



Plano de Gestão de Dados Acionável por Máquina, da teoria à prática: uma análise das ferramentas ARGOS e FioDMP

Viviane Veiga / Patrícia Henning / Erick Penedo / Simone Dib
Fiocruz/ICICT

João Cardoso / Filipa Pereira
FCT/FCNN

Aceleração dos avanços científicos em saúde requer reuso de dados de pesquisa

(Mons, 2020)

Benefícios do reuso de dados de pesquisa

Econômicos



€100 billion

Sociais



Salvar vidas

(Mons, 2020)



O reuso de dados
começa com um
adequado

Plano de Gestão de
dados

O que é um plano de gestão de dados?

Documento formal que descreve detalhadamente como os dados serão tratados ao longo de todo o ciclo de vida da pesquisa, e após sua conclusão.

O que é um plano de gestão de dados?

Os sistemas de PGD utilizam *templates* com uma série de questões pré-estabelecidas que abordam aspectos relacionados à coleta, produção, organização, compartilhamento e preservação dos dados

A importância do gerenciamento eficiente de dados

Diretrizes Éticas e Legais

Um plano de gestão de dados (PGD) eficiente é crucial para um gerenciamento ético e legal dos dados.

Etapas Automatizadas

Os pesquisadores também exigem que os PGDs automatizem as etapas para criá-los e atendam às suas necessidades de gerenciamento de dados.

Avanços Tecnológicos

Um bom DMP deve enfatizar a interoperabilidade semântica, tecnológica e política.

Dados compartilháveis

Deve apoiar na geração de dados compartilháveis

Estudos mostram que o PGD...

Não conduz o pesquisador em uma gestão de dados alinhada às questões éticas e legais e aos principais requisitos de gestão de dados de pesquisa dos domínios disciplinares

Não apoia os pesquisadores na gestão de dados de pesquisa e na geração de dados compartilháveis com alto potencial de reuso.

Não enfatiza a interoperabilidade semântica, tecnológica e política.

Não automatiza os passos para elaboração do PGD.

Não atende às demandas dos pesquisadores.

Não está alinhado aos Princípios FAIR

Estudos mostram que os PGDs:

Se tornou um documento **burocrático**, com **informação ambígua** ou **incompleta** e muitas das vezes **sem utilidade**

10 Princípios para PGDam

Pensando em soluções

Princípio 1	Integrar o PGD ao fluxo de trabalho de todas as partes integrantes do ecossistema de dados de pesquisa.
Princípio 2	Permitir que sistemas automatizados atuem em nome dos atores envolvidos.
Princípio 3	Elaborar políticas tanto voltadas para as pessoas quanto para as máquinas.
Princípio 4	Descrever os componentes do ecossistema de gerenciamento de dados para máquinas e seres humanos.
Princípio 5	Usar Identificadores Persistentes e Vocabulários Controlados.
Princípio 6	Seguir um modelo de dados comum para o PGD acionável por máquina.
Princípio 7	Disponibilizar o PGD para uso de pessoas e máquina.
Princípio 8	Apoiar a avaliação e o monitoramento da gestão de dados.
Princípio 9	Atualizar o PGD com versão de documento.
Princípio 10	Publicar o PGD.

Miksa et al. (2019)

Parceria



FCT

Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Pensando em soluções



A FCT (FCCN), em conjunto com a Universidade do Minho, conduziu um estudo prévio detalhando nove ferramentas de PGD em uso pela comunidade científica internacional. Após uma análise comparativa das funcionalidades e características das nove ferramentas, chegou-se à conclusão de que a ferramenta ARGOS era a opção que melhor se enquadrava para um serviço nacional a ser prestado pela FCT a todas os pesquisadores em Portugal. Vale a pena ressaltar que a ferramenta ARGOS é desenvolvida pelo projeto [OpenAIRE](#).

<http://repositorium.sdum.uminho.pt/handle/1822/67530>

Pensando em soluções



A Fiocruz, através do Ict, também realizou um estudo prévio de diversas ferramentas disponíveis para a elaboração de PGD. Após testes e análises dessas ferramentas de acesso livre, constatou-se que nenhuma delas atendia às particularidades e idiosincrasia da instituição. Diante dessa situação, a Fiocruz optou por desenvolver sua própria ferramenta, intitulada FioDMP. Por meio dessa decisão foi possível atender as especificidades da instituição, bem como as necessidades dos seus pesquisadores, possibilitando, ainda, a integração com outros sistemas da instituição, da comunidade científica e do campo da saúde. A FioDMP foi a primeira ferramenta a oferecer duas opções de formulário: simplificada e FAIR.

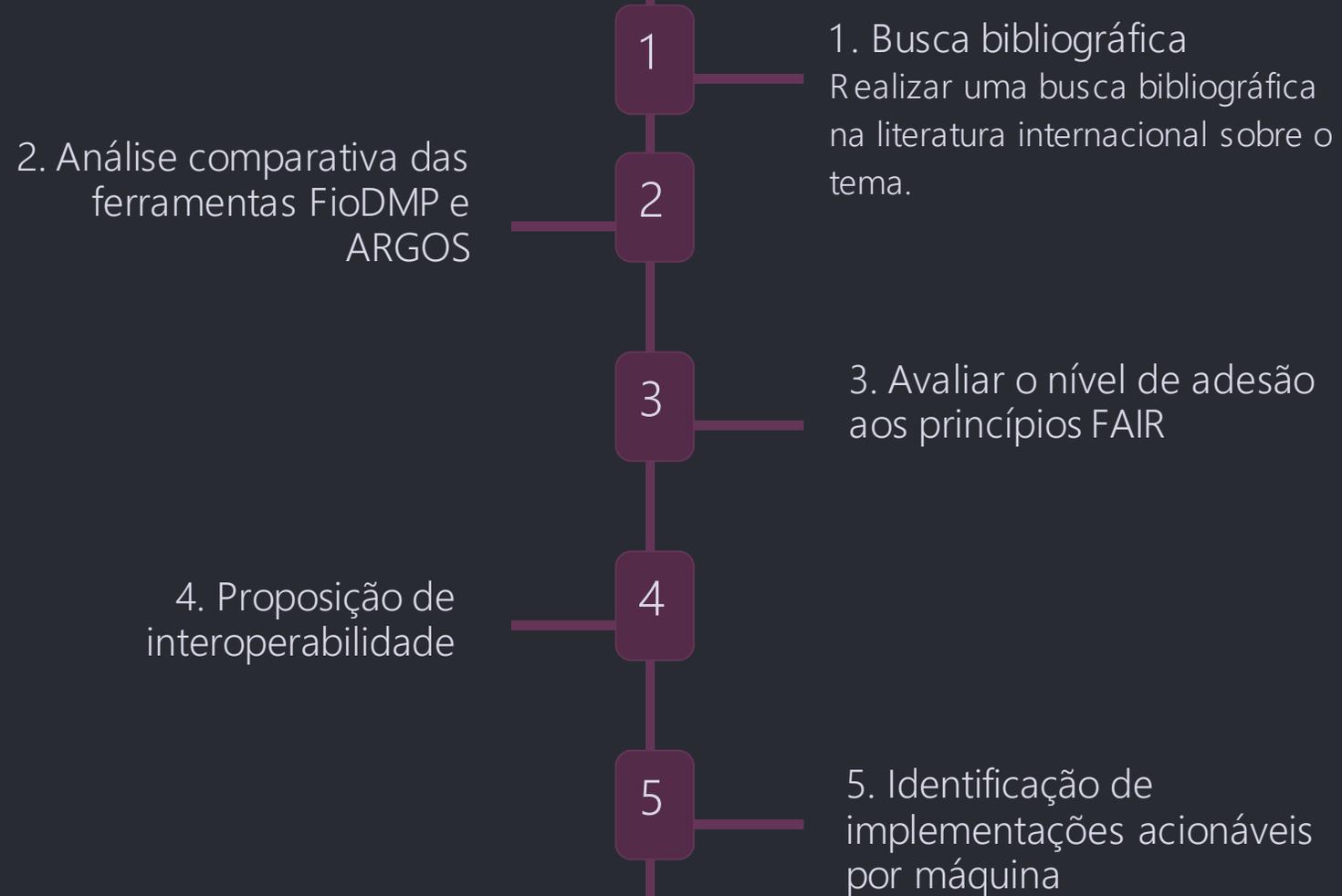
Pensando em soluções

Como propor melhorias tecnológicas para as ferramentas ARGOS e FioDMP, de modo a atender aos 10 princípios para PGDam?

Pensando em soluções

- interoperabilidade tecnológica, semântica e política entre os atores envolvidos na elaboração de PGD
- Alinhamento dos PGDs aos Princípios FAIR
- Dados compartilháveis e acionáveis e compreensíveis tanto por humanos quanto por máquinas
- Alto potencial de dados reusáveis

Metodologia



Para que serve um plano de gestão de dados?

Não conduz o pesquisador em uma gestão de dados alinhada às questões éticas e legais e aos principais requisitos de gestão de dados de pesquisa dos domínios disciplinares e dos princípios FAIR.

Não enfatiza a interoperabilidade semântica, tecnológica e política.

Não automatiza os passos para elaboração do PGD.

Não atende às demandas dos pesquisadores.

FioDMP

Argos

Modelo Simplificado e Modelo FAIR

Integração com DeCS/Mesh

Modelos com princípios FAIR

Integração com ORCID

Integração com Crossref

Login Único Fiocruz

Integração com Crossref

Diferentes Templates

Exportação em PDF, DOCX e JSON

Integração com Repositório Arca Dados*

Integração com o DMP Common Standards

Exportação de PGD em formato PGDma

Identificador Persistente

Preenchimento Simplificado

Integração com Repositório Zenodo

Integração com DataCite

Funcionalidades

FioDMP



1) **Crossref**: utiliza a interface de programação de aplicação. A API do Crossref responde perguntas relacionadas à afiliação/instituição do pesquisador e para indicação das agências financiadoras da pesquisa. Nesses casos, o sistema Crossref busca as informações, em seu próprio banco de dados. Conforme o pesquisador inicia a digitação aparecem as opções para ele selecionar;

Pensando em soluções



2) **Dataverse**: utiliza a API do Dataverse para integração com o repositório de dados da Fiocruz, o Arca Dados. Essa integração possui três funções:

a) ao finalizar o preenchimento do PGD, o pesquisador poderá clicar no botão ‘Depositar no Repositório’ que possibilitará a criação automática de um conjunto de dados, diretamente no repositório Arca Dados, contendo as informações preenchidas no PGD;

Pensando em soluções

2) Dataverse



b) o depósito dos formatos PDF, DOCX e JSON do PGD que foram gerados no final do seu preenchimento;

c) após esses procedimentos o Arca Dados gera um identificador único (DOI) associado ao conjunto de dados e retorna o link do DOI, referente ao conjunto de dados, ao citado PGD no FioDMP;

Pensando em soluções



3) **DeCS/Mesh**: utiliza a API do vocabulário controlado DeCS/Mesh para fornecer opções de palavras-chave. À medida que o usuário digita vão aparecendo as opções dos termos ligados ao vocabulário controlado. Caso o termo digitado não conste no vocabulário, o sistema permite ao usuário escrever o termo livremente;

4) **Login Único Fiocruz**: utiliza a API denominada 'Login único Fiocruz' que recorre ao sistema de autenticação institucional da Fiocruz, possibilitando que aos usuários da instituição façam *login* na plataforma FioDMP sem a necessidade de criarem uma nova conta, utilizando, apenas, o *e-mail* e senha institucional;

5) **Exportação em formatos DOCX, PDF ou JSON**: ao finalizar o PGD é possível gerar um arquivo contendo todas as informações preenchidas no FioDMP e exportá-lo em três diferentes formatos: DOCX, PDF e JSON. O formato JSON segue os princípios do *RDA DMP Common Standards*, que permite a integração com outros sistemas de PGDs, que também utilizem esse mesmo padrão. Os formatos DOCX e PDF apresentam um *design* personalizado da Fiocruz.

Pensando em soluções



Argos

De forma a potenciar as suas funcionalidades e a permitir uma maior adaptação aos requisitos da FCT, foram então consideradas algumas customizações e integrações, destacando-se:

- a) O incentivo à utilização de **identificadores persistentes**, designadamente de autor, projeto e financiador;
- b) A **adição de campos pré-definidos via API**, de forma a agilizar o preenchimento dos PGDs e a assegurar PGDam. Este processo foi facilitado, uma vez que o OpenAIRE recorre às API do Datacite, re3data.org, Crossref, ORCID, Cordis e EOSC.

c) **mapeamento entre questões** que compõem o *template* PGD, e os campos que compõem o perfil aplicativo DCS.

Resultando em:

- 1- Mapeamento entre questões e um campo específico do perfil aplicativo DCS;
- 2- Exportação de PGD em formato PGDma, em cumprimento com o perfil aplicativo DCS.

Concluindo sem concluir

A colaboração entre instituições e domínios, a integração de tecnologias avançadas e a busca constante por melhorias são fatores cruciais para o avanço de um sistema de PGD.

É necessário constante avaliação de planos de gestão de dados para garantir sua real utilidade em apoiar na geração de dados compartilháveis e com alto potencial de reuso



Viviane Veiga



Filipa Pereira



Patrícia Henning



Erick Penedo



João Cardoso



Simone Dib