

Metodologia para identificação de tipos de dados de pesquisa

A experiência da Fundação Oswaldo Cruz (Fiocruz)

Methodology for identifying data types for research: the experience of the Fundação Oswaldo Cruz (Fiocruz) / Metodología para la identificación de tipos de datos de investigación: la experiencia de la Fundação Oswaldo Cruz (Fiocruz)

Ivone Pereira de Sá

Doutora em Ciência da Informação pela Universidade Federal Fluminense (UFF).
Tecnologista da Casa de Oswaldo Cruz (COC) da Fundação Oswaldo Cruz (Fiocruz), Brasil.

ivone.sa@fiocruz.br

Aline Lopes de Lacerda

Doutora em História Social pela Universidade de São Paulo (USP).
Pesquisadora do Departamento de Arquivo e Documentação da COC/Fiocruz, Brasil.

alopeslacerda@gmail.com

Eliane Monteiro de Santana Dias

Mestra em Preservação e Gestão do Patrimônio Cultural das Ciências e da Saúde pela COC/Fiocruz. Gestora da Biblioteca de História das Ciências e da Saúde da COC/Fiocruz, Brasil.

eliane.dias@fiocruz.br

Marcus Vinícius Pereira da Silva

Doutorando em Ciência da Informação na UFF. Assessor em Gestão da COC/Fiocruz, Brasil.

marcus.silva@fiocruz.br

Renata Lourenço Mendes Kessler

Mestra em Ciência da Informação pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict). Tecnologista em Saúde Pública da COC/Fiocruz, Brasil.

renata.lourenco@fiocruz.br

Karina Praxedes

Mestra em Gestão de Documentos e Arquivos pela Universidade Federal do Estado do Rio de Janeiro (Unirio). Assistente Técnica do Sistema de Gestão de Documentos e Arquivos da COC/Fiocruz, Brasil.

karina.praxedes@fiocruz.br

RESUMO

O artigo discorre sobre o desenvolvimento e a aplicação de metodologia para identificação de tipos de dados de pesquisa produzidos e coletados nas pesquisas da Fundação Oswaldo Cruz (Fiocruz). Trata-se de um relato de experiência do grupo de trabalho interdisciplinar que aplicou a metodologia no âmbito do projeto-piloto de gestão de dados de pesquisa coordenado pela Vice-Presidência de Ensino, Informação e Comunicação da Fiocruz.

Palavras-chave: gestão de dados de pesquisa; metodologia; tipos de dados de pesquisa.

ABSTRACT

The article discusses the development and application of a methodology for identifying types of data produced and collected during research activities conducted by the Fundação Oswaldo Cruz (Fiocruz). This is an experience report by the interdisciplinary working group that applied the methodology in the scope of the pilot project of research data management, coordinated by Fiocruz's Vice-Presidency for Education, Information and Communication.

Keywords: research data management; methodology; research data types.

RESUMEN

El artículo trata sobre el desarrollo y la aplicación de una metodología para identificar tipos de datos de investigación producidos y recogidos en las investigaciones de la Fundação Oswaldo Cruz (Fiocruz). Es un relato de experiencia del grupo de trabajo interdisciplinario que aplicó la metodología en el ámbito del proyecto piloto de gestión de datos de investigación coordinado por la Vice-Presidencia de Educación, Información y Comunicación de la Fiocruz.

Palabras clave: gestión de datos de investigación; metodología; tipos de datos de investigación.

Introdução

No ano de 2017, a Fundação Oswaldo Cruz (Fiocruz), instituição pública de pesquisa em saúde ligada ao Ministério da Saúde do Brasil e internacionalmente conhecida como um grande centro de referência em pesquisa científica, iniciou as primeiras conversas sobre a elaboração de diretrizes institucionais para gestão, compartilhamento e abertura de dados produzidos ou coletados no contexto das diversas atividades de pesquisa em curso na instituição. A iniciativa insere-se no rol de ações da Fiocruz relacionadas ao Movimento Internacional de Acesso Aberto ao Conhecimento, a partir do qual foi lançado, em 2011, o repositório institucional Arca¹ – um espaço para depósito e acesso livre e gratuito da produção técnica e científica dos profissionais, docentes e discentes da instituição – e foi publicada, em 2014, a *Política de Acesso Aberto ao Conhecimento*² (Santos et al., 2014).

Atenta ao movimento em prol da ciência aberta, a Fiocruz, por meio da coordenação da Vice-Presidência em Educação, Informação e Comunicação (VPEIC), criou o Grupo de Trabalho de Ciência Aberta (GTCA), composto por profissionais com diferentes formações e trajetórias ligados às suas unidades, com vistas a definir estratégias e etapas para implementação de uma política de gestão, compartilhamento e abertura de dados para pesquisa da Fiocruz. Para tanto, atuou em quatro principais linhas de ação: pesquisa, capacitação, ferramentas e projetos-pilotos.

As pesquisas realizadas pelo GTCA embasaram o planejamento de estratégias e ações do grupo e alguns resultados já se encontram disponíveis em acesso aberto no repositório institucional Arca (Santos, 2017; Santos; Clinio, 2018; Guanaes, 2018). Os estudos também embasaram a elaboração do *Termo de referência: gestão e abertura de dados para pesquisa na Fiocruz* (Fiocruz, 2018), que pode ser considerado o ponto de partida para o debate da temática com a comunidade Fiocruz e orientou a criação da *Política de gestão, compartilhamento e abertura de dados para pesquisa*³ (Fiocruz, 2020a), aprovada em setembro de 2020, após submissão à consulta pública e validação pelo Conselho Deliberativo da Fiocruz.

1 Disponível em: <https://www.arca.fiocruz.br/>. Acesso em: 15 maio 2021.

2 Disponível em: https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/portaria_-_politica_de_acesso_aberto_ao_conhecimento_na_fiocruz.pdf. Acesso em: 18 maio 2021.

3 A Fiocruz adotou o termo “dados para pesquisa” para incorporar tanto os dados produzidos no contexto da pesquisa, comumente denominados dados de pesquisa, quanto aqueles coletados durante a atividade de pesquisa, como os dados administrativos produzidos por outras instituições dos setores público ou privado.

Para implementação da política, identificou-se a necessidade de desenvolvimento de duas ferramentas essenciais para gestão, compartilhamento e abertura dos dados provenientes das pesquisas, sendo elas um sistema de plano de gestão de dados⁴ (PGD) e um repositório institucional de dados para pesquisa. Atualmente, é importante atentar ao papel dessas ferramentas no atendimento à imposição crescente, por parte de agências de fomento nacionais e internacionais, quanto à abertura de dados das pesquisas por elas apoiadas. Em tal contexto, um único repositório de dados na Fiocruz permitirá maior autonomia institucional, segurança jurídica e garantias de preservação a longo prazo.

O projeto também está alinhado com o compromisso 3 do 4º Plano de ação nacional para o governo aberto, desenvolvido no âmbito da Parceria para Governo Aberto (do inglês, Open Government Partnership – OGP), que propõe “estabelecer mecanismos de governança de dados científicos para o avanço da ciência aberta no Brasil” (Brasil, 2018, p. 22). Nesse cenário, a Fiocruz foi convidada pela Rede Nacional de Ensino e Pesquisa⁵ (RNP) e pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), para participar como um dos estudos de caso do projeto-piloto para o desenvolvimento de infraestrutura tecnológica federada de repositórios de dados de pesquisa no Brasil.

O estudo intitulado *Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas para compartilhamento de dados no Brasil – relatório* (Rocha et al., 2018), realizado pela RNP e pelo Ibict, analisou o Dspace,⁶ CKAN e o Dataverse⁷ como softwares de repositórios de dados, e esse último foi escolhido após a análise de 56 critérios agrupados nas seguintes classes: ambiente do repositório; conjuntos de dados; descrição e documentação; produção de conjuntos de dados; armazenamento a longo prazo; acesso e uso; e desenvolvimento e manutenção do software (Rocha et al., 2018). A partir dessa análise, a Fiocruz criou um grupo de trabalho para avaliar a solução tecnológica do Dataverse e o seu suporte para armazenamento, compartilhamento e publicação de dados para pesquisa⁸ (Fiocruz, 2020b).

Em continuidade às ações para o desenvolvimento do repositório institucional de dados da Fiocruz, iniciou-se uma nova fase do projeto, na

4 Disponível em: <https://www.arca.fiocruz.br/handle/icict/37896>. Acesso em: 2 set. 2021.

5 A RNP é uma plataforma digital para educação, pesquisa e inovação no Brasil, que tem por missão promover o uso inovador de redes avançadas.

6 Disponível em: <https://duraspace.org/dspace/>. Acesso em: 18 maio 2021.

7 Disponível em: <https://dataverse.org>. Acesso em: 18 maio 2021.

8 Mais informações sobre este estudo podem ser obtidas no Relatório final piloto de repositório de dados de pesquisa fase 1 (Fiocruz, 2020b).

qual realizaram-se testes no software junto a pesquisadores da Fiocruz. Pesquisadores de diferentes áreas foram convidados a participar do piloto, que compreendia uma série de etapas e envolveu equipe multidisciplinar, incluindo profissionais da Casa de Oswaldo Cruz (COC): a) avaliar metadados mínimos disponíveis no software; b) desenvolver metodologia para mapear e identificar os tipos de dados para pesquisa; c) elaborar tabela de temporalidade dos dados para pesquisa; e d) elaborar plano de preservação digital. Nesse contexto, este artigo tem como objetivo apresentar a experiência da Fiocruz na construção de uma metodologia desenvolvida pela COC, uma das unidades técnico-científicas da instituição, para identificar os tipos de dados de pesquisa. A metodologia também visa contribuir para o alcance de outros objetivos, tais como a definição do prazo de guarda dos dados e a sua preservação digital.

Origem, trajetória e compromisso da Fiocruz com a divulgação do conhecimento

A origem da Fiocruz remonta à criação do Instituto Soroterápico Federal, em 1900, decorrente da chegada da peste bubônica no Brasil. Em 1908 o Instituto foi renomeado Instituto Oswaldo Cruz (IOC), em homenagem ao renomado pesquisador e então diretor da instituição. Inspirado no Instituto Pasteur de Paris, o IOC buscou combinar a pesquisa pura e aplicada com o ensino e, em 1912, foi autorizada a criação de um hospital para tratamento e pesquisa em doença de chagas. Além da atuação nas áreas de pesquisa, ensino e assistência, o instituto também exercia a atividade de produção de vacinas e outros produtos biológicos.

Em 1970 a Fiocruz foi criada como uma entidade pública, a partir da reunião de diferentes instituições ligadas ao Ministério da Saúde. Além do IOC, foram incorporadas à Fiocruz: o Serviço de Produtos Profiláticos, Instituto Fernandes Figueira (IFF), Instituto Nacional de Endemias Rurais (formado por centros de pesquisa localizados na Bahia, Recife e Minas Gerais), Instituto Evandro Chagas (localizado no Pará), Instituto de Leprologia do Serviço Nacional de Lepra e Fundação de Recursos Humanos para a Saúde, atual Escola Nacional de Saúde Pública Sérgio Arouca (Ensp).

No contexto de redemocratização do país e do Movimento pela Reforma Sanitária, no final da década de 1980, ocorreu o processo de ressignificação das noções de saúde e de pesquisa em saúde e emergiu o conceito de saúde coletiva. A expressão designa um campo de saberes e práticas referidas à saúde como fenômeno social e, como tal, incorpora pesquisas históricas, sociológicas,

antropológicas e epistemológicas, entre outras, em busca de teorização interdisciplinar, bem como de práticas intersetoriais em saúde que atendam às necessidades da sociedade (Campos, 2000). A Fiocruz, sob a gestão do médico sanitário Sérgio Arouca, teve um papel importante nas formulações de novos princípios, conceitos e noções para o campo da saúde, alguns deles incorporados na Constituição de 1988 e na criação do Sistema Único de Saúde (SUS).

Também durante a gestão de Sérgio Arouca (1985-1989) criaram-se unidades e setores relacionados ao novo e ampliado conceito de saúde e de pesquisas em saúde. Entre eles, para o objeto deste artigo, destacamos a Casa de Oswaldo Cruz, dedicada às ações de memória e preservação do patrimônio cultural das ciências e da saúde, e a Superintendência de Informação Científica, atual Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (Icict).

Atualmente a Fiocruz possui 17 unidades técnico-científicas de diferentes áreas de atuação e cinco escritórios, com presença em todas as regiões do país, além de um escritório em Maputo, capital de Moçambique, e tem como missão

produzir, disseminar e compartilhar conhecimentos e tecnologias voltadas para o fortalecimento e a consolidação do SUS e que contribuam para a promoção da saúde e da qualidade de vida da população brasileira, para a redução das desigualdades sociais e para a dinâmica nacional de inovação, tendo a defesa do direito à saúde e da cidadania ampla como valores centrais. (Fiocruz, 2010, p. 21)

A instituição possui hoje trinta áreas de pesquisa com distintas práticas científicas, como: história, saúde e ciência; educação não formal e divulgação das ciências e saúde; pesquisa e desenvolvimento de fármacos e medicamentos; epidemiologia, métodos estatísticos e quantitativos; microbiologia em saúde e ambiente; e doenças crônicas e não transmissíveis, medicina regenerativa.⁹ Cabe ressaltar que pesquisas também são desenvolvidas nos diversos programas de pós-graduação da instituição, especialmente nos mestrados e doutorados acadêmicos.

A Casa de Oswaldo Cruz e a contribuição da perspectiva arquivística

Criada em 1985, a Casa de Oswaldo Cruz tem entre suas atribuições desenvolver atividades de constituição, organização, preservação, guarda e difusão de acervos

9 Disponível em: <https://portal.fiocruz.br/areas-de-pesquisa>. Acesso em: 6 jan. 2021.

documentais; e propor, em suas áreas de atuação, políticas públicas e diretrizes institucionais nos âmbitos de saúde, ciência, tecnologia e cultura (COC, 2019). Responsável pela preservação e acesso aos arquivos permanentes das unidades e pela orientação, planejamento e supervisão de programas e atividades de gestão de documentos e arquivos nas unidades da Fiocruz, através do Sistema de Gestão de Documentos e Arquivos da Fiocruz (Sigda), a COC adquiriu larga experiência nessa área de atuação e, no contexto brasileiro, foi uma das pioneiras na construção de uma base de dados de acervos arquivísticos permanentes – a Base Arch.¹⁰

No contexto da gestão, compartilhamento e abertura de dados para pesquisa, considera-se que estes são produzidos em contextos institucionais específicos, são utilizados como evidências e comprovam os resultados obtidos. Entende-se, neste sentido, que o arcabouço teórico-metodológico arquivístico tem potenciais contribuições, principalmente, no que tange ao princípio da proveniência¹¹ e à presunção de autenticidade¹² para alicerçar práticas e procedimentos da gestão de dados para pesquisa. Uma visão integrada de todo o decurso da pesquisa científica, lançando mão de saberes e fazeres consolidados da arquivologia, abrangendo aspectos inerentes ao tratamento do dado ao longo do seu ciclo de vida, bem como processos e documentos produzidos na sua efetivação, podem consubstanciar a disponibilização qualificada, o (re)uso efetivo, a confiança e a preservação desses dados pelo tempo necessário.

Com relação à identificação dos tipos de dados, acatamos aquela definição de dado de pesquisa apresentada na publicação *Livro Verde – Ciência aberta e dados abertos*:

Unidades de informação criadas ou coletadas no curso da pesquisa científica, que são frequentemente formatadas de maneira a torná-las adequadas à comunicação, à interpretação e ao processamento por computador. (Santos, 2017, p. 122)

Atualmente, a instituição possui projetos de pesquisa e desenvolvimento tecnológico ligados às ciências biológicas, exatas, naturais, humanas e sociais.

¹⁰ A Base Arch é o repositório de informações sobre o arquivo permanente da Fiocruz, representativo da história das ciências biomédicas e da saúde pública no Brasil. Reúne documentos institucionais e pessoais, dos gêneros textual, iconográfico, cartográfico, sonoro, audiovisual, micrográfico e eletrônico, que abrangem o período de 1756 até os dias atuais. Disponível em: <http://basearch.coc.fiocruz.br/>. Acesso em: 10 jan. 2021.

¹¹ Sobre isso, recomendamos a leitura de Jorge (2018). Na subseção “2.2.2 - Dados para pesquisa: o foco no uso”, a autora indica que a confiabilidade de um dado é diretamente impactada pela proveniência.

¹² “Credibilidade de um documento enquanto documento, isto é, a qualidade de um documento ser o que diz ser e que está livre de adulteração ou qualquer outro tipo de corrupção” (Conarq, 2016, p. 10).

Cada área de conhecimento lida com técnicas e métodos específicos, o que ocasiona diferentes acepções de dado de pesquisa e pode demandar soluções diversas de gestão, incluindo especificações de acesso e armazenamento. Nesse sentido, no estudo tipológico desses dados, faz-se necessário entender os elementos cognitivos e as práticas sociais de cada uma das áreas de pesquisa da Fiocruz. Para tal fim, a COC criou um grupo formado por profissionais das áreas de arquivologia, biblioteconomia e história para desenvolver uma metodologia de identificação dos tipos de dados para pesquisa na Fiocruz, dada a necessária complementariedade das três áreas para o fim almejado.

Metodologia para identificação de tipos de dados de pesquisa

O primeiro desafio foi alinhar uma metodologia única para ser aplicada em áreas, projetos ou grupos de pesquisa com formações organizacionais, características e escopos distintos. Com vistas ao alinhamento dessas diferentes estruturas, utilizamos como base a metodologia proposta por Sá (2019) em estudo cujo objetivo foi elaborar diretrizes comuns de metadados para dados de pesquisa que atendam a diferentes domínios do conhecimento. A autora propõe a análise de distintas configurações organizacionais por meio do estudo de comunidades científicas e de prática (CoPS). Segundo Sá (2019, p. 161):

Entende-se que a produção dos membros da linha de pesquisa traduz a prática deles. Os artigos produzidos pelos integrantes das CoPS são artefatos que fazem parte de um processo dinâmico e ativo da pesquisa: são resultantes de atividades, conversas, negociações, utilização de métodos e procedimentos e são os resultados/produtos das pesquisas que refletem a experiência compartilhada e a interação entre os integrantes.

Sendo assim, para o desenvolvimento da metodologia de identificação de tipos de dados, as comunidades científicas e de prática foram definidas como grupos formais que trabalham em conjunto, de forma cooperativa e colaborativa, desempenhando atividades de pesquisa científica visando ao alcance de objetivos em comum e produzindo resultados. São pesquisadores que, reunidos, realizam a prática da pesquisa seguindo bases científicas.

Os grupos da Fiocruz selecionados para a aplicação do projeto-piloto da metodologia aqui apresentada desenvolvem pesquisas científicas por meio de fazeres e ações, que culminam em resultados/produtos de pesquisa, sejam artigos científicos, patentes, medicamentos, vacinas etc. Entende-se que tais

pressupostos sejam adequados também a projetos de grande porte, realizados por redes que envolvem pesquisadores de outras instituições.

Como a identificação dos tipos de dados de pesquisa é um tema raro e ainda pouco discutido na literatura da área de ciência da informação, ainda mais restrito no campo da arquivologia em nível internacional e, principalmente, em nível nacional, não foi identificado um modelo proposto e estabelecido. Sendo assim, a metodologia foi desenvolvida a partir do entendimento apresentado por Sá (2019) de que o dado de pesquisa é oriundo de um método, técnica ou procedimento utilizado para a realização da pesquisa científica como, por exemplo, técnicas histológicas, técnicas de laboratório clínico, testes sorológicos e hematológicos, tomografia, ultrassonografia, métodos epidemiológicos, coleta de dados, grupos focais, entre outros.

A partir desse entendimento, a metodologia está dividida em seis etapas, apresentadas a seguir.

Etapa 1: Caracterização das comunidades científicas e de prática da Fiocruz

Tem como objetivo caracterizar e cadastrar as comunidades científicas e de prática que produzem e coletam dados de pesquisa por meio da reunião de informações disponíveis na internet e/ou com o(s) responsável(is) pela ação/projeto, como: a) nome do projeto ou grupo de pesquisa; b) vinculação institucional; c) descrição do projeto ou grupo de pesquisa; d) histórico do projeto ou grupo de pesquisa; e) área do conhecimento; f) membros da equipe; g) contato.

Etapa 2: Levantamento da produção das comunidades científicas e de prática da Fiocruz

Tem como objetivo identificar e analisar a prática dessas comunidades através da produção científica, dos projetos de pesquisa e relatórios de seus integrantes. Esses documentos são considerados artefatos que fazem parte de um processo dinâmico e ativo e são os resultados das pesquisas que refletem a experiência compartilhada e a interação entre os membros da comunidade. A partir da análise do conteúdo informacional destes artefatos, torna-se possível identificar quais foram os métodos, as técnicas e os procedimentos utilizados que dão origem aos dados de pesquisa. O mapeamento da produção científica poderá ser realizado por grupos de pesquisa, por pesquisadores ou por projetos, conforme a configuração da comunidade científica e de prática, por meio dos currículos cadastrados na plataforma Lattes ou enviados pelos pesquisadores.

Nessa etapa são coletadas informações sobre: a) nome do projeto ou grupo de pesquisa; b) membros (nome, formação, vínculo institucional); c) área e subárea do conhecimento; d) produção científica (artigos e outros tipos de documentos).

Etapa 3: Identificação de procedimentos, técnicas e métodos adotados nas pesquisas científicas

Essa ação se divide em três atividades. A primeira visa analisar os relatórios de pesquisa e artigos científicos mapeados na etapa anterior para identificar os procedimentos metodológicos, etapas da pesquisa, materiais e métodos utilizados e extrair os termos relacionados aos métodos, técnicas e procedimentos.

Em seguida, é realizada a compatibilização dos termos com um vocabulário controlado. No caso da Fiocruz, optamos por utilizar o *Descritores em Ciências da Saúde (DeCS)*,¹³ principal vocabulário controlado na área da saúde.

Por fim, são coletadas informações sobre os dados de pesquisa produzidos a partir dos métodos, processos e técnicas identificados.

Nessa ação deve-se:

- identificar a(s) pessoa(s)-chave nas comunidades científicas e de prática;
- enviar lista de termos relacionados aos métodos, técnicas e procedimentos para a(s) pessoa(s)-chave validar(em) e complementar(em) com os tipos de dados de pesquisa que são produzidos;
- receber o retorno da listagem de métodos, técnicas e procedimentos com validação e inclusão de informações sobre os tipos de dados.

Etapa 4: Elaboração do formulário de coleta de dados

A etapa fornece elementos relevantes para a realização da visita técnica. É voltada para dirimir dúvidas que não foram sanadas nas etapas anteriores. Compreende as seguintes atividades:

- identificar e formular questões cujas informações não puderam ser obtidas na realização das etapas anteriores;

¹³ “Os *Descritores em Ciências da Saúde*, conhecido como DeCS, é um vocabulário estruturado e multilíngue que foi criado pela Bireme (Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde), no ano de 1986, desenvolvido a partir do *Medical Subject Headings (MeSH)* da U. S. National Library of Medicine, para servir como uma linguagem única na indexação de materiais bibliográficos e permitir o uso de terminologia comum para ser usado na pesquisa e recuperação de assuntos da literatura científica nas fontes de informação na área das ciências da saúde, proporcionando um meio consistente e único para a recuperação da informação” (*Descritores*, 2021).

- enviar por e-mail ou ferramenta de *survey* o formulário de coleta de dados para a(s) pessoa(s)-chave;
- receber o retorno do formulário de coleta de dados preenchido;
- analisar e sistematizar as informações registradas nos formulários.

Etapa 5: Realização de visitas técnicas

A visita técnica (pesquisa de campo) consiste em apurar, presencial ou remotamente, junto aos pesquisadores, informações que não puderam ser obtidas nas etapas anteriores (análise da produção científica e aplicação do formulário). Essa ação contempla as seguintes atividades:

- elaborar roteiro com as questões a serem apresentadas na visita técnica;
- elaborar critérios para a seleção da(s) pessoa(s) da equipe a ser(em) entrevistada(s);
- realizar a entrevista com gravação de áudio e análise *in loco* dos dados de pesquisa;
- realizar a transcrição das entrevistas;
- validar a entrevista e coletar assinatura de termo de autorização para uso e depósito enquanto registro documental.

O roteiro da entrevista foi dividido em sete blocos, sendo eles: aspectos gerais; características descritivas dos conjuntos de dados de pesquisa; características técnicas e organizacionais dos conjuntos de dados de pesquisa; documentação; qualidade; preservação e comunicação/divulgação.

Bloco 1: Aspectos gerais

1. Questão de pesquisa.
2. Resumo das funções e atividades realizadas na pesquisa.
3. Propósito científico da coleta de dados (função do dado na pesquisa). Por que os dados foram produzidos?
4. Domínio do conhecimento em que os dados foram produzidos.
5. A quem se destina, qual é o público-alvo dos dados na pesquisa?
6. Quais as formas de uso dos dados?
7. Como percebem o uso dos dados em novas pesquisas por outros pesquisadores?
8. Consideram que os conjuntos de dados possuem valor imediato e duradouro?
9. Consideram que os conjuntos de dados são passíveis de reprodução?

Bloco 2: Características descritivas dos conjuntos de dados de pesquisa

1. Todos os dados foram produzidos no formato digital (nato digital) ou existem dados analógicos? Em caso afirmativo, estes serão digitalizados?
2. Nome/título da coleção de dados (o que).
3. Autores/produtores dos dados (quem).
4. Resumo da coleção de dados.
5. Período em que os dados foram coletados: período de tempo que cobre (quando).
6. Escopo geográfico onde os dados foram coletados (onde).
7. Como os dados foram coletados? Descrição dos métodos, práticas e instrumentos utilizados na coleta dos dados (cópia do formulário de consentimento, configuração das entrevistas, critérios para seleção dos sujeitos das entrevistas, instruções dadas aos entrevistadores, instrumento de coleta, como os questionários/roteiros das entrevistas, lista de entrevistas).
8. Origem dos dados: computacionais, experimentais, observacionais.
9. Tipos (natureza) de dados: áudio, vídeo, imagem, texto, entre outros.
10. Status do fluxo de trabalho da pesquisa segundo sua fase: dados brutos, crus ou preliminares, derivados, canônicos ou referenciais.
11. Qual a quantidade de conjuntos de dados produzidos por tipos (natureza)?
12. O que identificam como peculiar (características muito específicas) nos dados produzidos?

Bloco 3: Características técnicas e organizacionais dos conjuntos de dados de pesquisa

1. Qual é o esquema de organização dos conjuntos de dados de pesquisa?
2. Organização do armazenamento dos conjuntos de dados: nomes dos arquivos que compõem a coleção. Como os conjuntos de dados, arquivos, pastas, diretórios e tabelas de banco de dados foram nomeados e como se relacionam? Utilizam algum padrão para nomenclatura?
3. Formato dos arquivos.
4. Foram realizadas conversões de formato de arquivo? Em caso afirmativo, por quê?
5. Tamanho dos arquivos.
6. Existe relacionamento entre os conjuntos de dados?
7. Unidades de medida utilizadas.
8. Como os dados foram processados? Métodos e procedimentos utilizados para o processamento de dados, por exemplo, softwares.

9. Foram realizadas modificações nos arquivos dos dados? Em caso afirmativo, essas modificações foram registradas e datadas?
10. Software (incluindo o número da versão) usado para preparar e ler a coleção de dados.
11. Propósito científico da coleta de dados (função do dado na pesquisa). Por que os dados foram produzidos?

Bloco 4: Documentação

1. Que tipos de documentos foram criados que podem apoiar o entendimento sobre como os dados foram produzidos, processados e de como estão estruturados, organizados, nomeados? Por exemplo, caderno de campo, questionários, roteiros de entrevistas, modelos de descrição etc.
2. Existem dados armazenados em bases de dados?
3. Quais metadados são utilizados para a descrição dos dados e para o registro da proveniência? Utilizam algum padrão ou esquema de metadados específico?
4. Fazem uso de taxonomias, vocabulários controlados?
5. Utilizaram dados de outras pesquisas ou de outras fontes?

Bloco 5: Qualidade

1. Quais verificações (ações de controle e inspeção) foram utilizadas para encontrar valores inválidos, respostas inconsistentes, registros incompletos etc.?
2. Quais verificações são e serão usadas para gerenciar as versões dos dados?

Bloco 6: Preservação

1. Consideram que todos os dados devem ser preservados?
2. Quais seriam os dados mais relevantes e que deveriam ser priorizados em um planejamento de preservação?
3. Existe alguma orientação específica da área em relação a preservação dos dados?

Bloco 7: Comunicação/divulgação

1. Foram realizadas ou estão previstas ações de divulgação da pesquisa?
2. Com relação à ciência cidadã, em que medida os conjuntos de dados de pesquisa podem ser utilizados?
3. Os conjuntos de dados de pesquisa podem ser utilizados como recurso educacional?

Resultados preliminares

Para testar e validar a metodologia foram conduzidos quatro pilotos dentro do campo da saúde pública, em diferentes áreas: saúde mental, genética, imunologia e epidemiologia. Destacam-se, dentro dos pilotos, pesquisas transdisciplinares que envolvem a saúde coletiva, psicologia, história, comunicação, entre outras. Observaram-se também pesquisas cujos dados levantados contribuem para a construção de políticas públicas em saúde e para o levantamento e registro de fatos que colaboram para a construção da memória e da história dentro do campo, o que reforça a necessidade da adoção de procedimentos metodológicos do campo da arquivologia.

Os resultados obtidos até agora demonstram uma diversidade de tipos de dados produzidos nas pesquisas analisadas, por meio de entrevistas, avaliações clínicas, análises laboratoriais e documentos históricos. Foram produzidos dados dos tipos numéricos, textuais, audiovisuais, sonoros e em diferentes formatos como, por exemplo, Comma-Separated Values (CSV); Data File Format (DAT); Portable Document Format (PDF). Os resultados apontam para a potencial reutilização dos dados produzidos pelos grupos e por eles acumulados, por parte de diferentes tipos de público, como pesquisadores, gestores da área da saúde, jornalistas e políticos.

Após a aplicação da metodologia de identificação de tipos de dados de pesquisa da Fiocruz nos quatro pilotos, concluiu-se que os resultados obtidos foram satisfatórios, considerando sua aplicação válida em qualquer tipo de pesquisa, podendo ainda, em cada caso, ser avaliada a necessidade de sua aplicação na íntegra ou de forma parcial, por exemplo, devido à falta de tempo ou urgência dos resultados, pode-se suprimir as etapas 2, 3 e 4 sem prejuízo nos resultados a serem obtidos.

A partir dos resultados obtidos, foi diagnosticada a necessidade de identificar prazos de guarda, ou seja, estabelecer a temporalidade dos dados de pesquisa, visto que há indicação de limitação de espaço de armazenamento de longo prazo no repositório de dados e para que haja uma maior qualidade na preservação, já que não são todos os dados que precisam de guarda permanente.

Por fim, indica-se ainda a necessidade de elaboração de um plano de preservação digital do repositório de dados da Fiocruz, alinhado com o Programa de Preservação Digital de Acervos da Fiocruz (Fiocruz, 2020c). Lançado em 2020, o programa define os padrões e requisitos mínimos para o desenvolvimento dos planos de preservação digital, incluindo a gestão de riscos, a gestão da propriedade intelectual, a seleção de estratégias de preservação, os sistemas de gestão e etapas de preservação a serem adotados, bem como a definição de custos e financiamento.

As atividades executadas pelo grupo de trabalho Tipos de Dados de Pesquisa da Fiocruz, nessa etapa, não esgotaram a riqueza de aspectos presentes no levantamento dos tipos de dados de pesquisa que podem e devem estar acessíveis a grupos mais amplos, contribuindo para o esclarecimento dessas necessidades já identificadas e de novas que venham a surgir.

Considerações finais

Observou-se, ao longo de todo o processo de construção da metodologia e de sua aplicação, que os dados de pesquisa são produzidos em determinados contextos institucionais e são elementos fundamentais para validar as evidências e comprovar os resultados obtidos, e, por conta disso, considera-se importante a aplicação das bases teóricas da arquivologia para a garantia de sua autenticidade e confidencialidade.

A Casa de Oswaldo Cruz/Fiocruz, que atua ao longo de 35 anos na organização de acervos arquivísticos e acompanha as tendências e desafios mundiais de implementação das teorias e das práticas arquivísticas, vem incorporando e desenvolvendo ferramentas e métodos que permitem o acesso às informações contidas nos diferentes registros informacionais. A metodologia para identificação de tipos de dados de pesquisa da Fiocruz, um trabalho original, vem somar essas ações e ampliar seu escopo de atuação e seu pioneirismo na reflexão das práticas arquivísticas sobre a importância do inter-relacionamento entre dados de pesquisa, informações e documentos arquivísticos, contribuindo assim para a gestão de dados de pesquisa, fomentando as ações de ciência aberta e servindo como modelo a ser adotado nas demais instituições governamentais de ciência, tecnologia e inovação.

Referências

- BRASIL. Ministério da Transparência e Controladoria-Geral da União. 4º Plano de ação nacional em governo aberto. Brasília, DF, 2018. Disponível em: https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/4o-plano-de-acao-nacional_portugues.pdf. Acesso em: 18 maio 2021.
- CAMPOS, Gastão Wagner de Sousa. Saúde pública e

- saúde coletiva: campo e núcleo de saberes e práticas. *Ciência & Saúde Coletiva*, v. 5, n. 2, p. 219-230, 2000.
- CASA DE OSWALDO CRUZ (COC). *Manual da organização*. Rio de Janeiro, 2019. Disponível em: http://coc.fiocruz.br/images/stories/PDFs/_Manual_org_alt2019.pdf. Acesso em: 2 maio 2021.
- CONSELHO NACIONAL DE ARQUIVOS (Conarq). *Glossário de documentos arquivísticos digitais*. Rio de Janeiro, 2016. Disponível em: https://www.gov.br/conarq/pt-br/assuntos/camaras-tecnicas-setoriais-inativas/camara-tecnica-de-documentos-eletronicos-ctde/2016_CTDE_Glossario_V7.pdf. Acesso em: 31 maio 2021.
- DESCRIPTORIOS em Ciências da Saúde: DeCS. São Paulo: Bireme/Opas/OMS, 2021. Disponível em: <http://decs.bvsalud.org>. Acesso em: 8 maio 2021.
- FUNDAÇÃO OSWALDO CRUZ (Fiocruz). *Programa de preservação digital de acervos da Fiocruz*. Rio de Janeiro: Fiocruz, 2020a.
- _____. *Relatório final piloto de repositório de dados de pesquisa fase 1*. Rio de Janeiro, 2020b. Disponível em: <https://www.arca.fiocruz.br/handle/icict/41414>. Acesso em: 2 maio 2021.
- _____. Vice-Presidência de Ensino, Informação e Comunicação. *Política de gestão, compartilhamento e abertura de dados para pesquisa: princípios e diretrizes*. Rio de Janeiro: Fiocruz/VPEIC, 2020c. 19 p. Disponível em: <https://www.arca.fiocruz.br/handle/icict/46408>. Acesso em: 19 maio 2021.
- _____. Grupo de Trabalho em Ciência Aberta. *Termo de referência: gestão e abertura de dados para pesquisa na Fiocruz*. Rio de Janeiro: Fiocruz/Presidência, 2018. 15 p. Disponível em: <https://www.arca.fiocruz.br/handle/icict/26803>. Acesso em: 2 maio 2021.
- _____. VI Congresso Interno da Fiocruz: a Fiocruz como instituição pública estratégica de Estado para saúde – Relatório final. Rio de Janeiro, 2010.
- GUANAES, Paulo Cezar Vieira (org.). *Marcos legais nacionais em face da abertura de dados para pesquisa em saúde: dados pessoais, sensíveis ou sigilosos e propriedade intelectual*. Rio de Janeiro: Fiocruz, 2018. Disponível em: <https://www.arca.fiocruz.br/handle/icict/28838>. Acesso em: 2 maio 2021.
- JORGE, Vanessa de Arruda. *Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública: o caso do vírus Zika*. 2018. 263 f. Tese (Doutorado em Ciência da Informação) – Universidade Federal do Rio de Janeiro; Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2018. Disponível em: <https://www.arca.fiocruz.br/handle/icict/32604>. Acesso em: 31 maio 2021.
- ROCHA, Rafael Port et al. *Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas – relatório 2018*. Disponível em: <http://hdl.handle.net/10183/185126>. Acesso em: 19 maio 2021.
- SÁ, Ivone Pereira de. *A Zika enquanto objeto de fronteira: diretrizes para identificação de metadados de dados de pesquisa*. 2019. 334 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal Fluminense, 2019. Disponível em: <https://app.uff.br/riuff/handle/1/13854>. Acesso em: 19 maio 2021.
- SANTOS, Paula Xavier dos; CLINIO, Anne. *Sumário executivo Livro Verde – Ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional*. Rio de Janeiro: Fiocruz, 2018. Disponível em: <https://www.arca.fiocruz.br/handle/icict/26809>. Acesso em: 2 maio 2021.
- _____. (coord.). *Livro Verde – Ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional*. Rio de Janeiro: Fiocruz, 2017. Disponível em: https://www.icict.fiocruz.br/sites/www.icict.fiocruz.br/files/LivroVerde-31-01-2018_versao_Final.pdf. Acesso em: 2 maio 2021.
- _____. et al. *Política de acesso aberto ao conhecimento: análise da experiência da Fundação Oswaldo Cruz/Fiocruz*. Recife: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde, Rio de Janeiro, v. 8, n. 2, p. 210-226, jun. 2014.

Recebido em 31/1/2021

Aprovado em 20/8/2021