



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz



Instituto de Comunicação e Informação
Científica e Tecnológica em Saúde

Especialização em Informação Científica e Tecnológica em Saúde

**Metadados e a ampliação do acesso aos dados
de pesquisa nos Repositórios digitais do campo da
saúde.**

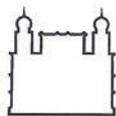
Por

ANDERSON SILVA DE ARAUJO

Projeto apresentado ao Instituto de Comunicação e Informação Científica e tecnológica em Saúde da Fundação Oswaldo Cruz como requisito parcial para obtenção do título de Especialista em Informação Científica e Tecnológica em Saúde

Orientadora:

Dr.^a Viviane Santos de Oliveira Veiga



Ministério da Saúde
FIOCRUZ
Fundação Oswaldo Cruz



CURSO DE ESPECIALIZAÇÃO EM INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

Metadados e a ampliação do acesso aos dados de pesquisa nos Repositórios digitais do campo da saúde.

Por

Anderson Silva de Araujo

Projeto apresentado ao Instituto de Comunicação e Informação Científica e Tecnológica em Saúde da Fundação Oswaldo Cruz como requisito parcial para obtenção do título de Especialista em Informação Científica e Tecnológica em Saúde.

Orientadora:

Dr.^a Viviane Santos de Oliveira Veiga

Rio de Janeiro, novembro de 2019

RESUMO

Este projeto visa propor orientações metodológicas para o uso de metadados na descrição de dados de pesquisa nos repositórios institucionais e temáticos do campo da saúde. Como proposta metodológica será realizado um levantamento bibliográfico na literatura especializada nacional e estrangeira; um Levantamento dos Repositórios Institucionais e Temáticos de dados de pesquisa; e um mapeamento dos padrões de metadados para descrição de dados de pesquisa. Apresenta alguns levantamentos preliminares que mostram a necessidade da área da informação apoiar o campo da saúde no uso de padrões de metadados. Espera-se que identificados os repositórios utilizados na área e seus metadados e os padrões de metadados adequados para descrever os dados de pesquisa do campo da saúde em seus diversos subdomínios possa-se propor orientações metodológicas para o uso de metadados que apoiem a ampliação do acesso aos dados de pesquisa em saúde no Brasil.

Palavras-chave: dados de pesquisa; repositório institucional; repositório temático; metadados.

SUMÁRIO

Sumário

1	INTRODUÇÃO	4
2	JUSTIFICATIVA	5
3	REFERENCIAL TEÓRICO	6
3.1	Os Repositórios digitais	6
3.2	Dados de pesquisa.....	7
3.3	Gestão de dados	10
3.4	Metadados	12
4	OBJETIVOS	15
4.1	OBJETIVO GERAL.....	15
4.2	OBJETIVOS ESPECÍFICOS	15
5	METODOLOGIA.....	16
6	RESULTADOS ESPERADOS	26
7	CRONOGRAMA	27
8	ORÇAMENTO	28
	REFERÊNCIAS BIBLIOGRÁFICAS	30

1 INTRODUÇÃO

As instituições, as agências de fomento e os próprios pesquisadores ligados na área da ciência da informação começam a compreender que os dados de pesquisa científica podem constituir uma fonte riquíssima de recursos informacionais para a pesquisa científica e para o ensino da ciência.

Os repositórios que trabalham com dados científicos estão se incorporando rapidamente à realidade da informação científica. Esses dados podem ser usados, reusados e compartilhados. Esses dados podem capacitar os pesquisadores a formular novos tipos de indagações, hipóteses no estudo de questões cruciais para a ciência e para a sociedade.

O uso dos Repositórios Institucionais e temáticos é fundamental para o acesso a esse tipo de material e ao desenvolvimento de pesquisas de alta qualidade e excelência. Eles têm um papel crucial como facilitador nos processos de compartilhamento dos dados, na garantia da sustentabilidade e acessibilidade dos dados em longo prazo. Permitindo que os dados possam ser reusados em outros projetos ao longo do tempo e do espaço.

É necessária uma organização de uma coleta e transmissão desses dados para que sejam acessíveis e até reutilizados em outras pesquisas.

Cabe aos profissionais envolvidos com esse tipo informação proporcionar formas de visibilidade. Os repositórios cumprem esses requisitos quando há uma boa gestão dos dados de pesquisa científica.

2 JUSTIFICATIVA

A produção de conhecimento sobre a coleta e tratamento dos dados de pesquisa científica nos repositórios institucionais e temáticos surgiu da necessidade de produzir material de orientação para profissionais que trabalham com repositórios. Esta literatura visa apoiar na constituição de diretrizes metodológicas que possibilitem a ampliação do acesso aos dados gerados pela instituição.

Desde 2017 tenho trabalhado no Repositório do Observatório do Cuidado no Laboratório de Informação Científica e Tecnológica em Saúde. Na alimentação do repositório tenho acesso aos dados de pesquisa, porém os dados de pesquisa na maioria das vezes, não são disponibilizados.

O compartilhamento de dados de pesquisa em repositórios perpassa tanto questões políticas como metodológicas.

Este projeto busca verificar os passos metodológicos necessários para se pensar o compartilhamento de dados de pesquisa em repositórios digitais.

Os pesquisadores devem garantir que os dados relevantes relacionados aos resultados e às publicações geradas pelas pesquisas estejam organizados, sejam arquivados, tornem-se acessíveis e possam ser citados, permitindo que os resultados e a metodologia de coleta possam ser verificados ou mesmo reutilizados no futuro.

O presente projeto justifica-se pela necessidade de fornecer orientações metodológicas que possibilitem a ampliação do acesso aos dados gerados e propor critérios para a padronização e inclusão de elementos em sua estrutura.

3 REFERENCIAL TEÓRICO

3.1 Os Repositórios digitais

A Iniciativa de Arquivos Abertos e o Movimento de Acesso Livre são iniciativas que vêm construindo as condições necessárias para permitir o acesso livre à produção científica, com o intuito de não somente alterar o processo de aquisição de informação científica, mas também a sua produção, disseminação e uso.

Em resumo um repositório digital pode ser definido como um arquivo digital que reúne uma coleção de documentos digitais. Os repositórios digitais que adotam o modelo OAI adotam o protocolo OAI-PMH (Open Archive Initiative – Protocol for Metadata Harvesting).

De um modo geral, os termos “repositórios institucionais” ou “temáticos” são adotados para caracterizar os repositórios digitais que reúnem respectivamente a produção científica de uma instituição ou de uma área do conhecimento.

Repositórios institucionais são coleções que capturam e preservam a produção intelectual de uma ou mais instituições, universidades ou comunidades (CROW, 2002) ou ainda um conjunto de serviços que uma instituição oferece aos membros de sua comunidade, com o objetivo de gerenciar e disseminar materiais digitais criados pela instituição e membros da comunidade (LYNCH, 2003).

O que caracteriza os repositórios institucionais é o fato de serem orientados para a informação produzida no ambiente das instituições, sendo desenvolvidos, implementados e mantidos por elas.

O conteúdo de um repositório institucional possui as seguintes características: definido institucionalmente; acadêmico; cumulativo e perpétuo; e aberto e interoperável. (CROW, 2002).

Já os repositórios temáticos armazenam documentos com uma delimitação concisa de sua cobertura designada por um assunto, área do

conhecimento ou temática específica. Os repositórios temáticos seriam então um conjunto de serviços oferecidos por uma sociedade, associação ou organização, para gestão e disseminação da produção técnico-científica em meio digital, de uma área ou subárea específica do conhecimento (KURAMOTO, 2006, p. 83).

3.2 Dados de pesquisa

De acordo com Dudziak (2019):

Dados de pesquisa são os materiais comumente registrados e aceitos na comunidade científica como necessários para validar os resultados de pesquisa e incluem: fatos e estatísticas recolhidas para posterior referência ou análise, documentos (texto, Word), planilhas (Excel, etc), cadernos de laboratório, cadernos de campo, diários, questionários, transcrições, fitas de áudio, fitas de vídeo, fotografias, filmes, sequências de proteínas ou genéticos, respostas de teste, slides, artefatos, amostras, coleção de objetos digitais adquiridos e gerados durante o processo de pesquisa, conteúdos de banco de dados (vídeo, áudio, texto, imagens), modelos, algoritmos, scripts, arquivos de log, software de simulação, metodologias e fluxos de trabalho, procedimentos operacionais, padrões e protocolos.

Os dados de pesquisa podem ser classificados como dados observacionais, dados computacionais, dados experimentais. Podem ser também identificados como dados brutos (raw data) ou preliminares, dados derivados, dados referenciais ou canônicos. Embora sejam considerados o alicerce do conhecimento científico, tecnológico e médico, dados de pesquisa não são fáceis de estruturar, organizar, descrever e disponibilizar, para que sejam compreensíveis agora e no futuro. (SAYÃO; SALES, 2015)

Do ponto de vista do pesquisador, a gestão do ciclo de vida dos dados de pesquisa envolve oito componentes: planejar o dado, criar e coletar, assegurar e aprimorar a qualidade do dado, descrever o dado a partir de metadados apropriados, preservar em repositórios adequados, possibilitar a descoberta, integrar com outros dados, analisar para reiniciar o ciclo.

O volume de dados de pesquisa aumenta exponencialmente. Sua gestão e preservação digital nos repositórios institucionais e temáticos requer o estabelecimento e o fortalecimento de políticas globais, governamentais e institucionais mandatórias capazes de garantir o acesso íntegro e integral ao

conjunto de dados que embasam as pesquisas científicas. Também a curadoria e preservação desses dados é um desafio enorme que envolve, além dos editores, instituições de fomento e de pesquisa, pesquisadores e bibliotecários. (DUDZIAK, 2019)

Para que as equipes de pesquisa desfrutem de todos os benefícios dos dados de pesquisa produzidos, as instituições devem criar curadores digitais qualificados e um gerenciamento eficaz do ciclo de vida da curadoria. Garantindo que dados importantes de pesquisa digital sejam adequadamente salvaguardados para uso futuro.

Ao preservar e compartilhar materiais digitais para que outras pessoas possam reutilizá-los, maximiza o impacto da pesquisa e inspira confiança entre os conselhos de pesquisa e órgãos de financiamento.

A preservação de dados digitais deve ser um aspecto essencial de todos os projetos de pesquisa. Alguns dados de pesquisa são únicos e não podem ser substituídos se destruídos ou perdidos.

Um programa de preservação de dados adequado à instituição deve ser usado para proteger esse enorme investimento de tempo e recursos. Sem boas práticas, o registro científico e o patrimônio documental criado em formato digital permanecerão em risco de obsolescência digital e também das fragilidades inerentes à mídia digital.

Segundo Veiga (2017, pg. 24):

Os dados de pesquisa não possuem uma formalização ou ritual estabelecido e trata-se de uma tipologia heterogênea em seu formato e em seu objeto, os problemas enfrentados para o seu compartilhamento são diversos e perpassam por outras questões como ética e integridade na pesquisa, dados sensíveis, entre outros. Apesar de se estar em um momento de discussão e busca para avançar no acesso aberto aos dados de pesquisa, ainda não foi alcançado o acesso aberto aos artigos científicos.

A diferença entre os dados de pesquisa e outros tipos de dados, como dados administrativos, governamentais e qualquer outro dado de coleta sem a finalidade de pesquisa, mas que podem se tornar dados de pesquisa se forem utilizados com a finalidade de investigação científica é o seu propósito. Os dados

que são coletados, observados ou criados com o propósito de análise para produzir e validar um resultado de pesquisa original é um dado de pesquisa. (UNIVERSITY LEICESTER, 2012).

Os princípios da ciência aberta se baseiam no acesso aberto aos dados de pesquisa e às publicações científicas, principalmente as financiadas com recursos públicos; ferramentas e métodos de pesquisa abertos; processos de investigação colaborativos; a implementação de uma ciência cidadã; e a inovação aberta (VEIGA, 2017, pg. 44).

A Budapest Open Access Initiative - BOAI (BUDAPEST, 2002) define o Acesso Aberto como o acesso totalmente irrestrito e gratuito por parte de qualquer cientista, acadêmico, professor, estudante ou outro interessado na literatura acadêmica revisada por pares por toda a extensão do globo desfazendo as barreiras que impedem o acesso a esta literatura.

BOAI abre espaço para a disponibilização abre também para qualquer pré-publicação não revisada que se tenha a intenção de disponibilizar on-line para comentar ou alertar colegas sobre importantes achados de pesquisa (BUDAPEST, 2002).

Por 'acesso aberto' a esta literatura, nos referimos à sua disponibilidade gratuita na internet, permitindo a qualquer usuário ler, baixar, copiar, distribuir, imprimir, buscar ou usar esta literatura com qualquer propósito legal, sem nenhuma barreira financeira, legal ou técnica que não o simples acesso à internet. A única limitação quanto à reprodução e distribuição, e o único papel do copyright neste domínio sendo o controle por parte dos autores sobre a integridade de seu trabalho e o direito de ser propriamente reconhecido e citado (BUDAPEST, 2002).

A Ciência Aberta pressupõe Dados Abertos. Existem várias definições para o que sejam "dados abertos", não existindo consenso na literatura. Inclusive a que define como conjuntos de dados cujos metadados sejam obrigatoriamente públicos. Qualquer pessoa pode descobrir, usando mecanismos de busca, se os dados existem, e como obtê-los. No entanto, os dados propriamente ditos não são obrigatoriamente públicos, inclusive, alguns só podem ser usados por grupos bem restritos de pesquisa, por exemplo por razões éticas ou privacidade (MEDEIROS, 2018), mas esta não é uma definição comumente utilizada na literatura da área.

3.3 Gestão de dados

Para a gestão de dados de pesquisa, o planejamento é uma fase essencial. O seu início se dá quando a pesquisa ainda está sendo delineada e deve considerar como os dados serão gerenciados durante o desenvolvimento do projeto e como eles serão compartilhados depois. Dessa forma é necessário formalizar as ações e compromissos que serão estabelecidos em relação aos dados desde os seus primeiros estágios (SAYÃO; SALES, 2015, pg. 15).

A gestão dos dados se torna fundamental para os pesquisadores, posto que, nos últimos anos, muitas agências financiadoras de pesquisa têm introduzido no seu elenco de exigências para financiamento de projetos de pesquisa que um plano de gestão e de compartilhamento de dados faça parte dos pedidos de auxílio.

Em 2008 o Conselho Internacional para a Ciência (Interdisciplinary Body of the International Council for Science – ICSU) reuniu-se em Maputo, Moçambique, onde foi criado o Sistema Mundial de Dados – World Data System – WDS , uma organização interdisciplinar que promove a gestão de longo prazo, e o acesso universal e equitativo aos dados científicos de qualidade, assim como os serviços de dados, produtos e informações em toda a gama de disciplinas das ciências naturais às ciências sociais e as humanidades (DUDZIAK, 2016).

O ICSU-WDS visa facilitar a pesquisa científica, ao coordenar e apoiar serviços de dados científicos confiáveis para o fornecimento, uso e preservação de conjuntos relevantes de dados, integrando gestores de dados e editoras científicas (DUDZIAK, 2016).

Em julho de 2013, fundou-se a Aliança de Dados de Pesquisa – Research Data Alliance – RDA, uma organização internacional para o desenvolvimento de infraestrutura e promoção de atividades orientadas a reduzir as barreiras de troca e compartilhamento de dados. Congrega grupos de trabalho e grupos de interesse dedicados a discutir e elaborar diretrizes e recomendações para a gestão de dados de pesquisa, incluindo certificação de repositórios, reprodutibilidade de dados, dados publicados e editoras científicas, big data, etc. (DUDZIAK, 2016).

Em novembro de 2013, as editoras científicas anunciaram a permissão da mineração de dados e textos com fins científicos na União Europeia. Os signatários da STM Text and Data Mining for non-commercial scientific research se comprometeram a conceder as licenças de direitos autorais necessários para permitir a mineração de textos e dados de conteúdo protegido por direitos autorais e outros objetos, em condições razoáveis, para fins de investigação científica não comercial na União Europeia, garantindo as devidas salvaguardas técnicas e tecnológicas (DUDZIAK, 2016).

Finalmente, em 2014, foi assinada a Joint Declaration of Data Citation Principles, a partir da qual as editoras endossaram os princípios da FORCE11 e reforçaram o pressuposto de que os dados da pesquisa são parte integrante do registro acadêmico.

O FORCE11 é uma comunidade de acadêmicos, bibliotecários, arquivistas, editores e financiadores de pesquisas que surgiram organicamente para ajudar a facilitar a mudança em direção a uma melhor criação e compartilhamento de conhecimento. A publicação digital de artigos aprimorados permitirá uma comunicação acadêmica mais eficaz, que também será ampliada para incluir, por exemplo, a publicação de ferramentas de software e a pesquisa de comunicação por meio de canais de mídia social.

O FORCE11 cresceu a partir do Workshop FORC realizado em Dagstuhl, Alemanha, em agosto de 2011.

Desta forma, além de assegurar que os pesquisadores obtenham o crédito adequado por seu trabalho, os princípios da citação de dados estabelecem que os dados de pesquisa devem estar facilmente acessíveis, ser devidamente preservados, citados e referenciados.

Isso significa que os dados passam a ser produtos citáveis, aos quais deve ser dada a devida atribuição e crédito como evidências de pesquisa. Além disso, é preciso atribuir aos dados um identificador único que garanta o acesso persistente e fácil localização por meio de metadados que descrevam especificidades dos dados, documentos e materiais associados, e permitam sua recuperação e verificação.

O DataCite é uma organização internacional sem fins lucrativos que fornece identificadores persistentes (DOIs) para dados de pesquisa. Os métodos de citação de dados devem ser suficientemente flexíveis para acomodar as práticas variantes entre as comunidades, mas não devem diferir tanto a ponto de comprometer a interoperabilidade das práticas de citação de dados.

Alguns repositórios oferecem aos autores a opção de hospedar os dados associados a um manuscrito de forma confidencial e fornecer acesso anônimo para revisores antes do lançamento público. Estes repositórios coordenam então a divulgação pública dos dados como a data de publicação da revista. É responsabilidade do autor comunicar-se com o repositório para assegurar que a liberação pública dos dados, materiais e métodos seja feita no momento da publicação on-line do artigo. Além dos repositórios de dados por disciplina, várias instituições criaram e mantêm repositórios institucionais e mesmo nacionais. (DUDZIAK, 2016)

Para obter informações sobre repositórios de dados é possível consultar o R3Data – Registry of Research Data Repositories – uma ferramenta de busca de repositórios de dados de pesquisa.

3.4 Metadados

Metadados é comumente definido como sendo dados sobre dados, mas na realidade metadado possui uma definição mais abrangente.

Sem uma descrição minuciosa do contexto tecnológico dos arquivos de dados, do contexto no qual os dados foram criados ou coletados, das medidas que foram feitas, dos detalhes espaciais e temporais, dos instrumentos usados, dos parâmetros e unidades e da qualidade dos dados e da sua proveniência, é improvável que os dados possam ser descobertos, interpretados, gerenciados e efetivamente usados e reusados. Os metadados cumprem essa tarefa, porque eles são a documentação dos dados. Os metadados que são usados para descreverem os dados, permitem que eles estejam autodocumentados agora e no futuro (SAYÃO; SALES, 2015, pg. 23).

Devido ao conhecimento especializado no tratamento e na gestão da informação, os bibliotecários são os profissionais mais adequados para manter a qualidade do conteúdo digital (produzidos pelos autores da instituição) e na elaboração de metadados de acordo com os padrões internacionais (HORWOOD et al., 2004).

A pesquisa científica produz e coleta dados que são variados e heterogêneos e que têm natureza, formatos diferentes e são coletados em volumes variados e passam por diferentes processos que dependem de cada disciplina e dos objetivos da pesquisa, portanto é necessário descrever, com algum grau de detalhe, as principais características desses dados, incluindo a natureza e origem, escopo e a escala dos dados que serão produzidos. Isto ajudar outros pesquisadores a compreenderem os dados, sua relação com os dados existentes e os possíveis riscos de disseminá-los.

É importante descrever a convenção adotada para dar nomes para os conjuntos de dados, arquivos e pastas.

Definir qual é o elenco de metadados que são necessários para que os dados possam ter significado e possam ser interpretados ao longo do tempo e do espaço.

A seguir segue um modelo proposto por Sayão e Sales (2015, pg. 36) como exemplo de metadados para dados de pesquisa genéricos:

Exemplo de metadados:

1. Título da coleção
2. Identificador persistente da
Coleção
3. Resumo da coleção
4. Autores
5. Citação da coleção
6. Data da publicação dos dados
7. Período de tempo que cobre
8. Escopo geográfico
9. Direitos associados aos dados
10. Assunto/palavras-chave
11. Origem dos dados (experimental,
observacional, crus, derivados...)
12. Tipo de dados (inteiro, imagem...)
13. Instrumentos usados
14. Aquisição dos dados
15. Tipo de arquivo
16. Método de processamento de

Dados (software)

17. Processamento de dados (scripts ou código)

18. Parâmetros da coleção de dados

Controle de qualidade

Os princípios FAIR, um acrônimo para Findable, Accessible, Interoperable e Reusable, está presente nas discussões e práticas contemporâneas da ciência de dados, desde o início de 2014, e tiveram sua aplicação consolidada em 2017, quando a Comissão Europeia passou a exigir a adoção de plano de gestão de dados, com base nesses princípios, por projetos financiados por seus recursos. Desde então, tais princípios passaram a serem norteadores da descoberta, do acesso, da interoperabilidade, do compartilhamento e da reutilização dos dados de pesquisa. (HENNING; et al, 2018)

O GO FAIR é uma iniciativa internacional que adota uma abordagem “bottom-up”, ou seja, com atuação dos seus membros de baixo para cima, iniciada com apoio dos governos da Holanda, Alemanha e França, com a participação dos Estados Unidos e, mais recentemente, com a adesão do Brasil. Tem por objetivo promover o desenvolvimento coerente da Internet global de serviços e dados FAIR, seguindo as diretrizes do European Open Science Cloud (EOSC). O GO FAIR tem por prioridade tornar todos os tipos de dados, que se encontram fragmentados e desconectados, mais facilmente localizáveis, acessíveis, interoperáveis e reutilizáveis, ou seja, FAIR, facilitando, dessa forma, o seu reconhecimento por máquinas e pessoas. (VEIGA; QUEIROZ, 2019)

A Rede de Implementação GO FAIR Brasil Saúde é uma rede temática responsável pela elaboração de estratégias de implementação dos princípios FAIR no campo da saúde. Um grupo de trabalho para estudar padrões de metadados de pesquisa. Sua coordenação está sob a responsabilidade do Instituto de Comunicação e Informação Científica e Tecnologia em Saúde (ICICT) da Fundação Oswaldo Cruz (Fiocruz), que conta com a participação de diversas instituições das áreas de Saúde Pública, Vigilância Sanitária, Informação e Comunicação em Saúde, História do Patrimônio Cultural das Ciências e da Saúde, Oncologia, Enfermagem e Educação Profissional em Saúde. (VEIGA; QUEIROZ, 2019)

4 OBJETIVOS

4.1 OBJETIVO GERAL

Propor orientações metodológicas no uso de padrões de metadados que possibilitem a ampliação do acesso aos dados de pesquisa nos Repositórios Institucionais e Temáticos no campo da saúde.

4.2 OBJETIVOS ESPECÍFICOS

Identificar os repositórios de dados de pesquisa no campo da saúde;

Mapear os metadados utilizados nos repositórios de dados do campo da saúde;

Propor diretrizes metodológicas no uso de padrões de metadados para a inclusão dos dados de pesquisa na estrutura dos Repositórios Institucionais e Temáticos do campo da saúde.

5 METODOLOGIA

Trata-se de uma pesquisa de cunho exploratório e bibliográfico onde será realizado um levantamento bibliográfico na literatura especializada nacional e estrangeira sobre os dados de pesquisa científica.

Com base no levantamento bibliográfico, se observará os conceitos e aplicações dos dados de pesquisa para serem gerenciados e compartilhados de forma a garantir o maior benefício possível para o avanço científico e tecnológico e então disponibiliza-los nos repositórios institucionais ou temáticos.

Levantamento Bibliográfico

Nesta fase inicial será realizado um levantamento na literatura sobre os dados de pesquisa científica na busca por assunto do Portal de periódicos da CAPES, e especificamente nas bases de dados web of Science e Scopus; na base de dados do PUBMED CENTRAL, na BRAPCI e no Google Acadêmico.

O Portal de Periódicos da CAPES é uma biblioteca virtual que reúne e disponibiliza a instituições de ensino e pesquisa no Brasil o melhor da produção científica internacional. Ele conta com um acervo de mais de 37 mil títulos com texto completo, 126 bases referenciais, 11 bases dedicadas exclusivamente a patentes, além de livros, enciclopédias e obras de referência, normas técnicas, estatísticas e conteúdo audiovisual. (CAPES, 2019)

O PubMed Central é um serviço da U. S. National Library of Medicine (NLM). Inclui cerca de 21 milhões de citações de artigos de periódicos. O maior componente é a Base de dados MEDLINE que indexa cerca de 5.000 revistas publicadas nos Estados Unidos e mais de 80 outros países. (FAMENA, 2012)

A Base BRAPCI é o produto de informação do projeto de pesquisa “Opções metodológicas em pesquisa: a contribuição da área da informação para a produção de saberes no ensino superior”, cujo objetivo é subsidiar estudos e propostas na área de Ciência da Informação, fundamentando-se em atividades planejadas institucionalmente. Com esse propósito, foram identificados os títulos de periódicos da área de Ciência da Informação (CI) e indexados seus artigos, constituindo-se a base de dados referenciais. (BRAPCI, 2019)

Google Acadêmico é um serviço de busca do **Google** voltado especialmente para estudantes, pesquisadores, cientistas, universitários e curiosos. A ferramenta funciona como um repositório de teses, artigos científicos, resumos, monografias, dissertações e livros. (EVEN3, 2019)

Levantamento dos Repositórios Institucionais e Temáticos

Nesta fase será realizado um mapeamento dos Repositórios de dados, Institucionais e Temáticos com o objetivo de coletar dados que reflitam a realidade destes repositórios delimitando os repositórios na área da saúde. Este mapeamento será realizado utilizando como fontes de informação o Re3Data - Registry of Research Data Repositories e o GoogleDataset Search.

O Re3Data - Registry of Research Data Repositories é um registro global de repositórios de dados de pesquisa que abrange repositórios de dados de pesquisa de diferentes disciplinas acadêmicas. Apresenta repositórios para armazenamento permanente e acesso a conjuntos de dados para pesquisadores, órgãos financiadores, editores e instituições acadêmicas. O re3data.org promove uma cultura de compartilhamento, maior acesso e melhor visibilidade dos dados da pesquisa. O registro entrou em operação no outono de 2012 e é financiado pela Fundação Alemã de Pesquisa (DFG) .

O Google Dataset Search é um site dedicado à pesquisa de conjuntos de dados para ajudar pesquisadores a descobrir dados disponíveis publicamente que podem ser de grande ajuda para seus projetos.

Levantamento de padrões de metadados do campo da saúde

Nesta fase será realizado um levantamento em várias fontes de informação que trabalham com a disseminação de padrões de metadados para dados de pesquisa, como o Fairsharing, o Digital Curation Centre (DCC) e o Research Data Alliance (RDA), além dos metadados utilizados nos repositórios levantados na fase anterior da pesquisa.

O FAIRSHARING é um recurso de suporte ao FAIR que fornece um registro informativo e educacional sobre padrões de dados, bancos de dados, repositórios e políticas, juntamente com ferramentas e serviços de pesquisa e visualização que interoperam com outros recursos que permitem estão alinhados

aos princípios FAIR. O FAIRsharing orienta os consumidores a descobrir, selecionar e usar padrões, bancos de dados, repositórios e políticas com confiança, e os produtores a tornar seus recursos mais detectáveis, mais amplamente adotados e citados

O GRUPO DO DCC para estudo dos padrões de metadados é um centro de especialização reconhecido internacionalmente em curadoria digital, com foco na criação de capacidades e habilidades para o gerenciamento de dados de pesquisa.

A Research Data Alliance (RDA) é uma iniciativa para construir conexões técnicas e sociais para viabilizar o compartilhamento aberto de dados científicos. A visão da RDA é tornar viável que pesquisadores possam compartilhar abertamente seus dados entre diferentes tecnologias, disciplinas e países, de forma a endereçar os grandes desafios da sociedade em escala global.

Análise dos resultados

Nesta fase será realizada uma análise dos dados levantados, tanto na literatura quanto nos diretórios de repositórios visando mapear os metadados utilizados nos repositórios de dados do campo da saúde e propor diretrizes metodológicas para a inclusão dos dados de pesquisa na estrutura dos Repositórios Institucionais e Temáticos do campo da saúde.

RESULTADOS PRELIMINARES

Nesta seção apresentamos alguns resultados preliminares desta pesquisa realizado junto com a equipe do GO FAIR Brasil Saúde. Uma busca realizada em algumas das fontes citadas acima pode nos dar uma visão geral se a metodologia proposta apoia o alcance dos objetivos.

Usando os filtros de assunto “Neuroscience” e “Health Science ” no Fairsharing temos seguintes resultados de plataformas de dados de pesquisa que são recomendados pela comunidade

Filtro assunto:	Endereço Eletrônico
Neuroscience.	
Bases:	
ModelDB	http://senselab.med.yale.edu/modeldb
OpenNeuro	http://www.openeuro.org
NeuroVault	http://neurovault.org
Open Connectome Project	http://opeconnecto.me
Donders Repository	http://data.donders.ru.nl
G-Node Data Infrastrucure	http://gin.q-node.org

Filtro Assunto: Health Science.	Endereço Eletrônico
Bases:	
Research Domain Criteria Database	http://rdocdb.nimh.nih.gov
Trials related to Mental Illness	http://data-archive.nimh.nih.gov/ndct
Swedish National Data Service	http://snd.ou.se/en
Ag Data Commons	http://data.nal.usda.gov

Um levantamento no Re3data mostra 2.423 repositórios de dados de pesquisa. Utilizando o filtro por área do conhecimento encontramos 124 repositórios no campo da Saúde pública, pesquisa em serviços de saúde e medicina social. A fase posterior consistirá em análise destas plataformas e seus metadados.

Abaixo apresentamos um levantamento preliminar sobre padrões de metadados no Fairsharing, RDA Grupo de metadados e no DDC:

General Research Data - Metadata Standards

CERIF - Common European Research Information Format

O CERIF é o padrão que a União Europeia recomenda a seus estados membros para registrar informações sobre atividades de pesquisa.

Data Package

A especificação do pacote de dados é um formato genérico de wrapper para troca de dados, consistindo em uma pasta contendo arquivos de dados e um arquivo descritor.

DataCite Metadata Schema

Uma lista independente de domínio das principais propriedades de metadados escolhidas para a identificação precisa e consistente de dados para fins de citação e recuperação.

DCAT - Data Catalog Vocabulary

DCAT é um vocabulário RDF projetado para facilitar a interoperabilidade entre catálogos de dados publicados na Web.

Dublin Core

Um padrão básico, independente de domínio, que pode ser facilmente entendido e implementado e, como tal, é um dos padrões de metadados mais conhecidos e mais amplamente utilizados.

OAI-ORE - Open Archives Initiative

Define padrões para a descrição e troca de agregações de recursos da Web.

Observations and Measurements

Este é um padrão que especifica a implementação XML para o modelo conceitual OGC e Observações e Medidas ISO (O&M), incluindo um esquema para Recursos de Amostragem.

PREMIS - The PREMIS (Preservation Metadata: Implementation Strategies

O Dicionário de dados define um conjunto de metadados que a maioria dos repositórios de objetos digitais precisaria gravar e usar para preservar esses objetos a longo prazo.

PROV

Uma especificação que fornece um vocabulário para trocar informações de proveniência.

RDF Data Cube Vocabulary

O vocabulário do Cubo de Dados é uma base essencial que suporta vocabulários de extensão para permitir a publicação de outros aspectos de fluxos de dados estatísticos ou outros conjuntos de dados multidimensionais.

Repository-Developed Metadata Schemas

Alguns repositórios decidiram que os padrões atuais não atendem às suas necessidades de metadados e, portanto, precisam criar seus próprios requisitos.

Este levantamento ainda está em andamento e será acrescido de novos documentos.

De forma preliminar foi feita uma busca de repositórios que tenham dados de pesquisa e relatórios de pesquisa.

Segue uma lista dos resultados encontrados:

ARCA (<https://www.arca.fiocruz.br/>) – Dá acesso aos dados de pesquisa e aos relatórios de pesquisa.

Repositório da Produção Científica e Intelectual da Unicamp (<http://www.repositorio.unicamp.br/>) – Dá acesso aos relatórios. Não há dados de pesquisa.

Infoteca-e (<https://www.infoteca.cnptia.embrapa.br/>) - Informação Tecnológica em Agricultura – Dá acesso aos relatórios de pesquisa e aos dados de pesquisa.

Portal do Livro Aberto em CT&I (<http://livroaberto.ibict.br/>) – Dá acesso aos relatórios de pesquisa. Em um primeiro momento não foram encontrados dados de pesquisa.

Dataset Search (<https://toolbox.google.com/datasetsearch>) – Ferramenta de busca de dados científicos pelo Google.

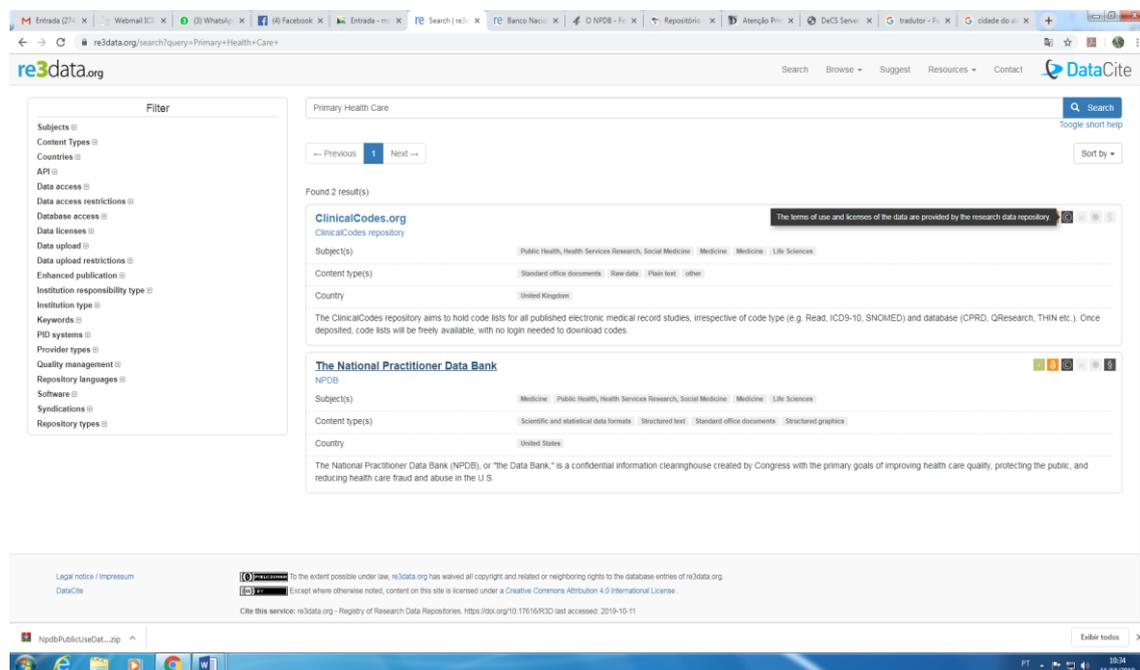
OpenEI (https://openei.org/wiki/Main_Page) - acesso aos dados sobre energia limpa e relatórios técnicos do Departamento de Energia

Repositório Alice (<https://www.alice.cnptia.embrapa.br/>) – Dá acesso aos relatórios de pesquisa e aos dados.

Re3Data (<https://www.re3data.org/>) – Registry of Research Data Repositories – uma ferramenta de busca de repositórios de dados.

Um levantamento realizado no Re3Data em 26 de setembro de 2019 tendo como estratégia de busca atenção primária encontramos 2 registros de repositórios: O Clinical Codes Org e o The National Practitioner Data Bank.

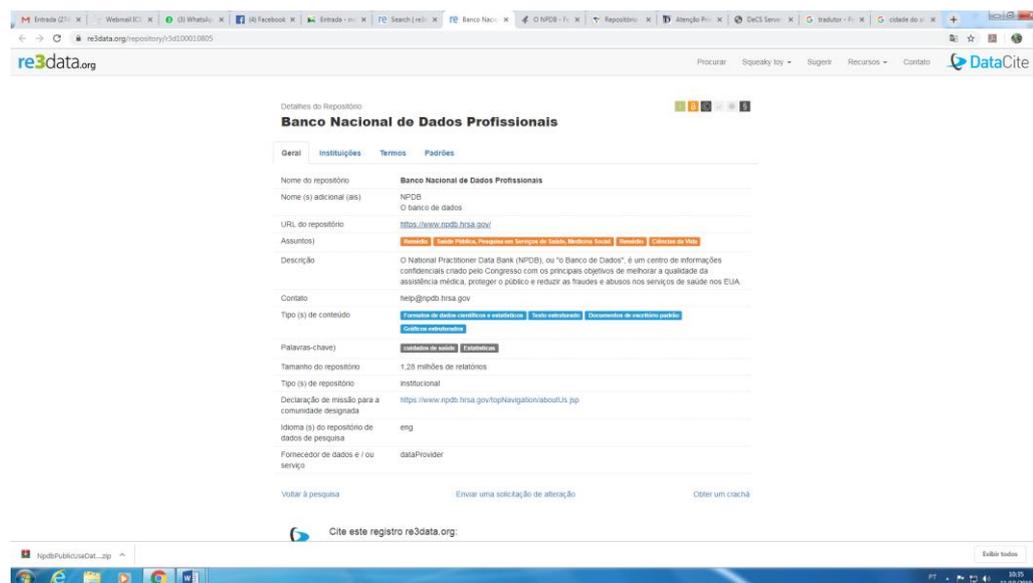
Figura 1 – Busca no Re3Data Atenção Primária



Fonte: Re3Data

O National Practitioner Data Bank infelizmente tem o acesso restrito.

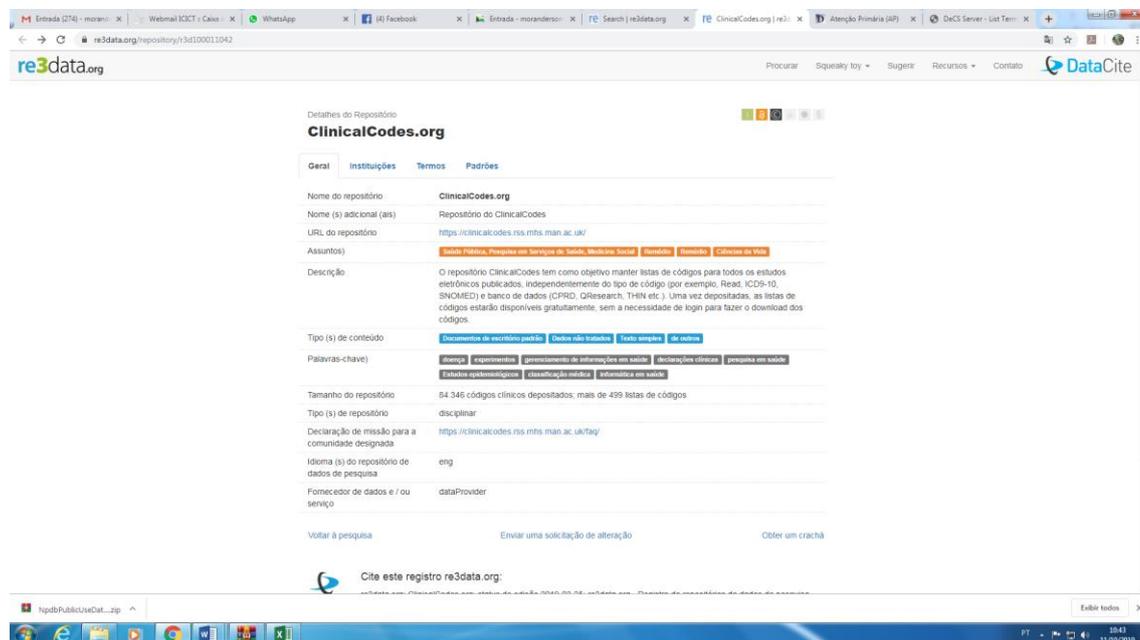
Figura 2 – Busca no Re3Data NPDB



Fonte: R3Data

O ClinicalCodes tem acesso aberto e uma boa base de dados: 84.346.

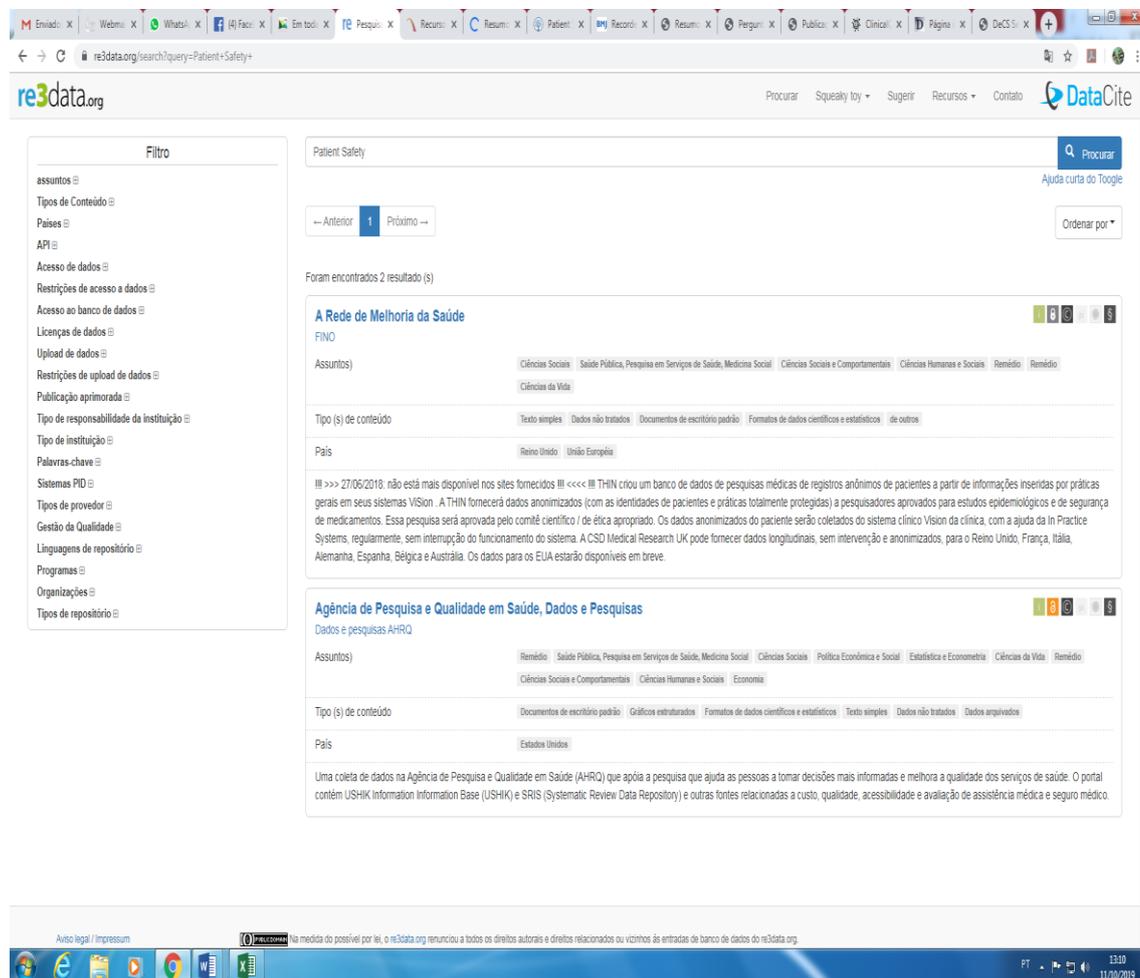
Figura 3: Busca no R3Data Clinical Codes



Fonte: R3Data

Estratégia de busca segurança do paciente encontramos 2 registros de repositórios: The Health Improvement Network e Agency for Healthcare Research and Quality, Data & Surveys.

Figura 4 - Busca no Re3Data Segurança do Paciente



Fonte: Re3Data

Infelizmente a The Health Improvement Network não está mais disponível. A Agency for Healthcare Research and Quality, Data & Surveys. Tem um excelente banco de dados. 15. 156 registros.

Figura 5 - Agency for Healthcare Research and Quality, Data & Surveys

The screenshot shows a web browser window displaying the search results for 'segurança do paciente' on the AHRQ website. The search bar contains the text 'segurança do paciente' and a 'Procurar' button. Below the search bar, the results are listed under the heading 'Procurar Resultados'. The first result is 'A segurança do paciente | Agência de Pesquisa e Qualidade em Saúde' with a URL starting with 'http://www.ahrq.gov/topics/la-seguridad-del-paciente.html'. The second result is 'The Medical Home: O que sabemos, o que precisamos saber? Uma revisão das evidências mais antigas sobre o...' with a URL starting with 'http://pcmh.ahrq.gov/page/medical-home-what-do-we-know-what-do-we-need-to-know-review-earliest-evidence-effectiveness-of-the-patient-centered-modelo-médico-em-casa'. The third result is 'The Medical Home: O que sabemos, o que precisamos saber? Uma revisão das evidências mais antigas sobre o...' with a URL starting with 'http://pcmh.ahrq.gov/page/medical-home-what-do-we-know-what-do-we-need-to-know-review-earliest-evidence-effectiveness-of-the-patient-centered-Modelo-médico-em-casa-23k-2019-04-01'. On the right side of the page, there is a 'Resultados por Domínio' section showing a bar chart with four categories: 'psnet.ahrq.gov (9744)', 'www.ahrq.gov (1778)', 'eficazhealthcare.ahrq.gov (1357)', and 'healthit.ahrq.gov (909)'. The browser's address bar shows 'search.ahrq.gov/search?q=segurança+do+paciente'.

Fonte: arhq.gov

6 RESULTADOS ESPERADOS

A gestão de dados de pesquisa tem sido pauta de agências de fomento, editores científicos e instituições de pesquisa. No campo da saúde a adequada gestão dos dados de pesquisa é fundamental para a reprodutibilidade da pesquisa e economia de recursos em pesquisa com o reuso dos dados.

Espera-se que esta pesquisa possa apoiar a identificação dos repositórios temáticos e institucionais de dados de pesquisa e, que a divulgação da existência destes dados estimule o compartilhamento dos dados de pesquisa em saúde.

Espera-se também que as orientações metodológicas apresentadas para o uso de padrões de metadados adequados para as diversas tipologias, formatos e finalidades possa apoiar na gestão dos dados de pesquisa em saúde e apoiar o compartilhamento e interoperabilidade dos dados.

Os padrões de metadados são importantes para que os registros sigam os princípios FAIR, isto é, sejam encontráveis, acessíveis, interoperáveis e reutilizáveis. Desta forma pesquisadores e outros usuários, humanos ou máquinas, poderão fazer uso dos dados. Desta forma esta pesquisa também prevê como resultado esperado a ampliação do acesso aos dados de pesquisa e suas citações.

7 CRONOGRAMA

O cronograma foi elaborado para o período de seis meses, incluindo dois meses para a análise de viabilidade antes da aprovação final. Todo o processo de implantação será acompanhado pelos gestores e profissionais a fim de verificar a evolução satisfatória das atividades determinada.

MESES DO ANO 2020						
AÇÃO	01	02	03	04	05	06
Levantamento Bibliográficos	X	X				
Levantamento dos Repositórios Institucionais e Temáticos de dados de pesquisa		X	X			
Levantamento dos padrões de metadados utilizados para descrição de dados de pesquisa				X		
Elaboração de orientações metodológicas para o uso dos metadados para descrição dos dados de pesquisa no campo da saúde					X	X

8 ORÇAMENTO

Não haverá a necessidade de um plano orçamentário para o desenvolvimento deste projeto, pois todos os recursos que nele foram envolvidos tais como: humanos, materiais e tecnológicos devem estar à disposição na própria instituição.

LISTA DE FIGURAS

Figura 1 – Busca Re3Data Atenção Primária	22
Figura 2 – Busca no Re3Data NPDB.....	22
Figura 3 – Busca no Re3Data Clinical Codes	23
Figura 4 – Busca no Re3Data Segurança do Paciente.....	24
Figura 5 – Agency for Healthcare Research and Quality, Data & Surveys.....	25

REFERÊNCIAS BIBLIOGRÁFICAS

BRAPCI. **Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (Brapci)**. 2019. Disponível em:

<http://www.brapci.inf.br/index.php/res/about>. Acesso em: 17 nov. 2019.

CAPES. Missão e objetivos. Disponível em:

https://www.periodicos.capes.gov.br/index.php?option=com_pcontent&view=pccontent&alias=missao-objetivos&Itemid=109. Acesso em: 17 nov. 2019.

CHILAG, J. **Non-conventional literature in agriculture** - an overview. *IAALD Q Bull* 1982; 27:2-7.

CROW, R. **The Case for Institutional Repositories**: A SPARC Position Paper, The Scholarly Publishing & Academic Resources Coalition. Washington: SPARC, 2002. Disponível em: https://rc.library.uta.edu/uta-ir/bitstream/handle/10106/24350/Case%20for%20IRs_SPARC.pdf. Acesso em: 11 nov. 2019.

DUDZIAK, Elisabeth. **Dados de Pesquisa agora devem ser armazenados e citados**. 2016. Disponível em: <http://www.sibi.usp.br/?p=6189>. Acesso em: 11 nov. 2019.

EVEN3. **Google acadêmico**: por que utilizá-lo como fonte para pesquisa e publicação de trabalhos. 2019. Disponível em: <https://blog.even3.com.br/google-academico-como-usar/>. Acesso em: 17 nov. 2019.

FAMENA. **PubMed**. Disponível em:

http://www.famema.br/site_ensino/ensino/biblioteca/docs/PUBMED-2012-OUTUBRO.pdf. Acesso em: 17 nov. 2019.

GL'99 Conference Program. **Fourth International Conference on Grey Literature**: New Frontiers in Grey Literature. GreyNet, Grey Literature Network Service. Washington D.C. USA, 4-5 Oct. 1999.

HENNING, P. C.; RIBEIRO, C. J. S.; SALES, L.; MOREIRA, J.; SANTOS, L. O. B. S. Desmistificando os princípios fair: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados fair. **Encontro Nacional de Pesquisa em Ciência da Informação**, n. XIX ENANCIB, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/103243>. Acesso em: 15 nov. 2019.

HORWOOD, Lyne et al. OAI compliant institutional repositories and the role of library staff. **Library management**, [s.l.], v. 25, n. 4-5, p. 170-176, 2004.

Disponível em:

<http://www.emeraldinsight.com/doi/abs/10.1108/01435120410533756>. Acesso em: 11 nov. 2019.

KURAMOTO, Hélio. Informação científica: proposta de um novo modelo para o Brasil. **Ci. Inf.** [online]. 2006, vol.35, n.2, pp.91-102. Disponível em: http://www.scielo.br/scielo.php?pid=S0100-19652006000200010&script=sci_abstract&tlng=pt. Acesso em: 11 nov. 2019.

LYNCH, Clifford A. Institutional Repositories: essential infrastructure for scholarship in the Digital Age. **ARL**, n. 226, p. 1-7, Feb. 2003. Disponível em: https://www.researchgate.net/publication/215993581_Institutional_Repositories_Essential_Infrastructure_For_Scholarship_In_The_Digital_Age. Acesso em: 11 nov. 2019.

MACHADO, Rejane Ramos. **Representação e recuperação dos relatórios de pesquisa da Fiocruz**: proposta de metodologia para ampliação do acesso. 2005. 111 f. Dissertação (Mestrado em Saúde Pública) - Escola Nacional de Saúde Pública Sérgio Arouca, Fundação Oswaldo Cruz, Rio de Janeiro, 2005. Disponível em: <https://www.arca.fiocruz.br/handle/icict/1317>. Acesso em: 11 nov. 2019.

MEDEIROS, C.B. Gestão de Dados Científicos – da coleta à preservação [online]. **SciELO em Perspectiva**, 2018. Disponível em: <https://blog.scielo.org/blog/2018/06/22/gestao-de-dados-cientificos-da-coleta-a-preservacao/>. Acesso em: 11 nov. 2019.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 11 nov. 2019.

UNIVERSITY LEICESTER. **Research data**: definitions, 2012. Disponível em: https://www2.le.ac.uk/services/researchdata/documents/UoL_ReserchDataDefinitions_20120904.pdf. Acesso em: 11 nov. 2019.

VEIGA, Viviane Santos de Oliveira. **Percepção dos pesquisadores portugueses e brasileiros da área de Neurociências quanto ao compartilhamento de artigos científicos e dados de pesquisa no acesso aberto verde**: custos, benefícios e fatores contextuais. 2017. 294 f. Tese (Doutorado em Ciências)-Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2017. Disponível em: <https://www.arca.fiocruz.br/handle/icict/26842>. Acesso em: 11 nov. 2019.

VEIGA, Viviane Santos de Oliveira; QUEIROZ, Claudete Fernandes de. **Rede GO FAIR Brasil Saúde**: uma rede de apoio à Gestão e Abertura de Dados de Pesquisa em Saúde no Brasil. In: ENCONTRO IBÉRICO EDICIC, 9., Barcelona, 2019. 1 p.