

REPOSITÓRIOS E PLANO DE GESTÃO DE DADOS: O CASO IPALLIANCE

MAFALDA LOPES*

MIGUEL PADILHA**

FRANCISCO VIEIRA***

YULIA KARIMOVA****

Resumo: A finalidade deste trabalho é demonstrar que a criação do Plano de Gestão de Dados contribui para a organização e sistematização dos dados produzidos e geridos no âmbito de projetos, facilitando e apoiando as equipas na tomada de decisão em questões de gestão direta e operacional de um dado projeto, como, p. ex., a escolha de repositório para depósito de dados. Tomando como estudo de caso o projeto IPAlliance, metodologicamente foi utilizado modelo colaborativo em que o data steward tem um papel central, assumindo função orientadora no processo de implementação de boas práticas, apoiando a gestão de dados, incluindo o depósito de dados. Conclui-se que os Planos de Gestão de Dados podem ser muito úteis para a estruturação dos projetos, promovendo processos de investigação e desenvolvimento, bem como a reutilização e replicação de dados de investigação em outros contextos.

Palavras-chave: Ciência aberta; Educação para a Saúde; IPAlliance; Plano de Gestão de Dados; Repositórios de dados.

Abstract: The aim of this work is to demonstrate that the creation of a Data Management Plan contributes to the organisation and systematisation of data produced and managed within projects, thereby facilitating and supporting teams in decision-making, concerning the direct management and operation of a given project, such as the selection of a repository for data deposition. Using the IPAlliance project as a case study, a collaborative model was methodologically employed in which the Data Steward plays a central role, assuming a guiding function in the implementation of best practices and supporting data management, including data deposition. It is concluded that Data Management Plans can be highly beneficial for structuring projects, promoting research and development processes, as well as the reuse and replication of research data in other contexts.

Keywords: Open science; Health education; IPAlliance; Data Management Plan; Data repositories.

1. OBJETIVOS

No contexto da Ciência Aberta e requisitos existentes para divulgação dos resultados de investigação, pretende-se discutir a importância da criação de um Plano de Gestão de Dados (PGD) como suporte ao depósito, publicação e partilha de dados para um projeto cofinanciado pelo PRR – Plano de Recuperação e Resiliência e pelo mecanismo NextGenerationEU.

* Escola Superior de Enfermagem do Porto. Email: lauralopes@esenf.pt. ORCID: <https://orcid.org/0000-0001-6157-7821>.

** Escola Superior de Enfermagem do Porto/CINTESIS@RISE. Email: mpadilha@esenf.pt. ORCID: <https://orcid.org/0000-0002-6631-776X>.

*** Escola Superior de Enfermagem do Porto/IMPGroup. Email: franciscovieira@esenf.pt. ORCID: <https://orcid.org/0000-0002-0183-1987>.

**** INESC TEC. Email: ylaleo@gmail.com. ORCID: <https://orcid.org/0000-0002-1015-6709>.

Neste enquadramento o presente trabalho focou-se em três aspetos principais: (1) a Ciência Aberta, (2) os Planos de Gestão de Dados (PGD) e (3) os repositórios de dados. De facto, o conceito de Ciência Aberta reflete um contexto de acesso aberto à informação, possibilitando aos investigadores a partilha de conhecimento em ambiente de livre acesso ao público e, assim, impactando significativamente a nível social e económico, promovendo a transparência na ciência (European Commission 2016; European Commission 2017; European Commission 2021). Por outro lado, os PGD visam especificar todas as questões relacionadas com a gestão de dados de investigação (GDI) no projeto, ajudam na sua organização (Karimova 2023) e são indicados como parte dos requisitos para o cumprimento das recomendações para a Ciência Aberta e princípios FAIR (*Findability, Accessibility, Interoperability, and Reusability*) (Wilkinson et al. 2016). Por outras palavras, os PGD apoiam os investigadores durante o projeto tanto na recolha, processamento, descrição e organização de dados como no depósito, preservação, disseminação e reutilização (Michener 2015). Por fim, os repositórios de dados visam responder à necessidade de disseminação de resultados do projeto, possibilitando a preservação de longo prazo (Vitale 2017). De facto, estes repositórios são plataformas que permitem armazenar, preservar, descrever e gerir conjuntos de dados, possibilitando, deste modo, a reutilização e disseminação de resultados de investigação (White 2014).

Desta forma, considerando os três tópicos referidos, pretende-se descrever e refletir sobre o processo de criação do PGD, no caso específico de um projeto colaborativo de inovação pedagógica, ponderando sobre o seu contributo para a melhoria da organização do projeto, dos dados recolhidos no decurso do projeto e, ainda, na facilitação e apoio a investigadores na tomada de decisão quanto à escolha do repositório para depósito de dados resultantes do desenvolvimento do projeto.

2. CONTEXTO

O estudo de caso aqui reportado teve por base a elaboração de um PGD para o projeto colaborativo de inovação pedagógica IPAlliance, uma aliança multidisciplinar e interprofissional de educação em saúde que visa criar oferta formativa que potencie o *upskilling* e *reskilling* de profissionais e estudantes na área da saúde. Esta aliança foi suportada pelos Programas Impulso Jovens STEAM e Impulso Adultos, no âmbito do PRR – Plano de Recuperação e Resiliência, financiado através do mecanismo comunitário NextGenerationEU.

Criada em 2021 entre a Escola Superior de Enfermagem do Porto, a Faculdade de Medicina da Universidade do Porto e a Escola Superior de Saúde do Instituto Politécnico do Porto, esta Aliança tem por foco a necessidade de oferta formativa interprofissional e de dotar o processo de aprendizagem de condições tecnológicas optimizadoras da aquisição de novas competências, considerando a centralidade da simulação na formação em saúde.

Esta parceria visa, ainda, promover a requalificação de competências clínicas dos profissionais de saúde através da formação em ambiente simulado. Mais especificamente,

promove a criação de um Centro de Educação em Saúde, responsável por criar as condições para a resposta cabal e eficaz ao que a Organização Mundial da Saúde (OMS) definiu como *Slow Motion Disaster*: doenças não transmissíveis, também conhecidas como doenças crónicas ou não comunicáveis, que, por sua vez, representam atualmente cerca de 86% das mortes na Europa, tendo um impacto substancial na deterioração da qualidade de vida dos cidadãos (Padilha et al. 2021).

O desenvolvimento dos centros de simulação fará com que o Porto Innovation District, que representa uma das maiores concentrações de talento, conhecimento e inovação na Península Ibérica, se torne uma referência na educação para a saúde na Europa.

O Porto Innovation District é uma área de um quilómetro quadrado, tradicionalmente designada por Polo Universitário da Asprela, localizada no Porto, Portugal, que compreende mais de 16 Instituições de Ensino Superior, 26 Unidades de I&D e Laboratórios de Investigação, 3 Centros Hospitalares e mais de 70 *startups* a serem incubadas num Parque de Ciência e Tecnologia. Em pouco mais de um quilómetro quadrado, trabalham mais de 14 000 pessoas, a maioria com formação superior e altamente especializada (Pereira 2020). O Innovation District acolhe, ainda, cerca de 38 000 estudantes em mais de 500 cursos superiores e lança no mercado mais de 10 000 licenciados por ano, o que corresponde a cerca de um terço do total de licenciados da região norte de Portugal.

2.1. Ciência Aberta

A Ciência Aberta permite a partilha do conhecimento entre a comunidade científica, a sociedade e as empresas, possibilitando ampliar o reconhecimento e o impacto social e económico da ciência. Mais do que a disponibilização em acesso aberto de dados e publicações, a Ciência Aberta representa a abertura do processo científico enquanto um todo, reforçando o conceito de responsabilidade social na ciência (Brunner e Ascher 1992; Crüwell et al. 2019).

A implementação de práticas de Ciência Aberta é, também, geradora de múltiplas oportunidades de inovação e permite impulsionar o desenvolvimento de novos produtos, serviços, negócios e empresas (Ministério da Ciência, Tecnologia e Ensino Superior 2016; Sveinsdottir, Davidson e Proudman 2021). A gestão de dados desempenha um papel fundamental na implementação da Ciência Aberta, sendo o alicerce sobre o qual repousam os princípios da transparência, replicação e acesso à ciência. A nível conceptual, «dados» são definidos no Modelo de Ciclo de Vida da Curadoria do Centro de Curadoria Digital (DCC)¹ como «qualquer informação em formato digital binário» (Higgins 2008), sendo tratados no sentido de qualquer informação digital que possa ser considerada numa perspetiva ampla.

¹ Disponível em: <https://www.dcc.ac.uk>.

As tipologias de dados de investigação são diversas, podendo ser experimentais, observacionais, operacionais, dados de terceiros, do sector público, dados de monitorização, dados processados ou dados reaproveitados. Harvey (2010) descreve que esta amplitude dos dados se refere a tudo o que é digital, baseando-se nas Diretrizes da UNESCO para a Preservação do Património Digital (UNESCO 2022; UNESCO 2020).

Em Portugal, está em curso a preparação de uma Política Nacional para a Ciência Aberta, trabalho este iniciado pelo Ministério da Ciência, Tecnologia e Ensino Superior². Neste particular, o sítio *Web* criado para a Ciência Aberta³ em Portugal descreve os pilares da política como sendo: (1) transparência nas práticas, metodologia, observação e recolha de dados, (2) disponibilidade pública e reutilização de dados científicos, (3) acesso público e transparência na comunicação científica, (4) utilização de ferramentas baseadas na *Web* para facilitar a colaboração científica. Na prática, trata-se de um apelo geral e aspiracional, recomendado, para que os investigadores partilhem os seus dados (OECD 2007). A política sugere que se elabore um plano de gestão de dados, propondo um modelo básico, i.e., tabela de conteúdos, e que se sigam as melhores práticas para qualquer disciplina científica em que a investigação seja efetuada (Sveinsdottir, Davidson e Proudman 2021).

O objetivo da Ciência Aberta é, assim, garantir a disponibilidade e acesso a publicações académicas, dos dados resultantes da investigação académica e da metodologia, incluindo códigos ou algoritmos que foram utilizados para gerar dados (National Academies... 2018). Estes dados da investigação devem, por sua vez, ser colocados em repositórios fiáveis e descritos de forma que facilitem a sua compreensão e reutilização. De facto, a Ciência Aberta contribui para o crescimento social e económico ao promover a verificabilidade, a reutilização e a transparência dos resultados (Vitale 2017). Além disso, os 17 Objetivos de Desenvolvimento Sustentável (ODS)⁴ adotados pelos Estados-Membros das Nações Unidas, em 2015, tendo por finalidade proteger o planeta, acabar com a pobreza e assegurar a paz e a prosperidade até 2030, exigem uma ampla colaboração no seio da comunidade mundial de dados e da comunidade científica em geral, bem como o desenvolvimento de infraestruturas de informação. A realização dos ODS está em consonância com as melhores práticas, recomendações e atividades de Gestão de Dados de Investigação da Research Data Alliance (RDA)⁵, uma das iniciativas de GDI para o paradigma da Ciência Aberta (Research Data Alliance 2019).

No panorama atual, para além das exigências de cumprimento dos ODS e respetivo mapeamento nas atividades do dia a dia das instituições, a grande diversidade de dados que caracterizam os projetos de investigação, levantando muitas questões relacionadas

² Disponível em: <https://www.portugal.gov.pt/pt/gc23/area-de-governo/ciencia-tecnologia-e-ensino-superior>.

³ Disponível em: <https://www.ciencia-aberta.pt>.

⁴ Disponível em: <https://ods.pt>.

⁵ Disponível em: <https://www.rd-alliance.org>.

com o acesso, a descrição, a reutilização, a gestão e a reutilização (Vines et al. 2014), torna ainda mais premente a criação de PGD individualizados. Na realidade, este poderá ser fator crítico de sucesso para uma boa gestão dos dados, aspetos cruciais para a qualidade dos dados de investigação (Eynden et al. 2011), uma vez que os dados devem ser descritos com o maior detalhe possível, a fim de serem localizáveis, acessíveis, interoperáveis e reutilizáveis (princípios FAIR) (Wilkinson et al. 2016).

2.2. Gestão de Dados de Investigação e Planos de Gestão de Dados

Segundo a literatura (Milotic 2018; Karimova 2023), a criação dos PGD permite apoiar os investigadores a refletir sobre um conjunto de questões relacionadas com GDI, tanto no depósito e partilha de dados como na conformidade de dados e projetos com princípios FAIR (Wilkinson et al. 2016) e com os requisitos de GDI no geral. Um repositório de dados de investigação é uma das principais ferramentas que permite aos investigadores o depósito, publicação e partilha de dados. Neste enquadramento, os investigadores são incentivados a aplicar os princípios FAIR para garantir o acesso, interoperabilidade, reutilização e verificabilidade dos resultados bem como a transparência da investigação, aumentando, simultaneamente, o seu valor (Nature 2018).

Assim, a GDI torna-se uma componente vital da investigação de alta qualidade e uma atividade regular para os investigadores, sendo os PGD considerados a forma mais eficaz que o investigador dispõe para comunicar as suas intenções em matéria de recolha, organização, armazenamento, utilização, manutenção e disponibilização dos dados resultantes de um projeto (Vitale e Sandy 2019). Os PGD são definidos como documentos que fornecem aos investigadores um mecanismo para declarar como irão gerir os dados associados ao ciclo de vida dos dados de um projeto (Smale et al. 2016) e após a sua execução. Logo, quanto mais cedo o PGD for redigido, mais fácil será criar a estratégia de gestão de dados e escolher o repositório adequado às necessidades dos investigadores e ao domínio de investigação a que o projeto se refere, permitindo disponibilizar os resultados de investigação e torná-los reutilizáveis e valiosos.

Os resultados do trabalho de Karimova (2023) demonstram, ainda, que em alguns casos, os investigadores só começam a refletir sobre repositórios e partilha de dados quando começam a criar o PGD. Por outras palavras, os investigadores não têm conhecimento, *a priori*, de quais os repositórios existentes, não tendo, assim, a necessidade de gestão de dados identificada. Desta forma, é enfatizada a necessidade e importância na criação do PGD com vista à gestão de dados como parte da definição da estratégia do projeto (Karimova 2023).

Como já referido, um dos requisitos da GDI é que os dados cumpram os princípios FAIR. Este requisito não é sinónimo de que os dados têm de estar abertos, promovendo, antes, a possibilidade de facilitar a pesquisa, disseminação e reutilização de dados, contribuindo para acelerar os avanços científicos e cumprindo o desígnio de que «os dados

devem estar tão abertos quanto possível e tão fechados quanto necessário» (European Commission 2017). Os princípios FAIR coincidem, pois, com o ciclo de vida dos dados de investigação e também promovem a validade, a fácil compreensão e usabilidade a longo prazo (Corti et al. 2019). Deste modo, a verificação desses princípios vai beneficiar a abertura dos dados, tornando-os facilmente encontrados, acessíveis, interoperáveis e reutilizáveis, com destaque para esta última qualidade (Wilkinson et al. 2016).

Neste contexto, estes princípios podem ser vistos como diretrizes para investigadores, editores e organizações, ajudando a aumentar a reutilização dos dados científicos. Como afirma Corti et al. (2019), estes dados devem ser FAIR não só para os seres humanos, mas também para as máquinas, permitindo a recuperação, o acesso automatizado e a consequente reutilização (Corti et al. 2019).

Assim, o fornecimento de metadados de qualidade, a seleção do repositório e a publicação dos dados são algumas das atividades que os investigadores devem realizar quando seguem as políticas da Ciência Aberta. No entanto, estas atividades consomem muito tempo e não são óbvias para os investigadores, exigindo conhecimento, esforço e experiência específicos. Para apoiar os investigadores, as instituições procuram soluções para desenvolver fluxos de trabalho, ferramentas, sistemas e serviços que visam satisfazer os requisitos GDI tanto dos investigadores como das agências de financiamento (Karimova 2023; Karimova et al. 2017).

Neste contexto, no âmbito do projeto IPAlliance, a colaboração entre um especialista de suporte de GDI (i.e., *data steward*) e os investigadores surgiu como uma oportunidade e permitiu dar início a uma nova tarefa: a criação do PGD, promovendo o conhecimento sobre a GDI e ajudando na organização do projeto em geral. De facto, «todos precisam de um plano de gestão de dados» (Nature 2018). O PGD não é apenas um documento interno de gestão, mas é visto, também, como um instrumento importante para a divulgação e disseminação de resultados e apoio na tomada de decisão no que respeita ao depósito de dados, sendo, por isso, considerado um documento vivo e útil para estruturar o curso das atividades do projeto, integrando-se com outros sistemas e fluxos de trabalho (Simms e Jones 2017) e abrangendo toda a implementação do projeto (Cox e Verbaan 2018).

3. MÉTODO

Cada vez mais os financiadores exigem que qualquer proposta de projeto inclua um documento com uma explicação detalhada da forma como os dados de investigação serão geridos durante o projeto, incluindo aspetos como: (1) a quantidade e que tipo de dados serão recolhidos, (2) como vão ser geridos durante e após o fim do projeto, (3) como será protegida a identidade dos participantes caso sejam recolhidos dados sensíveis e pessoais e (4) que dados serão partilhadas no final do projeto. Este documento é designado por

PGD, embora os diferentes financiadores lhe atribuam outras nomenclaturas como por exemplo DMP (Data Management Plan) (Cox e Verbaan 2018).

De facto, o estudo de caso que se apresenta neste trabalho refere-se ao projeto IPAlliance, que não contemplava como requisito obrigatório a criação do PGD. Encarou-se como uma oportunidade a colaboração com um *data steward*. Esta colaboração teve como objetivos, por um lado, aplicar algumas das recomendações da Ciência Aberta para uma sistematização da informação relativamente aos dados do projeto e, por outro lado, informar e alertar a equipa de investigação para a necessidade de gerir os dados.

Pretendeu-se, ao mesmo tempo, demonstrar que criação do PGD contribui para uma melhor organização do projeto, apoiando os investigadores na tomada de decisão nas questões relacionadas com a escolha do repositório para publicação dos dados resultantes do desenvolvimento do projeto.

Um primeiro passo consistiu em sensibilizar os investigadores para as diferentes componentes da Ciência Aberta, de modo a promover os seus benefícios e orientar os investigadores para as ferramentas e metodologias existentes (Medeiros et al. 2020). Para o estudo de caso aplicou-se o método colaborativo de criação de PGD, que prevê 9 fases de trabalho⁶: (1) entrevistas com os investigadores, (2) avaliação do tipo de dados, (3) análise de exemplo de PGD para o domínio de investigação, (4) análise do *Data Protection Impact Assessment*, caso necessário, (5) revisão das publicações do projeto, (6) análise dos dados e dos requisitos de descrição, (7) análise para o domínio das práticas de preservação, partilha e reutilização dos dados recolhidos, (8) desenvolvimento de uma versão *draft* do PGD, validação e melhoria do PGD e, finalmente, (9) aprovação da primeira versão do PGD com investigadores para publicação⁷.

O processo foi iniciado com a explicação da finalidade do PGD para a GDI em reunião com a equipa do projeto onde foi discutido e apresentado um modelo de PGD, realçando a importância da sua existência no contexto da execução dos projetos e reforçando a importância do foco nos dados recolhidos e na necessidade de descrever e documentar esses dados.

Considerando a importância e especificidade da GDI, todo o processo foi liderado por um *data steward*, técnico especializado responsável pela GDI, garantindo que os dados do projeto se tornassem acessíveis, utilizáveis, seguros e confiáveis. Na realidade, o *data steward* tem funções de supervisão dos dados do projeto, sendo responsável por garantir a qualidade e a adequação dos dados à finalidade do trabalho em curso, incluindo os metadados e qualidade dos mesmos para a descrição desses dados. Acumula, ainda, a responsabilidade partilhada de sensibilização, de garantia da acessibilidade, de promoção da utilização adequada e segura dos dados, bem como de divulgação (Cramer 2019),

⁶ Extended collaborative DMP-building method pode ser consultado em Karimova 2023.

⁷ Proposto por Karimova et al. 2017; Karimova, Ribeiro e David 2020; Karimova, Ribeiro e David 2021; Karimova 2023.

contribuindo, ainda, para criar processos e procedimentos de monitorização da utilização aos dados (Teperek et al. 2018).

Ao nível de criação do PGD, o *data steward* organizou o processo de trabalho em três fases. Numa primeira fase do trabalho, de início de implementação do PGD, foram realizadas duas entrevistas para identificar o âmbito do projeto, o tipo de dados que se previa serem recolhidos e solicitar o acesso à documentação de candidatura a financiamento do projeto.

Esta primeira fase incluiu, de acordo com a metodologia adotada, (1) a realização de entrevistas com os investigadores, (2) a avaliação do tipo de dados, (3) a análise de exemplo de PGD para o domínio de investigação, (4) a análise do *Data Protection Impact Assessment*.

Numa segunda fase, de desenvolvimento e redação do PGD, foi feita (5) a revisão das publicações do projeto, (6) a análise dos dados e dos requisitos de descrição.

Na terceira e última fase, de aplicação do método de trabalho procedeu-se, (7) à análise para o domínio das práticas de preservação, partilha e reutilização dos dados recolhidos, (8) ao desenvolvimento de uma versão *draft* do PGD, validação e melhoria do PGD e, finalmente, (9) à aprovação da primeira versão do PGD com investigadores para publicação.

Essas atividades promoveram, por um lado, o apoio à tomada de decisão para a escolha do repositório para depósito e publicação dos dados, ao mesmo tempo que permitiram sistematizar um fluxo de trabalho com o objetivo de descrever os dados recolhidos, tratados e divulgados durante e após a execução do projeto. Por outro lado, possibilitou o envolvimento da equipa do projeto nas questões de GDI, serviu como formação para futuros projetos e ainda culminou na publicação do PGD, e consequentemente no depósito dos dados.

4. RESULTADOS

A colaboração com os investigadores do projeto IPAlliance resultou na criação e publicação da primeira versão do Plano de Gestão de Dados (Lopes et al. 2023) e no suporte à escolha do repositório de dados de investigação mais apropriado; foi escolhido o Zenodo⁸.

Embora a colaboração tenha corrido como planeado, e com contínuo interesse dos investigadores, durante o processo de trabalho a equipa foi confrontada com alguns desafios, assim como: (1) a inexistência de infraestruturas nacionais para o depósito dos dados e publicação do PGD, (2) o imperativo de sensibilização da equipa do IPAlliance para a implementação de novos fluxos de trabalho, para além dos previstos para a execução do projeto, e (3) um aumento significativo do volume de trabalho. Tornou complexo o trabalho, ainda, a relativa novidade do tema ao nível institucional,

⁸ Disponível em: <https://zenodo.org>.

que leva à inexistência de práticas de gestão de dados claramente definidas, o que complexificou o desenvolvimento deste processo.

Finalmente, como foi mencionado acima, a falta de infraestruturas para publicação do PGD e depósito de dados também dificultou este processo final de escolha de repositório de dados. Verificou-se, de facto, que não existe nenhum repositório de dados nacional que permita a publicação de dados de investigação do domínio de saúde. Ainda assim, este processo permitiu fomentar a reflexão, planeamento e execução de práticas de gestão de dados a nível institucional e ao nível dos projetos específicos.

Como não existe a nível nacional um repositório que responda às nossas necessidades, foi escolhido o Zenodo. Este é um repositório generalista que permite aos investigadores, cientistas, projetos e instituições partilhar, preservar e apresentar resultados de investigação multidisciplinar (e.g. dados, *software* e publicações) caso não tenham os repositórios institucionais ou temáticos existentes nas comunidades de investigação. Como modelo de PGD, optou-se pelo modelo do DMPonline⁹, com as adaptações necessárias. Este *template* apresenta um modelo estruturado de questões que devem ser consideradas na criação de plano, e a que os investigadores devem responder de forma objetiva.

Durante a criação do plano, verificou-se que, no projeto IPAlliance, os dados serão recolhidos e processados ao longo do desenvolvimento do projeto. Posteriormente serão organizados em dados RAW, dados processados e dados públicos. Os dados RAW dividem-se em dois tipos de acesso — fechado e aberto. Os de acesso fechado incluem informações sensíveis, privadas, confidenciais e pessoais, e não podem ficar em acesso público. Os de acesso aberto são acessíveis a partir da Plataforma NAU, onde serão disponibilizados os Cursos MOOC¹⁰. Haverá, também, um tipo de dados que serão recolhidos a partir de dados processados pela plataforma da NAU, referentes à avaliação dos formandos que frequentam os cursos. Estes dados serão extraídos para ficheiros Excel e serão devidamente anonimizados.

Em suma, a primeira versão do PGD do IPAlliance foi publicada no Zenodo (Lopes et al. 2023). A criação do PGD permitiu o apoio e organização dos dados, motivar e envolver a equipa no planeamento e execução do projeto, identificar regras europeias e nacionais relativas às questões de proteção, armazenamento e acesso a dados pessoais, nomeadamente no que se refere aos requisitos da Ciência Aberta. De destacar, ainda, que este processo contribuiu:

- Para a tomada de decisão e esclarecimento quanto ao imperativo de criar uma política de gestão de dados na Escola Superior de Enfermagem;
- Para a criação de serviços de suporte à investigação para o cumprimento de requisitos de gestão de dados;

⁹ Disponível em: <https://dmponline.dcc.ac.uk>.

¹⁰ Disponível em: <https://www.nau.edu.pt/pt/cursos/?limit=21&offset=0&organizations=149>.

- Para a identificação de necessidades institucionais de adesão a um repositório nacional adequado para dados de investigação no domínio da saúde.

Este processo potenciou um trabalho multidisciplinar colaborativo, gerido pelo *data steward*, tendo este um papel significativo no apoio ao projeto e aos investigadores. Permitiu ainda orientar, alertar e promover o foco nos dados, contextualizando a temática relacionada com GDI e os princípios FAIR. No geral o uso de metodologia proposta por Karimova (2023) ajudou a organizar o projeto, a identificar e descrever os dados, a estruturar as tarefas a desenvolver ao longo do projeto, sendo assim um contributo importante que facilitou a realização do nosso objetivo de criação e publicação do PGD.

Por fim, a colaboração entre *data steward* e investigadores demonstrou a importância da criação de PGD na fase inicial do projeto e sua periódica atualização, requerendo constante monitorização e melhoria, de acordo com o desenvolvimento do projeto e acompanhamento até final do projeto.

CONCLUSÕES

No âmbito da Ciência Aberta, torna-se evidente a importância da Gestão de Dados de Investigação como um alicerce para o avanço do conhecimento. A evolução da ciência exige uma abordagem prática e eficaz na capacitação dos investigadores, promovendo a criação e implementação de Planos de Gestão de Dados que respondam às necessidades específicas de cada projeto. A formação prática informal em GDI é essencial para fornecer aos investigadores competências necessárias para gerir eficazmente os dados gerados no decorrer de um projeto.

Os resultados demonstram que, mesmo que o PGD não seja um requisito obrigatório, ele desempenha um papel crucial na definição da estratégia de gestão de dados. Destaca-se, ainda, que a gestão de dados deve começar na fase inicial do projeto, porque o PGD ajuda a organizar melhor todas as questões relacionadas com a gestão de dados no projeto, e, também, ajuda a prevenir e evitar as situações indesejáveis relacionadas com, por exemplo, recolha de dados sensíveis. O PGD é um valioso instrumento de apoio à tomada de decisão na divulgação de dados, garantindo a usabilidade e reutilização dos resultados produzidos.

Realce-se, ainda, as vantagens da escolha de um repositório de dados generalista que permite o armazenamento dos dados, a criação de uma comunidade para agregar outras publicações do IPAlliance, bem como novas versões do PGD, caso se verifique a necessidade. A comunidade no Zenodo¹¹ permite beneficiar, também, os investigadores, oferecendo-lhes a possibilidade de organizar, estruturar, publicar e promover os resultados de investigação de acordo com as recomendações da Ciência Aberta, promovendo

¹¹ Disponível em: <https://zenodo.org/communities/esep?q=&l=list&p=1&s=10&sort=newest>.

a colaboração e a partilha de conhecimento dentro da comunidade científica, fortalecendo o compromisso com os requisitos da Ciência Aberta e dos princípios FAIR.

Em resumo, a formação informal em GDI, que resultou na criação de um PGD, é uma tática *bottom-up* relevante para iniciar o processo de implementação de uma política de gestão de dados em projetos e, também, em instituições. Para a promoção desse desígnio a figura do *data steward* é crucial, assumindo o papel de orientador no processo de implementação de boas práticas, apoiando a gestão de dados, incluindo o depósito de dados. Além disso, esta figura promove a importância da GDI, demonstrando que os PGD podem ser úteis para a estruturação de projetos.

De facto, este estudo de caso medeia entre os «dados» (a essência da Ciência Aberta), a política de gestão de dados (estruturação que sustenta o PGD) e o Plano de Gestão de Dados (o instrumento intermediário), que organiza e prepara os dados para a partilha com a comunidade. É, pois, neste lento e periclitante balancear entre as dimensões micro (dados) e macro (conhecimento) que reside a importância da promoção da ciência aberta e implementação de uma política de gestão de dados na investigação.

REFERÊNCIAS

- BRUNNER, R. D., e W. ASCHER, 1992. Science and social responsibility. *Policy Sciences* [Em linha]. (25), 295-331 [consult. 2023-09-26]. Disponível em: <https://doi.org/10.1007/BF00138787>.
- CORTI, L., et al., 2019. *Managing and sharing research data: a guide to good practice*. [S.l.]: SAGE.
- COX A, e E. VERBAAN, 2018. Data Management Planning. Em: *Exploring Research Data Management*. Facet, 115-124.
- CRAMER, J. J., 2019. 6 Key Responsibilities of the Invaluable Data Steward [Em linha] [consult. 2023-09-26]. Dun & Bradstreet. Disponível em: <https://www.dnb.com/perspectives/master-data/6-key-responsibilities-of-data-stewards.html>.
- CRÜWELL, S., et al., 2019. Seven Easy Steps to Open Science. *Zeitschrift für Psychologie* [Em linha]. 227(4), 237-248 [consult. 2023-09-26]. ISSN 2151-2604. Disponível em: <https://doi.org/10.1027/2151-2604/a000387>.
- EUROPEAN COMMISSION, 2021. *Open research data and data management plans*. [S.l.]: European Commission.
- EUROPEAN COMMISSION, 2017. *Guidelines on the implementation of Open Access to scientific publications and research data in projects supported by the European Research Council under Horizon 2020*. [S.l.]: European Commission.
- EUROPEAN COMMISSION, 2016. *H2020 programme: guidelines to the rules on open access to scientific publications and open access to research data in horizon 2020*. [S.l.]: European Commission.
- EYNDEN, V., et al., 2011. *Managing and sharing data: best practice guide for researchers* [Em linha]. 3.ª ed. [consult. 2023-09-26]. Disponível em: <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>.
- HARVEY, R., 2010. Defining Data [Chapter 4]. Em: G. OLIVER, e R. HARVEY. *Digital Curation: A How-To-Do-It Manual*. Chicago: Neal-Schuman Publishers, pp. 45-52. No. 025.06.
- HIGGINS, S., 2008. The DCC Curation Lifecycle Model. *International Journal of Digital Curation* [Em linha]. 3(1), 134-140 [consult. 2023-09-26]. ISSN 1746-8256. Disponível em: <https://doi.org/10.2218/ijdc.v3i1.48>.

- KARIMOVA, Y., 2023. *Research data description in multiple domains: supporting researchers with data management plans*. Tese de doutoramento, Faculdade de Engenharia da Universidade do Porto.
- KARIMOVA, Y., C. RIBEIRO, e G. DAVID, 2021. Institutional support for data management plans: case studies for a systematic approach. *International Journal of Metadata, Semantics and Ontologies* [Em linha]. **15**(3), 178-191 [consult. 2023-09-26]. Disponível em: <https://doi.org/10.1504/ijmso.2021.123041>.
- KARIMOVA, Y., C. RIBEIRO, e G. DAVID, 2020. Institutional support for data management plans: five case studies. Em: Emmanouel GAROUFALLOU, e Maria-Antonia OVALLE-PERANDONES, eds. *Metadata and Semantic Research (MTSR 2020)* [Em linha]. Springer: Cham, pp. 308-319 [consult. 2023-09-26]. Disponível em: https://doi.org/10.1007/978-3-030-71903-6_29.
- KARIMOVA, Y., et al., 2017. Description + annotation: semantic data publication workflow with Dendro and B2NOTE. *International Journal of Metadata, Semantics and Ontologies* [Em linha]. **12**(4), 182-194 [consult. 2023-09-26]. ISSN 1744-263X. Disponível em: <https://doi.org/10.1504/ijmso.2017.093645>.
- LOPES, M., et al. 2023. IPAlliance – Integrated Platform for Lifelong Learning and Training of Healthcare Professionals: Data Management Plan. Em: *Zenodo* [Em linha] [consult. 2023-09-26]. Disponível em: <https://zenodo.org/records/8135761>.
- MEDEIROS, C. B., et al., 2020. *IAP input into the UNESCO Open Science Recommendation*. InterAcademy Partnership.
- MICHENER, W. K., 2015. Ten simple rules for creating a good data management plan. *PLOS Computational Biology* [Em linha]. **11**(10), e1004525 [consult. 2023-09-26]. ISSN 1553-7358. Disponível em: <https://doi.org/10.1371/journal.pcbi.1004525>.
- MILOTIC, T., 2018. Good enough practices in data management and how to translate these in data management plans (DMPs). Em: *Zenodo* [Em linha] [consult. 2023-09-26]. Disponível em: <https://doi.org/10.5281/zenodo.1421739>.
- NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE, POLICY AND GLOBAL AFFAIRS, BOARD ON RESEARCH DATA AND INFORMATION, e COMMITTEE ON TOWARD AN OPEN SCIENCE ENTERPRISE, 2018. *Open science by design: realizing a vision for 21st century research*. Washington, DC: The National Academies Press. ISBN 9780309476249.
- NATURE, 2018. Everyone needs a data-management plan. *Nature* [Em linha]. 2018-03-13. **555**(7696), 286 [consult. 2023-09-26]. Disponível em: <https://doi.org/10.1038/d41586-018-03065-z>.
- OECD, 2007. *OECD principles and guidelines for access to research data from public funding* [Em linha]. Paris: OECD Publishing [consult. 2023-09-26]. ISBN 9789264034020. Disponível em: <https://doi.org/10.1787/9789264034020-en-fr>.
- PADILHA, J. M., et al., 2021. Easiness, usefulness and intention to use a MOOC in nursing. *Nurse Education Today* [Em linha]. **97**, 104705 [consult. 2023-09-26]. ISSN 0260-6917. Disponível em: <https://doi.org/10.1016/j.nedt.2020.104705>.
- PEREIRA, I., 2020. *City branding e inovação: O Papel do branding local no desenvolvimento do ecossistema de inovação da Asprela (cidade do Porto)* [Em linha]. Dissertação de mestrado, FEP [consult. 2023-09-26]. Disponível em: <https://repositorio-aberto.up.pt/handle/10216/130962>.
- PORTUGAL. Ministério da Ciência, Tecnologia e Ensino Superior, 2016. *Ciência Aberta | Sobre Ciência Aberta* [Em linha] [consult. 2024-03-20]. Lisboa: MCTES. Disponível em: <https://www.ciencia-aberta.pt/sobre-ciencia-aberta>.
- RESEARCH DATA ALLIANCE, 2019. RDA for the Sustainable Development Goals. Introduction: Fit with the overall RDA vision and mission. *Research Data Alliance* [Em linha.] [consult. 2023-09-26]. Disponível em: <https://www.rd-alliance.org/groups/rda-sustainable-development-goals>.
- SIMMS, S. R., e S. JONES, 2017. Next-Generation data management plans: Global, machine-actionable, FAIR. *International Journal of Digital Curation* [Em linha]. **12**(1), 36-45 [consult. 2023-09-26]. ISSN 1746-8256. Disponível em: <https://doi.org/10.2218/ijdc.v12i1.513>.

- SMALE, N., et al., 2016. The History, advocacy and efficacy of data management plans. *BioRxiv* [Em linha], n.º 443499 [consult. 2023-09-26]. Disponível em: <https://doi.org/10.1101/443499>.
- SVEINSDOTTIR, T., J. DAVIDSON, e V. PROUDMAN, 2021. An Analysis of open science policies in Europe, v7 (version 7). Em: *Zenodo* [Em linha] [consult. 2023-09-26]. Disponível em: <https://doi.org/10.5281/zenodo.4725817>.
- TEPEREK, M., et al., 2018. Data stewardship addressing disciplinary data management needs. *International Journal of Digital Curation* [Em linha]. **13**(1), 141-149 [consult. 2023-09-26]. ISSN 1746-8256. Disponível em: <https://doi.org/10.2218/ijdc.v13i1.604>.
- UNESCO, 2022. *An Introduction to the UNESCO recommendation on open science* [Em linha] [consult. 2023-09-26]. Disponível em: <https://doi.org/10.54677/xoir1696>.
- UNESCO, 2020. *Open Science Recommendation* [Em linha] [consult. 2023-09-26]. Disponível em: <https://www.unesco.org/en/open-science>.
- VINES, T. H., et al., 2014. The Availability of Research Data Declines Rapidly with Article Age. *Current Biology* [Em linha]. **24**(1), 94-97 [consult. 2023-10-27]. ISSN 0960-9822. Disponível em: <https://doi.org/10.1016/j.cub.2013.11.014>.
- VITALE, C. H., 2017. The Current State of Meta-Repositories for Data [Chapter 11]. Em: L. R. JOHNSTON, ed. *Curating Research Data*. Chicago, Illinois: Association of College and Research Libraries, pp. 251-261. Volume One: *Practical Strategies for Your Digital Repository*.
- VITALE, C. H., e H. M. SANDY, 2019. Data management plans: A review. *DESIDOC Journal of Library & Information Technology* [Em linha]. **39**(6), 322-328 [consult. 2023-09-26]. ISSN 0976-4658. Disponível em: <https://doi.org/10.14429/djlit.39.06.15086>.
- WHITE, H. C., 2014. Descriptive metadata for scientific data repositories: a comparison of information scientist and scientist organizing behaviors. *Journal of Library Metadata* [Em linha]. **14**(1), 24-51 [consult. 2023-10-26]. ISSN 1937-5034. Disponível em: <https://doi.org/10.1080/19386389.2014.891896>.
- WILKINSON, M. D., et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* [Em linha]. **3**(1) [consult. 2023-09-26]. ISSN 2052-4463. Disponível em: <https://doi.org/10.1038/sdata.2016.18>.