FioDMP

FIODMP – MODELO SIMPLIFICADO

DATAFICAÇÃO POPULAR EM FAVELAS BRASILEIRAS: ARRANJOS SOCIOTÉCNICOS, CURADORIA E PRESERVAÇÃO DIGITAL DE DADOS RELACIONADOS À SAÚDE, AMBIENTE E VIOLÊNCIA

FioDMP-id: https://fiodmp.fiocruz.br/MCM3ZER · **Versão:** 1.009

CRIADOR DO PGD

Nome: Sofia Barbosa

Email: pesquisador1237@example.org

Afiliação: Casa de Oswaldo Cruz (COC)

SEÇÃO A

Informação Administrativa

Indique se o Plano de Gestão de Dados (PGD) será de acesso público (disponível para todos) ou privado (com acesso restrito a um grupo específico ou instituição):

Público

Informe o nome do criador do PGD:

Sofia Barbosa

Indique se você é o pesquisador principal responsável pela condução e coordenação do projeto de pesquisa:

Não

Informe o nome completo do principal responsável pelo projeto:

Sofia Barbosa

Informe a afiliação do responsável principal pelo projeto, selecionando a opção correspondente:

Sou da Fiocruz

Indique o nome da sua instituição.

Casa de Oswaldo Cruz (COC)

Informe o e-mail do pesquisador principal:

pesquisador1237@example.org

Informe o link do Currículo Lattes do responsável principal pelo projeto:

http://lattes.cnpq.br/0000000000000000

Informe o link do indicador persistente do responsável principal pelo projeto:

0000-0000-0000-0000

Indique a área do conhecimento correspondente a sua pesquisa

Ciências Sociais Aplicadas

Especifique a área das Ciências Sociais Aplicadas.

Ciência da Informação

Informe o título do projeto:

DATAFICAÇÃO POPULAR EM FAVELAS BRASILEIRAS: ARRANJOS SOCIOTÉCNICOS, CURADORIA E PRESERVAÇÃO DIGITAL DE DADOS RELACIONADOS À SAÚDE, AMBIENTE E VIOLÊNCIA

Indique a fase do projeto no momento da elaboração do Plano de Gestão de Dados (PGD):

Durante o planejamento do projeto

Informe o resumo do projeto:

Moradores, coletivos e organizações de favelas brasileiras têm produzido dados sobre saúde, ambiente e violência para tornar visíveis problemas sub-representados nas estatísticas oficiais, reivindicar direitos e subsidiar políticas públicas. Contudo, limitações técnicas e de governança podem comprometer a autenticidade, a integridade, a inteligibilidade e o acesso a esses registros, enquanto sua abertura indiscriminada pode gerar apropriação indevida, exposição e estigmatização. Nesse cenário, o projeto analisará como, nos arranjos sociotécnicos da datificação popular, as práticas documentárias configuram a produção, a organização, a preservação, a governança e a circulação desses dados, compreendidos

como documentos socialmente constituídos. De natureza aplicada e abordagem qualitativa, a pesquisa articulará revisão bibliográfica e documental, mapeamento nacional de iniciativas e estudos de caso orientados pela pesquisa-ação. Os princípios FAIR e CARE serão mobilizados criticamente para conciliar encontrabilidade, acessibilidade, interoperabilidade e reutilização com benefício coletivo, autoridade para controlar, responsabilidade e ética. Com base no DPC RAM e nos Níveis de Preservação Digital da NDSA, serão construídos e aplicados, de forma participativa, um instrumento de avaliação e diretrizes de preservação digital. Como resultados, serão produzidos um mapa georreferenciado das iniciativas de datificação, uma base de referências sobre o assunto e diretrizes de curadoria e preservação digital validadas com coletivos e especialistas. A pesquisa contribuirá para os estudos do documento e da datificação popular; para ampliar a integração e a eficiência do sistema de informação de saúde, sistematizando e difundindo as iniciativas de datificação popular em redes autorizadas de produção de conhecimento; e para o fortalecimento da Política de Saúde Digital do Sistema Único de Saúde.

Informe os termos (descritores ou palavras-chave) que representem o conteúdo do projeto:

Favelas

Curadoria de Dados

Preservação Digital

Informe a data de início do projeto:

03/2027

Informe o idioma do projeto:

Português

Informe se o projeto tem financiamento:

Sim

Especifique qual o financiador do projeto:

Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

SEÇÃO B

Descrição dos dados coletados e/ou reutilizados

Informe como os dados da pesquisa serão obtidos:

Coletados e/ou Produzidos

Reutilizados de outras fontes

Informe os tipos de dados que serão coletados e/ou produzidos:

Áudio: gravações sonoras (fala, música, ruídos, sinais acústicos), armazenadas como arquivos de som (ex.: WAV, MP3) ou fluxos digitais.

Dados numéricos: valores quantitativos em forma de números (inteiros, decimais, séries), geralmente em tabelas/planilhas/bancos de dados (ex.: medições, contagens, indicadores).

Dados raster (ou matriciais): dados espaciais em "grade" de células/pixels; cada pixel guarda um valor (ex.: altitude, temperatura, NDVI, classes). Típico de imagens de satélite e modelos contínuos.

Dados temporais espacializados: dados que combinam localização (onde) e tempo (quando), permitindo ver mudanças ao longo do tempo em um espaço (ex.: séries temporais de chuva por município, evolução de desmatamento por área).

Fotografias aéreas: imagens capturadas do ar (aviões, drones), geralmente com alta resolução e usadas para mapeamento, inspeção e análise do território.

Imagem: representação visual digital (2D) que pode ser foto, figura, microscopia, satélite etc.; pode ser "comum" (RGB) ou conter bandas/espectros e metadados.

Mapas de uso e cobertura da terra (pixels): mapas temáticos normalmente em formato raster, onde cada pixel é classificado em uma categoria (ex.: floresta, água, agricultura, urbano).

Modelos 3D: representações tridimensionais de objetos/ambientes (ex.: malhas 3D, nuvens de pontos, modelos BIM), usadas para simulação, visualização e medições.

Texto: informação em linguagem escrita (livre ou estruturada), como relatórios, descrições, transcrições, artigos, ou campos textuais em tabelas.

Vídeo: sequência de imagens (frames) ao longo do tempo, normalmente com áudio opcional (ex.: MP4), usada para registro de eventos, monitoramento e análises temporais.

Informe os formatos de dados que serão coletados e/ou produzidos:

CSV (Comma-Separated Values) - Formato simples de arquivo em que os valores dos dados são separados por vírgulas, amplamente utilizado em tabelas e planilhas.

DOC (Document Format) - Formato de documento do Microsoft Word, usado para criar e editar textos com formatação.

Excel (XLS/XLSX) – Formatos de arquivos para planilhas eletrônicas, usados em versões antigas (XLS) e recentes (XLSX) do Microsoft Excel.

JPEG (Joint Photographic Experts Group) - Formato de compressão de imagem amplamente utilizado, ideal para fotografias.

JSON (JavaScript Object Notation) - Formato leve e legível para troca de dados entre sistemas.

MP4 (MPEG-4 Video File) - Formato padrão de arquivo de vídeo do MPEG-4, amplamente utilizado para vídeos digitais.

PDF (Portable Document Format) - Formato que preserva a formatação de documentos, garantindo consistência em diferentes dispositivos e sistemas.

PNG (Portable Network Graphics) - Formato de imagem sem perdas, com suporte para transparência.

RDF (Resource Description Framework) - Padrão para descrever e interligar recursos na web, usando dados semânticos.

TIFF (Tagged Image File Format) - Formato de imagem sem perdas, ideal para impressões e arquivos de alta qualidade.

TXT (Text Format) - Arquivo simples de texto sem formatação avançada, utilizado para dados básicos.

XML (Extensible Markup Language) - Linguagem de marcação amplamente utilizado para a codificação de documentos estruturados, legível por humanos e máquinas.

Informe o volume aproximado para dados coletados e/ou produzidos:

Entre 10 GB e 100 GB

Indique quais softwares serão utilizados para a coleta ou produção de novos dados:

Ainda não definido.

Declare como será feito o versionamento dos dados:

Ainda não definido

Indique a(s) metodologia(s) que será(ão) utilizada(s) para a coleta ou produção de novos dados:

Abordagem qualitativa

Indique a abordagem qualitativa.

Ainda não definido

Selecione a origem principal dos dados que serão coletados no projeto.

Não sei

Computacionais - Dados originados de sistemas computacionais, como simulações, modelagem, sensores automáticos, logs de servidores ou qualquer dado gerado por meio de processamento digital.

Experimentais - Dados obtidos por meio de experimentos controlados, nos quais variáveis são manipuladas para investigar causas e efeitos.

Governamentais - Dados coletados, gerenciados e disponibilizados por entidades governamentais, incluindo estatísticas demográficas, informações econômicas, dados geo ambientais, registros de saúde e regulamentações.

Observacionais - Dados coletados diretamente da observação de fenômenos naturais ou comportamentos, sem intervenção ativa do pesquisador.

Redes Sociais - Informações geradas a partir das interações de usuários em plataforma de redes sociais, incluindo perfis, postagens, curtidas, compartilhamentos e outros tipos de engajamento on-line.

Indique quais repositórios ou bases de dados serão utilizados para obter os dados que foram coletados ou gerados por terceiros:

Ainda não definido.

Indique a origem dos dados a serem reutilizados:

Combinação de dados de terceiros e dados reutilizados de projetos anteriores do próprio pesquisador

Informe se a proveniência - histórico e origem dos dados - serão registrados durante o projeto.

Sim

Indique como a proveniência dos dados será documentada.

Ainda não definido.

Informe os tipos de dados que serão reutilizados:

Áudio: gravações sonoras (fala, música, ruídos, sinais acústicos), armazenadas como arquivos de som (ex.: WAV, MP3) ou fluxos digitais.

Dados numéricos: valores quantitativos em forma de números (inteiros, decimais, séries), geralmente em tabelas, planilhas ou bancos de dados (ex.: medições, contagens, indicadores).

Dados temporais espacializados: dados que combinam localização (onde) e tempo (quando), permitindo ver mudanças ao longo do tempo em um espaço (ex.: séries temporais de chuva por município, evolução de desmatamento por área).

Fotografias aéreas: imagens capturadas do ar (aviões, drones), geralmente com alta resolução e usadas para mapeamento, inspeção e análise do território.

Imagem: representação visual digital 2D que pode ser foto, figura, microscopia, satélite etc.; pode ser comum (RGB) ou conter bandas, espectros e metadados.

Mapas de uso e cobertura da terra (pixels): mapas temáticos normalmente em formato raster, onde cada pixel é classificado em uma categoria (ex.: floresta, água, agricultura, urbano).

Modelos 3D: representações tridimensionais de objetos ou ambientes (ex.: malhas 3D, nuvens de pontos, modelos BIM), usadas para simulação, visualização e medições.

Texto: informação escrita, livre ou estruturada, como relatórios, descrições, transcrições, artigos ou campos textuais em tabelas.

Vídeo: sequência de imagens ao longo do tempo, normalmente com áudio opcional (ex.: MP4), usada para registro de eventos, monitoramento e análises temporais.

Informe os formatos dos dados que serão reutilizados:

CSV (Comma-Separated Values) - Formato simples de arquivo em que os valores dos dados são separados por vírgulas, amplamente utilizado em tabelas e planilhas.

DOC (Document Format) - Formato de documento do Microsoft Word, usado para criar e editar textos com formatação.

Excel (XLS/XLSX) – Formatos de arquivos para planilhas eletrônicas, usados em versões antigas (XLS) e recentes (XLSX) do Microsoft Excel.

JPEG (Joint Photographic Experts Group) - Formato de compressão de imagem amplamente utilizado, ideal para fotografias.

MP4 (MPEG-4 Video File) - Formato padrão de arquivo de vídeo do MPEG-4, amplamente utilizado para vídeos digitais.

PDF (Portable Document Format) - Formato que preserva a formatação de documentos, garantindo consistência em diferentes dispositivos e sistemas.

PNG (Portable Network Graphics) - Formato de imagem sem perdas, com suporte para transparência.

RDF (Resource Description Framework) - Padrão para descrever e interligar recursos na web, usando dados semânticos.

TIFF (Tagged Image File Format) - Formato de imagem sem perdas, ideal para impressões e arquivos de alta qualidade.

TXT (Text Format) - Arquivo simples de texto sem formatação avançada, utilizado para dados básicos.

XML (Extensible Markup Language) - Linguagem de marcação amplamente utilizado para a codificação de documentos estruturados, legível por humanos e máquinas.

Informe o volume aproximado dos dados reutilizados:

Entre 1 GB e 10 GB

Informe se há restrições de uso para os dados que serão obtidos de outras fontes:

Sim

Indique o tipo de restrição.

Ainda não definido

Indique como os dados de terceiros serão processados.

Ainda não definido.

Indique quem gerou ou coletou os dados de terceiros:

Não sei

SEÇÃO C

Documentação e Qualidade dos Dados

Escolha o padrão de metadados que será adotado para descrever os conjuntos de dados.

DDI (Data Documentation Initiative) - Padrão de metadados para a documentação e descrição de conjuntos de dados, facilitando a descoberta e uso de dados de pesquisa.

Dublin Core - Padrão de metadados para a descrição de recursos de informação, como documentos, imagens, páginas web, vídeos e outros tipos de recursos digitais.

Formato ISO 19115-1:2014 - Padrão de metadados que define o esquema necessário para a descrição de informações e serviços geográficos, incluindo detalhes sobre a identificação, extensão, qualidade, esquemas espaciais e temporais, referência espacial e distribuição de dados geográficos digitais.

Indique quais metadados serão utilizados para identificar os dados, de forma a facilitar que outras pessoas os encontrem.

Os conjuntos de dados serão descritos com base no padrão Dublin Core, complementado pelo DDI para os dados qualitativos e pelo ISO 19115-1:2014 (Perfil MGB 2.0) para os dados geoespaciais. Serão registrados os seguintes metadados de identificação: título do conjunto de dados; autor/responsável; instituição (Fundação Oswaldo Cruz); data de produção; descrição/resumo; tipo de dado (texto, áudio, numérico, vetorial); formato do arquivo; idioma; versão; condições de acesso e licença; identificador persistente (DOI); e palavras-chave. Para o mapa georreferenciado das iniciativas, os metadados incluirão: nome da iniciativa; território; município; unidade federativa; região; coordenadas geográficas; tema (saúde, ambiente ou violência); atores; período de atuação; métodos de coleta; tipos e formatos de dados; sistemas e plataformas utilizadas; produtos; fontes de financiamento; parcerias; condições de acesso; práticas de preservação; e mecanismos de governança. Para as entrevistas e estudos de caso, os metadados incluirão: identificador da entrevista; iniciativa vinculada; território; data de realização; papel do entrevistado; duração; tema; formato do arquivo; situação de consentimento; e versão da transcrição. Palavras-chave: dataficação popular; práticas documentárias; favelas; governança de dados; curadoria digital; preservação digital; saúde; ambiente; violência; princípios fair; princípios care; geração cidadã de dados.

SEÇÃO D

Armazenamento e Backup durante o processo de pesquisa.

Defina, no mínimo, dois locais onde os dados serão armazenados:

Arca Dados - Repositório institucional de dados da Fiocruz.

Dispositivos de armazenamento externos (como pen drives, HDs externos ou SSDs portáteis).

Serviços de nuvem institucional – Plataformas de armazenamento em nuvem oferecidas por instituições de pesquisa ou empresas.

Serviços de nuvem privada – Plataformas de armazenamento em nuvem de propriedade ou gestão privada.

Informe quais medidas serão adotadas para garantir a segurança e a proteção dos dados durante a pesquisa:

Controle de acesso restrito – Limitação de acesso aos dados apenas para pessoas autorizadas.

Armazenamento em repositórios seguros – Utilização de plataformas confiáveis para guardar os dados.

Backup regular – Realização periódica de cópias de segurança para evitar perdas.

SEÇÃO E

Requisitos Legais e Éticos

Informe se a equipe está ciente das normas relacionadas à avaliação ética na pesquisa:

Sim

Informe se a pesquisa envolve o tratamento de dados que exige avaliação ética:

Sim

Indique se a equipe tem conhecimentos sobre a legislação sobre proteção e tratamento de dados pessoais:

Sim

Informe se a pesquisa envolve tratamento de dados pessoais.

Sim

Informe se o tratamento de dados pessoais está em conformidade com a Lei nº 13.709/2018.

Sim

Indique se há previsão de período de embargo relacionado ao direito autoral.

Sim para uma parte dos dados

Informe se a pesquisa envolve outras categorias de dados que exijam período de embargo para sua disponibilização?

Não

SEÇÃO F

Compartilhamento de Dados

Indique se a pesquisa prevê compartilhamento de dados.

Ainda não definido

Informe a licença a ser aplicada aos dados do projeto.

Ainda não definido.

Informe quais ferramentas de software serão necessárias para acessar e usar os dados:

Software especializado para análise de dados (exemplos: MATLAB, R, Python)

Sistema de Informação Geográfica (exemplos: QGIS, ArcGIS)

Ferramentas de edição de texto e planilhas (exemplos: Microsoft Word, Excel)

Navegador de internet para acessar repositórios online

Informe como será assegurado o registro de um identificador único e persistente (como DOI - Digital Object Identifier) para os conjuntos de dados:

Ainda não definido.

SEÇÃO G

Responsabilidades e Recursos Financeiros para Gestão de Dados de Pesquisa

Indique quais recursos que serão dedicados à gestão dos dados e à garantia de que sejam FAIR (localizáveis, acessíveis, interoperáveis, reutilizáveis):

Financiamento específico para a gestão de dados

Dedicação de tempo da equipe para curadoria e manutenção dos dados

Infraestrutura tecnológica (servidores, sistemas de armazenamento etc.)

Ferramentas de software para garantir a interoperabilidade dos dados

Capacitação da equipe em gestão de dados FAIR

Indique o(s) responsável(eis) pelas principais atividades de gestão de dados no projeto:

O pesquisador principal – supervisiona e toma as decisões finais relacionadas à gestão dos dados.

O especialista em dados – realiza a curadoria e manutenção dos dados ao longo do projeto.

A equipe de TI – gerencia a infraestrutura de armazenamento e garante a segurança dos dados.

A equipe de suporte – elabora a documentação e realiza o compartilhamento dos dados.