

A gestão de dados de pesquisa no contexto da COVID-19: análise de requisitos em situação de emergência em saúde pública

The management of research data in the context of COVID-19: analysis of requirements in emergency situations in public health

Luisi Maria Costa de Oliveira
Cícera Henrique da Silva
Viviane Veiga

Resumo

Este estudo tem como objetivo principal mapear e analisar as diretrizes para a gestão de dados de pesquisa no contexto da COVID-19. Utiliza levantamento documental para mapear as diretrizes para gestão de dados de pesquisa no contexto da COVID-19 e no anterior à pandemia. O trabalho empírico consiste na análise dos documentos emitidos pela Comissão Europeia a partir de um modelo de avaliação de políticas de dados de pesquisa que contempla as categorias grau de aplicação do documento e cobertura de critérios. Percebe-se que o cenário atual impõe maiores compromissos aos pesquisadores e ações mais imediatas com o gerenciamento de dados. Conclui-se que a gestão de dados é um fator preponderante para fornecer acesso aos dados de pesquisa em saúde, contribuindo para a geração de novos conhecimentos e benefícios para a sociedade.

Palavras-chave: *Gestão de Dados de Pesquisa, COVID-19, Saúde Pública.*

Abstract

This study has as main objective to map and analyze the guidelines for the management of research data in the context of COVID-19. It uses documentary survey to map the guidelines for management of research data in the context of COVID-19 and before the pandemic. The empirical work consists of analyzing the documents issued by the European Commission based on a model for evaluating research data policies that includes the categories coverage of criteria and the degree of application of the document. It is clear that the current scenario imposes greater commitments on researchers and more immediate actions with data management. It is concluded that data management is a major factor to provide access to health research data, contributing to the generation of new knowledge and benefits for society.

Keywords: *The management of research data, COVID-19, Public Health.*

1 Introdução

Os dados de pesquisa têm assumido protagonismo central no processo de comunicação na ciência, dado o seu potencial de apoiar a transparência, reprodutibilidade na pesquisa e possível reuso para nova produção de conhecimento. Nos últimos anos, instituições acadêmicas, financiadores da pesquisa em saúde e os principais periódicos da área emitiram declarações nas quais ratificam que as estratégias de gestão de dados de pesquisa (GDP) são preponderantes para o compartilhamento, acesso e reuso de dados, de modo a maximizar o conhecimento e os seus benefícios para a saúde, sobretudo em situação de emergência em saúde pública de interesse internacional (ESPII) (WELLCOME TRUST, 2012; 2016).

Em 30 de janeiro de 2020, a Organização Mundial da Saúde (OMS) declarou que o surto do novo coronavírus (Sars-Cov-2) constituía uma ESPII, motivando a comunidade científica a reforçar o compromisso com o compartilhamento e acesso aos dados de pesquisa de forma rápida e gratuita visando auxiliar em uma resposta global (WELLCOME TRUST, 2020).

Apesar disso, o cenário pandêmico da COVID-19 ainda testemunha muitos dos reconhecidos desafios para se garantir o acesso e uso de dados de maneira equitativa, ética e eficiente. O estudo de Pisani, Ghataure e Merson (2018) assinala que a necessidade de compartilhar os dados rapidamente para a comunidade de pesquisa em situações de emergência em saúde pública (ESP) pode dificultar a adequada gestão e curadoria dos dados, de modo a torná-los imediatamente localizáveis, acessíveis, interoperáveis e reutilizáveis (FAIR – *Findable, Accessible, Interoperable, Reusable*). Cox e Verbaan (2018) identificam que grande parte do esforço está sendo colocado nos próprios pesquisadores. No entanto, a maioria das ações realizadas no escopo da gestão de dados, como o compartilhamento e abertura, requer infraestruturas e incentivos que não são controlados por pesquisadores individuais e grupos de pesquisa, mas por financiadores ou pelo próprio ecossistema de pesquisa em geral (PISANI; GHATAURE; MERSON, 2018).

Diante dessas considerações iniciais, coloca-se a questão da pesquisa: como os principais financiadores da pesquisa em saúde orientam os pesquisadores sobre a GDP produzidos em situações de ESP, especificamente no contexto da COVID-19? O objetivo geral da pesquisa é mapear e analisar as diretrizes para a GDP no contexto da COVID-19. Os objetivos específicos são: a) identificar os principais documentos normativos para GDP emitidos pelos financiadores da pesquisa em saúde; b) avaliar esses documentos no que diz respeito ao grau de aplicação e

cobertura de critérios; c) comparar as diretrizes para gestão de dados de pesquisa antes e durante a pandemia de COVID-19. No momento em que o mundo vivencia uma pandemia de graves proporções, faz-se necessária uma reflexão da comunidade científica na utilização de estratégias de GDP que podem otimizar a resposta da ciência para sociedade, acelerando o progresso de pesquisas em vacinas e tratamentos. Verificou-se que no contexto da COVID os requisitos dos financiadores da pesquisa exigem o acesso imediato aos dados de pesquisa e que a GDP se tornou um instrumento relevante para o enfrentamento da pandemia.

2 Metodologia

A metodologia utilizada foi levantamento bibliográfico para a construção das bases teóricas da pesquisa e pesquisa documental para analisar e comparar as diretrizes para a GDP no contexto da COVID-19 e no anterior. O trabalho empírico consistiu na análise de documentos emitidos pela Comissão Europeia (CE), o segundo maior financiador da pesquisa em saúde (VIERGEVER; HENDRIKS, 2016). Justifica-se ainda a escolha pela CE tendo em vista a preponderância do órgão no cenário internacional da pesquisa no que diz respeito às iniciativas ligadas ao movimento de acesso aberto e ciência aberta.

Após buscas realizadas nos sítios eletrônicos da CE, os documentos mapeados foram sistematizados no editor da Microsoft *Excel* (2010) e analisados conjuntamente a partir de uma estrutura de modelo de avaliação de políticas de dados de pesquisa que contempla as categorias grau de aplicação do documento (recomendatório, mandatário ou não identificado) e cobertura de critérios (total, parcial ou nenhuma), desenvolvida com base nos modelos de Hodson e Molloy (2014), Horton e Digital Curation Centre (2016), Digital Curation Centre (2019) e Monteiro (2019). Após análise, foi realizado cotejamento entre as diretrizes para gestão de dados de pesquisa antes e durante a pandemia da COVID-19.

3 Fundamentação teórica

Nas últimas duas décadas, a GDP tornou-se um tópico relevante na agenda da pesquisa nacional e internacional. A discussão foi alavancada por fatores como o surgimento de uma ciência intensiva em dados (*e-Science*); o imperativo social e econômico para viabilizar o acesso aos resultados da pesquisa impulsionado pelo movimento da ciência aberta; o reconhecimento por órgãos financiadores de que dados bem gerenciados e preservados são

condições fundamentais para seu compartilhamento, acesso e reuso, visando maximizar o investimento público (BRINEY, GOBEN, ZILINSKI, 2017; COX; VERBAAN, 2018).

Esse cenário foi circunscrito pelo surgimento de uma série de recomendações e exigências emitida por financiadores da pesquisa de forma a reordenar as práticas de pesquisa e a atitude dos pesquisadores em relação à GDP e ao compartilhamento (SALES; SAYÃO, 2018). Maior atenção foi dada à questão após o Plano de Gestão de Dados (PGD) se tornar um documento obrigatório para a concessões de recursos de grande parte dos órgãos de fomento. Conforme explica Michener (2015), o PGD é um documento formal anexo às propostas de pesquisas, no qual os pesquisadores devem descrever os procedimentos adotados para o tratamento dos dados de pesquisa gerados desde os primeiros estágios até após a finalização do projeto.

Para a Organisation for Economic Co-operation and Development – OECD (2007, p. 13) dados de pesquisa são “registros factuais (resultados numéricos, registros textuais, imagens e sons) utilizados como fontes primárias para pesquisa científica e que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa”. Uma das principais características dos dados de pesquisa é sua heterogeneidade, visto que são provenientes de específicos domínios disciplinares, criados para diferentes propósitos, comunidades acadêmicas, tecnologias e processo (SAYÃO; SALES, 2015). Portanto, pensar em estratégias de GDP requer considerar as variadas tipologias que os dados de pesquisa podem assumir. Ela envolve um conjunto de atividades associadas ao ciclo de vida dos dados, envolvendo o projeto e criação, armazenamento, segurança, preservação, recuperação, compartilhamento e reutilização, considerações éticas, legais e estruturas de governança para os dados de pesquisa (PINFIELD; COX; SMITH, 2014).

Na pesquisa em saúde, torna-se fundamental considerar as questões éticas e legais quando compartilhamento e abertura de dados tornam-se o ímpeto para atender as aspirações de um cenário de ESP, como a da COVID-19. É preciso pensar em estratégias para GDP que estejam alinhadas a procedimentos de consentimento apropriados, aos processos de desidentificação e anonimização de dados e às legislações vigentes para a salvaguarda da privacidade dos sujeitos da pesquisa. Além disso, deve-se considerar o equilíbrio entre os possíveis ganhos e danos com o compartilhamento e abertura de dados de pesquisa em saúde.

4 Análise dos resultados

A CE é um órgão executivo da União Europeia (UE) responsável por elaborar as propostas legislativas europeias e por executar as decisões do Parlamento Europeu e do Conselho da UE.

No âmbito do Horizon 2020, o principal e maior programa de financiamento de pesquisa e inovação da EU, a questão do acesso e GDP foi ampliada por meio do projeto-piloto para dados de pesquisa abertos (*Open Research Data pilot*). As duas questões centrais para os participantes do piloto ORD é a elaboração do PGD e o acesso aos dados de pesquisa, respeitando a abordagem “tão aberto quanto possível, tão fechado quanto necessário”.

Os documentos recuperados para a análise das políticas de GDP foram: *AGA – Annotated Model Grant Agreement (2019)*; *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 (2017)*; *Guidelines on FAIR Data Management in Horizon 2020 (2016)* e o *Horizon 2020 projects working on the 2019 coronavirus disease (COVID-19), the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and related topics: Guidelines for open access to publications, data and other research outputs (2020)*. A avaliação conjunta desses documentos permitiu verificar que o grau de aplicação dos documentos é de caráter recomendatório e mandatório, variando de acordo com a participação ou não no piloto ORD. O quadro 1 sintetiza a análise de cobertura de critérios contemplados no modelo de avaliação.

Quadro 1: Análise de políticas para GDP quanto ao grau de aplicação e cobertura de critérios.

| CATEGORIAS | | | | | | | | | | | | | | | | | |
|-------------------|------------|-------------------------------|------------------|-----------------|------------------------|-------------------|---|-------------------------------|-----|---------|------------------------|--------------------------|---------------------------------|-------|----------|-------------|--------------------------|
| Grau de aplicação | | | | | Cobertura de critérios | | | | | | | | | | | | |
| Recomendatório | Mandatório | Recomendatório/ Mandatório | Não identificado | Aspectos gerais | | | | Planejamento e infraestrutura | | | | Disponibilização e reuso | | | | | |
| | | | | Definição | Escopo | Responsabilidades | Monitoramento de conformidade e sanções | Revisão | PGD | Suporte | Curadoria/ Preservação | Padrões de metadados | Repositórios e centros de dados | Custo | Retenção | Propriedade | Acesso/ Compartilhamento |
| | | | | ● | ● | ○ | ● | ● | ● | ● | ● | ● | ● | ○ | ● | ○ | ● |

Legenda: ■ Grau de aplicação identificado ● Cobertura total ○ Cobertura parcial ◯ Nenhuma cobertura

Fonte: Dados de pesquisa.

Verificou-se que a maior parte das recomendações e exigências relacionadas à GDP é orientada apenas aos beneficiários participantes do piloto ORD, como a elaboração do PDG, depósito de dados em repositórios e fornecimento de acesso aberto (quando possível). No entanto, projetos não participantes, incluindo projetos de ações de saúde, podem ser obrigados a atender tais exigências caso os dados gerados sejam relevantes para lidar com uma ESP.

Observou-se que antes da pandemia, as obrigações legais para atender as aspirações de um cenário de ESP exigiam o acesso aos dados de pesquisa em um período de até 30 dias após a

geração, no entanto, no contexto da COVID-19 tal acesso deve ser imediato. Verificou-se também que o PGD é um documento obrigatório para pesquisas relacionadas à COVID-19, mesmo não havendo menção sobre a participação dessas pesquisas no piloto ORD, devendo ser submetido à Comissão em um tempo de prazo menor ao previsto anteriormente.

5 Considerações finais

Antes do cenário pandêmico, a CE já havia declarado suas expectativas para o compartilhamento e acesso aos dados relevantes para atender ESP, no entanto, a COVID-19 provocou algumas mudanças visando acelerar a disponibilização dos dados para a comunidade científica. Percebe-se que a pandemia COVID-19 tem demonstrado as fragilidades do sistema atual de comunicação na ciência e acelerado processos de mudança para o estabelecimento de políticas e normativas que favoreçam a GDP de forma socialmente responsável. Conclui-se que a GDP é uma prática relevante para o enfrentamento da pandemia, contribuindo para uma resposta mais rápida para a sociedade.

Referências

- BRINEY, Kristin; GOBEN, Abigail, ZILINSKI, Lisa. Institutional, funder, and journal data policies. *In*: JOHNSTON, Lisa R (ed.). **Curating research data: practical strategies for your digital repository**. Chicago: Association of College and Research Libraries, 2017. cap. 2, p. 61-78.
- COX, Andrew; VERBAAN, Eddy. **Exploring research data management**. London: Facet publishing, 2018.
- DIGITAL CURATION CENTRE. **Overview of funders' data policies**. Disponível em: <<http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>>. Acesso em: 17 set. 2020.
- HODSON, Simon; MOLLOY, Laura. **Current best practice for research data management policies: a memo for the Danish e-Infrastructure Cooperation and the Danish Digital Library**. [S.l.]: CODATA, 2014. Disponível em: <<https://zenodo.org/record/27872#.X8FRs81KjIU>>. Acesso em: 20 set. 2020.
- HORTON, L.; DIGITAL CURATION CENTRE. **Overview of UK Institution RDM Policies**. Version 6, aug. 2016. Disponível em: <<https://www.dcc.ac.uk/guidance/policy/institutional-data-policies>>. Acesso em: 15 set. 2020.
- MICHENER, William K. Ten Simple Rules for Creating a Good Data Management Plan. **PLoS Comput Biol.**, v. 11, n. 10, oct. 2015. DOI: 10.1371/journal.pcbi.1004525. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4619636/>>. Acesso em: 13 out. 2020.
- MONTEIRO, Gabriela. **Mapeamento e análise das políticas institucionais de financiadores da pesquisa brasileira: cenário dos dados científicos abertos**. 2019. 146 f. Dissertação (Mestrado em Gestão da Informação)– Centro de Ciências Humanas e da Educação, Universidade do Estado de Santa Catarina, Santa Catarina, 2019.
- ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD). **OECD principles and guidelines for access to research data from public funding**. 2007. Disponível em: <<https://www.oecd.org/sti/inno/38500813.pdf>>. Acesso: 15 nov. 2020.
- PINFIELD, S; COX, A. M.; SMITH, J. Research data management and libraries: relationships, activities, drivers and influences. **PLoS ONE**, [S.l.], v. 9, n. 12, p. 1-28, dec. 2014. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114734#pone.0114734-Cox>>. Acesso em: 15 nov. 2020.
- PISANI, Elizabeth; GHATAURE, Amrita; MERSON, Laura. **Data sharing in public health emergencies: A study of current policies, practices and infrastructure supporting the sharing of data to prevent and respond to epidemic and pandemic threats**. [London: Wellcome Trust], 2018. Disponível em: <https://wellcome.figshare.com/articles/Data_sharing_in_public_health_emergencies_A_study_of_current_policies_practices_and_infrastructure_supporting_the_sharing_of_data_to_prevent_and_respond_to_epidemic_and_pandemic_threats/5897608>. Acesso em: 16 nov. 2020.

SALES, Luana; SAYÃO, Luis. A ciência invisível: revelando os dados da cauda longa da pesquisa. **Encontro Nacional de Pesquisa em Ciência da Informação**, n. XIX ENANCIB, 2018. Disponível em: <<http://hdl.handle.net/20.500.11959/brapci/103678>>. Acesso em: 28 ago. 2020.

SAYÃO, Luis; SALES, Luana. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: Comissão de Energia Nuclear, 2015.

VIERGEVER, Roderik F.; HENDRIKS, Thom C. C. The 10 largest public and philanthropic funders of health research in the world: what they fund and how they distribute their funds. **Health Research Policy and Systems**, [S. l.], v. 14, n.18, feb. 2016. Disponível em: <<https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-015-0074-z>>. Acesso em: 6 jan. 2020.

WELLCOME TRUST. **Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak**. 2020. Disponível em: <<https://wellcome.org/coronavirus-covid-19/open-data>>. Acesso em: 20 nov. 2020.

_____. **Sharing research data to improve public health: full joint statement by funders of health research**. [2012]. Disponível em: <<https://wellcome.org/what-we-do/our-work/sharing-research-data-improve-public-health-full-joint-statement-funders-health>>. Acesso em: 20 nov. 2020.

_____. **Statement on data sharing in public health emergencies**. 2016. Disponível em: <<https://wellcome.org/press-release/statement-data-sharing-public-health-emergencies>>. Acesso em: 20 nov. 2020.