



Plano de gestão de dados FAIR: uma proposta para a Fiocruz

FAIR data management plan: a proposal for Fiocruz

Viviane Veiga*

Patricia Henning**

Simone Dib***

Erick Penedo****

Jefferson da Costa Lima*****

Luiz Olavo Bonino da Silva*****

Luís Ferreira Pires*****

* Doutora em Ciências pelo Programa de Pós-Graduação em Informação e Comunicação em Saúde da Fundação Oswaldo Cruz (Fiocruz). Professora do Programa de Pós-Graduação em Informação e Comunicação em Saúde da Fundação Oswaldo Cruz (Fiocruz) e Coordenadora da Rede de Bibliotecas da Fundação Oswaldo Cruz (Fiocruz). Endereço: Av. Brasil, 4365 - Manguinhos, Rio de Janeiro - RJ, 21040-360. Telefone: (21) 3865-3201 E-mail: viviane.veiga@icict.fiocruz.br

** Doutora em Ciências pelo Programa de Pós-Graduação em Informação e Comunicação em Saúde da Fundação Oswaldo Cruz (Fiocruz). Professora do Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Endereço: Av. Pasteur, 458, 4o andar - Urca, Rio de Janeiro - RJ, 22290-240, Brasil. Telefone: (21) 2542 8542. Email: henningpatricia@gmail.com

*** Mestre em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação, convênio UFRJ/IBICT (2013). Possui especialização em Gestão do Conhecimento e Inteligência Empresarial pela Universidade Federal do Rio de Janeiro (2003), bibliotecária da Fundação Oswaldo Cruz e atua na Biblioteca de Saúde Pública, Endereço: Av. Brasil, 4365 - Manguinhos, Rio de Janeiro - RJ, 21040-360, Brasil. Telefone: (21) 3865-3201 E-mail: simone.dib@icict.fiocruz.br

**** Graduação em Sistemas de Informação pela Universidade do Grande Rio. Analista de Informação e Comunicação no Instituto de Comunicação e Informação Científica e Tecnológica em Saúde da Fundação Oswaldo Cruz (Fiocruz). Endereço: Av. Brasil, 4365 - Manguinhos, Rio de Janeiro - RJ, 21040-360, Brasil. Telefone: (21) 3865-3201 E-mail: erick.penedo@icict.fiocruz.br

***** Mestre pela Escola de Matemática Aplicada da Fundação Getúlio Vargas (EMAP/FGV) em 2016, especialização em Sistemas de Informação pela Universidade Federal do Rio de Janeiro (NCE/2012) e graduação em Matemática pela Universidade do Estado do Rio de Janeiro (2004). Coordenador adjunto da Plataforma de Ciência de Dados aplicada à Saúde (ICICT/Fiocruz). Endereço: Av. Brasil, 4365 - Manguinhos, Rio de Janeiro - RJ, 21040-360. Telefone: (21) 3882-9229 E-mail: jefferson.lima@icict.fiocruz.br.

***** Doutor em Ciência da Computação pela Universidade de Twente na Holanda. Coordenador Internacional de Tecnologia do Escritório de Apoio e Coordenação Internacional GO FAIR e Professor Associado do grupo BioSemantics no Centro Médico da Universidade de Leiden na Holanda. Endereço: Av. Pasteur, 458, 4o andar - Urca, Rio de Janeiro - RJ, 22290-240. Telefone: (+31) 624619131. E-mail: luiz.bonino@go-fair.org

***** Doutor em Engenharia Elétrica pela Universidade de Twente na Holanda. Professor Associado da Faculdade de Engenharia Elétrica, Matemática e Ciência da Computação e membro do grupo de Serviços e Cibersegurança (SCS) da Universidade de Twente na Holanda. Endereço: 5, Drienerlolaan, 7522 NB nschede, Holland. Telefone: (+31) 053 489 9111. E mail: l.ferreirapires@utwente.nl

RESUMO

Este artigo traz para discussão o papel dos planos de gestão de dados como instrumento facilitador da gestão dos dados durante todo o ciclo de vida da pesquisa. A abertura de dados de pesquisa é pauta prioritária nas agendas científicas, por ampliar tanto a visibilidade e transparência das investigações, como a capacidade de reprodutibilidade e reuso dos dados em novas pesquisas. Nesse contexto, os princípios FAIR, um acrônimo para ‘Findable’, ‘Accessible’, ‘Interoperable’ e ‘Reusable’ são fundamentais por estabelecerem orientações basilares e norteadoras na gestão, curadoria e preservação dos dados de pesquisa direcionados para o compartilhamento e o reuso. O presente trabalho tem por objetivo apresentar uma proposta de template de Plano de Gestão de Dados, alinhado aos princípios FAIR, para a Fundação Oswaldo Cruz. A metodologia utilizada é de natureza bibliográfica e de análise documental de diversos planos de gestão de dados europeus. Concluímos que a adoção de um plano de gestão nas práticas científicas de universidades e instituições de pesquisa é fundamental. No entanto, para tirar maior proveito dessa atividade é necessário contar com a participação de todos os atores envolvidos no processo, além disso, esse plano de gestão deve ser machine-actionable, ou seja, acionável por máquina.

Palavras-chave: Plano de Gestão de Dados; Dado de Pesquisa; Princípios FAIR; PGD Acionável por Máquina; Ciência Aberta.

INTRODUÇÃO

Estamos vivenciando uma mudança de paradigma das atividades científicas onde a produção do conhecimento deve ser “[,,] compartilhada abertamente, tão cedo quanto possível, desde o início da sua descoberta” (NIELSEN, 2011). Este é o lema da Ciência Aberta mencionado por Nielsen (2011) no livro *Reinventing Discovery*, que em seus diferentes estágios destaca a importância da abertura dos dados de pesquisa, seu compartilhamento e reuso. Essa temática vem sendo discutida com prioridade no meio acadêmico, nos institutos de pesquisa e nas principais agências de fomento nacionais e internacionais por envolver questões complexas tais como as particularidades intrínsecas dos diferentes domínios disciplinares; as questões legais, éticas e de integridade pertinentes ao compartilhamento e reuso dos dados; a

ABSTRACT

This article proposes to discuss the role of data management plans as a tool to facilitate data management during researches life cycle. Today, research data opening is a primary agenda at scientific agencies as it may boost investigations' visibility and transparency as well as the ability to reproduce and reuse its data on new researches. Within this context, FAIR principles, an acronym for Findable, Accessible, Interoperable and Reusable, is paramount, as it establishes basic and guiding orientations for research data management, curatorship and preservation with an intent on its sharing and reuse. The current work intends to present to the Fundação Oswaldo Cruz a new Data Management Plan template proposal, aligned with FAIR principles. The methodology used is bibliographical research and documental analysis of several European data management plans. We conclude that the adoption of a management plan on universities and research institutions scientific activities is paramount. However, to be fully benefited from this activity, all actors involved in the process must participate, and, on top of that, this plan must be machine-actionable.

Keywords: Data Management Plan; Research Data; FAIR Principles; DMP Machine-Actionable; Open Science.

necessidade de infraestrutura tecnológica e técnica que garanta a interoperabilidade e a preservação dos dados; a sensibilização e capacitação dos atores envolvidos e, a preocupação com financiamentos e sustentabilidade a médio e longo prazo.

Destaca-se a importância da inserção do pesquisador e dos demais atores envolvidos no processo de gestão e abertura de dados, tendo em vista não apenas as exigências das agências de fomento e das políticas das revistas científicas, mas também, pelo altruísmo científico implícito no processo e pelo reconhecimento que lhes serão atribuídos em suas carreiras e vida profissional.

Nesse contexto, um Plano de Gestão de Dados (PGD) se faz necessário por ser considerado um instrumento das boas práticas de gestão possibilitando que os dados sejam acessados, preservados e reusados em novas pesquisas. Outro aspecto importante a ser considerado é que o PGD possa ser legível e acionável por máquina, permitindo a integração automatizada com outros sistemas de informação, facilitando o preenchimento e otimizando o tempo do pesquisador.

Com o intuito de possibilitar que os dados sejam mais facilmente encontráveis, acessíveis, interoperáveis e reusáveis, surgem em 2014 os princípios FAIR, um acrônimo para ‘Findable’, ‘Accessible’, ‘Interoperable’ e ‘Reusable’, mundialmente considerados os elementos norteadores das boas práticas de todo o processo de gestão de dados de pesquisa. Nesse sentido, foi dada ênfase à compreensão do ecossistema dos princípios FAIR, de autoria do European Commission Expert Group on FAIR Data (2018) e dos artigos gerados pelo Grupo de Trabalho denominado “DMP Common Standards”¹ do Research Data Alliance (RDA)², o que nos permitiu um maior aprofundamento nos elementos teóricos necessários para a elaboração desta proposta. Utilizou-se como abordagem metodológica a análise documental de diversos planos de gestão de dados de instituições europeias com o objetivo de oferecer à Fundação Oswaldo Cruz (Fiocruz) um modelo de Plano de Gestão de Dados alinhado aos princípios FAIR, bem como discutir evidências sobre estudos recentes sobre PGD *machine-actionable* ou seja acionáveis por máquina.

OS PRINCIPIOS FAIR

Os princípios FAIR surgem como resultado da conferência internacional ‘Jointly designing a data FAIRPORT’³ realizada em 2014, com o intuito de contribuir para o aumento do compartilhamento e reuso dos dados de pesquisa. No entanto, sua publicação só aconteceu dois anos depois no periódico *Scientific Data, da Nature*, por Wilkinson et al. (2016). A partir de então, esses princípios passaram a ser reconhecidos internacionalmente, sendo considerados os elementos norteadores para a descoberta, acesso, interoperabilidade e reuso dos dados, por oferecerem orientações basilares no tratamento, reuso e publicação de conjunto de dados, sob condições claramente entendidas por humanos e máquinas (WILKINSON et al., 2016).

Apresentamos abaixo os princípios de forma simplificada:

¹ Disponível em: <<https://www.rd-alliance.org/groups/dmp-common-standards-wg>> Acesso em: 06 ago 2019.

² Disponível em: <<https://www.rd-alliance.org/>> Acesso em: 06 ago 2019.

³ Disponível em: <<https://www.lorenzcenter.nl/lc/web/2014/602/info.php3?wsid=602>> Acesso em: 06 ago 2019.

Localizável (Findable)

- a) dados e metadados precisam ter um identificador único persistente;
- b) os dados devem ser descritos com metadados ricos;
- c) ter o identificador persistente para o conjunto de dados descrito nos metadados;
- d) metadados e dados devem ser recuperáveis por meio de repositórios confiáveis;

Acessível (Accessible)

- a) dados e metadados devem ser recuperados pelo seu identificador usando protocolos de comunicação padrão;
- b) os protocolos devem ser gratuitos, abertos e suportar autenticação e autorização;
- c) os metadados devem estar acessíveis mesmo quando os dados não estiverem mais disponíveis.

Interoperável ('Interoperable')

- a) dados e metadados devem ser codificados usando padrões de representação acordados;
- b) dados e metadados devem usar vocabulários alinhados aos princípios FAIR e incluir referências relevantes.

Reutilizável (Reusable)

- a) dados e metadados precisam estar associados a atributos relevantes;
- b) dados e metadados devem ser liberados com licenças de uso claramente definidas;
- c) metadados e dados devem estar associados à sua proveniência de forma detalhada;
- d) dados e metadados devem atender aos padrões da comunidade (HUDSON, 2018).

Recentemente, o *Expert Group on FAIR Data* (2018) lançou em novembro de 2018 o Relatório Final e Plano de Ação denominado '*Turning FAIR into Reality*'. Esse documento é visto como um instrumento condutor para a aplicação dos princípios FAIR em diferentes contextos e dimensões. Além de discorrer sobre conceitos, benefícios, habilidades, métricas e padrões, o documento apresenta o Ecossistema dos dados FAIR, formado por cinco componentes: políticas, planos de gestão de dados, identificadores persistentes, padrões e repositórios. Ademais, esclarece a relação do PGD com os princípios FAIR, o que nos concedeu o embasamento teórico para a condução desta proposta (EUROPEAN COMMISSION EXPERT GROUP ON FAIR DATA, 2018).

CIÊNCIA ABERTA E DADOS ABERTOS NA FIOCRUZ

Historicamente, a Fundação Oswaldo Cruz (Fiocruz) exerce um papel importante no percurso da implantação da Ciência Aberta no Brasil. Esse papel está expresso nas diversas práticas que vêm sendo estudadas, desenvolvidas e gradativamente implementadas ao longo dos últimos 10 anos. O acesso aberto à produção científica, por exemplo, consolidou-se por meio do Repositório Institucional da Fiocruz – ARCA⁴, lançado oficialmente em 2011 e fortalecido com a Política de Acesso Aberto ao Conhecimento⁵, instituída em 2014, garantindo à sociedade o acesso gratuito, público e aberto ao conhecimento produzido por essa instituição (FUNDAÇÃO OSWALDO CRUZ, 2014).

Consequentemente, estudos específicos vêm sendo realizados pelo Grupo de Trabalho em Ciência Aberta da Fiocruz desde o início de 2017, voltados para assessorar a elaboração da política de compartilhamento e abertura de dados de pesquisa na Fiocruz. Essa política é considerada um desdobramento do movimento pelo acesso aberto ao conhecimento científico, ora em curso.

Esses estudos deram origem a três documentos-base criados para subsidiar a implementação da abertura dos dados de pesquisa da instituição, fortalecendo a cultura da ciência aberta focada em práticas científicas mais colaborativas e participativas:

- a) *Livro Verde: Ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional, que sistematiza a experiência de oito países e da União Europeia na implantação da abertura de dados para pesquisa* (SANTOS, 2017).
- b) *Marcos legais nacionais em face da abertura de dados para pesquisa em saúde: dados pessoais, sensíveis ou sigilosos e propriedade intelectual contendo uma compilação de atos normativos identificados no ordenamento jurídico nacional com coerência com a abertura de dados para pesquisa em saúde* (GUANAES, 2018).
- c) *Termo de Referência: gestão e abertura de dados para pesquisa na Fiocruz: apresenta um conjunto de princípios e diretrizes voltados para uma proposta inicial de abertura de dados de pesquisa na Fiocruz* (FUNDAÇÃO OSWALDO CRUZ, 2019).

A necessidade de criação de um Plano de Gestão de Dados alinhado aos princípios FAIR para essa instituição foi identificada com precisão no documento *Termo de Referência: gestão e abertura de dados para pesquisa na Fiocruz*, especificamente:

No Princípio 2 – Gestão e abertura de dados de pesquisa, sobretudo, nas seguintes diretrizes:

- “D 2.3 Adotar o plano de gestão de dados como ferramenta de apoio à visão estratégica da Fiocruz que leve em consideração as especificidades da pesquisa e a sua tipologia de dados [...]” (p. 7) e

⁴ Disponível em: <<https://www.arca.fiocruz.br/>>. Acesso em: 06 ago 2019.

⁵ Disponível em: <https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/portaria_-_politica_de_acesso_aberto_ao_conhecimento_na_fiocruz.pdf>. Acesso em: 11 ago 2019.

- “D 2.9 Adotar os princípios FAIR na gestão de dados e utilizar indicadores e métricas para avaliar o grau de aplicação dos princípios FAIR (FAIRness) [...]” (p. 8).

É importante destacar que um modelo de Plano de Gestão de Dados de pesquisa institucional deve estar coerente com os princípios FAIR e atender aos imperativos inerentes aos diferentes domínios disciplinares. Com isso, a instituição se insere em mais uma das práticas da Ciência Aberta.

PLANO DE GESTÃO DE DADOS

O avanço das tecnologias amplia cada vez mais a possibilidade de se trabalhar com dados periodicamente capturados que demandam processamento em larga escala, exigindo planejamento e gerenciamento seguro de forma a garantir a sua interpretação e reuso em novas pesquisas.

O Plano de Gestão de Dados é uma ferramenta para apoiar na solução de problemas de gestão por ser um documento formal, dinâmico e vivo que descreve detalhadamente como os dados de pesquisa serão tratados durante todo o ciclo de vida da pesquisa e possivelmente após a sua conclusão. Na Europa e nos Estados Unidos, essa é uma prática já formalizada junto às agências de fomento. A Comissão Europeia, por exemplo, desde o início de 2017, passou a exigir um plano de gestão de dados de todos os candidatos a financiamentos, de acordo com as recomendações do *Guidelines on FAIR Data Management in Horizon 2020* (EUROPEAN COMMISSION, 2016). Com isso consegue garantir maior transparência nos seus investimentos, melhor controle na gestão dos dados gerados, facilitando, quando possível, o seu compartilhamento.

No Brasil, essa exigência junto às agências de fomento ainda não é uma realidade. Apresenta-se de forma acanhada somente em alguns projetos financiados pela Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP).⁶ No entanto, percebe-se que o entendimento sobre o cenário mencionado está em estágio inicial no nosso país, sendo reconhecido o esforço e o ineditismo no Brasil na elaboração desta proposta de *template* de plano de gestão de dados de pesquisa, baseado nos princípios FAIR.

Proposta de Plano de Gestão de Dados para a Fiocruz

A criação de um PGD alinhado aos princípios FAIR, que respeite os imperativos inerentes aos diferentes domínios disciplinares do campo da saúde, é um desafio necessário que está sendo enfrentado pela Fiocruz para garantir o sucesso da gestão, compartilhamento e reuso dos seus dados de pesquisa. Esse desafio foi assumido pelo Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT), indo ao encontro do fortalecimento do novo paradigma das práticas de gestão de dados contidas no âmbito da Ciência Aberta mundial.

A proposta do ICICT/Fiocruz de um PGD foi baseada no documento intitulado “Practical Guide to The International Alignment of Research Data Management” elaborado pelas ‘Science Europe’ e ‘Netherlands Organisation for Scientific Research’

⁶ Disponível em: <<http://www.fapesp.br/gestaodedados>>. Acesso em: 06 ago 2019.

(SCIENCE EUROPE, 2019) por considerá-lo, entre todos os PGD analisados, o que mais se alinha aos princípios FAIR.

Além disso, foi necessário adaptar esse documento às necessidades dos pesquisadores da instituição, sendo ele dividido em sete etapas, contendo um total de 35 perguntas conforme apresentadas no quadro 1.

Quadro 1 – Proposta de *template* de Plano de Gestão de Dados do ICICT/Fiocruz

Seção A: Informação Administrativa	
A1.	Informe o nome completo do pesquisador principal
A2.	Informe o e-mail de Contato
A3.	Qual o seu telefone?
A4.	Informe o link do Currículo Lattes do Pesquisador Principal
A5.	Informe o link do ORCID do Pesquisador Principal.
A6.	Informe o título do seu projeto
A7.	Descreva o resumo do projeto
A8.	A qual Unidade da Fiocruz o seu projeto está vinculado?
A9.	Qual a data de início do projeto?
A10.	Indique a previsão de término do projeto
A11.	O projeto tem financiamento?
A12.	Especifique o financiamento
A13.	Qual a versão deste Plano de Gestão de Dados?
A14	Quem será o responsável pelo preenchimento e atualização do Plano de Gestão de Dados?
Seção B: Descrição dos dados coletados ou reuso de dados existentes	
B1.	Como os dados serão coletados, produzidos ou como os dados existentes serão reutilizados?
B2.	Quais os tipos de dados que serão coletados ou produzidos?
B3.	Quais os formatos de dados que serão coletados ou produzidos?
B4.	Qual o volume aproximado dos dados coletados ou produzidos?
Seção C: Documentação e Qualidade dos Dados	
C1.	Indique os metadados adotados
C2.	Indique o padrão de metadados adotado
C3.	Indique os documentos que acompanharão os dados
C4.	Indique as medidas adotadas para a organização e controle de qualidade dos dados
Seção D: Armazenamento e Backup durante o processo de pesquisa	
D1.	Como os dados serão armazenados e como serão feitas as cópias de segurança durante a pesquisa?
D2.	Como a segurança dos dados e a proteção dos dados sensíveis serão tratadas durante a

	pesquisa?
Seção E: Requisitos Legais, Éticos e de Códigos de Conduta	
E1.	Se os dados pessoais forem tratados, como será assegurado o cumprimento da legislação relativa aos dados pessoais e à proteção dos dados?
E2.	Como as questões legais referentes aos direitos de propriedade intelectual serão gerenciadas? Qual legislação se aplica?
E3.	Indique a licença que será aplicada.
E4.	Como as possíveis questões éticas e os códigos de conduta serão levados em consideração?
Seção F: Compartilhamento de Dados e Preservação a longo prazo	
F1.	Como e quando os dados serão compartilhados? Existem possíveis restrições ao compartilhamento e/ou à abertura de dados ou motivos para embargo?
F2.	Como os dados para preservação serão selecionados e onde os dados serão preservados a longo prazo (por exemplo, em um repositório de dados)?
F3.	Quais métodos ou ferramentas de software serão necessários para acessar e usar os dados?
F4.	Como será assegurado o registro de um identificador único e persistente (como um DOI - Digital Object Identifier) para cada conjunto de dados?
Seção G: Responsabilidades e Recursos Financeiros para a Gestão de Dados de Pesquisa	
G1.	Quem será o responsável pela gestão dos dados?
G2.	Quais recursos (ex.: financeiros, tempo) serão dedicados à gestão de dados e à garantia que eles sejam FAIR (localizável, acessível, interoperável, reutilizável)?
G3.	Para projetos colaborativos, explique como será a coordenação da gestão de dados e as responsabilidades de cada um dos parceiros.

Fonte: Elaborado pelos autores (2019)

O quadro 1 apresenta as 35 perguntas referentes à proposta de template do Plano de Gestão de Dados do ICICT/Fiocruz, dentre as quais 14 delas são de cunho administrativo e as 21 restantes são aquelas referentes à gestão dos dados de pesquisa propriamente. Essas perguntas estão inseridas em um formulário online, de acesso restrito aos pesquisadores da instituição, contendo textos explicativos de ajuda para o seu preenchimento, quando necessária. O objetivo deste formulário é simplificar ao máximo o procedimento de criação do PGD, levando em conta as múltiplas demandas dos pesquisadores e as possíveis barreiras relacionadas ao entendimento das perguntas. Contém perguntas fechadas, facilitando e minimizando erros de preenchimento. É possível atualizar as respostas durante todo o ciclo de vida da pesquisa e em cada atualização gera-se uma nova versão do plano. Vale ressaltar que neste momento este template do PGD está em fase de validação institucional. Estratégias de capacitação no preenchimento do *template*, tanto para os pesquisadores como para os bibliotecários e outros atores que serão os facilitadores no processo, estão em fase de elaboração. Também estão sendo avaliadas diversas ferramentas para automatizar o PGD, tendo em vista as tendências internacionais da área.

PGD Acionável por Máquina

Considerando a complexidade dos dados de pesquisa, em especial os provenientes de *big data* e aqueles que contêm informações sensíveis e sigilosas, a cooperação entre os atores envolvidos no processo de gestão se faz cada vez mais necessária. Segundo Miksa, Cardoso e Borbinha (2018, p.2733) muitas vezes “[...] os pesquisadores simplesmente não têm conhecimento suficiente, nem tempo para preparar um PGD, nem para realmente implementá-lo”.

As ferramentas tradicionais disponíveis para automatizar os planos de gestão de dados utilizadas pela comunidade científica internacional, como DMPonline⁷ ou DMPTool⁸, por exemplo, não atendem satisfatoriamente às demandas das pesquisas baseadas em dados complexos dos dias de hoje. Essas oferecem um formulário online com as questões referentes aos PGD, com possíveis explicações, com campos para respostas gerando, ao final, um relatório em .doc ou .pdf. A comunidade científica deseja muito mais do que um documento estático com finalidade administrativa. Portanto, o desafio do momento está em desenvolver um modelo de PGD acionável por máquina, o que permite a participação de forma contínua e dinâmica de atores envolvidos em todo o processo de gestão. Tais atores são aqueles definidos por Miksa et al (2019) como as agência de financiamento, os revisores éticos, os especialistas legais, os pesquisadores, editores, administradores do repositório, provedores de infraestrutura, o *staff* de apoio à pesquisa, os bibliotecários de dados e o administrador institucional. Esses mesmos autores definiram os 10 princípios que um PGD necessita para ser acionável por máquina, apresentados no quadro 2.

Quadro 2 - Princípios para um PGD acionável por máquina

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.

Fonte: Miksa et al (2019)

⁷ Disponível em: <<https://dmponline.dcc.ac.uk/>> Acesso em: 06 ago 2019.

⁸ Disponível em: <<https://dmptool.org/>> Acesso em: 06 ago 2019.

Esses princípios e práticas agregam maior valor ao PGD permitindo que partes sejam geradas e compartilhadas automaticamente durante a geração e atualização. Dessa forma, torna-se um instrumento valioso para a gestão dos dados de pesquisa possibilitando que as informações sejam descobertas, compartilhadas e reutilizadas por outros serviços de informação e/ou pesquisadores.

O Grupo de Trabalho “DMP Common Standard”, da *Research Data Alliance (RDA)*⁹, está trabalhando na implementação de um padrão de PGD acionável por máquina denominado *machine-actionable DMP* - (maDMP). Diversos estudos foram realizados por meio de consultas abertas, em que diferentes atores interessados foram ouvidos buscando chegar a um consenso sobre a definição, os requisitos e as informações necessários para um modelo (maDMP). O grupo sugere a ferramenta *DMPlanner*¹⁰, disponível ainda na versão inicial.

Entende-se que um melhor gerenciamento de dados é possível mediante o aprimoramento da infraestrutura dos PGDs utilizando plataformas mais apropriadas conectadas a outros sistemas de informação. Este parece ser o caminho para facilitar o reuso dos dados de pesquisa.

O ICICT/Fiocruz já percebeu a necessidade de inserir seus bibliotecários, profissionais da informação e pesquisadores nesse novo cenário que se apresenta e para isso vem estudando a literatura da área e estudando possibilidades de parcerias para o desenvolvimento do seu PGD acionável por máquina.

CONSIDERAÇÕES FINAIS

Diante das tendências que se evidenciam é possível afirmar que a gestão dos dados de pesquisa alinhada aos princípios FAIR, embora em estágio inicial, em países em desenvolvimento, traz grandes benefícios tanto para a ciência, pesquisa e inovação, quanto para a sociedade como um todo.

Além disso, garante maior transparência à utilização dos recursos públicos, assim como amplia a capacidade de reprodutibilidade dos dados em novas investigações e promove maior agilidade, qualidade e rapidez na geração de novos conhecimentos.

Nesse sentido, a adoção de um Plano de Gestão de Dados nas práticas científicas serve para atender às exigências das agências de fomento e das revistas científicas, bem como representa um instrumento para o pesquisador criar estratégias de coleta, armazenamento, validação, segurança, preservação e compartilhamento dos seus dados, durante todo o ciclo de vida da pesquisa. Por outro lado, o PGD acionável por máquina (DMPma) ainda não é uma realidade consolidada.

A Fiocruz, percebendo a importância da sua inserção nessa realidade, está envidando esforços para aderir ao crescente desafio da contemporaneidade acadêmica e científica, elevando seus dados de pesquisa a um grau de importância tal que demande de todos os pesquisadores e atores envolvidos na produção do conhecimento uma atenção especial na gestão e curadoria dos dados. Desta forma, apoia e fomenta as práticas de gestão e abertura de dados de pesquisa alinhados aos princípios FAIR, dando sua contribuição por meio desta proposta de *template* de PGD,

⁹ Disponível em: <<https://www.rd-alliance.org/>> Acesso em: 06 ago 2019.

¹⁰ Disponível em: <<http://arcticcc.github.io/dmplanner/>> Acesso em: 06 ago 2019.

assim como dos estudos que vem realizando voltados para o desenvolvimento de PGD acionável por máquina.

Artigo recebido em 09/07/2019 e aprovado em 21/11/2019.

REFERÊNCIAS

EUROPEAN COMMISSION. *Guidelines on FAIR data management in Horizon 2020, Version 3.0.* 2016. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Acesso em: 06 ago. 2019.

EUROPEAN COMMISSION EXPERT GROUP ON FAIR DATA. *Turning FAIR data into reality: final report and action plan.* 2018. Disponível em: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf. Acesso em: 06 ago. 2019.

FUNDAÇÃO OSWALDO CRUZ. *Política de acesso aberto ao conhecimento.* Portaria da Presidência. 2014. Disponível em: https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/portaria_-_politica_de_acesso_aberto_ao_conhecimento_na_fiocruz.pdf. Acesso em: 06 ago. 2019.

FUNDAÇÃO OSWALDO CRUZ. Grupo de Trabalho em Ciência Aberta. *Termo de referência: gestão e abertura de dados para pesquisa na Fiocruz.* Rio de Janeiro: FIOCRUZ/Presidência, 2018. Disponível em: <https://www.arca.fiocruz.br/handle/icict/26803>. Acesso em: 06 ago. 2019.

GUANAES, Paulo Cezar Vieira (org.). *Marcos legal nacionais em face da abertura de dados para pesquisa em saúde: dados pessoais, sensíveis ou sigilosos e propriedade intelectual.* Rio de Janeiro: Fiocruz, 2018. 123 p. Disponível em: <https://www.arca.fiocruz.br/handle/icict/28838>. Acesso em: 06 ago. 2019.

HUDSON, S. FAIR data, FAIR services and the FAIR data action plan. In: INTERNATIONAL CONFERENCE “DATA ANALYTICS AND MANAGEMENT IN DATA INTENSIVE DOMAINS” (DAMDID/ RCDL’2018), 20., 2018, Moscow, Russia. *Electronic proceedings [...]*. Disponível em: <http://ceur-ws.org/Vol-2277/abstract02.pdf>. Acesso em: 06 ago. 2019.

MIKSA, T.; CARDOSO, J.; BORBINHA, J. Framing the scope of the common data model for machine-actionable data management plans. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA. *Electronic proceedings [...]*. 2018. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8622618&tag=1>. Acesso em: 06 ago. 2019.

MIKSA, T.; SIMMS, S.; MIETCHEN, D.; JONES, S. Ten principles for machine-actionable data management plans. *PLOS-Computational Biology*, 2019. Disponível em: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006750>. Acesso em: 06 ago. 2019.

NIELSEN, M. *Reinventing discovery: the new Era of Networked Science*, 2011. Disponível em: <https://press.princeton.edu/titles/9517.html>. Acesso em: 06 ago. 2019.

SANTOS, P. X. (coord.). *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. 11 p. Disponível em: <https://www.arca.fiocruz.br/bitstream/icict/24117/2/Livro-Verde-07-06-2018.pdf>. Acesso em: 06 ago. 2019.

SCIENCE EUROPE. *Practical guide to the international alignment of research data management*, 2018. p. 40. Disponível em: https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf. Acesso em: 06 ago. 2019.

WILKINSON, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, v. 3, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 06 ago. 2019.