



Implementação de serviços institucionais de administração de dados

Implementation of Institutional Data Stewardship Services

Paloma Marín-Arraiza*

Silvana Aparecida Borsetti Gregorio Vidotti**

RESUMO

As tarefas de gestão de dados de pesquisa ao longo do processo de pesquisa têm se tornado muito importantes devido à alta produção de dados e à exigência da sua preservação. Tanto bibliotecas quanto seções de apoio à pesquisa de diversas instituições de ensino e pesquisa têm começado a implementar serviços para a gestão de dados e a profissionalização desta gestão. Com um caráter qualitativo, e após um levantamento bibliográfico em bases de dados abertas, contextualiza-se a gestão de dados de pesquisa, analisam-se os perfis profissionais e determinam-se três fases para a implementação institucional destes serviços: elaboração de uma política, estabelecimento de uma unidade de informação e integração de profissionais da gestão de dados.

Palavras-chave: Administração de Dados; Dados de Pesquisa; Gestão de Dados de Pesquisa; Política de Dados; Serviços Institucionais.

ABSTRACT

The tasks of managing research data throughout the research process have become very important due to the high production of data and the requirement for its preservation. Both libraries and research support sections of various research institutions have started to implement services for data management and the professionalization of this management. With a qualitative character, and after a bibliographic search in open databases, research data management is contextualized, professional profiles are analyzed, and three phases are determined for the institutional implementation of these services: the elaboration of a policy, the establishment of an information unit and the integration of data management professionals.

Keywords: Data Stewardship; Research Data; Research Data Management; Data Policy; Institutional Services.

INTRODUÇÃO

Nos últimos anos, a prática científica tem se tornado, em parte, uma prática orientada por dados, conhecida como o Quarto Paradigma da Ciência, em que os dados e o trabalho com eles ganham uma grande relevância, desde a fase de concepção da

* Mestre em Informação e Comunicação Científica pela Universidade de Granada (Espanha). Information manager para identificadores persistentes na biblioteca da Universidade Técnica de Viena (TU Wien). Endereço: Resselgasse, 4, 1040, Viena (Áustria). Telefone: + 43 (1) 58801 4404. E-mail: paloma.arraiza@tuwien.ac.at e paloma.arraiza@unesp.br

** Doutora em Educação pela Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Professora do Programa de Pós-Graduação em Ciência da Informação da UNESP. Endereço: Av. Hygínio Muzzi Filho, 737, Campus universitário, 17525 900, Marília, SP. Telefone: (14) 34021370 E-Mail: silvana.vidotti@unesp.br

pesquisa e coleta de dados até a fase de compartilhamento e/ou publicação de dados, bem como sua posterior reutilização.

Em muitos casos, os dados de pesquisa requerem um tratamento de anonimização antes de serem publicados, por exemplo, quando lidamos com dados sensíveis, como dados pessoais, dados médicos, dados que podem inferir comportamentos humanos, entre outros. Outras vezes é possível aplicar períodos de embargo aos dados, por exemplo, se são dados oriundos de pesquisas patenteáveis. Assim, o *Open Research Data Pilot* da Comissão Europeia propõe a sentença “[...] tão aberto quanto possível, tão fechado quanto necessário” (EUROPEAN COMMISSION, 2016, tradução nossa).

Diversos países, como os Países Baixos, França ou Qatar, têm estabelecido políticas nacionais de ciência aberta e gestão de dados (COÖPERATIE SURF, 2018; FRENCH MINISTER OF HIGHER EDUCATION, RESEARCH AND INNOVATION, 2018; QATAR MINISTRY OF INFORMATION & COMMUNICATIONS TECHNOLOGY, 2015). No Brasil, existe uma política de dados abertos como parte da Lei de Acesso à Informação, porém, o foco não são os dados científicos senão os dados governamentais disponíveis na Infraestrutura Nacional de Dados Abertos (INDA). Também, há estudos que apontam diretrizes para o desenvolvimento de uma política nacional brasileira de gestão de dados de pesquisa (GDP) (COSTA, 2017).

Por sua vez, muitas instituições (como universidades e centros de pesquisa) têm começado a elaborar estratégias para a gestão dos dados por elas produzidos e a redigir políticas institucionais. Por exemplo, a Fundação Oswaldo Cruz (FIOCRUZ) possui um termo de referência para a gestão e abertura dos dados de pesquisa da instituição (FUNDAÇÃO OSWALDO CRUZ, 2018). Depois, muitas instituições optam por estabelecer centros de informação ou centros especializados na GDP. Em alguns casos, essas tarefas são realizadas pelas bibliotecas universitárias ou de pesquisa ou as bibliotecas têm uma alta participação. Alguns exemplos são a Universidade de Glasgow² na Escócia, a Universidade Técnica de Darmstadt³ na Alemanha, o Instituto Federal de Tecnologia de Zurique⁴ na Suíça ou a Universidade de Estocolmo⁵ na Suécia. Um último passo é a incorporação de pessoal especializado na gestão de dados, desde bibliotecários de dados até os administradores de dados (*data stewards*). A figura do administrador de dados desenvolve os processos de governança de dados, de acordo com as políticas de dados vigentes (seja no nível de faculdade, universidade, estadual ou nacional), bem como a custódia dos dados e sua adequação aos padrões estabelecidos em cada área do conhecimento ou instituição. No contexto acadêmico, o administrador de dados atua em cada faculdade e atende as necessidades de uma área do conhecimento específica na qual é especialista.

Neste contexto, este trabalho analisa a implementação de serviços institucionais de GDP e descreve as fases da possível implementação destes serviços. Além disso, contrasta os perfis e áreas de atuação de três profissionais da gestão de dados

¹ “[...] as open as possible as close as necessary”.

² Universidade de Glasgow: <https://www.gla.ac.uk/myglasgow/datamanagement/rdmatglasgow/>

³ Universidade Técnica de Darmstadt: https://www.tu-darmstadt.de/tudata/tudata/digitale_forschungsdaten_an_der_tu/index.de.jsp

⁴ Instituto Federal de Tecnologia de Zurique: <https://ethz.ch/services/en/service/a-to-z/research-data.html>

⁵ Universidade de Estocolmo: <https://www.su.se/english/staff/services/research/research-data/research-data-group>

presentes no contexto científico-acadêmico: o bibliotecário de dados, o gestor de dados e o administrador de dados.

PROCEDIMENTOS METODOLÓGICOS

A metodologia utilizada para esta pesquisa é de carácter qualitativo com uma abordagem exploratória. Inicia-se com uma análise bibliográfica dos termos “*data stewardship*” e “gestão de dados de pesquisa” para contextualizar melhor o estado da arte. No levantamento, incluem-se resultados em português, espanhol e inglês presentes nos registros de *Google Scholar*, *Dimensions*, *BASE* e *SciELO*. Estes registros foram selecionados devido ao seu carácter de infraestrutura aberta e por oferecer uma abrangência maior na distribuição geográfica das publicações. A busca não foi delimitada temporalmente, tendo em vista que a atualidade da temática pesquisada faz com que os resultados sejam posteriores a 2012. Cabe assinalar que o termo “*data stewardship*” não tem representação na literatura não anglófona (p.ex. *administración de datos* em português ou *administración de datos* em espanhol) e na anglófona começou se popularizar apenas em 2014.

Após este levantamento bibliográfico, analisam-se serviços de gestão de dados já oferecidos por algumas instituições a partir da informação disponível nos seus sites *web*.

DADOS DE PESQUISA E GESTÃO DE DADOS DE PESQUISA

Segundo o glossário de termos sobre gestão de dados do Ponto de Coordenação Nacional Holandês de Gestão de Dados de Pesquisa (*Landelijk Coördinatiepunt Research Data Management*), dado é “[...] um conjunto de valores de variáveis qualitativas ou quantitativas que podem ser medidas, coletadas, relatadas e analisadas⁶”. Se esses dados são oriundos de uma pesquisa e servem para validar os resultados, então são denominados “dados de pesquisa”.

Como destaca Borgman (2008), os dados de pesquisa possuem um forte carácter disciplinar. Assim, para a ecologia, os dados de pesquisa incluem medições do tempo, leituras de sensores ou registros históricos; para a medicina, raios-x, resultados de exames, relatórios de patologias; para a química, estruturas de proteínas; para a astronomia, estudos espectrais; para a biologia, espécimes; e para a física, eventos e objetos. Nas ciências sociais, há sondagens de opinião, pesquisas públicas, entrevistas, experiências de laboratório ou de campo, registros demográficos, registros de censo, registros de voto e indicadores econômicos. Nas humanidades, os dados são os registros de atividades humana como jornais, fotografias, cartas, diários, livros, certidões de nascimento, óbito ou casamento, mapas, entre outros.

Além da identificação disciplinar, Sales e Sayão (2019) propõem uma taxonomia para os dados de pesquisa, na qual eles seriam classificados quanto a origem, grau de processamento, natureza, abordagem da pesquisa, nível de sensibilidade, materialidade, perenidade e abertura. Assim, podemos classificar os dados segundo a forma de coleta e uso para a pesquisa (origem); o grau de processamento (brutos, pré-processados ou processados); a abordagem (qualitativos ou quantitativos); a natureza ou o formato dos dados (números, textos, artefatos), o nível de

⁶ Data: https://www.edugroepen.nl/sites/RDM_platform/RDM%20Glossary/D/Data.aspx

sensibilidade (não sensíveis, confidenciais ou pessoais), a materialidade (digitais ou físicos), a perenidade (dados canônicos ou episódicos) e a abertura (grau de abertura ou licença).

Entender a origem dos dados e a sua disciplina serve de base para realizar um adequado processo de GDP. A GDP é entendida como “[...] um conjunto de práticas para lidar com a informação coletada e criada durante a pesquisa”⁷ (HIGMAN; BANGERT; JONES, 2019, p. 2). A definição se amplia da seguinte forma:

A gestão de dados de pesquisa diz respeito à organização de dados, desde a sua entrada no ciclo de investigação até a disseminação e arquivo de resultados de valor. A gestão de dados de pesquisa faz parte do processo de investigação e visa tornar o processo de investigação o mais eficiente possível, satisfazendo as expectativas e requisitos da universidade, das agências de apoio à pesquisa, da investigação e da legislação. (WHYTE; TEDDS, 2011, p. 1, tradução nossa)

Porém, para atingir o objetivo do reuso de dados e a reprodutibilidade da pesquisa, os dados de pesquisa não são suficientes. Para tornar reprodutível uma pesquisa é necessário fornecer todos os dados, bem os códigos para executar a coleta e análise novamente, recriando os resultados (BARBA, 2018). Como afirmam Chen et al. (2019), as práticas para a reprodutibilidade vão além do compartilhamento e abertura dos dados.

Dentro dessas práticas ao longo do ciclo de pesquisa, encontram-se a elaboração de planos de gestão de pesquisa (sejam em formato PDF ou em formato legível por máquinas (MIKSA et al., 2019)), a formatação dos dados seguindo padrões comuns na área, o processo de anonimização de dados sensíveis, a escolha de um repositório para o depósito ou publicação de dados (se possível um repositório confiável ou certificado, p.ex. Core Trust Seal⁸) ou o registro de um identificador persistente para o conjunto de dados e sua futura disseminação.

Ainda, cada conjunto de dados, em determinadas disciplinas, terá um processo diferente de gestão, diferentes requerimentos para a infraestrutura de arquivamento (DALLMEIER-TIESSEN et al., 2014) e diferentes modelos de citação (CALLAGHAN et al., 2012). Além disso, dada a relevância da publicação de dados de pesquisa e os correspondentes metadados, surgiu a iniciativa FAIR (*Findable, Accessible, Interoperable, Reusable*) (WILKINSON et al., 2016), que apresenta quinze princípios para a adequada publicação dos dados, que garantem que os dados sejam legíveis por máquinas e por humanos.

A administração de dados (*data stewardship*) “[...] envolve todas as atividades necessárias para garantir que os dados da investigação digital sejam encontráveis, acessíveis, interoperáveis e reutilizáveis (FAIR) ao longo prazo, incluindo a gestão, arquivo e reutilização de dados por terceiros” (JANSEN et al., 2019, p. 38, tradução nossa). O profissional encarregado das tarefas de administração de dados é o administrador de dados (denominado às vezes custodiador de dados ou *data custodian*). Um administrador de dados é responsável por assegurar a qualidade, a

⁷ Original em inglês: “Research data management (RDM) can be defined as a set of practices to handle information collected and created during research”.

⁸ CoreTrustSeal: <https://www.coretrustseal.org/>

integridade e as modalidades de acesso aos dados e metadados seguindo as leis vigentes, a política institucional de gestão de dados e as permissões individuais.

A PROFISSIONALIZAÇÃO DA GESTÃO DE DADOS: BIBLIOTECÁRIOS DE DADOS, GESTORES DE DADOS E ADMINISTRADORES DE DADOS

Devido à crescente demanda por uma adequada GDP no contexto dos projetos de pesquisa, tanto instituições quanto grupos de pesquisa têm visto a necessidade de contratar ou capacitar pessoal para o desenvolvimento destas tarefas. No contexto de gestão de dados surgem diferentes profissões, algumas delas relacionadas ao ambiente acadêmico⁹. A seguir, no Quadro 1, comparam-se as competências e tarefas dos três profissionais de dados mais vinculados ao âmbito acadêmico.

Quadro 1. Comparação das competências dos bibliotecários acadêmicos de dados, administradores de dados e gestores de dados

| | Bibliotecário acadêmico de dados | Administrador de dados | Gestor de dados |
|-------------------------------|---|---|---|
| Competências e tarefas | <p>Desenho e desenvolvimento conceitual de sistemas (conhecimento de tecnologia, metadados, protocolos, infraestruturas digitais de pesquisa, governança, curadoria digital, etc.) (RICE; SOUTHALL, 2016)</p> <p>Domínio do ecossistema de pesquisa (KENNAN, 2016; XIA; WANG, 2014)</p> <p>Conhecimento e aplicação de fluxos de trabalho em infraestruturas digitais (RICE; SOUTHALL, 2016; XIA; WANG, 2014)</p> <p>Conhecimento de tecnologias da informação e familiaridade com as linguagens de programação para dados (R e Python) (BRADLEY-RIDOUT, 2018; FEDERER, 2018; KENNAN,</p> | <p>Assessoramento específico aos pesquisadores sobre GDP (VERHEUL <i>et al.</i>, 2019)</p> <p>Desenvolvimento de políticas de gestão de dados no nível de faculdade (VERHEUL <i>et al.</i>, 2019)</p> <p>Informação e formação aos investigadores sobre GDP (VERHEUL <i>et al.</i>, 2019)</p> <p>Aconselhamento sobre a escrita de um plano de GDP (JANSEN <i>et al.</i>, 2019; VERHEUL <i>et al.</i>, 2019; MONS, 2018)</p> <p>Conhecimento dos padrões de metadados e ontologias presentes na área do conhecimento (VERHEUL</p> | <p>Concepção de planos de gestão de dados específicos para cada projeto.</p> <p>Proposta de soluções para a adaptação de dados existentes aos princípios FAIR.</p> <p>Contribuição na estratégia geral em torno dos planos de gestão de dados para o departamento.</p> <p>Concepção de modelos de metadados para os dados do projeto com base nas ontologias e normas existentes.</p> <p>Co-desenvolvimento de software e ferramentas para o projeto.</p> |

⁹ Fora do ambiente acadêmico encontramos exemplos de profissionais dos dados como o jornalista de dados, o analista de inteligência de negócio ou o analista de dados, entre outros.

| | | | |
|----------------------------|--|---|---|
| | <p>2016)</p> <p>Conhecimento de métricas para infraestruturas de pesquisa (FEDERER, 2018; RICE; SOUTHALL, 2016)</p> <p>Conhecimento de GDP (FEDERER, 2018; SNIPES, 2018; KENNAN, 2016; RICE; SOUTHALL, 2016; XIA; WANG, 2014; TENOPIR <i>et al.</i>, 2014)</p> <p>Conhecimento de catálogos de sistemas e dados (qualidade de metadados, semântica e <i>common standards</i>) (KENNAN, 2016)</p> <p>Habilidades comunicativas e didáticas (KENNAN, 2016)</p> | <p><i>et al.</i>, 2019; MONS, 2018)</p> <p>Conhecimento dos aspectos legais e éticos da gestão de dados (JANSEN <i>et al.</i>, 2019; DIJKERS, 2019; MONS, 2018)</p> <p>Trabalho direto com novos fluxos de trabalho e novos desenvolvimentos de infraestrutura para a área (DIJKERS, 2019; VERHEUL <i>et al.</i>, 2019)</p> | <p>Análise dos dados dentro do projeto.</p> |
| Formação | <p>Biblioteconomia com especialização em gestão de dados.</p> <p>Formação próxima a gestão de dados com grande conhecimento do mundo da biblioteconomia.</p> | <p>Doutorado na área do conhecimento da faculdade na qual desenvolvem suas atividades e profunda compreensão de GDP e das políticas a respeito.</p> | <p>Formação (preferência mestrado) na área do grupo de pesquisa e amplos conhecimentos na ciência de dados e GDP.</p> |
| Unidade de trabalho | <p>Biblioteca universitária ou de pesquisa ou centro de informação</p> | <p>Faculdade</p> | <p>Grupo de pesquisa</p> |

Fonte: Elaboração própria a partir de diversos autores e de descrição de vagas de emprego para gestores de dados¹⁰.

Além das atividades de GDP desenvolvidas diretamente pelos profissionais mencionados, várias instituições têm inserido competências em GDP nos graus curriculares de forma transversal em todas as áreas de estudo (TAMMARO; CASAROSA, 2014). Um exemplo são os cursos de GDP implementados na

¹⁰ As principais vagas de emprego foram “Data Manager/Data Steward” do Barcelona Supercomputing Centre (<https://www.biocat.cat/es/data-managerdata-steward-rezre3>) e “Data Steward for Rare Diseases Project” do Radboud University Medical Center (<https://www.euraxess.bg/jobs/398281>).

Universidade de Bielefeld na Alemanha para graduandos quanto pós-graduandos, onde desenvolve-se competências de organização, técnicas e legais para GDP (WILJES; CIMIANO, 2019). A Universidade Ártica de Noruega combina a GDP com os conceitos de ciência aberta e integridade na pesquisa nas formações oferecidas. O grupo alvo são os doutorandos, devido a que costumam ser mais receptivos aos desenvolvimentos (WANG, 2019).

Desta forma, os futuros graduados poderão aplicar este conhecimento nas pesquisas e em suas áreas de atuação profissional. Parte destas atividades de GDP, podem ser adquiridas mediante oficinas práticas (*hands-on workshop*), como as oferecidas pela iniciativa comunitária “The Carpentries”¹¹ nas suas diferentes seções (Data Carpentries, Software Carpentries e Library Carpentries). Especificamente, a seção Library Carpentries oferece uma introdução ao trabalho com dados para bibliotecários e explica ferramentas como R ou Python, para o processamento; Jupyter Notebooks¹², para a apresentação de documentos de trabalho com dados (cadernos eletrônicos); OpenRefine¹³, para a limpeza dos dados; e Git¹⁴, para o controle das versões.

No âmbito acadêmico, os profissionais da gestão de dados estão inseridos em um contexto que abrange diferentes atores internos e externos à instituição (grupos de pesquisa, departamentos de tecnologias da informação, bibliotecas e outras seções da instituição, agências de apoio à pesquisa, empresas da iniciativa privada, etc.). Por isso é importante consolidar os serviços de gestão de dados.

FASES PARA A IMPLEMENTAÇÃO DE SERVIÇOS INSTITUCIONAIS DE GESTÃO DE DADOS

Para implementação dos serviços institucionais de gestão de dados descritos na seção anterior, identificam-se três fases:

- a) Fase 1: Elaboração de uma política institucional de gestão de dados.
- b) Fase 2: Estabelecimento de uma unidade de informação.
- c) Fase 3: Integração dos especialistas em gestão de dados.

As fases apresentadas não são obrigatórias e podem ser fusionadas. Por exemplo, é possível que uma instituição estabeleça uma unidade de informação e depois crie uma política de GDP como parte da estratégia de comunicação e engajamento dos pesquisadores. Também não é necessário que aconteçam de forma consecutiva, podem ser simultâneas. A ideia final é contextualizar como a implementação pode ser desenvolvida.

Fase 1: Elaboração de uma política institucional de gestão de dados

Elaborar uma política de GDP serve para suprimir a deficiência legal ou a falta de política de algumas instituições (CLARE et al., 2019). Assim, muitas instituições

¹¹ The Carpentries: <https://carpentries.org/>

¹² Jupyter Notebooks: <https://jupyter.org/>

¹³ Openrefine: <http://openrefine.org/>

¹⁴ Git: <https://git-scm.com/>

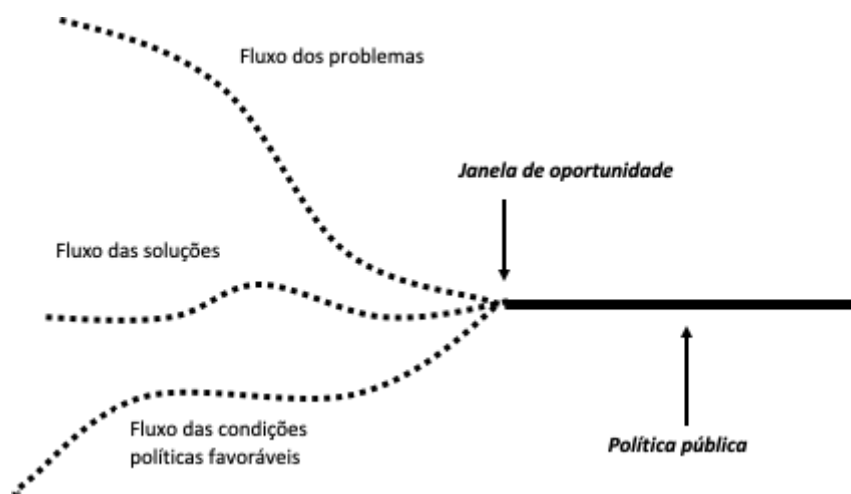
ajustam-se aos requerimentos governamentais (p.ex. conformidade com políticas nacionais de GDP) ou às agências de pesquisa (p.ex. conformidade com requerimentos de dados FAIR ou open).

Possuir uma política de GDP é um indicativo do compromisso institucional com esta tarefa, porém, apesar da importância que a gestão de dados está adquirindo, muitas instituições ainda carecem deste tipo de política.

Para analisar como uma instituição pode inserir a elaboração de uma política de gestão de dados na sua agenda, partimos do Modelo de Fluxos Múltiplos (KINGDON, 1984). Este modelo permite focar nas configurações de uma agenda institucional para oferecer alternativas para a formulação de políticas. Segundo este modelo, a formulação em si é a convergência de três fluxos: o fluxo dos problemas, o fluxo das soluções e o fluxo das condições políticas favoráveis.

Ao convergir estes três fluxos, abre-se uma janela de oportunidade que representa o ambiente favorável para o lançamento ou elaboração de uma nova política (Figura 1).

Figura 1. Diagrama do Modelo de Fluxos Múltiplos



Fonte: Adaptado¹⁵ de Kingdon (1984).

No caso do tratamento da gestão de dados, os fluxos contextualizam-se da seguinte maneira:

- a) Fluxo dos problemas: crise de reprodutibilidade das pesquisas provocada pela inadequada gestão dos dados (PENG, 2015) e a consequente necessidade de um trabalho com os dados (formatação, padronização, identificação, disponibilização em formatos adequados, respeito às licenças, etc.). Este problema surge principalmente na comunidade científica;

¹⁵ Adaptação realizada pela Profa. Ângela Maria Grossi de Carvalho para a disciplina “Políticas Públicas de Informação e Tecnologia” do PPGCI da UNESP.

- b) Fluxo das soluções: implementação de infraestrutura para a custódia dos dados e formação e contratação de profissionais especialistas na gestão desses dados, bem como a capacitação interna dos grupos de pesquisa;
- c) Fluxo das condições políticas favoráveis: compromisso institucional com a gestão de dados. Às vezes, este compromisso vem de esferas políticas superiores, como agências de apoio à pesquisa ou ministérios de ciência e pesquisa. Para o caso da GDP, diversas organizações internacionais (p.ex. GOFAIR¹⁶, Research Data Alliance - RDA¹⁷ e Committee on Data International Science Council - CODATA¹⁸) têm contribuído com o favorecimento das condições. Especificamente, no caso da RDA e seus grupos de trabalho, potencia-se a participação das comunidades científicas, tendo o objetivo de obter soluções rápidas aos problemas do tratamento dos dados de pesquisa (MÉNDEZ, 2018).

Para que uma instituição possa aproveitar esta janela de oportunidade e elaborar uma política de gestão de dados, pode começar pela análise e redigir uma política a partir do zero ou adaptar modelos desenvolvidos para este fim.

Um exemplo deste tipo de modelo foi elaborado como parte do projeto Leaders Activating Research Networks (LEARN), que inclui epígrafes sobre a jurisdição, os direitos autorais, o tratamento com os dados, as responsabilidades, deveres e direitos da instituição e dos pesquisadores, bem como a validade da política (LEARN, 2017). Ademais, este projeto também disponibiliza diretrizes para o desenvolvimento de políticas neste contexto.

Fase 2: Estabelecimento de uma unidade de informação

Após o desenvolvimento de uma política institucional, há a necessidade de uma unidade de informação onde os pesquisadores da instituição possam receber as informações necessárias, considerando a fase do ciclo de pesquisa onde eles estejam (pré-projeto, projeto ou pós-projeto). Estas unidades são conhecidas também como pontos de referência ou pontos de informação. Em algumas instituições, estas unidades constituem um centro de GDP (p.ex. o Center for Research Data Management¹⁹ da Universidade Técnica de Viena - TU Wien).

Este tipo de unidade abrange competências de trabalho tanto técnicas quanto de organização e, de forma generalizada, atua como nexos entre diferentes seções da instituição que desenvolvem tarefas necessárias para cada fase do ciclo de pesquisa. Por exemplo, o departamento jurídico para o esclarecimento dos aspectos legais da pesquisa; o comitê de ética para a autorização da coleta de dados ou do tratamento dos dados sensíveis; a seção de tecnologias da informação para a disponibilização do *software*; a biblioteca para o arquivamento dos dados e o registro de identificadores persistentes. Este conjunto de atores dentro da própria instituição requer uma correta estruturação de fluxos de trabalho para oferecer informações claras e objetivas aos pesquisadores.

¹⁶ GOFAIR: <https://www.go-fair.org>

¹⁷ RDA: <https://rd-alliance.org/>

¹⁸ CODATA: <http://www.codata.org/>

¹⁹ Site do Centre for Search Data Management da TU Wien: <https://www.tuwien.at/en/research/rtd-support/research-data/overview/>

Nesse contexto, surgem diferentes propostas de elaboração de marcos de trabalho para a gestão dos dados de pesquisa. Entre elas, destaca-se o modelo DIAMANT - Design and Information Architecture for Data MANagement Technologies do projeto PODMAN²⁰, que vincula os serviços de diferentes seções dentro de uma instituição com as fases do projeto de pesquisa e a consequente necessidade de serviços aos pesquisadores em cada fase. A unidade de informação serve como nexos de contato entre a fase do projeto, a função do processo de gestão de dados e as seções institucionais (Figura 2).

Figura 2. Modelo DIAMANT para a supervisão de tecnologias para a gestão de dados do projeto PODMAN



Fonte: Blask e Förster (2019, p.3) (tradução e reedição da figura nossa).

Além disso, a unidade de informação deve encaminhar o pesquisador à seção competente para resolver o seu problema. Destarte, a ideia do modelo é elaborar

²⁰ O projeto PODMAN foi financiado pelo Ministério Alemão de Educação e Pesquisa: <https://fdm.uni-trier.de/>

processos de gestão de dados otimizados para a perspectiva funcional (funções da gestão de dados), para a perspectiva de organização (seções envolvidas na gestão) e para a perspectiva de governança (relacionamentos entre função e seção), de forma que o usuário final (pesquisador) se beneficie da integração do processo de gestão de dados no processo de pesquisa (BLASK; FÖRSTER, 2019). Possuir esses tipos de estratégias bem definidas também contribui com a maturidade dos serviços oferecidos pela instituição (PENG, 2018).

Com frequência, as bibliotecas universitárias ou de pesquisa desenvolvem parte das tarefas de GDP e ampliam seu portfólio de serviços nessa direção (SCHIRRWAGEN *et al.*, 2019; ALCALÁ PONCE DE LEÓN; ANGLADA I DE FERRER, 2018; TENOPIR *et al.*, 2014). Portanto, em muitas instituições, a unidade de informação encontra-se na biblioteca. Também existem universidades, como a Universidade Técnica de Darmstadt, a Universidade de Estocolmo ou o Instituto Tecnológico Federal de Zurique, que criam grupos multidisciplinares para a gestão de dados. Estes grupos estão formados por empregados da biblioteca, do departamento de apoio à pesquisa e do departamento de informática. No Brasil, a FIOCRUZ segue esta mesma linha com o estabelecimento do Grupo de Trabalho em Ciência Aberta, onde tratam-se temas como os termos de referência para a GDP ou a implementação de planos de gestão de dados acionáveis por máquinas (em inglês: machine-actionable DMP).

Porém as formas de comunicar a informação relativa ao processo de GDP e os fluxos de trabalho são variáveis, já que dependem da estrutura interna da instituição, do nível de conhecimento sobre GDP do pesquisador e da área do conhecimento. É importante centralizar a informação para não tornar as tarefas de GDP excessivamente demoradas e tediosas para os pesquisadores, para que eles possam destinar o tempo às tarefas de pesquisa.

Fase 3: Integração dos especialistas em gestão de dados

O estabelecimento de uma unidade de informação destinada à GDP deve levar consigo a integração de especialistas em gestão de dados no pessoal institucional. Dado que as bibliotecas costumam ser as seções onde se estabelece esta unidade, o profissional é um bibliotecário de dados. Estes profissionais desenvolvem principalmente um trabalho de formação de pesquisadores e de gestão da informação sobre GDP. Alguns exemplos de instituições que trabalham com bibliotecários de dados são a Universidade Livre de Berlin na Alemanha, a Universidade de Edimburgo na Escócia ou o Colégio Universitário de Dublin na Irlanda.

Não obstante, dependendo do número de pesquisadores da instituição ou da distribuição geográfica dela (p.ex. universidades estaduais multicampi), uma pequena equipe pode ser insuficiente e gerar uma deficiência de pessoal (YORK; GUTMANN; BERMAN, 2018). Para poder oferecer uma atenção direta em cada faculdade, estabelece-se a figura do administrador de dados, que, como descrito anteriormente, assessora ao longo do processo de pesquisa sobre planos de GDP, padrões de metadados, tratamento e infraestrutura de dados.

Um exemplo de integração da figura do administrador de dados é a Universidade Técnica de Delft (TU Delft) nos Países Baixos. Esta universidade possui um administrador de dados em cada uma das oito faculdades, tendo o administrador de dados a mesma categoria profissional que um pós-doutor. A coordenação das atividades dos administradores de dados é realizada desde a biblioteca, de forma

similar ao que acontece com os bibliotecários especializados (*liaison librarian*) em muitas bibliotecas universitárias. Simultaneamente, a biblioteca possui uma equipe de apoio à GDP que trabalha em conjunto com o centro de GDP da universidade (VERHEUL et al., 2019). Desta maneira, é possível abranger todos os aspectos da GDP (técnico/funcional, legal, de organização e de coordenação). Ainda, a metodologia de trabalho desenvolvida é aberta e publicada por meio do blog “Open Working”²¹.

A implementação de serviços de administração de dados também envolve o trabalho conjunto com pesquisadores para atuar a favor de uma adequada GDP. Estes pesquisadores, denominados *data champions* pela TU Delft e a Universidade de Cambridge (SAVAGE; CADWALLADER, 2019)—as duas instituições europeias que têm promovido publicamente o trabalho com *data champions*—, servem como exemplo de boa GDP e ajudam na sua promoção. Isto é um indicador de que a tarefa de GDP não consiste exclusivamente em uma responsabilidade institucional, mas também dos pesquisadores.

Segundo Jansen et al. (2019), o controle do dados de pesquisa, a reutilização de dados, a aplicação direta dos princípios FAIR, a responsabilidade sobre os direitos de propriedade intelectual e o compartilhamento dos dados, entre outros, são responsabilidades diretas dos pesquisadores. Não obstante, podem exercer estas responsabilidades com a ajuda de um gestor de dados dentro do projeto e com o apoio e o assessoramento de um administrador de dados no âmbito de faculdade.

CONSIDERAÇÕES FINAIS

Dada a relevância da preservação e do adequado tratamento dos dados oriundos da pesquisa e o requerimento de muitas agências de apoio à pesquisa para a publicação desses dados (ou pelo menos dos metadados em situações de dados sensíveis), as tarefas de GDP têm se tornado cruciais para instituições de pesquisa.

Resulta, portanto, importante analisar como são desenvolvidas estas tarefas em instituições, quais as fases para implementar serviços de apoio à GDP e quais perfis profissionais surgem para este fim.

Há três perfis profissionais do trabalho com dados no âmbito científico-acadêmico, cada um com uma área de atuação diferente. O bibliotecário de dados atua como uma figura central dentro da biblioteca e é encarregado da supervisão geral dos dados de pesquisa e dos sistemas da informação disponíveis na instituição. O administrador de dados localiza-se em cada faculdade ou centro de pesquisa e assessora os pesquisadores de determinada área do conhecimento nas tarefas de GDP. Consequentemente, deve ser um especialista na respectiva área: os administradores de dados possuem doutorado na área ou estão prestes a finalizá-lo. O gestor de dados trabalha dentro do grupo de pesquisa e é encarregado da análise e preparação final dos dados seguindo os princípios FAIR e os padrões comuns da área.

Para que estes profissionais dos dados possam se integrar nos fluxos de trabalho institucionais, as instituições devem estabelecer, previamente, um contexto para os serviços relacionados com a GDP.

Distinguímos três fases no estabelecimento destes serviços: a elaboração de uma política institucional de gestão de dados que estabeleça a regras dentro da instituição

²¹ Open Working - TU Delft: <https://openworking.tudl.tudelft.nl/>

e defina o compromisso institucional com a GDP; o estabelecimento de uma unidade de informação que ajude a regular os fluxos de trabalho e informação dentro da instituição para fornecer melhores serviços de GDP; e a integração dos especialistas em gestão de dados, com foco nos administradores de dados, que trabalhem no âmbito de faculdade e possam servir de suporte direto para os pesquisadores.

O objetivo final destes serviços é facilitar as tarefas relacionadas com a GDP para que os pesquisadores tenham o foco na pesquisa. Além disso, os dados já começam a ser considerados produtos de pesquisa avaliáveis; portanto, é necessário o seu correto tratamento para disponibilizá-los como parte da produção científica.

Com estas fases oferece-se um norteamento para instituições que estejam incluindo a GDP e os conceitos de administração de dados como parte do seu portfólio de serviços.

Artigo recebido em 09/07/2019 e aprovado em 25/10/2019.

REFERÊNCIAS

ALCALÁ PONCE DE LEÓN, M. A.; ANGLADA I DE FERRER, L. From open access to open data: collaborative work in the university libraries of Catalonia. *Liber Quartely*, v. 28, p. 1-14, 2018. Disponível em: <http://doi.org/10.18352/lq.10253> . Acesso em: 01 jul. 2019.

BARBA, L. A. *Terminologies for reproducible research*: ArXiv:1802.03311 [cs], ArXiv: 1802.03311, 9 fev. 2018. Disponível em: <http://arxiv.org/abs/1802.03311>. Acesso em: 01 jul. 2019.

BLASK, K.; FÖRSTER, A. Designing an information architecture for data management technologies: introducing the DIAMANT model. *Journal of Librarianship and Information Science*, p. 1-9, 2019. Disponível em: <http://journals.sagepub.com/doi/10.1177/0961000619841419> . Acesso em: 01 jul. 2019.

BORGMAN, C. L. Data, disciplines, and scholarly publishing. *Learned Publishing*, v. 21, n. 1, p. 29-38, 2008. Disponível em: <http://doi.wiley.com/10.1087/095315108X254476> . Acesso em: 01 jul. 2019.

BRADLEY-RIDOUT, G. Preferred but not required: examining research data management roles in Health Science Librarian Positions. *Journal of the Canadian Health Libraries Association / Journal de l'Association des bibliothèques de la santé du Canada*, v. 39, n. 3, p. 138-145, 2018. Disponível em: <https://journals.library.ualberta.ca/jchla/index.php/jchla/article/view/29368> . Acesso em: 01 jul. 2019.

CALLAGHAN, S.; DONEGAN, S.; PEPLER, S.; THORLEY, M.; CUNNINGHAM, N.; KIRSCH, P.; AULT, L.; BELL, P.; BOWIE, R.; LEADBETTER, A.; LOWRY, R.; MONCOIFFÉ, G.; HARRISON, K.; SMITH-HADDON, B.; WEATHERBY, A.; WRIGHT, D. Making data a first class scientific output: data citation and publication by NERC's environmental data centres. *International Journal of Digital Curation*, v. 7, n. 1, p. 107-113, 2012. Disponível em: <http://www.ijdc.net/article/view/208> . Acesso em: 01 jul. 2019.

CATAR. Qatar Ministry of Information & Communications Technology. *Data management policy*. 2015. Disponível em: http://www.motc.gov.qa/sites/default/files/data_management_policy.pdf. Acesso em: 01 jul. 2019.

- CHEN, X. *et al.* Open is not enough. *Nature Physics*, v. 15, n. 2, p. 113-119, 2019. Disponível em: <https://doi.org/10.1038/s41567-018-0342-2>. Acesso em: 01 jul. 2019.
- COÖPERATIE SURF. *A national strategy for research data management*. 2018. Disponível em: <https://www.lcrdm.nl/en>. Acesso em: 01 jul. 2019.
- COSTA, M. M. *Diretrizes para uma política de gestão de dados científicos no Brasil*. 2017. Disponível em: <http://repositorio.unb.br/handle/10482/24895>. Acesso em: 01 jul. 2019.
- DALLMEIER-TIESSSEN, S.; LAVASA, A.; HERTERICH, P.; RUEDA, L.; KOTARSKI, R.; NEWBOLD, E. A comparative analysis of disciplinary data management workflows. In: IEEE/ACM JOINT CONFERENCE ON DIGITAL LIBRARIES 2014, London, United Kingdom. *Electronic proceedings [...]*. Disponível em: <http://ieeexplore.ieee.org/document/6970180/>. Acesso em: 01 jul. 2019.
- DIJKERS, M. P. A beginner's guide to data stewardship and data sharing. *Spinal Cord*, v. 57, n. 3, p. 169-182, 2019. Disponível em: <http://www.nature.com/articles/s41393-018-0232-6>. Acesso em: 01 jul. 2019.
- EUROPEAN COMMISSION. *Guidelines on FAIR data management in Horizon 2020*. 2016. 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: 01 jul. 2019.
- FEDERER, L. Defining data librarianship: a survey of competencies, skills, and training. *Journal of the Medical Library Association*, v. 106, n. 3, 2018. Disponível em: <http://jmla.pitt.edu/ojs/jmla/article/view/306>. Acesso em: 01 jul. 2019.
- FRANÇA. French Minister of Higher Education, Research and Innovation. *National plan for open science*. [S.l: s.n.], 2019. Disponível em: https://libereurope.eu/wp-content/uploads/2018/07/SO_A4_2018_05-EN_print.pdf. Acesso em: 01 jul. 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.
- HIGMAN, R.; BANGERT, D.; JONES, S. Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights the UKSG Journal*, v. 32, p. 18, 2019. Disponível em: <http://insights.uksg.org/articles/10.1629/uksg.468/>. Acesso em: 01 jul. 2019.
- JANSEN, P.; VAN DEN BERG, L.; VAN OVERVELD, P.; BOITEN, J.-W. Research data stewardship for healthcare professionals. In: KUBBEN, P.; DUMONTIER, M.; DEKKER, A. (ed.). *Fundamentals of clinical data science*. Cham: Springer International Publishing, 2019. p. 37-53. Disponível em: https://doi.org/10.1007/978-3-319-99713-1_4. Acesso em: 01 jul. 2019.
- KENNAN, M. A. Data management: knowledge and skills required in research, scientific and technical organisations. In: IFLA CONFERENCE 2016 2016, Columbus. *Proceedings [...]* 2016.
- KINGDON, J. W. *Agendas, alternatives, and public policies*. Harlow: Pearson, 1984.
- LEARN. *Model policy for Research Data Management (RDM) at research institutions/institutes*: LEARN Toolkit of best practice for research data management. 2017. Disponível em: <http://discovery.ucl.ac.uk/1546518/>. Acesso em: 01 jul. 2019.
- MÉNDEZ, E. RDA, la otra RDA: la Research Data Alliance. Cuando 'FAIR' es algo más que 'justo'. *Anuario ThinkEPI*, v. 12, p. 378, 2018. Disponível em:

<https://recyt.fecyt.es/index.php/ThinkEPI/article/view/thinkepi.2018.63>. Acesso em: 01 jul. 2019.

MIKSA, T.; SIMMS, S.; MIETCHEN, D.; JONES, S. Ten principles for machine-actionable data management plans. *PLOS Computational Biology*, v. 15, n. 3, p. e1006750, 2019. Disponível em: <http://dx.plos.org/10.1371/journal.pcbi.1006750> . Acesso em: 01 jul. 2019.

MONS, B. *Data stewardship for open science: implementing FAIR principles*. 2018. Disponível em: <https://www.taylorfrancis.com/books/9781498753180>. Acesso em: 01 jul. 2019.

PENG, G. The State of assessing data stewardship maturity: an overview. *Data Science Journal*, v. 17, 2018. Disponível em: <http://datascience.codata.org/articles/10.5334/dsj-2018-007/>. Acesso em: 01 jul. 2019.

PENG, R. The reproducibility crisis in science: a statistical counterattack. *Significance*, v. 12, n. 3, p. 30-32, 2015. Disponível em: <http://doi.wiley.com/10.1111/j.1740-9713.2015.00827.x>. Acesso em: 01 jul. 2019.

RICE, R.; SOUTHALL, J. *The data librarian's handbook*. London: Facet Publishing, 2016. Acesso em: 01 jul. 2019.

SALES, L. F.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. *Conhecimento em Ação*, v. 4, n. 1, p. 31-48, 2019. Disponível em: <https://revistas.ufrrj.br/index.php/rca/article/view/26337/14573>. Acesso em: 01 jul. 2019.

SAVAGE, J. L.; CADWALLADER, L. Establishing, developing, and sustaining a community of data champions. *Data Science Journal*, v. 18, p. 23, 2019. Disponível em: <http://datascience.codata.org/articles/10.5334/dsj-2019-023/>. Acesso em: 01 jul. 2019

SCHIRRWAGEN, J.; CIMIANO, P.; AYER, V.; PIETSCH, C.; WILJES, C.; VOMPRAS, J.; PIEPER, D. Expanding the research data management service portfolio at Bielefeld University: according to the three-pillar principle towards data FAIRness. *Data Science Journal*, v. 18, 2019. Disponível em: <http://datascience.codata.org/articles/10.5334/dsj-2019-006/>. Acesso em: 01 jul. 2019.

SNIPES, G. Everyone's a data librarian now. *Journal of New Librarianship*, v. 3, n. 1, p. 28-31, 2018. Disponível em: <http://www.newlibs.org/article/3240-everyone-s-a-data-librarian-now>. Acesso em: 01 jul. 2019.

TAMMARO, A. M.; CASAROSA, V. Research data management in the curriculum: an interdisciplinary approach. *Procedia Computer Science*, v. 38, p. 138-142, 2014. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1877050914013830>. Acesso em: 11 mar. 2019. Acesso em: 01 jul. 2019.

TENOPIR, C.; SANDUSKY, R. J.; ALLARD, S.; BIRCH, B. Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, v. 36, n. 2, p. 84-90, 2014. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0740818814000255>. Acesso em: 01 jul. 2019.

VERHEUL, I.; IMMING, M.; RINGERMA, J.; MORDANT, A.; PLOEG, J.-L. Van der; PRONK, M. *Data Stewardship on the map: a study of tasks and roles in Dutch research institutes*. Zenodo, 2019. Disponível em: <https://zenodo.org/record/2669149>. Acesso em: 01 jul. 2019.

WANG, Y. Engaging with RDM through a PhD course on academic integrity and open science at UiT The Arctic University of Norway. In: CLARE, C. et al. Engaging researchers with data management: the cookbook. [S.l.]: Open Book Publishers, 2019.

WHYTE, A.; TEDDS, J. *Making the case for research data management*: DCC briefing papers. Edinburgh: Digital Curation Centre. 2011. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers>. Acesso em: 01 jul. 2019.

WILJES, C.; CIMIANO, P. Teaching research data management for students. *Data Science Journal*, v. 18, p. 38, 13 ago. 2019. Disponível em: <http://datascience.codata.org/articles/10.5334/dsj-2019-038/>. Acesso em: 20 out. 2019.

WILKINSON, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, v. 3, p. 160018, 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 01 jul. 2019.

XIA, J.; WANG, M. Competencies and responsibilities of social science data librarians: an analysis of job descriptions. *College & Research Libraries*, v. 75, n. 3, p. 362-388, 2014. Disponível em: <http://crl.acrl.org/index.php/crl/article/view/16367>. Acesso em: 01 jul. 2019.

YORK, J.; GUTMANN, M.; BERMAN, F. What do we know about the stewardship gap. *Data Science Journal*, v. 17, 2018. Disponível em: <http://datascience.codata.org/articles/10.5334/dsj-2018-019/>. Acesso em: 01 jul. 2019.