitui o desdobramento de uma tese sobre visualização de dados e apresenta dados preliminares sobre os Programas de Pós-Graduação e suas instituições no âmbito dessa área do conhecimento.

Justifica-se a escolha desta temática por seu ineditismo e por sua importância. Para a Ciência da Informação, analisar o PDA representa mais uma estratégia de conhecer o desenvolvimento dessa área, sobretudo por ser uma área nova, ainda em processo de consolidação. Pelos resultados, tornar-se-á possível conhecer o panorama das pesquisas dessa área, de modo que as instituições e os pesquisadores fortaleçam os investimentos despendidos e ampliem as práticas de pesquisa.

2 Objetivos

A partir da temática delimitada para este estudo, a proposta direciona-se à representatividade da área da Ciência da Informação, pois investiga-se se essa área pode ser identificada por meio da análise ao conjunto de dados do PDA.

O objetivo central do estudo é apresentar o panorama científico da produção acadêmica referente às teses e dissertações na Pós-Graduação brasileira, na área da Ciência da Informação.

3 Procedimentos metodológicos

A metodologia utilizada se caracteriza como descritiva, porque expõe os detalhes sobre o objeto investigado, com foco em suas características e considerando variáveis a ele relacionadas (GIL, 2010). Tem abordagem quantitativa, visto que os dados são analisados em sua objetividade, por meio de técnicas estatísticas que mensuram as variáveis (CRESWELL, 2007).

Quanto aos procedimentos técnicos, aplicou-se um estudo bibliométrico que, neste estudo, se refere a um levantamento de informações sobre uma área do conhecimento, com o propósito de identificar aspectos característicos dessa área. Esse levantamento é realizado considerando vários atributos, o que facilita a identificação

de tendências (QUEVEDO-SILVA *et al.*, 2016).

O primeiro CTD refere-se às teses e dissertações defendidas entre 1987 e 2012. São disponibilizados dados ou variáveis referentes a: resumo, palavras-chaves, nome do discente, nome da instituição de ensino, unidade federativa, entre outros. O segundo catálogo abrangeu o período de 2013 a 2016 e contém 56 tipos de dados, sendo adicionados códigos de identificação para instituição, discente, docente, data de entrada e saída do programa, o endereço para o texto completo da dissertação ou tese na Plataforma Sucupira, dentre outros detalhes. Por fim, no terceiro CTD, os dados referem-se aos anos de 2017 a 2020 e possuem 58 variáveis. Nesse catálogo, foram adicionados mais dois campos em relação ao segundo, que mostram a existência de estudos vinculados a essa produção (como artigos de periódicos, de eventos, dentre outros) e, caso haja, qual é a identificação em outra base de dados (CAPES, 2022).

A partir da identificação e caracterização desses catálogos, delimitou-se o recorte de análise. Assim, foram selecionados os dados do código “60700009”, referentes à Ciência da Informação. Nesse processo, os dados foram extraídos em correlação com as variáveis ou categorias a serem mapeadas.

A coleta de dados foi realizada de modo manual, com *download* de 8 arquivos em formato CSV. Devido ao tamanho dos arquivos, foi utilizado o programa Openrefine para abrir os mesmos e filtrar os trabalhos da área do conhecimento “60700009”. Após o filtro aplicado, foi criado um arquivo com teses e dissertações da área selecionada. Assim, foi analisada a quantidade de teses e dissertações defendidas, considerando oito categorias a que elas foram distribuídas: 1 - ano de defesa, 2 - modalidade de programa, 3 - modalidade de programa por anos, 4 - grau acadêmico do curso, 5 - grau acadêmico por anos, 6 - tipos de produção, 7 – programas e 8 – instituições.

Para análise dos dados, recorreu-se ao uso das técnicas de Estatística Descritiva (para descrever e resumir conjuntos de dados), cujos resultados foram expostos no formato de gráficos e confrontados com estudos correlatos à temática. No tratamento

dos dados para visualização gráfica, adotou-se a linguagem Python e bibliotecas, como Pandas, Seaborn, Matplotlib e Plotly.

4 Resultados

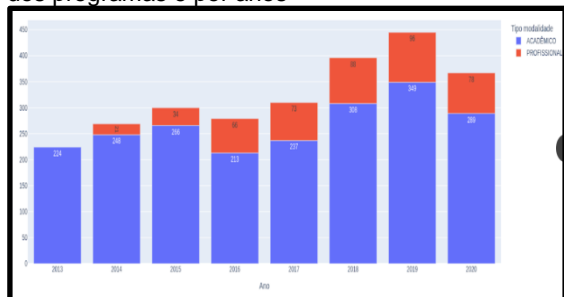
Foram identificados um total de 2590 teses e dissertações no período de 2013 a 2020. Ao distribuir esse quantitativo por ano de defesa, os dados indicaram o ano de 2019 como o mais produtivo (445 estudos), seguido pelo ano de 2018 (396) e 2020 (367). Os demais anos apresentaram defesas oscilante entre 224 (em 2012) e 310 (em 2017).

Esse resultado evidencia, em linhas gerais, um crescimento gradativo de defesas na área. Para Lança *et al.* (2018), essa área vem crescendo como campo de pesquisa em virtude de um aumento da produção dos programas. Assim, “[...] ganha cada vez mais espaço devido às suas contribuições relacionadas à gestão do conhecimento científico e tecnológico [...]” (LANÇA *et al.* 2018, p. 4555). Nota-se decréscimo (em relação ao ano anterior) apenas nos anos: 2016 e 2020. Neste último ano, o mundo vivia a dura realidade de uma pandemia.

A segunda categoria diz respeito à modalidade de programa expressa nos dados, que classifica os Programas como Acadêmico ou Profissional. Os resultados identificaram 2134 teses e dissertações na modalidade acadêmica e 456 na profissional, cabendo ressaltar que na Ciência da Informação ainda não há programas com doutorado na modalidade profissional.

Para a terceira categoria, referente a modalidade de programa por anos, foram mapeadas as defesas por anos, conforme mostra o Gráfico 1.

Gráfico 1 – Distribuição dos estudos por modalidades dos programas e por anos



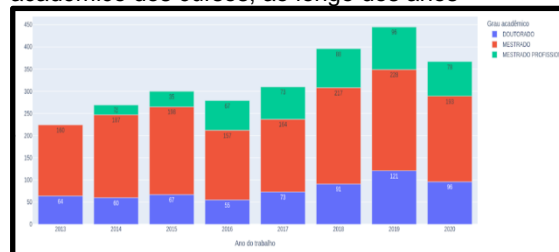
Fonte: Dados da pesquisa (2022).

Percebe-se, um crescimento gradativo, similar à distribuição das defesas por ano

(sem considerar a divisão por modalidades). Na modalidade acadêmica, houve diminuição em dois anos (2016 e 2020), e apenas um ano (2020) para a modalidade profissional. Nota-se, também, um crescimento gradativo quando se individualiza por modalidade (acadêmica e profissional), o que contribuiu, de modo proporcional, para o crescimento da área como um todo. O número menor de programas profissionais pode ser pelo fato de que essa modalidade é mais recente, embora sejam bastante adequados à Ciência da Informação, por direcionar-se à resolução de problemas empíricos vivenciados nas instituições (LANÇA *et al.*, 2018).

Por sua vez, na quarta e quinta categorias, os estudos foram distribuídos por grau acadêmico do curso e grau acadêmico por anos. Os dados mostram que o Mestrado acadêmico produziu 1504 defesas, o Doutorado acadêmico 627 e o Mestrado Profissional 459, com crescimento gradativo ao longo dos anos. Para o Mestrado e Doutorado acadêmicos, constata-se decréscimo de defesas nos anos 2016 e 2020, e, para o Mestrado Profissional, a diminuição ocorreu, apenas, em 2020, conforme exposto no Gráfico 2.

Gráfico 2 – Distribuição dos estudos por grau acadêmico dos cursos, ao longo dos anos



Fonte: Dados da pesquisa (2022).

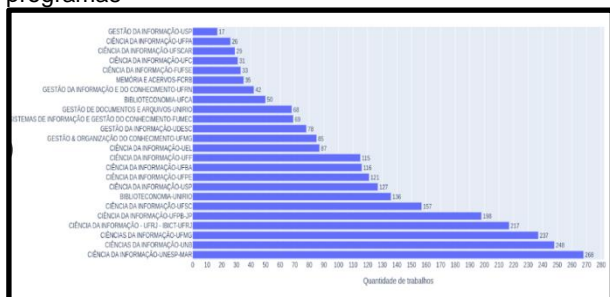
Os dados do Gráfico 2 possibilitam a mesma revelação dos dados contidos no Gráfico 1, de que os programas e cursos da área se fortaleceram, ao longo dos anos, haja vista o crescimento do número de pesquisas. Assim, nota-se uma relação de proporcionalidade entre os crescimentos gradativos dessas duas variáveis com o crescimento gradativo do número de estudos como um todo, sem considerar as variáveis. Costa, Oliveira e Araújo (2019) também levantaram um número expressivo de estudos, constatando que há forte impacto acadêmico deles no âmbito da ciência, tanto nacional quanto internacional. Martins *et al.*

(2015) destacaram que o aumento dos estudos está relacionado às demandas e oportunidades do mercado, sendo que o Mestrado predomina, sobretudo, por sua tradição como o primeiro nível de Pós-Graduação, como também por ser um curso de curta duração.

Na categoria seis, a distribuição dos estudos tipos de produção revelou a predominância do tipo dissertação (1907 estudos) em detrimento às teses (627), produto/processo/técnica (48), projeto técnico (6), editoria (1) e relatório final de pesquisa (1). Esse resultado pode ser explicado por existir uma quantidade maior de cursos de Mestrado. Ressalta-se que as teses e dissertações representam o principal produto da Pós-Graduação, pois relatam o estado da arte sobre um determinado assunto, apresentam um rico material metodológico e norteiam a condução de novas descobertas (COSTA; OLIVEIRA; ARAÚJO, 2019).

Para a sétima categoria, verificou-se a distribuição dos estudos por Programas de Pós-Graduação. Foi possível mapear a existência de 24 programas, os quais apresentam diversidade no número de estudos defendidos, conforme ilustrado no Gráfico 3.

Gráfico 3 – Distribuição dos estudos defendidos por programas



Fonte: Dados da pesquisa (2022).

Resultados semelhantes são alcançados quando se analisa a oitava categoria, sobre a distribuição dos estudos por instituições, sendo as três mais produtivas: Universidade Federal de Minas Gerais (322 estudos), Universidade Estadual Paulista (268) e Universidade de Brasília (248), e as três menos produtivas, as universidades: Federal do Pará (26), Federal de São Carlos (29) e Federal do Ceará (31).

Para a sétima e oitava categorias (por programas e por instituições), além da diversidade quanto à quantidade de estudos

defendidos, constatou-se que estão presentes em todas as regiões brasileiras. Esse resultado também foi constatado por Moreira e Ferneda (2020), ao relatarem que alguns programas reúnem mais da metade do total de produção científica da área.

5 Conclusão

Com este estudo foi possível apresentar um panorama inicial da produção acadêmica defendida na Pós-Graduação brasileira na área da Ciência da Informação à luz do PDA. Os resultados revelam um crescimento gradativo entre a produção geral, os programas e os cursos; a prevalência das dissertações; e uma distribuição heterogênea entre os programas e as instituições localizadas nas diferentes regiões brasileiras.

Concluiu-se que o PDA possibilita a visualização dos estudos em diferentes áreas do conhecimento. Por meio do PDA foi possível evidenciar que a Ciência da Informação vem se fortalecendo ao longo do tempo. Essa conclusão evidencia o crescente prestígio da área e dá transparência quanto aos investimentos públicos despendidos.

A ausência de estudos correlatos para fortalecer a discussão dos resultados constituiu a limitação da pesquisa. Para pesquisas posteriores, sugere-se analisar as temáticas dos estudos desenvolvidos, como também mapear e comparar mais de uma área do conhecimento.

Referências

BRASIL. **Lei n. 12.527**, de 18 de novembro de 2011. Regula o acesso a informações [...]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 08 set. 2022.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). **Plano de dados abertos**. Brasília, set. 2022. Disponível em: <https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/PlanodeDadosAbertosdaCAPES20202022.pdf>. Acesso em: 08 set. 2022.

COSTA, Belkiz; OLIVEIRA, Marlene; ARAÚJO, Ronaldo. Impactos das teses e

dissertações do Programa de Pós-Graduação em Ciência da Informação da UFMG. **Informação em Pauta**, Fortaleza, v. 4, n. 2, p. 11-31, jul./dez. 2019. Disponível em: <http://www.periodicos.ufc.br/informacaoempauta/article/view/42444/99878>. Acesso em: 08 set. 2022.

CRESWEL, John. **Projeto de pesquisa: método qualitativo, quantitativo e misto**. 2. ed. Porto Alegre: Artmed, 2007.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

GONÇALVES, Alexandre Leopoldo *et al.* Análise de agrupamentos sobre textos. *In*: CONGRESSO INTERNACIONAL DE CONHECIMENTO E INOVAÇÃO, 8., 2018, Guadalajara. **Anais [...]**. Guadalajara: Universidade de Guadalajara, 2018. p. 1-14. Disponível em: <https://proceeding.ciki.ufsc.br/index.php/ciki/article/view/589/246>. Acesso em: 08 set. 2022.

LANÇA, Tamie Aline *et al.* Produção científica dos Programas de Pós-Graduação em Ciência da Informação na Plataforma Lattes. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais [...]**. Londrina: UEL, 2018. p. 4555-4575.

MACEDO, Daiane *et al.* Uma ferramenta para recomendação de visualização de dados governamentais abertos. *In*: WORKSHOP DE COMPUTAÇÃO APLICADA EM GOVERNO ELETRÔNICO, 8., 2020, Niterói. **Anais [...]**. Niterói: UFF, 2020. p. 1-12. Disponível em: <https://sol.sbc.org.br/index.php/wcge/article/view/11261/11124>. Acesso em: 07 set. 2022.

MARTINS, Ana Carolina de Melo. Mestrado Profissional na área de Ciência da Informação no Brasil. **Revista ACB**, Florianópolis, v. 20, n. 3, p. 411-422, set./dez. 2015. Disponível em: <https://revistaacb.emnuvens.com.br/racb/article/view/1082>. Acesso em: 08 set. 2022.

MOREIRA, Jonathan Rosa; FERNEDA, Edberto. Produção científica nos Programas de Pós-Graduação nas áreas de informação no Brasil. **Informação & Informação**, Londrina, v. 25, n. 4, p. 1-20, jul./dez. 2020. Disponível em: <https://uel.br/revistas/uel/index.php/informacao/article/view/39670/pdf>. Acesso em: 08 set. 2022.

QUEVEDO-SILVA, Filipe *et al.* Estudo bibliométrico: orientações sobre sua aplicação. **Revista Brasileira de Marketing**, São Paulo, v. 15, n. 2, p. 246-262, abr./jun. 2016. Disponível em: <https://www.redalyc.org/pdf/4717/471755312008.pdf>. Acesso em: 08 set. 2022.

SILVA, Fabiano Couto Corrêa da. Visualização de dados abertos no setor público. **Revista Ibero-americana de Ciência da Informação**, Brasília, v. 11, n. 2, p. 523-540, maio/ago. 2018. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/8341/9632>. Acesso em: 07 set. 2022.

TORINO, Emanuelle; TREVISAN, Gustavo Lunardelli; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Dados abertos CAPES: um olhar à luz dos desafios para publicação de dados na web. **Ciência da Informação**, Brasília, v. 48, n. 3, p. 38-46, set./dez. 2019. Disponível em: <https://brapci.inf.br/index.php/res/download/139033>. Acesso em: 08 set. 2022.

WOLSKI, Luciano Zamperetti *et al.* Mineração de texto e clusterização em estudos bibliométricos. *In*: CONGRESSO INTERNACIONAL DE CONHECIMENTO E INOVAÇÃO, 11., 2021, Maringá. **Anais [...]**. Maringá: Unicesumar, 2021. p. 1-15. Disponível em: <https://proceeding.ciki.ufsc.br/index.php/ciki/article/view/1036>. Acesso em: 07 set. 2022.