

GESTÃO DE DADOS DE PESQUISA: UM PANORAMA DA ATUAÇÃO DA UNIÃO EUROPEIA

MÁRCIA TEIXEIRA CAVALCANTI*
LUANA FARIA SALES**

RESUMO

Se, na década passada, as discussões que se iniciaram pelos anos de 1970 ainda ocupavam o espaço acadêmico sobre qual o posicionamento que a comunidade científica seguiria em relação aos meios e modos de divulgação de sua produção, mais um elemento surge para aumentar o debate: a necessidade de compartilhamento dos dados da pesquisa. Cada vez mais os dados de pesquisa passam a ter um papel significativo dentro do cenário científico mundial. A proposta deste trabalho é realizar um mapeamento dos repositórios de dados de pesquisa existentes nos países membros da União Europeia, inclusive realizando um levantamento que quantifique o número de repositórios de dados existentes nestes países. Foi acessado o link de cada país relacionado, disponível no documento do *Openaire* intitulado “*Overview of Open Access in the EU member states*”, fazendo uma revisão sistemática sobre a

* Doutorado em Ciência da Informação pelo IBICT/UF RJ (2014); Mestrado em Ciência da Informação, pelo IBICT/UF RJ (2002); Graduação em Ciências Sociais, pela Universidade Federal do Rio de Janeiro/UF RJ (1991); Graduação em Letras Português/Literatura, pela Sociedade de Ensino Superior Estácio de Sá/UNESA (2006). II. Temas de Pesquisa: Atuação na área de Ciência da Informação, Ciências Sociais, com ênfase nos temas: memória, patrimônio, centros de documentação, ética na ciência, gestão do conhecimento. III. Participação em grupos de pesquisa do CNPq: Estudos em Epistemologia, Ética e Política de Informação/UFF (pesquisador); Informação, Memória e Sociedade/IBICT (pesquisador); Desenvolvimento de métodos e técnicas para preservação do conhecimento gerado pelas pesquisas do IEN (pesquisador) marciacavalcanti@gmail.com

** Doutora em Ciência da Informação pelo Programa de Pós-Graduação do IBICT/UF RJ (2011-2014). Mestre em Ciência da Informação pelo convênio UFF/IBICT (2004-2006), Graduação em Biblioteconomia e Documentação pela Universidade Federal Fluminense (2003). Atualmente é Analista em C & T da CNEN, atuando como Bibliotecária-Chefe do Instituto de Engenharia Nuclear e líder da área Temática Gestão do Conhecimento Nuclear. Atua ainda como professora da Universidade Federal do Estado do Rio de Janeiro ? UNIRIO, ministrando disciplinas relacionadas à Organização do Conhecimento. Tem experiência na área de Ciência da Informação, com ênfase em Organização do Conhecimento, Representação e Recuperação de Informações, atuando principalmente nos seguintes temas: Taxonomias, Ontologias, Vocabulários Controlados, Tesouros, Terminologia e Software de Tesouro. Possui interesse em tópicos ligados à Comunicação Científica, Tecnologia de Informação e Gestão do Conhecimento e desenvolve pesquisas especificamente nas temáticas de e-Science, curadoria digital de dados de pesquisa, biblioteca digital, metadados, repositórios institucionais, repositórios de dados, Sistemas CRIS e objetos digitais. luanafsales@gmail.com

existência e a quantidade de repositório de dados de pesquisa e como descrevem seu funcionamento. Como resultado, o artigo apresenta uma sistematização do panorama atual da gestão de dados de pesquisa e dos repositórios existentes e um panorama geral das políticas implementadas e em desenvolvimento na UE.

PALAVRAS-CHAVE: Dados de pesquisa. Acesso aberto. Curadoria digital de dados.

ABSTRACT

RESEARCH DATA MANAGEMENT: AN OVERVIEW OF PERFORMANCE IN THE EUROPEAN UNION

If, in the past decade, the discussions that began by 1970 still occupied the academic space on which the position that the scientific community would follow in relation to the dissemination of ways and means of production, another element appears to grow the debate: the need of sharing research data. Now, more and more research data have a significant role in the global scientific scene. The purpose of this study is to realize a mapping of the main actors involved in the establishment of national policies of archiving and open access from the countries member of European Union (EU) for research data, because Brazil has been basing in these standards and guidelines. It was accessed the link of each country, available on the "Overview of Open Access in the EU member states", making a systematic review of the existence and amount of research data repository and describe how it works. As a result, the article presents a systematic overview of the current research data management and existing repositories and an overview of the policies implemented and developed in the EU.

KEYWORDS: Research data. Open access. Digital curation of data.

1 CONSIDERAÇÕES INICIAIS

O aumento de periódicos especializados produzidos pelas sociedades científicas e comprados pelas bibliotecas, no final do século XIX, ocorreu como consequência do processo de profissionalização dentro dos diferentes campos de conhecimento. Antes de seu surgimento, as notícias relacionadas à ciência eram veiculadas de forma privada através de cartas trocadas entre os cientistas, como também em atas das reuniões científicas, em folhetins, volantes e jornais cotidianos (FREITAS, 2006; STUMPF, 1996).

Os periódicos foram, desde seus primórdios, importantes canais de publicação de notícias científicas. No século XIX, expandiram-se e especializaram-se, vindo a realizar importantes funções no mundo da ciência. Ao publicarem textos, os estudiosos registram o conhecimento (oficial e público), legitimam disciplinas e campos de estudos, veiculam a comunicação entre os cientistas e propiciam ao cientista o reconhecimento público pela prioridade da teoria ou

da descoberta (FREITAS, 2006, p. 54).

Do século XIX aos dias de hoje, não somente mudou o suporte em que esses periódicos são publicados como também a forma como eles são acessados. No entanto, a importância da publicação científica – como um canal de registro do conhecimento que está sendo produzido e do reconhecimento público de sua autoria – permanece.

O processo definitivo de mudança para o novo veículo de registro e comunicação da ciência só foi concluído no século passado, quando as revistas adquiriram credibilidade para, inclusive, substituir os livros. Os artigos, até aquela época, eram considerados como formas provisórias de comunicação, sendo sempre a forma monográfica de livros impressos a preferida para o registro definitivo da ciência. A visão de que cada observação ou experimento forma uma unidade por si mesmo, só começou a ter aceitação no século XVIII (STUMPF, 1996).

Stumpf, no final da década de 1990, descreve o processo de transformação que a publicação científica está passando com a possibilidade do uso das redes de telecomunicações para a transmissão eletrônica. Segundo ela, naquele momento, os editores científicos passam a introduzir nas revistas os artigos que já foram avaliados pelos pares, e os leitores, em seus próprios computadores pessoais, acessam a *home page* da revista onde podem visualizar a lista de autores, de artigos, resumos e os textos completos, mas esse acesso pode ser vetado àqueles que não pagaram pela assinatura, quando esta é necessária (STUMPF, 1996).

Na década de 1990, começa em várias universidades prestigiosas do mundo um movimento para que suas publicações sejam de acesso livre, como forma de democratizar os resultados da pesquisa. Em 1998, surge o SciELO como um projeto piloto e pioneiro nesta linha de ação, com apenas dez periódicos que nem sequer eram de “primeiro mundo” (PACKER, 2009). Isso foi alguns anos antes da famosa declaração da *Open Society* de Budapeste, das conferências de Berlim e de outras iniciativas nesse sentido. Dez periódicos-pareciam nada, quando o mercado das empresas editoriais comerciais publicavam dezenas de milhares de periódicos por assinatura (SPINAK, 2013).

Segundo Costa (2008), desde meados da década de 1990, a questão do acesso aberto (*open access*) relacionado à literatura científica vem sendo discutida, assim como a criação de repositórios

institucionais e a edição de periódicos científicos de acesso aberto. A autora chama a atenção para o fato de que o acesso aberto tanto pode se relacionar à literatura científica quanto ao conhecimento, e traça a diferença entre ambos os aspectos que, segundo ela, se mostram de extrema importância para a definição de políticas governamentais e institucionais.

É importante ressaltar que acesso aberto ao conhecimento científico refere-se tanto aos aspectos formais quanto informais do processo de comunicação científica (LEITE, 2006). Envolve, ainda, questões relacionadas não somente à comunicação da pesquisa, mas, também, a ações outras, tais como as interações existentes no contexto do ensino. Este, por outro lado, envolve tanto a instrução dada a grandes grupos em salas de aula, quanto a indivíduos, em sessões de orientação do trabalho de pesquisa. Isso, como observa Leite (2006), abrange aspectos da comunicação informal.

[...]

Levando em conta que a publicação formal de resultados de pesquisa conforma etapa fundamental do processo de comunicação científica, o acesso aberto à literatura científica guarda relação com as questões da comunicação. Esta, por outro lado, é, igualmente, componente crucial da gestão do conhecimento. Em ambos os contextos, entretanto, é preciso que se tenha em mente que os padrões de comunicação e de comportamento informacional diferem, para cada divisão do conhecimento, tal como ocorre a suas disciplinas e subdisciplinas (COSTA, 2008, p. 222-223).

Mueller, em 2006, considerava o movimento pelo acesso livre ao conhecimento científico como o fato mais importante relacionado à comunicação científica. Mas pondera que, ao mesmo tempo em que ele se mostrava como um enorme desafio para a comunidade científica, também provocaria uma mudança radical no sistema tradicional de comunicação deste conhecimento (MUELLER, 2006).

Na década passada, assim como atualmente, as discussões que se iniciaram pelos anos de 1970, lembrando que o Projeto Gutenberg¹ foi lançado em 1971, ainda ocupavam o espaço acadêmico sobre qual o posicionamento que a comunidade científica (incluindo aqui pesquisadores, instituições, editoras, agências) seguiria em relação aos meios e modos de divulgação da

¹ O Project Gutenberg é a primeira e a maior coleção unitária de livros eletrônicos (ou *ebooks*) gratuitos. Disponível em: <<https://bibletrasufrj.wordpress.com/2011/09/27/837/>>. Acesso em: 27 maio 2016.

sua produção, do seu conhecimento. Ressalta-se que mais um elemento surge para aumentar o debate: a necessidade de compartilhamento dos dados que subsidiam a pesquisa.

Cada vez mais os dados de pesquisa passam a ter um papel significativo dentro do cenário científico mundial. O acesso a um artigo e aos dados a ele relacionados torna possível a replicação dessas pesquisas e o desenvolvimento de novos estudos a partir do reuso dos dados de pesquisa. Esse processo de valorização do acesso aberto aos dados de pesquisa surge no longo caminho já pavimentado pelo acesso aberto aos periódicos científicos e neste cenário algumas questões merecem ser discutidas: todos os dados devem ser compartilhados? E os dados sensíveis, que geram patentes? Quem é o responsável pelo depósito dos dados a serem compartilhados e como fazê-lo?

Assim, a proposta deste trabalho é realizar um mapeamento dos repositórios de dados de pesquisa existentes nos países membros da União Europeia (UE), realizando um levantamento que quantifique o número de repositórios de dados existentes nesses países e mostre outras realizações importantes ao que tange à gestão de dados de pesquisa.

A escolha por trabalhar com esses países como objeto de reflexão se pautava no fato do Brasil já vir adotando, com sucesso, em suas linhas de atuação do movimento de acesso aberto a publicações científicas, os padrões e diretrizes já adotados na comunidade europeia. Muitos destes países ainda estão na etapa inicial de discussões sobre o tema, ao passo que outros não só já arquivam e compartilham periódicos científicos como também os dados de pesquisa.

2 O QUE É ACESSO ABERTO (OPEN ACCESS)?

As primeiras iniciativas sobre acesso aberto ocorreram na década de 1990, provavelmente porque foi nesta época que a tecnologia tornou possível esse tipo de demanda. Além de ser o período no qual ocorreu a crise dos periódicos, por conta do aumento abusivo do valor das assinaturas que inviabilizava a manutenção destas pelas instituições e mais ainda pelos pesquisadores individuais.

A *Open Archives Initiative* (OAI), criada em 1999, na Convenção de Santa Fé, no Novo México (EUA), define os princípios básicos do acesso aberto à produção científica. A partir desse marco,

importantes declarações de apoio ao acesso aberto promoveram a disseminação dessa filosofia, como por exemplo, as declarações de Budapeste, Bethesda e Berlin (SILVA; ALCARÁ, 2008, p. 1).

A *Budapeste Open Access Initiative* (BOAI)² prega que a convergência de uma tradição (a vontade dos cientistas em publicar os frutos de suas pesquisas) com a nova tecnologia (a internet), irá possibilitar a distribuição de uma literatura revista pelos pares e com acesso totalmente livre e irrestrito para todos. A remoção das barreiras de acesso, segundo eles, vai acelerar a pesquisa, diminuir as diferenças entre ricos e pobres, além de unir a humanidade na busca pelo conhecimento.

Argumentam que o acesso aberto encontrava-se, em fevereiro de 2002, limitado a pequenas coleções, mas que, mesmo assim, diferentes iniciativas vêm demonstrando que ele é economicamente viável, dá poder ao leitor para encontrar e fazer uso da literatura que o interessa, além de dar ao autor visibilidade e impacto. E com isso apelam a todos os interessados, pessoas e instituições que ajudem a remover as barreiras e abrir o acesso ao restante da literatura. Para conseguir isso, recomendam duas estratégias complementares: o autoarquivamento e o acesso aberto aos periódicos.

Após a BOAI, duas outras iniciativas ocorreram: A *Bethesda Statement on Open Access Publishing*, em abril de 2003, e A *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, em outubro de 2003. Ambas visavam incitar as discussões sobre acesso aberto entre os pesquisadores em suas áreas, buscando estimulá-los a publicar seus trabalhos seguindo esta filosofia, além de reforçar a necessidade de avaliação da produção disponível em acesso aberto, pretendendo assegurar os padrões de qualidade (SILVA; ALCARÁ, 2008, p. 2).

No Brasil, em maio de 2004, é emitida uma declaração de apoio ao acesso aberto, elaborada pelos participantes do 2º Simpósio Internacional de Bibliotecas Digitais, em Campinas. No ano seguinte, já são três iniciativas que se tornam públicas em apoio ao movimento de acesso aberto: em setembro de 2005, são emitidas a “Declaração de Salvador Compromisso com a Equidade”, pelo acesso equitativo e aberto, e a “Declaração de Salvador sobre Acesso Aberto: Perspectiva do Mundo em Desenvolvimento”, que

² Disponível em: <<http://www.budapestopenaccessinitiative.org/read>>. Acesso em: 27 maio 2016.

pede do governo a exigência do acesso aberto às pesquisas realizadas com dinheiro público, ambas elaboradas pelos participantes do Seminário Internacional Acesso Aberto para Países em Desenvolvimento, evento paralelo ao 9º Congresso Mundial de Informação em Saúde e Bibliotecas: Compromisso com a Equidade. Já em dezembro é emitida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia/IBICT a “Declaração de São Paulo em apoio ao Acesso Aberto” (FAUSTO, 2015).

Essas iniciativas vão influenciar as discussões sobre o direito do cidadão, seja ele pesquisador ou não, a ter acesso ao conhecimento produzido na sociedade em que vive, principalmente quando esse conhecimento é gerado a partir de dinheiro oriundo dos cofres públicos. A partir destas iniciativas, fica visível que as discussões sobre o acesso aberto à produção científica começam a atingir um número cada vez maior de instituições, pesquisadores, governos e agências de fomento. E também passa a ter uma definição mais clara sobre o que acesso aberto realmente significa, pois existem diferenças entre ele e o acesso livre de algumas barreiras.

A comunicação de resultados de pesquisa acadêmica em periódicos arbitrados passou por relevante transformação com o advento da Internet. O modelo dominante de negócios das publicações impressas – onde os leitores pagam para ter acesso aos artigos – passou a dividir espaço com periódicos eletrônicos e novas formas de acesso. Os elevados custos de subscrição de periódicos impulsionaram o movimento de acesso aberto (AA) iniciado nos anos 90, e o surgimento de publicações online livres de barreiras de acesso e da maior parte das barreiras de copyright existentes anteriormente (NASSI-CALÒ, 2015b).

A produção publicada na forma de acesso aberto não desconsidera os direitos autorais, a revisão por pares, a venda e os demais serviços característicos da produção que segue o modo já consagrado de publicar, mas com total ausência de barreiras de acesso e a não cobrança para acessá-la. Já o acesso livre de algumas barreiras pressupõe manter qualquer direito que não vise impedir o acesso.

Para garantir de modo seguro o acesso e o uso da obra, os autores e editoras contam com as licenças *Creative Commons*, pois elas indicam o nível de liberdade do conteúdo, de acordo com os níveis que são colocados. No site da organização³, encontram-se

³ Disponível em: <<https://br.creativecommons.org/>>. Acesso em: 27 maio 2016

disponíveis todas as seis licenças e suas respectivas características, inclusive o significado dos símbolos adicionados ao material licenciado.

Os nossos instrumentos fornecem a todos, desde criadores individuais até grandes empresas, uma forma padronizada de atribuir autorizações de direito de autor e de direitos conexos aos seus trabalhos criativos. Em conjunto, estes instrumentos e os seus utilizadores formam um corpo vasto e em crescimento de bens comuns digitais, um repositório de conteúdos que podem ser copiados, distribuídos, editados, remixados e utilizados para criar outros trabalhos, sempre dentro dos limites da legislação de direito de autor e de direitos conexos (CREATIVE..., s.d.).

Vale chamar a atenção, também, para o fato de que existem pelo menos cinco formas diferentes de publicações que podem ser consideradas de acesso aberto: via dourada – periódicos que fornecem seus conteúdos de forma gratuita; via verde – artigos arquivados em repositórios institucionais, mas que foram aceitos para publicação em periódico acadêmico; híbrida – o periódico comercial permite que o autor pague para que seu artigo fique disponível na forma de acesso aberto; embargo – periódicos assinados que liberam seus artigos após o término do período de embargo; por tempo limitado – os periódicos abrem seus arquivos para o acesso aberto durante um breve período de tempo (SPINAK, 2013).

Como se pode ver, de forma bem geral, as questões que envolvem as discussões sobre acesso aberto são instigantes e ainda se está longe de chegar a um consenso sobre como seria a melhor forma de posicionamento. Mas estas discussões não deixam qualquer dúvida sobre o fato de que o acesso livre ao conhecimento e à informação aumenta a visibilidade da pesquisa realizada, trazendo como consequência uma menor redundância de pesquisas futuras, pois se espera um aumento no número de pesquisadores que terá acesso aos seus resultados. A diminuição da redundância pressupõe o aumento de pesquisas inéditas, que, por sua vez, irá alimentar outras pesquisas que dependem destas para seu desenvolvimento, além de permitir uma melhor aplicação de recursos, principalmente em países que se encontram em estágios muito iniciais do investimento na ciência. E, como falado inicialmente, dentro deste cenário, cada vez mais o acesso aos dados da pesquisa realizada passa a ter um papel significativo com relação ao progresso científico.

3 O ACESSO ABERTO A DADOS DE PESQUISA

O movimento de acesso aberto (*open access*) à literatura científica já é conhecido de todos que, de alguma forma, estão relacionados ao mundo acadêmico e da pesquisa. Mas pouco se ouve falar, aqui no Brasil, e poderíamos até mesmo classificar como um assunto interdito, o acesso aberto aos dados de pesquisa. Mas esse movimento de dados abertos (*open data*), ainda que se tenha visto várias iniciativas na Europa e nos EUA, também não sensibilizou completamente a comunidade acadêmica internacional. Diversos artigos vêm sendo escritos, seminários e workshops realizados, além da criação de repositórios de dados, como forma de mostrar a importância acadêmica, científica, econômica e as diversas possibilidades de desenvolvimento nas mais diferentes áreas a partir do reconhecimento da necessidade de se adotar o paradigma de dados abertos.

O movimento de acesso aberto aos dados de pesquisa tem como objetivo disponibilizar, sem qualquer restrição (copyright, patente) e de forma gratuita, determinados dados gerados durante a pesquisa. Como a maioria das pesquisas são financiadas por agências de fomento públicas, o compartilhamento de dados de pesquisa acabou se tornando um tema importante na agenda política de diferentes países.

Dados de pesquisa podem ser entendidos como um conjunto de informações brutas, primárias, produzido durante o desenvolvimento da pesquisa e que precisa ser analisado e interpretado para a conclusão do trabalho. A *Organisation for Economic Co-operation and Development* (OECD), em sua publicação *OECD Principles and Guidelines for Access to Research Data from Public Funding*, de 2007, define dados de pesquisa como “registros fatuais (numéricos, documentos textuais, imagens e sons) utilizados como fontes primárias para a investigação científica, e que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa” (ORGANISATION..., 2007, p. 13, tradução nossa).

Os dados científicos podem ser classificados a partir de várias perspectivas. Por exemplo, a *National Science Foundation* (NSF) dos Estados Unidos apresenta uma categorização das origens dos dados: dados de observação (registros históricos, como medidas de precipitação, que não podem ser reproduzidos e necessitam de preservação permanente), dados computacionais (resultantes de

simulações, que teoricamente podem ser reproduzidos se for preservada informação sobre o modelo e a sua execução) e dados experimentais (que também não são facilmente reproduzíveis) (RODRIGUES, 2010, p. 11).

Os tipos de dados gerados pela pesquisa se dividem, de acordo com os seus propósitos e processos, em observacionais, experimentais, computacionais, derivados, mas também podem receber outras classificações. E a natureza desses dados é bem ampla, incluindo numérico, multimídia, textual, software, visualização, artefato, processo e simulação, dentre outros. Por sua variedade, compreende-se que esses dados mereçam atenção minuciosa com relação a sua preservação ao longo do tempo.

Para Arano (2011), a atualidade está sendo revolucionada pelo desenvolvimento das tecnologias de comunicação e informação. Os recentes avanços nas formas de digitalização crescente de dados em conjunto com a crescente virtualização dos espaços de trabalho têm mudado tanto a forma de processar, acessar e distribuir estes dados, como também a forma como as próprias pessoas se relacionam. Assim, no âmbito científico e acadêmico, está ocorrendo uma migração dos espaços físicos de trabalho para os virtuais, cujo conceito de *e-science* começa a se consolidar e manifestar a necessidade de gerar infraestruturas científicas que deem suporte adequado para trabalhar e preservar grandes volumes de dados em forma virtual e colaborativa, as chamadas *e-infrastructure*.

A implementação de uma infraestrutura científica requer um longo processo de planejamento e desenvolvimento, onde são necessários muitos recursos humanos e financeiros. Os benefícios mais evidentes são para a comunidade acadêmica e científica, já que permite a diferentes investigadores encontrar e reunir grandes quantidades de dados, trabalhar com dados externos resultantes de outras pesquisas, gerando novos conhecimentos a partir deles, e, também, facilitar o trabalho em um ambiente de colaboração, preservando a integridade e autoria de dados. Desta forma, os dados são usados, reutilizados e combinados, aumentando a produtividade e a capacidade de correlação em uma escala nunca antes vista. Também existem outros grupos beneficiados, tais como provedores de dados, organismos públicos de financiamento, o setor de inovação empresarial e industrial, o setor político e o público em geral. Todos eles têm a possibilidade, através do uso da infraestrutura científica e com diferentes objetivos, de acessar, depositar e utilizar os dados publicados em uma infraestrutura

científica (ARANO et al, 2011, p. 386, *tradução nossa*)⁴.

O compartilhamento de dados de pesquisa é um tema que vem não apenas incitando o debate no campo do acesso aberto como também se consolidando dentro de todas as esferas que envolvem a pesquisa científica.

O compartilhamento de dados para reutilização ou reprodução de experimentos exige conhecer sua origem e entender como foram produzidos, associando à informação métodos, algoritmos ou técnicas adotados, e ainda ter acesso ao software necessário para processá-los, o que torna o processo bastante complexo. Sem isso, pode não ser possível reproduzir o experimento original ou reutilizar o dado em uma outra pesquisa (MARQUES, 2014, p. 56).

O objetivo primordial do acesso aberto aos dados de pesquisa já realizados é colaborar para o desenvolvimento da pesquisa científica em escala mundial, pois o acesso a eles contribuiria para dinamizar novas pesquisas a partir da eliminação de etapas que já tenham sido realizadas por outros pesquisadores e que se tornaram conhecidas com a publicação dos dados coletados. O impacto também ocorre na economia de recursos com a diminuição de pesquisas duplicadas e com o direcionamento destes recursos para novas pesquisas.

A disponibilização de resultados de pesquisa em geral de forma aberta é parte do movimento *Open Data*, cujo objetivo principal é aumentar a transparência, reprodutibilidade, disponibilidade e interoperabilidade de dados básicos de pesquisa. Nos últimos anos, instituições de pesquisa, agências de fomento, o setor privado, governos e a sociedade civil vem debatendo as formas mais

⁴ La implementación de una infraestructura científica requiere de un largo proceso de planificación y desarrollo, donde se necesitan gran cantidad de recursos humanos y económicos. Los beneficios más evidentes son para la comunidad académica y científica, ya que permite a diferentes investigadores encontrar y reunir gran cantidad de datos, trabajar sobre datos ajenos resultantes de otra investigación, generar nuevo conocimiento a partir de ellos y, además, facilitar el trabajo en un entorno colaborativo, preservando la integridad y autoría de los datos. De esta manera los datos se usan, se reutilizan y se combinan, incrementando la productividad y la capacidad de correlación a una escala nunca vista con anterioridad. Asimismo, existen otros colectivos beneficiados tales como los proveedores de datos, las entidades públicas de financiación, el sector de innovación empresarial e industrial, el sector político y el público en general. Todos ellos tienen la posibilidad, a través del uso de la infraestructura científica y con diferentes objetivos, de acceder, depositar y utilizar los datos publicados en una infraestructura científica.

eficientes de operacionalizar repositórios capazes de armazenar os dados no formato legível por computadores (*machine readable*) e grandes avanços tem sido feitos desde então (NASSI-CALÓ, 2015a).

Segundo Spinak (2015), o movimento de dados abertos propõe encontrar a melhor forma de se preservar e trocar os dados resultados de experimentos; dar visibilidade à produção científica mundial; aumentar o retorno do investimento feito na pesquisa ao incentivar a troca e o reuso dos dados primários; gerar políticas sistemáticas focadas na competitividade da pesquisa baseada na gestão de dados.

Em 2014, o *Nature Publication Group* lança o primeiro periódico de acesso aberto a dados de pesquisa, o *Scientific Data*, cujo foco é a descrições de conjuntos de dados de pesquisa, com o objetivo de promover a partilha e o reuso de dados de forma mais ampla. É um periódico revisado por pares e os autores pagam uma taxa de publicação.

Scientific Data é um novo periódico de acesso aberto, *online*, para descrição de conjuntos de dados cientificamente valiosos. Nossos artigos, conhecidos como Data Descriptors, combinam conteúdo narrativo tradicional com curadoria, descrições estruturadas (metadados) dos dados publicados, para fornecer uma nova estrutura para a partilha de dados e reuso que, acreditamos, irá acelerar o ritmo da descoberta científica (NATURE, 2014, *tradução nossa*).⁵

O periódico se fundamenta em seis princípios: **crédito** – e reconhecimento dado aos cientistas que investem em elaborar conjuntos de dados valiosos disponíveis e reutilizáveis; **reuso** – padronizações e descrições detalhadas tornam os dados de pesquisa mais fáceis de encontrar e reutilizar; **qualidade** – avaliação crítica para verificar o rigor experimental e a totalidade de sua descrição; **descoberta** – facilidade em localizar os conjuntos de dados que são relevantes para suas pesquisas; **aberto** – os cientistas trabalham melhor quando eles podem facilmente se conectar e colaborar com seus pares; **serviço** – o periódico se compromete a fornecer um serviço de excelência tanto para autores quanto para leitores (NATURE, 2014, *tradução nossa*).

⁵ Scientific Data is a new open-access, online-only journal for descriptions of scientifically valuable datasets. Our articles, known as Data Descriptors, combine traditional narrative content with curated, structured descriptions (metadata) of the published data to provide a new framework for data-sharing and -reuse that we believe will ultimately accelerate the pace of scientific discovery.

A criação deste periódico pode ser vista como uma demonstração da tomada de consciência dos diversos setores que dependem do avanço científico, não apenas como a disponibilização dos dados oriundos de pesquisas científicas, mas também como uma preocupação crescente em como preservá-los pelo maior tempo possível.

Uma pesquisa realizada por pesquisadores do Canadá avaliou a preservação de dados pelos autores dos artigos a que deram origem, publicados entre dois e 22 anos atrás. Os resultados indicam que a perda de dados é maior quanto mais antiga é a publicação. A probabilidade de que os dados de um artigo tenham sido conservados pelo autor cai de um fator de 17% ao ano. Somase a isso a dificuldade de localizar os autores, uma vez que publicações mais antigas não incluem endereços eletrônicos ou estão desatualizados. A probabilidade de contatar um autor cai na razão de 7% ao ano. Desta forma, estima-se que 80% dos dados não deverão estar disponíveis decorridos 20 anos de sua geração (SCIENTIFIC ELECTRONIC LIBRARY ONLINE, 2014).

A criação de repositórios de dados é um dos caminhos para a preservação de dados gerados pela pesquisa científica. Os pesquisadores vêm sendo, cada vez mais, solicitados a arquivarem seus dados brutos, ou dados primários, em repositórios de dados de acesso público:

O armazenamento de dados científicos em repositórios e sua reutilização é uma das preocupações do recém-lançado Programa FAPESP de Pesquisa em *eScience*, expressão que resume o desafio de pesquisa para organizar, classificar e garantir acesso ao gigantesco volume de dados gerados continuamente em todos os campos de pesquisa, a fim de extrair novos conhecimentos e fazer análises abrangentes e originais (MARQUES, 2014, p. 56).

Um repositório, no sentido mais lato da palavra, é um local de armazenamento, de guarda e arquivamento de objetos. Atualmente, quando falamos de repositórios, entendemos que estes objetos sejam objetos digitais, definidos como “um item armazenado em uma biblioteca digital, que consiste em dados, metadados e um identificador” (ARMS, 2000, *tradução nossa*).

No contexto do acesso aberto, surgiram os repositórios institucionais (RI), que são “ferramentas de disseminação da informação técnico-científica que permitem o armazenamento, organização, recuperação e disseminação de documentos

acadêmicos, administrativos e científicos de uma instituição de forma integrada” (NUNES, 2012, p. 11). Um RI se diferencia de outros ambientes (base de dados, bibliotecas digitais e repositórios temáticos) por ser dedicado à produção da própria instituição (documentos criados pela instituição e seus membros). Já os repositórios digitais (RD):

São bases de dados online que reúnem de maneira organizada a produção científica de uma instituição ou área temática. Os RDs armazenam arquivos de diversos formatos. Ainda, resultam em uma série de benefícios tanto para os pesquisadores quanto às instituições ou sociedades científicas, proporcionam maior visibilidade aos resultados de pesquisas e possibilitam a preservação da memória científica de sua instituição. Os RDs podem ser institucionais ou temáticos. Os repositórios institucionais lidam com a produção científica de uma determinada instituição. Os repositórios temáticos com a produção científica de uma determinada área, sem limites institucionais (INSTITUTO..., 2012).

Lynch (2003) define um RI universitário como um conjunto de serviços que a universidade oferece aos membros de sua comunidade para a gestão e disseminação de materiais digitais criados pela própria instituição. A Instituição teria um compromisso organizacional para a gestão desses materiais, incluindo sua preservação em longo prazo, bem como a organização e acesso ou distribuição. Para ser eficaz, necessita que exista uma colaboração entre bibliotecários, tecnólogos da informação, arquivistas, professores e administradores universitários. Esta definição se estende para repositórios institucionais de maneira geral, independente de estarem vinculados a uma universidade.

O princípio da reprodutibilidade das pesquisas é o dinamismo mais importante para a criação dos repositórios de dados de pesquisa. Uma quantidade não desprezível de descobertas científicas acaba não sendo confirmada após sua publicação, por problemas que incluem erros e fraudes, mas que também se estendem a falsos resultados positivos ou negativos obtidos de boa fé. O problema assombra pesquisadores e revistas científicas, obrigados a cancelar a publicação de trabalhos cujos resultados soavam promissores, e tornou-se um pesadelo para empresas farmacêuticas e de biotecnologia (MARQUES, 2014, p. 57).

No desenvolvimento deste artigo, o repositório será classificado de acordo com o material que nele é depositado, no caso, repositório de dados científicos, ou repositório de dados de

pesquisa.

Este tipo de repositório é mais raro no contexto brasileiro. As iniciativas existentes de repositórios desse tipo no país surgem no âmbito de consórcios internacionais como é o caso do Projeto Genoma Humano ou como inovação proposta pelo usuário, isto é, os próprios pesquisadores, que em alguns casos isolados, sentem a necessidade de organizar e compartilhar seus dados com pares e alunos e começam a organizar banco de dados para tratar seus dados, como por exemplo, na área ambiental e de geociências (SALES, 2014, p. 62).

Por serem repositórios com conteúdos que possuem características próprias e necessitam de um tratamento específico, incluindo também conjuntos específicos de metadados para organizar estes conjuntos de dados, acabam por se diferenciarem dos repositórios digitais de maneira geral.

Sendo assim, na prática, a escolha da tecnologia a ser aplicada no desenvolvimento do repositório de dados digitais está relacionada à forma como se configurou o processo de curadoria dos dados, à que tipo de dado será armazenado, ao processo de pesquisa (*workflow*) que resultará nos dados e às possibilidades de reuso (SALES, 2014, p. 63).

Para que os dados científicos cumpram tudo aquilo que é esperado deles, é preciso que estejam organizados e estruturados. Os atributos a que os conjuntos de dados (*datasets*) devem obedecer para que possam ser compartilhados e reutilizados pelos pesquisadores se relacionam à forma como se processa a curadoria e às condições normativas de acesso.

O compartilhamento de dados de pesquisa se constitui em uma encruzilhada em que o pesquisador precisa contrabalançar suas expectativas e as questões éticas e legais relacionadas ao compartilhamento de dados chamados sensíveis:

Os dados sensíveis são dados que podem ser usados para identificar um indivíduo, espécie, objeto, processo ou local que introduz um risco de discriminação, dano ou atenção indesejada. Perante a lei e a administração ética em pesquisa da maioria das instituições, os dados sensíveis não podem normalmente ser compartilhadas nessa forma, com poucas exceções

(AUSTRALIAN..., 2016, p. 7, *tradução nossa*).⁶

Como já dito, o acesso aberto a dados de pesquisa pressupõe que determinados conjuntos de dados devam ser compartilhados de forma gratuita e sem qualquer restrição. Mas como proceder com relação a conjuntos de dados que não podem ser compartilhados? Como preservar e fazer a curadoria destes dados? Estas questões não se relacionam a dados protegidos por direitos autorais, pois o acesso a esses dados já está definido a partir dos direitos que recaem sobre eles. Elas estão relacionadas aos dados que foram produzidos dentro do contexto de compartilhamento, mas que, ainda assim, deveriam ter acesso restrito ou embargo.

Contudo, é preciso **atentar para as restrições** que determinam se um conjunto de dados pode ser disponibilizado abertamente – parcial ou completamente – para compartilhamento com outros pesquisadores, posto que nem todos os dados gerados por pesquisa podem ser livremente distribuídos. Para dados de natureza sensível, por exemplo, que envolvem seres humanos, processos patenteáveis, espécies e ambientes ecológicos em risco, etc., em que o acesso público não é apropriado, você deve indicar que medidas especiais devem ser tomadas para acesso aos dados (SAYÃO, 2015, p. 21).

Diversas formas são possíveis para se impedir que estes dados sejam acessados, ou tenham um acesso restrito, ainda que eles tenham sido depositados em um repositório de dados. Uma das primeiras medidas a ser tomadas parte do próprio pesquisador, ainda durante o processo da pesquisa, que é não gravar esses dados em local que tenha acesso à internet, e, posteriormente, também depositá-los em um repositório que siga este mesmo procedimento. Mas se o objetivo é que esses dados sejam publicados, outro caminho é modificá-los para que fiquem anônimos (anonimiza-los). E até mesmo publicá-los com acesso restrito, devendo ser especificadas as restrições colocadas.

4 O ACESSO ABERTO A DADOS NA EUROPA EM NÚMEROS

⁶ Sensitive data are data that can be used to identify an individual, species, object, process, or location that introduces a risk of discrimination, harm, or unwanted attention. Under law and the research ethics governance of most institutions, sensitive data cannot typically be shared in this form, with few exceptions.

A partir do entendimento de que a criação de repositórios de dados é um dos caminhos para a gestão e preservação dos dados gerados pelas pesquisas e, posteriormente, para seu compartilhamento através do acesso aberto, realizou-se um levantamento sobre a situação dos países membros da Comissão Europeia em relação tanto à criação de repositórios quanto ao seu funcionamento em acesso aberto, com o intuito de conhecer como estes países vêm se posicionando frente ao novo paradigma que surge.

Além disso, é interessante observar como estes países se posicionam frente ao projeto intitulado *Horizon 2020*, o maior programa europeu de financiamento à pesquisa e à inovação disponível ao longo de sete anos (2014-2020), que conta com financiamento público e busca atrair financiamento privado, objetivando levar um maior número de ideias do laboratório para o mercado⁷. A aplicação do programa visa impulsionar o crescimento econômico e gerar empregos, e conta com o apoio político dos líderes europeus, que concordam que a pesquisa científica é um investimento no futuro.

A publicação da Comissão Europeia *Fact sheet: Open Access in Horizon 2020*, de 9 de dezembro de 2013, define o acesso aberto como a prática de fornecer acesso *on-line* à informação científica de forma gratuita ao usuário final, sendo que informação científica pode se referir tanto a artigos de revistas científicas revisadas pelos pares como também aos dados resultantes da pesquisa científica.

Em 30 de outubro de 2015, foi publicado o *Guidelines on Data Management in Horizon 2020*, pois o projeto prevê uma ação piloto (*Open Research Data Pilot*) para o acesso aberto aos dados de pesquisa. Assim, o documento pretende fornecer diretrizes para o gerenciamento desses dados e se dirige àqueles que sejam beneficiados com investimentos do projeto, indicando como eles podem cumprir sua parte em relação à qualidade, compartilhamento e segurança dos dados gerados. Essa ação visa melhorar e ampliar o acesso e o reuso desses dados. O *Open Research Data Pilot* aplica-se a dois tipos de dados:

- Os dados de pesquisa, incluindo os metadados associados, necessários para validar os resultados apresentados em publicações científicas;

⁷ Disponível em: <<https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>>. Acesso em: 27 maio 2016

- Outros dados curados e/ou crus, incluindo os metadados associados, conforme especificado no plano de gestão de dados/DMP (EUROPEAN..., 2016).

Através de uma pesquisa *on-line* em novembro de 2015, chegamos até a página “*Overview of Open Access in the EU member states*”, disponível no site do OpenAIRE⁸. Esta página disponibiliza um resumo sobre o acesso aberto e os repositórios existentes em cada um dos países membros da UE, e entendemos que essas informações foram coletadas a partir do projeto DRIVER⁹, que tinha como objetivo, em sua fase inicial, constituir uma infraestrutura coesa de repositórios digitais de acesso livre. O projeto ajudou a estabelecer e desenvolver repositórios em cada um dos países da Europa, estimulou o *open access* e promoveu o desenvolvimento de políticas nacionais sobre o tema.

De acordo com o levantamento realizado, dos trinta e três países membros da UE, apenas três não possuem repositório de acesso aberto, seja para publicações ou para dados de pesquisa, sendo a França e o Reino Unido os países que passam a casa do milhão em números de publicações de acesso aberto. Abaixo, apresentamos uma tabela disponível no site com o número total de repositórios informados por cada país até o momento da pesquisa.

QUADRO 1 – European open access repositories landscape

⁸ Disponível em: <<https://www.openaire.eu/member-states-overview>>. Acesso em: 27 maio 2016

⁹ Disponível em: <<http://www.driver-repository.eu/Driver-About/About-DRIVER.html>>. Acesso em: 27 maio 2016

numbers below come from the OpenAIRE aggregated space (updated weekly).

Country	#repositories	# OA publications
Austria	5	44.771
Belgium	12	223.636
Bulgaria	2	2.167
Croatia	3	131.836
Cyprus	2	1.465
Czech Republic	7	128.336
Denmark	10	88.296
Estonia	1	9.101
Finland	4	143.355
France	24	1.077.070
Germany	90	671.846
Greece	5	20.964
Hungary	5	6.407
Iceland	2	21.970
Ireland	11	75.361
Italy	29	162.474
Latvia	2	1.772
Lithuania	1	25.247
Luxembourg	1	5.352
Malta	0	0
Netherlands	22	662.941
Norway	4	29.272
Poland	11	34.300
Portugal	36	195.965
Romania	0	0
Serbia	3	4.189
Slovakia	0	0
Slovenia	5	50.781
Spain	51	766.484
Sweden	18	155.043
Switzerland	13	207.931
Turkey	10	26.531
United Kingdom	53	3.831.686

Adaptado de: <<https://www.openaire.eu/member-states-overview>>. Acesso em: 27 maio 2016.

No Quadro 1, a relação de repositórios por país é encontrada a recomendação de que, ao clicar diretamente sobre o nome do país, seremos remetidos a informações referentes ao ambiente da pesquisa nacional, aos repositórios de acesso aberto e

às informações relacionadas e aos detalhes de contatos. Assim, o procedimento adotado foi acessar o link de cada país para fazer uma revisão sistemática sobre a existência e a quantidade de repositório de dados de pesquisa, como descrevem seu funcionamento e identificando as principais práticas de gestão de dados de pesquisa em cada país.

Da lista de países apresentados na tabela acima, apenas nove dentre os trinta e três declararam possuir repositórios e alguma política relacionada ao acesso aberto a dados de pesquisa. Por se basear nas declarações fornecidas pelos próprios países, que estão disponíveis quando acessamos o link de cada um na página <https://www.openaire.eu/member-states-overview>, é possível que todos os demais países apresentados não possuam nenhum repositório ou política mandatória, pois não foi possível identificar nenhuma informação relacionada ao tema nas páginas visitadas. Ou, por algum motivo, essas informações foram omitidas. Para confirmar as informações fornecidas, e até mesmo identificar outras, foi realizada uma pesquisa na página da re3data.org¹⁰.

A Áustria possui cinco repositórios de dados, mas informa que tem um projeto para criação de mais repositório que se iniciou em 2014, cujas informações se encontram em <http://e-infrastructures.at/en/startpage/>. O link nos remete à página *e-Infrastructures Austria!*, um projeto que coordena o estabelecimento e o desenvolvimento de infraestrutura de repositórios de recursos digitais na pesquisa e na ciência em toda a Áustria. Tem como um de seus subprojetos a construção de infraestrutura de repositórios para dados de pesquisa e outros objetos complexos. O projeto se iniciou em janeiro de 2014 e finda em dezembro de 2016, sendo coordenado pela Universidade de Viena com mais vinte e cinco instituições parceiras, tendo como cliente o *Federal Ministry of Science, Research and Economics BMWFV*. Segundo os dados disponíveis no re3data.org, existem vinte e três repositórios registrados.

A Bélgica possui quatro repositórios de dados de pesquisa atualmente, sendo três deles sustentados pelo *Flanders Marine Institute VLIZ: Open Marine Archive (OMA)*; *Ocean Biographic Information System*; *World Register of Marine Species (WoRMS)*; *Institut d'Astronomie et d'Astrophysiqu*. Mas, além desses, *Registry of Research Data Repositories* enumera mais vinte e dois em que as instituições belgas estão envolvidas. Os governos locais belgas e o governo federal são a favor dos dados abertos, e criaram mandatos

¹⁰ Disponível em: <<http://service.re3data.org/about>>. Acesso: 5 jun. 2016

que garantem o livre acesso aos dados governamentais. Mas existem vários outros projetos e iniciativas de apoio, execução, além de outras iniciativas que fornecem o suporte técnico. Segundo os dados disponíveis no re3data.org, existem vinte e oito repositórios registrados.

A Croácia possui três repositórios e participa do projeto *SERSCIDA – Support for Establishment of National/Regional Social Sciences Data*, através da *Faculty of Humanities and Social Sciences*, que é um projeto estratégico para cooperação e troca de conhecimentos entre os países membros da UE associados ao *Council of European Social Sciences Data Archives (CESSDA)* e ao *Western Balkan Countries (WBC)*, para arquivamento de dados em ciências sociais. Segundo os dados disponíveis no re3data.org, não existem repositórios registrados no país.

A República Tcheca relaciona tres repositórios de dados de pesquisa: *CKAN Czech Republic* (o link para esta página não é válido); *Czech Social Science Data Archive, Institute of Sociology of the Academy of Sciences of the Czech Republic*; *LINDAT-Clarín Repository, Centre for Language Research Infrastructure in the Czech Republic, Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University in Prague*. Segundo os dados disponíveis no re3data.org, existem sete repositórios registrados.

Mesmo que ainda não se tenha um nível pleno de consciência dos benefícios do acesso aberto para os pesquisadores na Itália, várias ações e políticas estão sendo implementadas nesse sentido, mas a consciência sobre o acesso aberto a dados de pesquisa ainda está em fase inicial. Ainda assim, o Ministério da Educação da Itália criou, em 2014, uma política mandatária de acesso aberto para publicações e conjuntos de dados originados do programa de financiamento para jovens cientistas, estabelecendo um período para que esse material seja depositado nos respectivos repositórios. No site do re3data.org, identificamos o registro de vinte e sete repositórios de dados de pesquisa no país.

A Lituânia declara que alguns repositórios de dados se encontram em fase de desenvolvimento e informa que tem iniciativas como o *National Open Access Research Data Archive MIDAS*, projeto que tem como objetivo estabelecer a infraestrutura para o arquivamento de dados de pesquisas nacionais e garantir o acesso a eles; e o *Lithuanian Data Archive for HSS/LIDA*, um serviço de dados em Ciências Sociais que permite a busca, navegação, análise e download de dados de pesquisa na área. Segundo os dados disponíveis no re3data.org, existem quatro repositórios registrados.

Em meados de 2008, em Portugal, foi criado o Repositório

Científico de Acesso Aberto de Portugal/RCAAP. A iniciativa teve como objetivo aumentar a visibilidade, acessibilidade e divulgação dos resultados da pesquisa científica no país. Em 2010, o RCAAP produziu o relatório *Open data Repositories: State of the Art*, que marca o início de suas intervenções no domínio da curadoria de dados de pesquisa, sua organização em repositórios e de seu acesso. O documento tinha o objetivo, entre outros, de informar e orientar sobre a criação de um projeto piloto de repositório de dados científicos, que se encontra em fase de desenvolvimento. Segundo os dados disponíveis no re3data.org, existem dois repositórios registrados.

Na Sérvia, a importância da preservação e reutilização de dados brutos de pesquisa é reconhecida por três grupos, que adquiriram suas experiências através da participação nos seguintes projetos: *Support for Establishment of National/Regional Social Sciences Data, SERCIDA; Comprehensive Knowledge Archive Network, Serbian Public Data in the LOD cloud; Spatial data sets for Serbia*. São listados os oito repositórios de acesso aberto existentes, com uma *url* válida para cada um deles. Dos oito repositórios, foi possível identificar o *Digital Repository of the Institute for Philosophy and Social Theory*, da Universidade de Belgrado, como um repositório de dados de pesquisa. Segundo os dados disponíveis no re3data.org, não existem repositórios registrados no país.

Na Suíça, em 2008, se realizou um estudo nacional do qual se originou a publicação intitulada "*Development of a model for a central long-term preservation of digital primary and secondary research data*", além de um projeto centralizado para as universidades, centros de pesquisa e bibliotecas universitárias suíças de gerenciamento e preservação em longo prazo dos dados de pesquisa primários e secundários. Devido a divergências, algumas instituições seguiram com projetos individuais com relação a gestão desses dados. Mas a partir de 2010, com um novo programa de financiamento, começou a se desenvolver um projeto que combina os esforços isolados destas instituições. Segundo os dados disponíveis no re3data.org, existem quarenta e seis repositórios registrados.

No Quadro 1 foram listados os trinta e três países membros da UE e o número de repositórios que cada um possui, a partir dessa listagem foi realizada uma pesquisa na página da re3data.org para buscar, dentre os que não declararam ter, ou estar criando, repositórios de dados de pesquisa, o número de repositórios de acesso aberto que eles têm registrados:

QUADRO 2 – Número de repositórios de acesso aberto registrados

no re3data por país

País	Nº. de repositórios	País	Nº. de repositórios
Bulgária	0	Luxemburgo	5
Cyprus	1	Malta	0
Dinamarca	18	Países Baixos	37
Estônia	5	Noruega	15
Finlândia	7	Polônia	4
França	78	Romênia	2
Alemanha	249	Eslovaquia	1
Grécia	9	Eslovenia	3
Hungria	5	Espanha	14
Islândia	1	Suécia	17
Irlanda	8	Turquia	1
Letônia	0	Reino Unido	227

FONTE: <<http://service.re3data.org/search>>. Acesso: 5 jun. 2016

CONSIDERAÇÕES FINAIS

Como visto, mesmo que a Europa esteja muito adiantada nas discussões sobre compartilhamento de dados de pesquisa, e ainda que alguns países já tenham seus repositórios criados, é possível perceber, a partir dos números apresentados nas tabelas acima, que muitos destes países se encontram como no Brasil, ou seja, ainda em fase inicial de implementação de estruturas adequadas para o compartilhamento de dados.

Do lado das comunidades científicas, está a crescer a sensibilidade para as consequências da verdadeira explosão na produção de dados científicos, quer pelo crescimento global das atividades de investigação, quer como consequência dos novos métodos e instrumentos de pesquisa e registo que originam cada vez maiores volumes de dados. Para além da produção, recolha, análise e interpretação dos dados, onde naturalmente os investigadores concentram o seu trabalho, tem vindo a afirmar-se uma outra área de intervenção, relacionada com a curadoria dos dados científicos (RODRIGUES, 2010, p. 8).

No entanto, os países que já possuem seus repositórios de dados e suas políticas de gestão de dados criadas podem contribuir consideravelmente com o compartilhamento de suas experiências, como, por exemplo, o Reino Unido, através das práticas bem-sucedidas gerenciadas pelo *Digital Curation Centre*, em Edimburgo – Escócia, que vem desenvolvendo guias, manuais, políticas, padrões de metadados para dados, planos de gestão de dados, modelos de curadorias, entre outros produtos e serviços.

Nas discussões sobre qual caminho o Brasil deve seguir em

relação às políticas a serem implementadas, as normas a serem adotadas e as decisões a serem tomadas há sempre questionamentos se ele deve acompanhar os rumos tomados pela UE ou os caminhos adotados pelos norte-americanos. Apesar da proximidade continental com a América do Norte, tradicionalmente a Ciência da Informação brasileira tende a se posicionar do lado Europeu, principalmente no que tange ao movimento do acesso aberto. Nesse caso, possivelmente, pela proximidade linguística com Portugal – país que vem avançando fortemente nos estudos teóricos e empíricos que envolvem o presente tema.

Sendo assim, o presente estudo veio oferecer um panorama da situação europeia atual a fim de subsidiar a comunidade brasileira de argumentos nas discussões da escolha sobre qual caminho seguir. Faz-se necessário, ainda, um estudo similar sobre o panorama norte-americano para que seja possível escolher um caminho alternativo, quem sabe mesclar ambos, adotando as iniciativas de sucesso em cada um deles.

REFERÊNCIAS

ARMS, W. *From the 1999 manuscript of Digital Libraries*. s.l.: M.I.T. Press, 2000. Disponível em: <<http://www.cs.cornell.edu/wya/DigLib/MS1999/Glossary.html>>. Acesso em: 13 jan. 2016.

ARANO, S. et al. La comunidad «Recursos y datos primarios» de la Universitat Pompeu Fabra: los repositorios institucionales como infraestructuras científicas: estudio de caso. *Revista Española de Documentación Científica*, v. 34, n. 3, p. 385-407, jul./set., 2011.

AUSTRALIAN NATIONAL DATA SERVICE. *Publishing and sharing sensitive data*. ANDS Guides, Updated 16 May 2016. Disponível em: <http://www.ands.org.au/__data/assets/pdf_file/0010/489187/Sensitive-Data-Guide-2016.pdf>. Acesso em: 20 maio 2016.

COSTA, S. Abordagens, estratégias e ferramentas para o acesso aberto via periódicos e repositórios institucionais em instituições acadêmicas brasileiras. *Liinc em Revista*, v.4, n.2, p. 218-232, set. 2008.

CREATIVE COMMONS CORPORATION (Org.). *Sobre as licenças*. Disponível em: <<https://br.creativecommons.org/licencas/>>. Acesso em: 20 dez. 2015.

EUROPEAN COMMISSION. Directoral-general for research and innovation. *Guidelines on data management on horizon 2020*, 15 fev. 2016. Disponível em: <http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/o_a_pilot/h2020-hi-oa-data-mgt_en.pdf> Acesso em: 27 maio 2016.

- FAUSTO, S. Evolução do Acesso Aberto: breve histórico. *SciELO em Perspectiva*, [Internet], 2013. Disponível em: <<http://blog.scielo.org/blog/2013/10/21/evolucao-do-acesso-aberto-breve-historico/>>. Acesso em: 30 nov. 2015.
- FREITAS, M. H. Considerações acerca dos primeiros periódicos científicos brasileiros. *Ciência da Informação*, Brasília, DF, v. 35, n. 3, p. 54-66, set./dez. 2006.
- INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). *Sobre Repositórios Digitais*, [Internet], 2012. Disponível em: <<http://www.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais>>. Acesso em: 15 dez. 2015.
- LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL Bimonthly Report*, v. 26, 2003. Disponível em: <https://muse.jhu.edu/login?auth=0&type=summary&url=/journals/portals_and_the_academy/v003/3.2lynch.html>. Acesso em: 12 jan. 2016.
- MARQUES, F. Ciência transparente. *Revista FAPESP*, abril, 2014.
- MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. *Ciência da Informação*, Brasília, DF, v. 35, n. 2, p. 27-38, ago. 2006.
- NASSI-CALÔ, L. The BMJ exige compartilhamento de dados para publicar ensaios clínicos. *SciELO em Perspectiva*, São Paulo, 23 set. 2015a. Disponível em: <<http://blog.scielo.org/blog/2015/09/23/the-bmj-exige-compartilhamento-de-dados-para-publicar-ensaios-clinicos/>>. Acesso em: 30 nov. 2015.
- _____. Acesso Aberto na América Latina livre de periódicos predatórios. *SciELO em Perspectiva*, São Paulo, 13 nov. 2015b. Disponível em: <<http://blog.scielo.org/blog/2015/11/13/periodicos-predatorios-o-lado-obscurito-do-acesso-aberto/>>. Acesso em: 19 jan. 2016.
- NATURE PUBLICATION GROUP. Scientific Data, [Internet], 2014. Disponível em: <<http://www.nature.com/sdata/about/principles>> Acesso em: 20 maio 2016.
- NUNES, R. R. *Diretrizes para formulação de políticas mandatórias para consolidação dos repositórios institucionais brasileiros*. Dissertação (Mestrado) – Universidade Federal Fluminense, Niterói, 2012.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). Principles and guidelines for access to research data from public funding. OECD, 2007.
- RODRIGUES, E. et al. Os repositórios de dados científicos: estado da arte. *Relatório D-24: RCAAP*, 2010. Disponível em: <http://projeto.rcaap.pt/index.php?option=com_remository&Itemid=2&func= tartdown&id=271&lang=pt>. Acesso em: 25 dez. 2015.
- SALES, L. F.. *Integração semântica de publicações científicas e dados de*

pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares. Rio de Janeiro, 2014. Tese (Doutorado em Ciência da Informação). Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

SAYÃO, L. F.; SALES, L. F. *Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores*. Rio de Janeiro: CNEN/IEN, 2015.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE. Movimento Open Data se consolida internacionalmente. *SciELO em Perspectiva*, São Paulo, 14 jul. 2014. Disponível em: <<http://blog.scielo.org/blog/2014/07/14/movimento-open-data-se-consolida-internacionalmente/>>. Acesso em: 30 nov. 2015.

SILVA, T. E.; ALCARÁ, A. R. Políticas de acesso aberto à informação científica: iniciativas governamentais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO. 9., 2008, São Paulo. *Anais eletrônicos...* São Paulo: USP, 2008. Disponível em: <<http://www.ancib.org.br> > Acesso em: 10 dez. 2015.

SPINAK, E. Os artigos em acesso aberto chegaram para ficar: em menos de 10 anos aproximam de 50% do nível mundial. *SciELO em Perspectiva*. São Paulo, 28 ago. 2013. Disponível em: <<http://blog.scielo.org/blog/2013/08/28/os-artigos-em-acesso-aberto-chegaram-para-ficar-em-menos-de-10-anos-aproximam-de-50-do-nivel-mundial/>>. Acesso em: 19 jan. 2016.

_____. Semana Internacional dos Dados Abertos – o que há de novo?. *SciELO em Perspectiva*. São Paulo, 7 jan. 2015. Disponível em: <<http://blog.scielo.org/blog/2015/01/07/semana-internacional-dos-dados-abertos-o-que-ha-de-novo/>>. Acesso em: 30 nov. 2015.

STUMPF, I. R. C. Passado e futuro das revistas científicas. *Ciência da Informação*, Brasília, DF, v. 25, n. 3, 1996.