

OS PRINCÍPIOS *FAIR* COMO DIRETRIZES PARA MAXIMIZAR O USO E (RE)USO DOS DADOS DE PESQUISA: TRAÇANDO UM PARALELO COM A GESTÃO DE DOCUMENTOS ARQUIVÍSTICOS DIGITAIS

The FAIR principles as guidelines to maximize the use and (re)use of the research of the data: tracing a parallel with the management of digital archival documents

Bárbara Carvalho Diniz¹, Guilherme Ataíde Dias², Alzira Karla Araújo da Silva³

(1) Universidade Federal da Paraíba (UFPB), Jardim Cidade Universitária, s/n – Castelo Branco, João Pessoa – PB, 58051-900, dinizbarbara6@gmail.com

(2) Universidade Federal da Paraíba (UFPB), Jardim Cidade Universitária, s/n – Castelo Branco, João Pessoa – PB, 58051-900, guilhermeataide@gmail.com

(3) Universidade Federal da Paraíba (UFPB), Jardim Cidade Universitária, s/n – Castelo Branco João Pessoa – PB, 58051-900, alzirakarlaufpb@gmail.com

Resumo: Aborda o que são os princípios *FAIR*, seu surgimento e sua aplicação para maximizar o uso e reuso de dados de pesquisa. Buscou-se traçar um paralelo com a gestão de documentos arquivísticos digitais e como ambos podem contribuir na gestão de dados de pesquisa. É por isso que o objetivo principal deste artigo consiste em compreender como os princípios *FAIR* e a gestão de documentos arquivísticos digitais podem contribuir para tal gerenciamento. Metodologicamente, a pesquisa é de natureza aplicada, possui caráter exploratório, e seu procedimento técnico é de uma pesquisa bibliográfica. Foram então estudados os princípios e diretrizes do *FAIR*, como também da gestão de documentos arquivísticos digitais. Estabelece-se as semelhanças e os pontos em comum das ferramentas usadas para gerir dados e documentos digitais. A partir do paralelo traçado conclui-se que por mais que os objetos digitais (dados e documentos) possuam diferenças, existem semelhanças em relação a sua curadoria, procedimentos em comum em relação ao encontro, acesso e reuso de dados e documentos.

Palavras-chave: Princípios *FAIR*. Gestão de dados de pesquisa. Gestão de Documentos Arquivísticos Digitais.

Abstract: It addresses what FAIR principles are, their emergence, and their application to maximize the use and reuse of research data. It sought to draw a parallel with the management of digital archival documents and how both can contribute in the management of research data. This is why the main goal of this article is to understand how FAIR principles and the management of digital records can contribute to such management. Methodologically, the research is of an applied nature, has an exploratory character, and its technical procedure is a bibliographical research. The principles and guidelines of FAIR were studied, as well as the management of digital archival documents. It establishes the similarities and points in common of the tools used to manage data and digital documents. From the parallel traced it is concluded that although the digital objects (data and documents) have differences, there are similarities with respect to their curatorship, common procedures in relation to finding, accessing and reusing data and documents

Keywords: FAIR principles. Managing search data. Management of digital archival documents.

1 Introdução

Os princípios *FAIR* foram desenvolvidos em uma conferência sobre Ciências da Vida. Nessa conferência foram propostas e aprovadas pela comunidade científica um conjunto de diretrizes que tratassem da encontrabilidade, acessibilidade, interoperabilidade e reuso da vasta quantidade de dados produzidos pela ciência. Os princípios propostos têm como objetivo fazer com que um conjunto de dados possa ser encontrável (*findable*), acessível (*accessible*), interoperável (*interoperable*) e reusável (*reusable*) (WILKINSON et. al, 2016). Os princípios *FAIR* surgiram a partir dos fenômenos da *e-science*, da ciência

aberta e principalmente da necessidade do gerenciamento de dados de pesquisa. É primordial entender o que são dados de pesquisa para saber como tratá-los e preservá-los. Conhecer como surgem, suas fases e sua natureza é essencial para compreender os princípios *FAIR*. A OECD (2007) conceitua dados de pesquisa como registros, sendo eles: escores numéricos, registros textuais, imagens e sons que são utilizados como fontes primárias da pesquisa científica de forma a validar seus resultados. Um conjunto de dados de pesquisa é a constituição da representação do sujeito ou objeto investigado. Tal termo não se refere a cadernos de laboratório, análises

preliminares, rascunhos de artigos científicos, planos para pesquisas futuras, revisão por pares ou comunicações com colegas ou objetos físicos (por exemplo, amostras de laboratório, bactérias e teste em animais como ratos). Tais princípios e diretrizes estão voltados, principalmente, para os dados de pesquisa legíveis por computador, pois neste formato é que reside o potencial de melhorias na distribuição eficiente dos dados e sua aplicação para pesquisa.

Dataone (2018) ressalta que o ciclo de vida de dados compreende as fases do planejamento, coleta, asseguramento, descrição, preservação, descobrimento, integração e análise de dados. Isso tudo é cíclico, pois os dados podem ser úteis não só para a pesquisa que os gerou, mas para outras pesquisas com diferentes intenções do original. Para que a gestão de dados seja aplicada corretamente, é necessário termos um plano de gestão de dados. O planejamento então é essencial para a gestão, deve-se começar a partir do início da pesquisa, considerando como os dados serão gerenciados e como serão compartilhados futuramente. Ou seja, o plano de gestão de dados é um documento formal no qual se estabelece o compromisso de como tais dados serão tratados durante o desencadeamento da pesquisa, como também após a sua conclusão (SAYÃO; SALES, 2015). Em paralelo aos princípios *FAIR*, ciclo de vida de dados e o plano de gestão de dados, tem-se a Gestão Arquivística de Documentos Digitais. Por mais que o conceito do que são dados de pesquisa, seja diferente do significado de documento arquivístico digital, os conjuntos de dados e os documentos digitais possuem algumas semelhanças na forma em como devem ser gerenciados.

A preocupação sobre serem encontráveis, interoperáveis, acessíveis, reusáveis e preservados existem em ambos objetos digitais. Paes (2005, p.26) conceitua documento como “registro de uma informação, independentemente de sua natureza do suporte que a contém.” Logo complementa que o documento de arquivo é aquele que é produzido e/ou recebido por instituição pública ou privada no exercício de suas atividades, que tem como objetivo

constituir como elemento de prova ou informação. Também pode ser produzido ou recebido por indivíduos ao longo de sua existência (PAES, 2005). O Arquivo Nacional (2005, p.75) traz a seguinte definição sobre documento digital: “documento codificado em dígitos binários, acessível por meio de sistema computacional.” Um dos fatores que preocupa o gerenciamento de documentos digitais é em relação a sua preservação, visto que os meios digitais são bastante voláteis afetando questões como fragilidade do suporte, obsolescência de produtos de *software* ou programas de computador, formatos de arquivo, falta de confiabilidade e do espaço de armazenamento.

Em síntese, constata-se que apesar de diferentes, conjunto de dados e os documentos arquivísticos digitais possuem características em comum. Ambos os gestores procuram exercer cuidados em relação a sua curadoria. Este trabalho procura explorar como é feito o uso e reuso de dados de pesquisa a partir do princípio *FAIR* e como é feita a gestão de documentos arquivísticos digitais, mostrando então suas semelhanças e como estes podem contribuir para a gestão dos dados de pesquisa.

2 Objetivos

Objetiva-se compreender como os princípios *FAIR* e a gestão de documentos arquivísticos digitais podem contribuir para o gerenciamento de dados científicos. Em relação aos seus objetivos específicos tem como intenção: conhecer os princípios orientadores *FAIR*; entender as diretrizes e princípios sobre gestão de documentos arquivísticos digitais e; estabelecer as semelhanças entre os princípios *FAIR* e a gestão de documentos arquivísticos digitais.

3 Procedimentos Metodológicos

Esta pesquisa classificada como de natureza aplicada, pretende gerar conhecimentos sobre os princípios *FAIR* associados a gestão de documentos arquivísticos digitais, visando sua aplicação na prática para solucionar os problemas de gerenciamento de dados. Caracteriza-se como exploratória, pois traz à tona o problema do gerenciamento de dados de pesquisa, e levanta estudos bibliográficos para a construção de soluções que possam

colaborar com a gestão e preservação dos dados de pesquisa. Em relação aos seus procedimentos técnicos é classificada como uma pesquisa bibliográfica, porque “é desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos.” (GIL, 2002, p.44).

4 Discussões

Primeiramente abordaremos sobre o que são os princípios *FAIR* e como se desenvolve a gestão de documentos arquivísticos digitais. Trata da comparação entre o gerenciamento de dados de pesquisa e o de documentos arquivísticos digitais.

4.1 Princípios *FAIR*

No início deste trabalho foi mencionado que os princípios *FAIR* surgiram e foram desenvolvidos em uma conferência de Ciências da Vida. Os pesquisadores presentes propuseram e elaboraram quatro princípios para maximizar o uso e reuso dos dados, são eles: *Findability* (encontráveis), *Accessible* (acessível), *Interoperable* (interoperável) e *Reusable* (reusável). Wilkinson et al (2016) relata que tais princípios servem para orientação dos produtores e editores a como navegar em torno dos obstáculos dos dados. Estes princípios ajudam a maximizar o valor adquirido nas publicações acadêmicas contemporâneas em formato digital. Em resumo, todos os objetos de pesquisa digitais dos dados para a análise de conduta acadêmica, também devem aplicar os princípios *FAIR*, pois todos os componentes do processo da pesquisa deverão estar disponíveis para garantir a transparência, reprodução e reuso dos dados. A iniciativa *FORCE 11* (2018) publicou diretrizes para seguir os princípios *FAIR*. A seguir um quadro especificando o que é necessário para se cumprir estes princípios.

Quadro 1 – Diretrizes dos princípios *FAIR*

PRINCÍPIO	ESPECIFICAÇÕES
Encontrável (<i>Findable</i>)	Reencontrável em qualquer ponto no tempo; dados digitais devem ser persistentes, com ênfase em seus metadados; minimamente conter metadados acionáveis por máquina que permitam distingui-lo de

	outros objetos de dados; uso de identificador único e persistente.
Acessível (<i>accessible</i>)	Devem ser acessíveis, pois devem ser obtidos tanto por máquinas, quanto por humanos; mediante autorização apropriada; por meio de protocolo bem definido; assim máquinas e humanos serão capazes de julgar se o objeto é acessível.
Interoperáveis (<i>interoperable</i>)	Metadados devem ser acionados por máquinas; os (meta) formatos de dados devem utilizar vocabulários compartilhados e ontologias; os metadados dentro dos objetos de dados devem ser sintaticamente reproduzíveis e semanticamente acessíveis.
Reusáveis (<i>reusable</i>)	Os objetos de dados devem estar em conformidade com os princípios da encontrabilidade (encontrável) e interoperável; metadados devem ser ricos para que com o mínimo de esforço humano possam ser ligados ou integrados com outras fontes de dados; deve-se referir as suas fontes com metadados ricos para permitir a citação apropriada.

Fonte: Adaptado do *FORCE 11* (2018)

As diretrizes são claras quando enfatizam o uso de metadados e identificadores persistentes, autorizações e licenças para publicação e acessibilidade, como também o uso de vocabulários compartilhados e ontologias para o acesso em geral aos dados.

4.2 Gestão de Documentos Arquivísticos Digitais

O Conselho Nacional de Arquivos (2011) relata que a gestão de documentos arquivísticos trata de procedimentos e técnicas da produção, tramitação, uso, avaliação e arquivamento dos documentos produzidos e recebidos por entidades ou

indivíduos no decorrer de sua existência em fase corrente e intermediária. Os documentos então podem passar por até três fases, são elas: corrente, intermediária e permanente. A fase corrente se refere aos documentos usados com frequência e que são mantidos perto do setor que os produziu. Na fase intermediária os documentos são pouco utilizados e são mantidos longe de seus produtores aguardando prazos prescricionais, em que dependendo de avaliação pode ser eliminado ou recolhido para o arquivo permanente. Por fim, a fase permanente custodia os documentos que devem ser preservados por seu valor histórico, científico, cultural e informativo. Com o advento das Tecnologias da Informação, os documentos começaram a ser produzidos e armazenados em meio eletrônico. Para isso, é necessário que a gestão de documentos possa também ser aplicada em meio digital. A seguir o quadro 2 demonstra os procedimentos e ações da gestão de documentos arquivísticos digitais elaboradas pelo Conselho Nacional de Arquivos (2011).

Quadro 2 – Gestão de documentos arquivísticos digitais (procedimentos e ações)

PROCEDIMENTO	AÇÕES
<p>Captura Consiste no procedimento que declara um documento como documento arquivístico, este procedimento incorpora o documento ao sistema.</p>	Registro, classificação, indexação, atribuição de restrição de acesso e arquivamento.
<p>Avaliação Uma atividade essencial para a gestão de documentos, pois por meio da avaliação é possível racionalizar o grande número de documentos nas fases corrente e intermediária, fazendo então que só os documentos essenciais cheguem a fase permanente.</p>	Retenção dos documentos, eliminação, transferência e recolhimento.
<p>Pesquisa, localização e apresentação dos documentos O sistema deve possuir funções para recuperação e acesso aos documentos e as informações contidas</p>	A partir de parâmetros definidos que são selecionadas entre as informações coletadas no momento do registro do documento e entre os metadados associados a

neles, como forma de facilitar o dia a dia das atividades das instituições.	ele.
<p>Segurança: controle de acesso, trilhas de auditoria e cópias de segurança O sistema deve conter controles de acesso e procedimentos de segurança que possam garantir a integridade dos documentos.</p>	Controle de acesso, uso e rastreamento, trilha de auditoria, cópias de segurança e segurança da infraestrutura.
<p>Armazenamento O armazenamento deve garantir autenticidade e o acesso aos documentos pelo tempo estabelecido pela tabela de temporalidade (define o tempo de guarda dos documentos, seja por tempo determinado ou guarda permanente) e sua destinação.</p>	Fatores mais importantes para a seleção das opções de armazenamento: volume e estimativa dos documentos, segurança dos documentos, características do suporte e do ambiente, frequência de uso e custo relativo das opções de armazenamento dos documentos.
<p>Preservação Devem se manter acessíveis e utilizáveis pelo tempo necessário para sua guarda. Deve ter garantido sua longevidade, funcionalidade e seu acesso contínuo.</p>	Prevenção de obsolescência tecnológica e de danos físicos ao suporte que se encontra armazenado, técnicas como migração, <i>refreshing</i> , conversão, emulação, encapsulamento e preservação a tecnologia.

Fonte: Conselho Nacional de Arquivos (2011).

4.2 Pontos em comum

No apêndice A é demonstrado detalhadamente em um quadro quais os pontos em comum entre os princípios *FAIR* e a gestão de documentos arquivísticos digitais. Em resumo, notar que a gestão de dados de pesquisa a partir do uso dos princípios *FAIR* e a gestão de documentos digitais arquivísticos apresentam algumas diferenças em sua forma de administrar seus objetos digitais, contudo, ambas partilham de ferramentas e ações semelhantes na forma de recuperar, acessar e reusar suas informações.

5 Considerações Finais

Os princípios *FAIR* trazem como pilares serem encontráveis, acessíveis, interoperáveis e o reutilizáveis como preocupações da gestão de dados de

pesquisa. Estes princípios focam potencialmente em administração para o uso e reuso dos dados, não abordando a questão da preservação, que também é uma preocupação da gestão dos dados de pesquisa. A gestão de documentos arquivísticos digitais tem como prerrogativas não apenas a administração, mas assuntos que envolvem a política institucional e preservação dos documentos. Não é intenção fazer uma crítica aos princípios *FAIR*, pois estes foram criados para maximizar o uso e reuso de dados. A preservação aparentemente não é seu foco. Contudo, os princípios *FAIR* aliados aos princípios da gestão de documentos arquivísticos digitais demonstram qualidades essenciais na administração e preservação de dados de pesquisa. Quem tende a ganhar com isto é a ciência, pois terá acesso e preservação para o desenvolvimento presente e futuro.

Referências

- ARQUIVO NACIONAL (BRASIL). **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005. 232 p.
- CONSELHO NACIONAL DE ARQUIVOS (BRASIL). **e-ARQ Brasil: Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos/ Câmara Técnica de Documentos Eletrônicos. 1.1 versão**. Rio de Janeiro: Arquivo Nacional, 2011. 136p. Disponível em: <<http://www.siga.arquivonacional.gov.br/images/publicacoes/e-arq.pdf>>. Acesso em: 24 set. 2018.
- COMISSÃO EUROPEIA. **Guidelines on FAIR Data Management in Horizon 2020**. Versão 3.0. [S.l.: s.n], 2016. 12 p. Disponível em: <http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf>. Acesso em: 24 set. 2018.
- DATAONE. **Data Life Cycle**. 2018. Disponível em: <<https://www.dataone.org/data-life-cycle>>; Acesso em: 24 set. 2018.
- FORCE 11. **Guiding principles for Findable, Accessible, Interoperable and Reusable data publishing version b.1.0**. 2018. Disponível em: <<https://www.force11.org/fairprinciples>>. Acesso em: 24. set. 2018.
- GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 176 p.
- OECD. **OECD Principles and Guidelines for Access to Research Data from Public Funding**. Paris: OECD Publications, 2007, 24p. Disponível em: <<http://www.oecd.org/science/scitech/38500813.pdf>>. Acesso em: 24 set. 2018.
- PAES, Marilena Leite. **Arquivo: teoria e prática**. 3. ed. Rio de Janeiro: FGV Editora, 2005. 228p.
- SAYÃO, Luiz Fernando; SALES, Luana Farias. **Guia de Gestão de Dados de Pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. 90p. Disponível em: <http://carpedien.ien.gov.br/bitstream/ien/1624/1/GUIA_DE_DADOS_DE_PESQUISA.pdf>. Acesso em: 24 set. 2018.
- WILKINSON, Mark D. et al. *The FAIR Guiding Principles for scientific data management and stewardship*. **Scientific Data**. Nature. 2016. Disponível em: <<https://www.nature.com/articles/sdata201618#article-info>>. Acesso em: 24 set. 2018.

APÊNDICE A – SEMELHANÇAS ENTRE OS PRINCÍPIOS *FAIR* E A GESTÃO DE DOCUMENTOS ARQUIVÍSTICOS DIGITAIS

Princípios <i>FAIR</i>	Gestão de documentos arquivísticos digitais	Pontos em comum
Dados encontráveis	Encontrabilidade de documentos.	Pontos em comum sobre encontrabilidade. (serem encontráveis)
Encontrável a partir dos metadados, uso de identificador persistente, uso de palavras-chave, nomenclatura apropriada.	Uso de identificador, descrição informativa = metadados, recursos de indexação, uso de palavras-chave, documentos encontrados a partir dos metadados.	Uso de identificador, encontrável a partir de metadados, uso de palavras-chave.
Dados acessíveis	Acessibilidade de documentos.	Pontos em comum sobre acessibilidade.
Mediante autorização apropriada, protocolo bem definido, controle de acesso, especificar onde se depositam os dados, metadados, documentação e códigos.	Controle de acesso dos documentos, atributos de segurança, administrador do sistema determina quem pode ter acesso aos documentos.	Acesso mediante a autorização apropriada, controle de acesso.
Dados interoperáveis	Interoperabilidade de documentos.	Pontos em comum sobre interoperabilidade.
Metadados acionados por máquinas, metaformatos utilizam vocabulários controlados e ontologias, metadados dentro dos objetos de dados deve ser reproduzíveis e acessíveis, especificar dados, vocabulários, padrões e metodologias deverá seguir.	Consulta, recuperação, importação e exportação de documentos e seus metadados, operações de interoperabilidade devem respeitar a legislação vigente e a política de segurança, capaz de interoperar com outros sistemas por meio de padrões abertos.	Não foram encontrados pontos em comum neste aspecto.
Dados reusáveis	Reuso de documentos.	Pontos em comum sobre reuso.
Devem estar em conformidade com os princípios de encontrabilidade (encontráveis) e interoperabilidade (interoperáveis), bem descritos para que possam ser integrados a outros dados, referência a fonte e metadados, processo de garantia de qualidade de dados, cumprimento de tempo da validade de dados.	Os documentos arquivísticos devem possuir valores de autenticidade, confiabilidade, organicidade, acessibilidade e unicidade. No plano de classificação e destinação deve estar atribuído o tempo de guarda do mesmo para fins administrativos ou históricos.	Deve ser estabelecido o tempo de guarda e validade dos dados/documentos. Garantia de qualidade, confiabilidade da informação.

Fonte: Adaptado do Conselho Nacional de Arquivos (2011), Comissão Europeia (2016) e *Force 11* (2018)