

JESSICA FERNANDA DOS SANTOS LIMA RAMOS

Repositórios de dados de pesquisa em câncer: uma análise a partir do Re3data

Dissertação de mestrado
Fevereiro de 2021



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO – UFRJ

ESCOLA DE COMUNICAÇÃO – ECO

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA - IBICT

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – PPGCI

JÉSSICA FERNANDA DOS SANTOS LIMA RAMOS

Repositórios de dados de pesquisa em câncer: uma análise a partir do Re3data

Rio de Janeiro

2021

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO – UFRJ

ESCOLA DE COMUNICAÇÃO – ECO

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBICT

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – PPGCI

JÉSSICA FERNANDA DOS SANTOS LIMA RAMOS

Repositórios de dados de pesquisa em câncer: uma análise a partir do Re3data

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação, convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro / Escola de Comunicação, como requisito parcial à obtenção do grau de Mestre em Ciência da Informação.

Orientador: Dr. Luis Fernando Sayão
Coorientadora: Dra. Viviane Santos de Oliveira Veiga

Rio de Janeiro

2021

Ficha elaborada por Jéssica Fernanda dos Santos Lima Ramos CRB 7 / 6965

R175d Ramos, Jéssica Fernanda dos Santos Lima.

Repositórios de dados de pesquisa em câncer: uma análise a partir do Re3data / Jéssica Fernanda dos Santos Lima Ramos. – Rio de Janeiro, 2021.
103 f.: il.

Orientador: Luis Fernando Sayão.

Coorientadora: Viviane Santos de Oliveira Veiga.

Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2021.

1. Dados de pesquisa. 2. Neoplasias. 3. Gestão da informação em saúde.
4. Repositório de dados. 5. Disseminação da informação. I. Sayão, Luis Fernando. II. Veiga, Viviane Santos de Oliveira. III. Título.

JÉSSICA FERNANDA DOS SANTOS LIMA RAMOS

Repositórios de dados de pesquisa em câncer: uma análise a partir do Re3data

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação, convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro / Escola de Comunicação, como requisito parcial à obtenção do grau de Mestre em Ciência da Informação.

Rio de Janeiro, __/__/____.

Prof. Dr. Luis Fernando Sayão (Orientador)
PPGCI/IBICT-ECO/UFRJ

Profa. Dra. Viviane Santos de Oliveira Veiga – (Coorientadora)
Fundação Oswaldo Cruz (FIOCRUZ)

Profa. Dra. Luana Farias Sales Marques
PPGCI/IBICT-ECO/UFRJ

Profa. Dra. Cícera Henrique da Silva
Fundação Oswaldo Cruz (FIOCRUZ)

A minha mãezinha, com amor.

AGRADECIMENTOS

Agradeço primeiramente a Deus – fortaleza e refúgio - pelo cuidado, força e por mais uma vitória.

À minha mãezinha Olga, por seu auxílio, confiança, dedicação, incentivo e por sempre sonhar comigo.

À minha família, por todo amor, carinho, confiança, alegria e amparo que recebo dia após dia.

À meus orientadores: Luis Sayão e Viviane por todo empenho, paciência, incentivo e auxílio na construção deste trabalho. Obrigada por não “soltarem a minha mão” nesse processo de construção de pesquisa, vocês são maravilhosos!

Aos membros da banca examinadora, doutoras que se dispuseram a avaliar e contribuir com esta pesquisa que marca a conclusão de um grande sonho que é o mestrado.

Dedico este projeto a todos os professores que fizeram parte da minha trajetória, colaborando para a minha formação acadêmica. Em especial à professora Luana Sales, pelo incentivo nos momentos de angústia e pela contribuição na construção da pesquisa.

RESUMO

A relevância dos dados de pesquisa em saúde vem pouco a pouco ganhando destaque nas grandes mídias. Observa-se que o tema vem transcendendo o campo científico e tecnológico, pois se expande para além do nicho formado por estudiosos do campo científico o qual o autor dos dados e sua equipe fazem parte. Na atualidade, é por meio do reuso dos dados de pesquisa arquivados em sistemas de gestão de dados e localizados por meio de plataformas de gestão que se obtém avanços científicos e tecnológicos. No âmbito da saúde, o reuso dos dados repercute em benefício da saúde pública mundial. Trazendo benefícios em relação à identificação de doenças, no desenvolvimento de tratamentos e vacinas e, também, no aperfeiçoamento de medicamentos, transparência a gestão pública e melhoria nas políticas de saúde. O presente estudo tem como objetivo analisar os repositórios de dados de pesquisa em câncer registrados no *Registry of Research Data Repositories* (Re3data), a fim de discutir a luz da Ciência da Informação, como os dados de pesquisa em câncer estão representados nos repositórios de dados de pesquisa em escala global. Trata-se de um estudo exploratório e empírico, onde foram feitas pesquisas bibliográficas e documentais a respeito dos assuntos: dados de pesquisa em saúde, oncologia e infraestruturas de acesso e gestão de dados de pesquisa, os repositórios de dados. A análise de dados foi qualitativa, onde os resultados obtidos foram analisados tendo em vista: apresentar o “estado da arte” sobre os dados de pesquisa em câncer em contexto internacional, identificando os repositórios específicos para armazenamento de dados de pesquisa em câncer registrados no Re3data e, por fim, oferecendo um panorama quanto aos tipos de repositórios, seus assuntos, nacionalidades, tecnologias e também quanto a tipologia dos dados envolvidos no armazenamento dos dados de pesquisa em câncer em repositórios digitais.

Palavras-chave: Dados de pesquisa. Neoplasias. Gestão da Informação em Saúde. Repositório de dados. Disseminação de informação.

ABSTRACT

The relevance of health research data is gradually gaining prominence in the mainstream media. It is observed that the theme has been transcending the scientific and technological field, as it expands beyond the niche formed by scholars in the scientific field to which the data author and his team are part. Currently, it is through the reuse of research data stored in data management systems and located through management platforms that scientific and technological advances are obtained. In the field of health, the reuse of data has repercussions for the benefit of global public health. Bringing benefits in relation to the identification of diseases, in the development of treatments and vaccines and, also, in the improvement of medicines, transparency in public management and improvement in health policies. The present study aims to analyze the repositories of cancer research data registered in the Registry of Research Data Repositories (Re3data), in order to discuss the light of Information Science, how the cancer research data are represented in the data repositories research on a global scale. It is an exploratory and empirical study, where bibliographical and documentary researches were made on the subjects: health research data, oncology and research data management and access infrastructures, data repositories. The data analysis was qualitative, where the results obtained were analyzed with a view to: presenting the “state of the art” on cancer research data in an international context, identifying the specific repositories for storing cancer research data registered in Re3data and, finally, offering an overview as to the types of repositories, their subjects, nationalities, technologies and also as to the type of data involved in the storage of cancer research data in digital repositories.

Keywords: Search data. Neoplasms. Health Information management. Search data repositories. Information dissemination.

LISTA DE ABREVIATURAS E SIGLAS

BIREME	Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde
GFZ	Centro de Pesquisa Alemão para Geociências
CDD	Classificação Decimal de Dewey
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DataONE	Data Observation Network For Earth
DANT	Doenças e Agravos não transmissíveis
DeCS	Descritores em Ciências da Saúde
DFG	Fundação Alemã de Pesquisa
KIT	Instituto de Tecnologia Karlsruhe
INCA	Instituto Nacional de Câncer José Alencar Gomes da Silva
LISA	Library and Information Science Abstracts
LILACS	Literatura Latino-Americana e do Caribe em Ciências da Saúde
ORCID	Open Researcher and Contributor ID
PD&I	Pesquisa, Desenvolvimento e Inovação
Re3data	Registry of Research Data Repositories
SEBICT	Serviço de Biblioteca e Informação Científica e Tecnológica
SUS	Sistema Único de Saúde
TIC	Tecnologia de Informação e Comunicação
UICC	União Internacional para o Controle do Câncer

LISTA DE FIGURAS

Figura 1 - Ciclo de vida dos dados.....	42
Figura 2 – Composição do plano de gestão de dados: recomendação...	46
Figura 3 – Ocorrência do termo de busca “câncer” em português.....	57
Figura 4 – Ocorrência do termo de busca “câncer” em inglês.....	57
Figura 5 – Ocorrência do termo de busca “câncer” em espanhol.....	58
Figura 6 – Classificações hierárquicas no Re3data.....	62
Figura 7 –Formulário de indicação de repositórios.....	63

LISTA DE QUADROS

Quadro 1 – Amostra de pesquisa.....	59
Quadro 2 – Dimensões de análise da amostra e elementos de caracterização.....	60
Quadro 3 – Áreas temáticas de abrangência.....	65
Quadro 4 – Cadeias de repositórios de dados de pesquisa.....	67
Quadro 5 – Uploads de conjuntos de dados.....	74
Quadro 6 – Tipos de conteúdo de repositórios de dados de pesquisa em câncer....	76
Quadro 7 – Estratégia de busca na BVS Prevenção e Controle de Câncer.....	89
Quadro 8 – Estratégia de busca na LILACS.....	89
Quadro 9 – Estratégia de busca no Medline via PubMed.....	90
Quadro 10 – Estratégia de busca no Web of Science.....	90
Quadro 11 – Repositórios com o termo “câncer” no resumo.....	92
Quadro 12 – Universo de pesquisa.....	99

LISTA DE GRÁFICOS

Gráfico 1 –	Tipos de repositórios de dados de pesquisa em câncer.....	61
Gráfico 2 –	Categorias de repositórios de dados de pesquisa em câncer.....	66
Gráfico 3 –	Abrangência geográfica dos repositórios de dados de pesquisa...	70
Gráfico 4 –	Publicações ampliadas em de repositórios de dados de pesquisa em câncer.....	75

SUMÁRIO

1	INTRODUÇÃO.....	15
1.1	Problema de pesquisa.....	21
1.2	Objetivos.....	22
1.2.1	OBJETIVO GERAL.....	22
1.2.2	OBJETIVOS ESPECÍFICOS.....	22
1.3	Justificativa.....	22
1.4	Estrutura do trabalho.....	25
2	FUNDAMENTAÇÃO TEÓRICA.....	27
2.1	Câncer e informação.....	27
2.2	Dados de pesquisa.....	37
2.3	Gestão de dados de pesquisa.....	44
2.3.1	REPOSITÓRIOS.....	46
3	METODOLOGIA.....	54
4	REPOSITÓRIOS DE DADOS DE PESQUISA EM CÂNCER NO R3 DATA.....	61
4.1	Tipos de repositórios de dados de pesquisa em câncer.....	61
4.2	Classificação de assunto.....	62
4.3	Abrangência geográfica.....	68
4.4	Tecnologias de repositórios de dados de pesquisa em câncer.....	71
4.5	Tipologia dos dados de pesquisa em câncer.....	76
5	CONSIDERAÇÕES FINAIS.....	79
	REFERÊNCIAS.....	83
	APENDICE A – LEVANTAMENTO BIBLIOGRÁFICO PRELIMINAR..	89
	APENDICE B – REPOSITÓRIOS LOCALIZADOS NO RE3DATA.....	92
	APENDICE C – PLANO DE GESTÃO DE DADOS DE PESQUISA.....	98

1 INTRODUÇÃO

Atualmente, a discussão sobre dados de pesquisa e questões que envolvem seu uso, geração e aplicabilidade virou tendência. Dados de pesquisa são produzidos amplamente e em escalas imensuráveis por áreas do conhecimento heterogêneas, que vão desde as Ciências Exatas e da Terra até as Artes. Na área da Saúde não é diferente, em um processo onde os dados são a base para obtenção de conhecimento em benefício da saúde pública, é inegável a importância dos dados de saúde coletados.

Dados de saúde são dados sensíveis e complexos que podem ser empregues não só para proporcionar transparência a administração pública, mas também podem ser usados na vigilância sanitária e epidemiológica, assim como no atendimento de baixa, média e alta complexidade. Os dados de pesquisa em saúde estimulam o trabalho cooperativo entre pesquisadores e podem ser reutilizados por áreas de pesquisa distintas. Somada a estas e outras razões, percebe-se que é irrefutável a importância e a urgência em conduzir estudos que se debrucem sobre o acesso, a disponibilização, a conservação, a preservação e o gerenciamento desses dados. Todavia, antes de debruçarmos sobre tais questões é preciso primeiramente compreender como os dados de saúde, que são produzidos amplamente e em escalas imensuráveis em âmbito internacional, estão representados em infraestruturas de acesso e gestão de dados de pesquisa.

De acordo com a Árvore de especialidades do conhecimento desenvolvida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (BRASIL, 2019), a área da Saúde possui nove subáreas das quais seis se subdividem formando novas especialidades. O fato de existirem subáreas dentro da grande área da Saúde, não impede que as subáreas se relacionem e trabalhem em parceria. Pelo contrário, o fato de existir cooperação é fator determinante para o desenvolvimento da Área como um todo. Nesse contexto, onde os dados são elementos revolucionários, pois operam nos mais variados domínios do conhecimento transformando saberes, o que se conhece sobre as particularidades dos dados de pesquisa em câncer abrigados em repositórios de dados?

Segundo a Organização Pan-Americana de Saúde (2018), o câncer é a segunda principal causa de morte no mundo, em uma escala onde um em cada seis óbitos decorrem da doença. O Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), estima para cada ano do triênio 2020-2022, a ocorrência de 625 mil casos novos de câncer (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2020, p. 25). Como consequência do alto índice de incidência da doença,

gastos com programas nacionais e regionais para o controle e prevenção de câncer geram grande e significativo impacto econômico.

O custo anual total da patologia câncer em 2010, de acordo com dados da Organização Pan-Americana de Saúde (2018), foi orçado em aproximadamente US\$ 1,16 trilhão. Estima-se que 70% das mortes ocasionadas pela doença ocorrem em países de baixa e média renda (ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 2018). Em decorrência de inferior poder econômico, além de problemas como: desemprego, desastres naturais, educação precária e fome, somam-se aos problemas enfrentados por países de baixa e média renda uma crise na saúde. Como exemplo da crise está o fato de somente um em cada cinco países, de baixa e média renda, possuem os dados necessários para conduzir uma política para o câncer (ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 2018).

A Constituição brasileira de 1988 considera a saúde um direito de todos e dever do Estado (BRASIL, 1988, art. 196). Nesse cenário está inserido o Sistema Único de Saúde (SUS), um sistema público de saúde brasileiro instituído pela Constituição Federal de 1988, no artigo 196 da Lei nº. 8.080/1990 (BRASIL, 2020). Financiado com recursos inerentes a União, Estados e Municípios e de outras fontes suplementares de financiamento, o SUS “é o único sistema de saúde pública do mundo que atende mais de 190 milhões de pessoas, sendo que 80% delas dependem exclusivamente dele para qualquer atendimento de saúde” (BRASIL, 2020). Quanto à amplitude dos serviços oferecidos pelo SUS à população, além de englobar serviços de atendimento de baixa média e alta complexidade, o SUS também oferece serviços de: urgência e emergência; atenção hospitalar; assistência farmacêutica e vigilâncias epidemiológica, sanitária e ambiental (BRASIL, 2020).

Presente em todo o território nacional, o SUS conta com sistemas de informação em saúde assistenciais e epidemiológicos alimentados por profissionais de saúde. Os sistemas de informação em saúde são importantes ferramentas de apoio à gestão para o planejamento e acompanhamento permanente de quadros de saúde. É por meio dos conjuntos de dados inseridos nos Sistemas que é possível analisar o perfil epidemiológico da população brasileira para que estratégias de prevenção e promoção de saúde, assim como novas tecnologias e procedimentos sejam incorporados. Isto é pesquisa.

A pesquisa científica é peça fundamental para manutenção e garantia do direito a saúde, pois viabiliza uma série de achados e inovações que contribuem para solução e

prevenção de problemas de saúde. Os achados e inovações obtidos por meio de pesquisa científica contribuem para a melhoria da qualidade de vida da população tanto no aspecto social quanto ao que se refere ao aspecto tecnológico e econômico, além de contribuir no desenvolvimento de pesquisas para a educação e atenção à saúde, servindo como estratégia de cidadania e evidenciando o papel decisivo da informação no ciclo que resulta na construção do conhecimento.

É através da informação que se faz ciência e ciência é conhecimento. Le Coadic (1996, p. 27) corrobora com o enunciado quando afirma que “[...] a informação é o sangue da ciência. Sem informação, a ciência não pode se desenvolver e viver. Sem informação a pesquisa seria inútil e o conhecimento não existiria”. Desta maneira, a busca pelo conhecimento só se inicia por meio do acesso ao conteúdo que está estabelecido e fundamentado, ou seja, pelo acesso a informação produzida por pesquisadores predecessores. Se comparado ao passado, atualmente a importância da informação é ainda mais eminente. O INCA (2020, p. 7), reforça a afirmativa ao destacar que na atualidade a informação é tida como o principal ativo da sociedade. Isto quando é baseada em dados válidos, confiáveis e atualizados. Dessa forma “ter a informação correta, no momento oportuno, faz a diferença para um planejamento estruturado e coerente com a realidade, permitindo ações eficazes e efetivas” (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2020, p. 7).

Dados são elementos que, entre outras medidas, quando bem indexados podem ser reinterpretados e transformados em itens informativos. No entanto, apesar da sua importância como vestígio do trabalho científico e do reconhecimento da urgência em conhecer, usar e preservar os dados oriundos da pesquisa em câncer, há escassez no que diz respeito a publicações que relacionem as temáticas. Em vista do constante aumento de casos da doença e sua ocorrência mundial, conduzir estudos que unam as temáticas traz contribuições para a memória científica, acesso, indexação e preservação dos dados coletados, assim como para o avanço na resolução de problemas de saúde. Contribuindo também para a ampliação da discussão sobre aprimoramento e políticas de recursos informacionais, que apoiam todo o ciclo da gestão de dados produzidos no decorrer da pesquisa. E que, além disso, são utilizados como ferramentas para integração, armazenamento e compartilhamento de dados de pesquisa em câncer.

Neste contexto de preservação, tratamento e acesso a fontes de informação estão às bibliotecas, que tradicionalmente são as instituições responsáveis por dar acesso aos

conteúdos informacionais. As bibliotecas na atualidade diferem das bibliotecas do passado em virtude do advento das Tecnologias da Informação e Comunicação (TIC), que vieram para refrear o problema consequente da excessiva difusão de materiais bibliográficos. As TIC possibilitaram a criação de diferentes recursos informacionais, recursos que deram novas opções de acesso à informação a usuários que por algum motivo não podem frequentar as bibliotecas a fim de saciar sua necessidade informacional.

Atualmente as bibliotecas são instituições “sem paredes”, pois o acesso aos itens não se dá exclusivamente na unidade de informação em questão. Hoje as bibliotecas mediam o acesso à informação que está na “palma da mão” do usuário. Apesar de oferecer auxílio para acesso a informação seja por meio de: bases de dados, bibliotecas virtuais, plataformas de pesquisa ou repositórios; as bibliotecas continuam sendo referência, espaço de pesquisa e aprendizado. As bibliotecas continuam insubstituíveis, mas com novas atribuições, se destacando e preservando a finalidade para a qual foram criadas. Isto pelo fato de se adequarem as mudanças advindas das novas tecnologias, sem perder características que fazem dela ambiente indispensável, que auxilia a busca por meio de orientação aos usuários e oferecendo novos e customizados serviços de informação. Para acompanhar novas tecnologias provenientes de avanços tecnológicos e guiar usuários e pesquisadores na busca por informação, atualmente conhecer plataformas digitais para depósito e obtenção de dados é tarefa que faz parte do dia-a-dia do profissional da informação. Profissional capacitado tanto no que diz respeito a orientação quanto a plataformas seguras para depósito e gerenciamento de dados tão sensíveis quanto os dados de pesquisa em câncer, quanto para a recuperação de dados de pesquisa em câncer que possuem potencial informacional, poupando tempo e recursos no desenvolvimento de novas pesquisas.

Cabe destacar que a preservação dos dados e o armazenamento dos mesmos, na estrutura ideal a qual se enquadra, não são ações importantes somente por permitir que outros pesquisadores reusem os dados disponibilizados. Tal atitude preserva os dados para que tenha potencial de reuso e acesso no futuro pelo próprio autor, que pode ou não o deixar “visível” a outros pesquisadores. Apesar de recomendada a divulgação e compartilhamento de dados é sabido que há, em alguns casos, restrições para a comunicação dos mesmos, seja por patentes ou por questões éticas.

Mesmo que isoladamente os dados de pesquisa não permitam compreensão sobre determinado fato ou assunto, ainda assim é importante preservá-los, pois são

considerados objetos informacionais. Objetos estes que quando contextualizados e representados de forma que possam ser interpretados, por pessoas ou lido por máquinas, adquirem valor informacional, isto porque possibilitam o reuso. A atribuição de metadados ao conjunto de dados é o que permite sua contextualização, fato que evidencia a ideia dos dados de pesquisa como elementos construídos por relações, sejam elas sociais, profissionais, tecnológicas ou científicas. Relações essas capazes de transformar estruturas cognitivas.

Quando se fala de dados de pesquisa em saúde, há de se considerar que o compartilhamento e reuso dos dados pode alterar a forma como a saúde evolui. Resultados de pesquisa científica utilizando dados de pesquisa vêm sendo divulgados, evidenciando a utilidade dos dados de pesquisa em saúde para descoberta e cura de patologias. Como exemplo do uso de dados de pesquisa em saúde na atividade científica está o caso “bebidas quentes e câncer”, onde por meio de pesquisa que reuniu 23 pesquisadores de 16 países, descobriu-se que o consumo de bebidas muito quentes pode facilitar o desenvolvimento de um tipo de câncer de esôfago (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2016). Outro caso onde o uso de dados de pesquisa em saúde foi benéfico à saúde pública foi para demonstrar a eficácia das leis de Ambientes Livres de Tabaco e Lei Antifumo. A descoberta se deu através dos dados de pesquisa de pesquisadores do INCA, do *Imperial College London*, do *Erasmus Medical Centre*, da *The International Union Against Tuberculosis and Lung Diseases* e da Universidade de São Paulo. Os pesquisadores “compilaram dados de mortalidade infantil em todos os municípios brasileiros de 2000 a 2016 e levantaram as diversas legislações de controle do tabagismo nos 26 estados brasileiros e no DF” (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2019a). Como resultado, verificou-se que em 16 anos, a restrição de fumar em ambientes públicos evitou a morte de 15 mil crianças no Brasil e caso a legislação fosse mais restritiva outras 10 mil vidas de menores poderiam ter sido salvas.

A apresentação de resultados de pesquisa oriundos da reutilização de dados de pesquisa em saúde evidencia a importância da pesquisa e seu potencial de gerar novos conhecimentos e recursos tecnológicos em prol da saúde pública. Seja na aplicação de novas técnicas em unidades de saúde, descoberta de cura para doenças ou para resolução de problemas sanitários, percebe-se que o reuso dos dados e o trabalho cooperativo entre pesquisadores de diferentes partes do mundo só trazem benefícios. Benefícios não só para a saúde, mas para o avanço social, ambiental, tecnológico,

econômico e da Ciência. Isto porque, entre outras vantagens, através do reuso de dados, pode-se averiguar a reprodutibilidade de técnicas e experimentos, assim como verificar e validar resultados de pesquisas científicas.

O acesso à informação científica também beneficia pacientes e acompanhantes, já que promove maior autonomia quanto à tomada de decisão compartilhada. O acesso e leitura de pesquisas médicas pode aproximar o corpo clínico dos pacientes, reduzindo a incerteza quanto a intervenções médicas e uso de medicamentos. Além de prover uma base para compreensão de orientações de saúde e fomentando responsabilidade em relação às escolhas e decisões sobre as suas próprias possibilidades de tratamento médico.

1.1 Problema de pesquisa

Dados de pesquisa são elementos específicos e que se caracterizam de forma diferenciada de acordo com o domínio, o objetivo da pesquisa e a interpretação de pesquisadores e grupos de pesquisa, que se debruçam sobre os mesmos. Dados de pesquisa em câncer, assim como qualquer outro domínio da saúde, possuem especificidades. Especificidades cujas a gestão de dados deve refletir essas peculiaridades.

A pesquisa oncológica objetiva descobrir a melhor maneira de tratar e prevenir o câncer. Identificando ações, tecnologias e medicamentos que possam abolir a doença ou oferecer melhora na qualidade de vida dos enfermos por meio de procedimentos mais precisos e menos invasivos.

Dados de pesquisa em câncer são fontes de informação com papel relevante na construção de conhecimento médico sobre uma pandemia com perspectiva anual de proliferação de casos. Além disso, envolvem questões legais, éticas e pessoais. Desta forma, dados de pesquisa em câncer, não podem ser tratados como dados de qualquer outro domínio da saúde.

A pesquisa na área do câncer, assim como os dados em câncer, tem características e metodologias próprias que precisam de reflexão específica para a gestão de seus dados. Nesse processo, o primeiro passo para um melhor aproveitamento dos dados gerados e utilizados é conhecer exatamente e profundamente quais são esses dados. Sendo assim, é necessário investigar como os dados de pesquisa em câncer estão registrados em repositórios digitais.

Desta maneira, processos de aprimoramento de ações de preservação, curadoria, comunicação e disponibilização de dados de pesquisa em câncer poderão ser aplicados e seus benefícios poderão ser sentidos não só para o avanço profissional ou da Ciência, mas para o progresso social, ambiental, econômico e principalmente da saúde pública mundial.

O levantamento bibliográfico prévio acerca de “dados de pesquisa” e “câncer” (apêndice A), realizado na Biblioteca Virtual em Saúde Prevenção e Controle de Câncer e nas bases de dados: LILACS, Medline via PubMed e *Web of Science*, revelou escassez quanto ao número de publicações que relacionassem as temáticas.

Frente à carência de estudos que apresentam as características dos dados de pesquisa oncológicos contidos em plataformas de gestão de dados, com o objetivo de

caracteriza-los para que sejam devidamente indexados e recuperados de forma eficaz, mantendo sua capacidade de reuso no futuro. Motivada, também, pelo crescimento de casos de compartilhamento e reuso de dados da área da saúde e pela falta de conhecimento acerca do panorama atual no subdomínio do câncer, a presente pesquisa se propõe a estudar a seguinte questão: como os dados de pesquisa em câncer estão representados nos repositórios de dados de pesquisa em escala global?

1.2 Objetivos

Os objetivos gerais e específicos que nortearam o presente estudo são apresentados nas próximas seções secundárias.

1.2.1 OBJETIVO GERAL

O objetivo geral do presente estudo é mapear e analisar os repositórios de dados de pesquisa em câncer registrados no Registry of Research Data Repositories - Re3data.

1.2.2 OBJETIVOS ESPECÍFICOS

No que se refere aos objetivos específicos, para alcançar o objetivo geral, este trabalho pretende:

- a) apresentar o estado da arte sobre dados de pesquisa em câncer;
- b) identificar os repositórios específicos para armazenamento de dados de pesquisa em câncer;
- c) identificar quais tipos de dados de pesquisa a área de Oncologia utiliza e produz para o desenvolvimento de novos conhecimentos.

1.3 Justificativa

Por ser um território plural, extenso e por possuir uma população com comportamentos e hábitos diversificados, a prevenção e controle de câncer no Brasil é um grande desafio para a saúde pública. De acordo com o INCA (2019) "câncer é o

nome dado a um conjunto de mais de 100 doenças que têm em comum o crescimento desordenado de células, que invadem tecidos e órgãos”.

Assim como as doenças cardiovasculares, o câncer compõe a lista de doenças e agravos não transmissíveis (DANT), que estão entre as principais responsáveis pelo adoecimento e óbito da população mundial (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2018b, p. 25). Esse impacto afeta principalmente os países de baixo e médio desenvolvimento, especialmente por mortes prematuras (WORLD HEALTH ORGANIZATION, 2013). Como já mencionado, segundo dados do INCA (2020, p. 25), estima-se no Brasil o aparecimento de 625 mil novas ocorrências de câncer, para cada ano do triênio 2020-2022. Por estar entre as principais causas de adoecimento e óbito da população mundial, é necessária adoção e expansão de ações que estimulem práticas de integração e compartilhamento de dados de pesquisa em câncer produzidos e/ou coletados em instituições de pesquisa e ensino em saúde.

O compartilhamento e estudo de dados de pesquisa em câncer contribuirão para que novas descobertas, decorrentes de pesquisa científica, sejam incorporadas em unidades de saúde com o objetivo de reduzir o avanço das taxas de incidência da doença. Além disso, o compartilhamento e integração dos dados também contribuirão para a redução do abismo entre o tempo que um achado científico é descoberto e sua aplicação em unidades de saúde. A propósito, também permitirão maior democratização de acesso aos dados tanto à comunidade científica quanto à sociedade em geral. Trazendo como benefício, maior transparência quanto ao uso de investimentos públicos em pesquisa científica em saúde, sobretudo, o avanço mais rápido, colaborativo e eficaz da Ciência, visando o bem-estar físico, social e mental da população.

A Ciência da Informação insere-se no contexto, pois é um campo voltado para os problemas de comunicação do conhecimento entre seres humanos. Ela investiga as propriedades e o comportamento da informação, tendo como um de seus objetivos tornar o conhecimento acessível (BORKO, 1968). Entre os processos inerentes a Ciência da Informação estão à produção, coleta, organização, armazenagem, interpretação, transmissão, transformação e utilização da informação (BORKO, 1968, p. 3).

Visualiza-se a temática: gestão e compartilhamento de dados de pesquisa sobre câncer, relevante para os estudos desenvolvidos no percurso do campo da Ciência da Informação, visto que elucida o caráter interdisciplinar da Área. Que ao contrário do

que era popular no passado, não só tem capacidade de relação com a Biblioteconomia ou a Informática, mas pode se aproximar de outras áreas do conhecimento. Fortalecendo, também, a ideia do Campo como forma de prover e aperfeiçoar o processo de transmissão da informação. Este fato reitera a ideia da Ciência da Informação como um campo que veio para transpor e provocar sua amplitude por meio da relação com outras Ciências. Não se limitando somente a seus paradigmas, mas dirigindo sua pesquisa para um desvendamento do mundo. Pinheiro e Loureiro (1995, p. 4) corroboram com a ideia ao declarar que a Ciência da Informação “se constrói por abordagens estratégicas voltadas para a solução ou trato de problemas”, característica que torna viável a relação da Área com qualquer outra área do conhecimento que em determinado momento desponte um paradigma informacional.

Ademais, ao revelar os dados de pesquisa produzidos e utilizados para pesquisa em câncer, torna-se exequível investigar mais profundamente as propriedades dos dados, não só considerando o contexto onde esses dados são coletados e/ou gerados, mas também o aprimoramento de serviços e desenvolvimento de produtos. Fortalecendo, também, a ideia do Campo como forma de prover e aperfeiçoar o processo de transmissão da informação, contribuindo para que os dados de pesquisa, ao serem reusados e reinterpretados forneçam novas descobertas terapêuticas e tecnológicas.

É por meio da reinterpretação e reuso dos dados de pesquisadores predecessores que é viável o surgimento de novas descobertas terapêuticas e tecnológicas, que repercutem em avanços sociais, econômicos e de saúde. Pesquisadores predecessores e suas equipes, ao disponibilizarem seus dados, contribuem para que no futuro, novas descobertas científicas sejam aplicadas em unidades de saúde. Facilitando o acesso ao volume cada vez mais crescente da produção do conhecimento no mundo e colaborando para um melhor entendimento dos fenômenos de saúde, sociais, espaciais, históricos, econômicos e culturais.

A motivação pessoal para elaboração deste trabalho se deu através da trajetória profissional e acadêmica da autora. Foi durante a graduação em Biblioteconomia e Gestão de Unidades de Informação, que a mesma foi apresentada as temáticas: informação oncológica e gestão de dados de pesquisa, por meio de estágio e iniciação científica, respectivamente. A experiência com a temática informação oncológica veio por meio de estágio em Biblioteconomia, na Biblioteca do Hospital Central I do Instituto Nacional de Câncer José Alencar Gomes da Silva. Já a experiência com gestão

de dados de pesquisa se deu por meio de iniciação científica, na participação como bolsista em Biblioteconomia no Serviço de Biblioteca e Informação Científica e Tecnológica (SEBICT), do Instituto de Engenharia Nuclear. Tais experiências motivaram a autora a ingressar no Curso de Especialização em Informação Científica e Tecnológica em Saúde na Fundação Oswaldo Cruz, que contribuiu para a reflexão sobre a importância e aprimoramento de técnicas para gestão, organização, preservação e análise de informação em saúde.

1.4 Estrutura do trabalho

Quanto à estrutura deste estudo, o trabalho é composto por cinco seções primárias textuais.

A primeira seção “Introdução” contextualiza a pesquisa e está dividida em seções secundárias. Na seção secundária “Objetivo geral” é apresentada a ação que conduz à resolução do problema de pesquisa. Na seção secundária “Objetivos específicos” são apresentadas, de forma detalhada, as ações propostas para chegar o objetivo geral. Na seção secundária “Justificativa” encontram-se os motivos que levaram a elaboração presente estudo, sua importância para o campo da Ciência da Informação, assim como sua relevância social e científica.

A segunda seção “Fundamentação teórica” apresenta o referencial teórico e assuntos relevantes que serviram como base para a construção da pesquisa. Além de oferecer novas informações, a leitura destes materiais possibilitou uma visão geral sobre o problema de pesquisa e a descoberta de trabalhos citados nas seções secundárias indicadas a seguir. Na seção secundária “Câncer e Informação” questões como o que é o câncer, as causas da patologia e sua mortalidade são respondidas, além disso a seção secundária declara o poder da informação como elemento capaz de contribuir para a melhoria das condições de saúde populacional. Desta forma, demonstra a possibilidade de relação entre as temáticas câncer e informação, onde só há desenvolvimento no campo da Saúde por meio do acesso e análise de evidências científicas, que por sua vez são elementos informativos. A seção secundária “Dados de pesquisa” apresenta os dados de pesquisa, seu significado e importância, assim como seus benefícios no ciclo da informação. A seção secundária “Gestão de dados de pesquisa” relaciona os dados de pesquisa a infraestruturas de gestão e compartilhamento dos dados. A seção terciária “Repositórios” discorre sobre o que são os repositórios, apresenta suas características e

os revela como infraestruturas digitais utilizadas para gestão e compartilhamento de dados de pesquisa na atualidade.

Já a terceira seção “Metodologia” indica os procedimentos metodológicos adotados para a elaboração deste estudo, assim como seu delineamento, técnicas de coleta e análise de dados.

A quarta seção “Repositórios de dados de pesquisa em câncer no R3data” apresenta os resultados obtidos após identificação dos repositórios de dados de pesquisa em câncer indexados no R3data.

A sexta seção “Considerações finais”, apresenta as conclusões da pesquisa. Por meio dela o leitor é convidado a uma reflexão, a partir, da comparação entre o momento atual onde enfrentamos uma pandemia e a função da informação em saúde no que diz respeito ao enfrentamento da COVID-19. A ideia é lembrar que o câncer também é uma pandemia e assim como tem sido feito em relação a COVID-19, dados de saúde podem e devem ser usados em processos de pesquisa que conduzem a novas descobertas terapêuticas e tecnológicas.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção primária são apresentadas as referências que serviram para embasar esta pesquisa.

2.1 Câncer e Informação

Quanto a sua etimologia, a palavra câncer advém do grego *karkínos* e foi utilizada pela primeira vez por Hipócrates, considerado o pai da medicina (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2011, p. 17; WIKIPEDIA, 2019a). Também conhecido como neoplasia ou tumor, o termo câncer não se refere a uma nova patologia, mas sim a uma doença que atinge o ser humano há mais de 3 mil anos antes de Cristo. Um dos mais antigos registros sobre a doença foi obtido em amostras de múmias egípcias, fato que comprova o quão vetusto é a comparência da doença na vida humana (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2011, p. 17).

O vocábulo câncer remete a um grupo de mais de 100 doenças, que têm como característica geral o crescimento contínuo e desordenado de células, que propendem a invadir tecidos e células do corpo (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2019b). Segundo o Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), a maioria das células do corpo cresce, se multiplica e morre de forma ordenada, constituindo-se como um processo contínuo considerado natural (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2011). Quanto ao crescimento de células cancerosas:

O crescimento das células cancerosas é diferente do crescimento das células normais. As células cancerosas, em vez de morrerem, continuam crescendo incontrolavelmente, formando outras novas células anormais. Diversos organismos vivos podem apresentar, em algum momento da vida, anormalidade no crescimento celular – as células se dividem de forma rápida, agressiva e incontrolável, espalhando-se para outras regiões do corpo – acarretando transtornos funcionais. O câncer é um desses transtornos (INCA, 2011, p.13-14).

Presente no grupo de doenças e agravos não transmissíveis (DANT), que são as principais responsáveis pelo adoecimento e óbito da população mundial, o câncer é a segunda causa de mortes no mundo e pode surgir em qualquer parte do corpo

(ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 2018; INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2018b, p. 26). Segundo a União Internacional para o Controle do Câncer (UICC), no ano de 2019, aproximadamente 9 milhões de pessoas morreram em decorrência do câncer e caso a patologia não seja controlada, a previsão de mortes em razão da doença aumentará para 13,2 milhões por ano até 2030 (UNION FOR INTERNATIONAL CANCER CONTROL, 2019).

O câncer é uma questão complexa de saúde pública mundial em função de sua extensão epidemiológica e percentual de prevenção, isto quando se considera que cerca de um terço dos novos casos poderiam ser evitados (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2011). No Brasil não é diferente, talvez a situação seja até pior, já que a complexidade da doença é ampliada por tratarmos de um território plural onde habita uma população com comportamentos e hábitos diversificados (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2018b, p. 26). Tais características corroboram para o fato da prevenção e controle de câncer no Brasil tratar-se de grande desafio para a saúde pública nacional.

Apesar da possibilidade de aparecimento do câncer em qualquer parte do corpo, alguns órgãos são mais afetados que outros, sendo que cada órgão pode ser acometido por diferentes tipos de tumores que podem ser mais ou menos agressivos. O câncer de pulmão é o mais comum de todos os tumores malignos, com registro de aumento de 2% ao ano na incidência mundial (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2011, p. 13). Já o câncer de mama é o segundo mais frequente no mundo, sendo mais comum entre as mulheres (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2019c). O câncer de pele não melanoma é o mais frequente no Brasil, correspondendo a quase 30% dos tumores malignos registrados em território nacional (INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2018c).

Quanto às causas para o surgimento do câncer, fatores externos e internos são apontados como incitadores para o seu desenvolvimento. O Descritor em Ciências da Saúde (CENTRO LATINO-AMERICANO E DO CARIBE DE INFORMAÇÃO EM CIÊNCIAS DA SAÚDE, 2019), que é um vocabulário controlado das Ciências da Saúde criado pelo Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME), apresenta o termo “fatores de risco” como:

Aspecto do comportamento individual ou do estilo de vida, exposição ambiental ou características hereditárias ou congênitas que, segundo evidência epidemiológica, esta sabidamente associado a uma condição de saúde considerada importante de se prevenir (CENTRO LATINO-AMERICANO E DO CARIBE DE INFORMAÇÃO EM CIÊNCIAS DA SAÚDE, 2019).

Sedentarismo, tabagismo, alimentação, consumo de bebidas alcoólicas, radiação e vírus estão entre os fatores de risco para o câncer, fatores estes que podem agir em conjunto ou em sequência (FUNDAÇÃO DO CÂNCER, 2019; INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2018a). Muitos destes fatores de risco podem ser evitados ao adotarmos hábitos saudáveis e ao evitarmos ambientes sociais onde há contato com elementos prejudiciais a saúde, como por exemplo, ao que ocorria antes da Lei Rio Sem Fumo.

A Lei Rio Sem Fumo ou Lei 5517/09, de 17 de agosto de 2009, vigora até hoje e criou ambientes de uso coletivo livres de tabaco no Rio de Janeiro (BRASIL, 2009). Antes as pessoas que optavam por não fumar tinham que contar com a sorte ao frequentar ambientes de uso coletivo, como por exemplo: restaurantes e transportes públicos. Isso porque não havia leis que regulassem sobre o que diz respeito ao consumo de derivados do tabaco e, conseqüentemente, ao fumo passivo em ambientes de uso coletivo. Por meio de evidências científicas que comprovam o risco a saúde e provam que não existem níveis seguros de exposição ao tabaco e seus derivados, foi possível criar uma lei que protege os cidadãos não fumantes do tabagismo passivo em ambientes de uso coletivo (BRASIL, 2019).

O exemplo da Lei Rio Sem Fumo exemplifica a popular afirmação que diz “a informação orienta a prática científica e clínica”. Aliás, não só exemplifica, mas também expande. Isto porque, no exemplo da Lei mencionada, vimos à informação não só como orientadora da prática científica e clínica. No caso apresentado, a informação foi fundamental para mudanças de comportamentos sociais que foram impulsionadas e reguladas por leis, contribuindo para a saúde coletiva.

E é como elemento dotado de capacidade para propiciar a melhoria das condições de saúde, que a temática câncer se alinha a informação. Aliás, não só nessa perspectiva, pois a informação tem capacidade imensurável de transformar e reorganizar culturalmente o homem, incluindo sua saúde e hábitos. Fato que traz a tona, a informação, como a mais potente e avassaladora força de transformação invisível

conhecida (ARAÚJO, 1991, p. 37 apud TARGINO, 2000, p. 5). Sobre o poder avassalante da informação no cotidiano:

[...] impõe-se como a mais poderosa força de transformação do homem, aliando-se aos modernos meios de comunicação para conduzir o desenvolvimento científico e tecnológico das nações, por meio da transferência de informações ou difusão de novas ideias e tecnologias. Insere-se no processo desenvolvimentista, configurando e fortalecendo a relação informação versus avanço social (TARGINO, 2000, p.5).

Mas afinal, o que é informação? Informação é:

1. Ato ou efeito de informar (-se); informe. 2. Fatos conhecidos ou dados comunicados acerca de alguém ou algo. 3. instrução. 4. Tudo aquilo que, por ter alguma característica distinta, pode ser ou é apreendido, assimilado ou armazenado pela percepção e pela mente humanas. 5. Qualquer sequencia de elementos que produz determinado efeito e, também., transmite e armazena capacidade de produzir tal efeito: informação genética. 6. Fato de interesse específico, conhecido graças a observação, pesquisa e análise (FERREIRA, 2004, p.748).

Em visão que reflete o entendimento do termo de acordo com os princípios da Ciência da Informação, Buckland (1991) esclarece que informação é estar informado, é a redução da ignorância e da incerteza. Para Vitorino e Piantola (2011, p. 101), a informação é condição de permanência e instrumento de mudança, destacando-se como componente fundamental para o exercício da cidadania no contexto democrático.

A informação é considerada como um elemento fundamental para avaliação e monitoramento das ações realizadas no setor saúde e está presente em dados demográficos e epidemiológicos disponibilizados por pesquisas científicas e tecnológicas, sistemas de informação, bases de dados, aparato informático, processamento de dados, conjunto de indicadores e relatórios vinculados à gestão (ARAÚJO; CARDOSO, 2007 apud LOFEGO; PINHEIRO, 2013, p.1).

Lofego e Pinheiro (2013, p. 1) acrescem a importância da informação ao afirmar que, no contexto da saúde, “a informação é considerada como um elemento fundamental para avaliação e monitoramento das ações realizadas e está baseada em indicadores, alimentados pelos sistemas de informação”. Já no contexto câncer, Lofego e Pinheiro (2013, p. 2) afirmam que “a comunicação e a informação são consideradas importantes estratégias na condução das políticas públicas do Programa, especialmente no que diz respeito ao estímulo à adesão de mulheres e profissionais de saúde”. Nessa perspectiva,

o uso crítico da informação no contexto da saúde e, conseqüentemente, do câncer se justifica a fim de saciar a necessidade informacional de pesquisadores e do corpo clínico em benefício da melhoria das condições de saúde da sociedade como um todo.

Já sabido o uso crítico da informação em saúde com o intuito de satisfazer a necessidade informacional, cabe destacar que a pesquisa bibliográfica contribui para a formação de profissionais especializados promovendo o desenvolvimento de pesquisa e produção de informação epidemiológica.

Nessa concepção, os dados gerados:

“se transformam em conhecimento para os gestores, coordenadores e profissionais de saúde na recondução de procedimentos ou reorientação das ações, tanto no nível central quanto local, mas são pouco apropriados pela população” (LOFEGO; PINHEIRO, 20013, p. 4).

Vitorino e Piantola (2011, p. 101), enfatizam que “o acesso a esse grande volume de dados apenas se constituirá como benefício – a informação – quando seus usuários os utilizarem para resolver ou minimizar questões informacionais”. Sendo assim, torna-se evidente que a cidadania e o desenvolvimento profissional, por meio de acesso a informação, não se constrói somente pelo acesso material à informação. Mas deve compreender também a capacidade de interpretação da realidade e de construção de significados pelos indivíduos (VITORINO; PIANTOLA, 2011, p. 101). Para isso, há de se adquirir habilidades para encontrar, avaliar e utilizar de modo apropriado à informação de que se necessita, é necessário ser competente informacional para fazer um bom uso dos instrumentos disponíveis.

Segundo Vitorino e Piantola (2011, p. 101), "o indivíduo competente informacional é aquele capaz de encontrar, avaliar e usar a informação eficazmente na solução de problemas e na tomada de decisão". Eisenberg (2008 apud VITORINO; PIANTOLA, 2011, p. 102) assegura que a competência informacional consiste numa “série de habilidades e conhecimentos que nos permitem encontrar, avaliar e usar a informação de que precisamos, assim como filtrar a informação de que não necessitamos”. As aptidões inerentes à noção de competência em informação asseguram o usuário a navegar pela informação disponível em qualquer tempo. Inclusive na busca de informação, por meio de novas tecnologias (VITORINO; PIANTOLA, 2011, p. 101). Desse modo o indivíduo competente em informação, motivado por razões e conceitos particulares, tem autonomia na busca por respostas e questões dirigidas para

resoluções de problemas de qualquer âmbito, gerando consequências éticas, científicas e sociopolíticas.

Em busca de auxílio a fim de atenuar sua necessidade informacional, as bibliotecas especializadas são os lugares para os quais pesquisadores em saúde e o corpo clínico se dirigem. Apesar de serem especialistas, estes usuários, sozinhos, não conseguem localizar as informações que precisam para dar continuidade em seus trabalhos. Assim, o levantamento de fontes de informação fica a cargo do bibliotecário e o especialista médico se ocupa de analisar e interpretar os conceitos das ideias contidas nos trabalhos recuperados (LIMA, 1973, p.149-150).

De acordo com Figueiredo (1979, p. 10), as bibliotecas especializadas, na forma como se conhece hoje em dia, começaram a surgir no começo do século XX, em resposta ao avanço crescente nas áreas de ciência e tecnologia. Quanto à razão de existirem bibliotecas em hospitais:

Para obter a necessária informação básica é preciso uma combinação de esforços, técnicas e recursos financeiros – enfim, a criação de uma unidade dentro da instituição, com a responsabilidade de localizar e tornar acessível a qualquer material bibliográfico que possa contribuir para o avanço científico dessa instituição (LIMA, 1973, p. 142).

Segundo Rothstein (apud GROGAN, 1995, p. 1), serviço de referência é a assistência pessoal prestada pelo bibliotecário aos leitores em busca de informação. É um serviço de respostas que atende o anseio de conhecer do usuário.

Serviço de referência em hospitais é um serviço de respostas, onde devido à urgência e ao fato de não conseguir acompanhar sozinho a contínua evolução da ciência e de sua aplicação em seu campo de trabalho o corpo clínico recorre ao bibliotecário[...] O espaço em que as descobertas são feitas é na biblioteca e o hospital atua como centro de treinamento e aperfeiçoamento profissional, onde os usuários irão aplicar o que descobriram na biblioteca. Os usuários testam, reformulam e transformam estas descobertas em novos conhecimentos (LIMA, 1973, p. 141-143).

Já a biblioteca em um hospital:

A biblioteca é aquele departamento de um hospital investido da responsabilidade e autoridade para assegurar ao pessoal docente, clínico, pesquisador, auxiliar e administrativo o acesso à informação, com finalidade de habilitá-lo a prover o melhor cuidado possível aos pacientes, dentro de suas limitações de recursos (JOHNSON, 1967 apud LIMA, 1973, p.143).

Nessa perspectiva destaca-se que:

Em qualquer situação ou localização, no entanto, os objetivos das bibliotecas especializadas, conforme estabelecido pela Especial Libraries Association são: adquirir, organizar, manter, utilizar, e disseminar materiais relacionados às atividades da organização a que pertencem (FIGUEIRERO, 1979, p. 11).

Além de proporcionar aos usuários acesso as bases de dados especializadas, as bibliotecas atuam como agentes que disseminam conhecimentos necessários aos estudos e tomadas de decisões das instituições a que pertencem (MIRANDA, 2007, p. 88). Nas bibliotecas, o bibliotecário fornece as informações que o usuário solicita sem exigir do mesmo que se torne autossuficiente (GROGAN, 1995, p. 9), mas afirmando seu papel de educador ao ensinar o usuário a fazer pesquisas em uma base de dados, por exemplo.

O bibliotecário é o profissional que atua como mediador entre o usuário e o problema da informação apresentado. A expressão “bibliotecário de referência” foi empregada por volta de 1988 por Melvil Dewey (GROGAN, 1995, p. 12), o bibliotecário de referência é o profissional responsável por prestar assistência ao usuário. De acordo com Accart (2012, p. 84) “é um profissional com educação superior, treinado em situações de trabalho diferentes que lhe permitem resolver as dificuldades encontradas”. Na prática atuar na função de referência é empregar seu conhecimento técnico a determinadas tarefas, com o objetivo de satisfazer ao máximo a busca do usuário (ACCART, 2012, p. 81). Mesmo quando não soluciona o problema, o bibliotecário do serviço de referência é quem apresenta ao usuário uma gama de opções, indicando fontes onde assuntos relacionados possam ser encontrados. Isto porque, sozinhos, os catálogos de biblioteca e de acesso em linha não satisfazem a necessidade informacional do usuário. Os catálogos, muitas vezes, remetem a informações bibliográficas inexatas ou defeituosas, impedindo que o mesmo localize “obras alternativas” que abordam o mesmo assunto buscado inicialmente. Somente o bibliotecário possui habilidade e domínio dos instrumentos bibliográficos e de referência (GROGAN, 1995, p. 10). Mesmo um usuário bem familiarizado com o material e com o assunto, não obtém tanto êxito ao localizar informações, principalmente por não possuírem muita paciência acabam desanimando ou desistindo da busca. O bibliotecário de referência na Biblioteca Nacional da Suíça:

Possui conhecimentos em quase todos os campos do conhecimento, não necessariamente profundos, mas antes gerais, sabe onde pode

encontrar informações suplementares (em livros, bases de dados ou com colegas). Um pouco ‘Sherlock Holmes’, não para enquanto não encontrar uma resposta, seja ela positiva ou negativa (ACCART, 2012, p. 135).

Na visão do usuário o profissional de referência tem a função de conselheiro e formador, mediando à necessidade de informação oriunda do usuário e as fontes de informação (ACCART, 2012, p. 127).

É válido salientar que as formas de trabalho de bibliotecários em hospitais não se limitam somente a sua presença executando “serviços tradicionais“ em bibliotecas. Hoje em dia, segundo Finamor e Lima (2018, p. 110), bibliotecários são integrantes de equipes multiprofissionais dos hospitais, apresentando-se a equipe de forma rotineira, utilizando e criando serviços de informação personalizados. Em ambientes hospitalares, bibliotecários membros de equipes multiprofissionais atuam também, coletando informações de pacientes para o apoio e a tomada de decisão dos médicos, assim como na elaboração de pesquisas médicas, coleta e localização de evidências médicas (FINAMOR; LIMA, 2018, p. 110).

O bibliotecário ao integrar equipes multiprofissionais leva um novo olhar sobre os dados obtidos, se destacando no domínio da saúde ao poupar o tempo de profissionais de saúde e acelerando o processamento de protocolos e tomada de decisão médica com base em evidências (FINAMOR; LIMA, 2018, p. 111). Seja através da facilitação ao acesso a literatura médica para responder perguntas clínicas, assim como: pela coleta de informações relevantes; pesquisa; organização; interpretação; armazenamento; recuperação; disseminação e tratamento da informação de forma sistemática ou estruturada (FINAMOR; LIMA, 2018, p. 110-111). Desta forma, o profissional bibliotecário mostra-se primordial no que diz respeito aos avanços científicos alcançados na área da saúde, tendo responsabilidade no processo onde dados científicos, ao serem reinterpretados e reusados, transformam-se em conhecimento e/ou tecnologia que beneficiam a saúde populacional.

Contudo, quais são as fontes de informação utilizadas para obtenção de informação em saúde? Para obter respostas a suas questões de pesquisa, pesquisadores, cientistas e o corpo clínico devem realizar busca por informações biográficas e documentais. As fontes de informação primárias, secundárias e terciárias, que se diferem pela originalidade dos conteúdos dos documentos, registram informações frutos do conhecimento de autores predecessores. Para Rodrigues e Blattmann (2014, p. 10),

fontes de informação são tudo o que gera ou transmite informação, ou seja, “[...] qualquer meio que responda a uma necessidade de informação por parte de quem necessita”. Posteriormente às fontes de informação já citadas, Molino, Melo e Ribeiro (2017, p. 7) destacam os benefícios da internet como atenuadora para a promoção da saúde:

Sabendo que a internet é uma poderosa ferramenta comunicadora em saúde, principalmente no contexto de promoção da saúde, sítios de saúde com boa qualidade são importantes tanto para profissionais de saúde quanto para a população geral, ajudando-os a entender a organização do sistema público de saúde, aprender sobre os programas de medicamentos e seus direitos, e melhorar a relação profissional-paciente (MOLINO, MELO; RIBEIRO, 2017, p. 7).

Apesar da informação registrada em papel não ter sido abandonada no percurso da evolução do conhecimento científico, sendo constantemente utilizada por pesquisadores no desenvolvimento de seus trabalhos, a tecnologia ampliou a variedade de fontes de informação utilizadas antes do surgimento das TIC. Bibliotecas virtuais, repositórios, bases de dados, plataformas digitais e periódicos eletrônicos estão entre as “novas fontes de informação” utilizadas na atualidade. Dentre essas fontes de informação, fruto da contemporaneidade, no campo da saúde destaca-se as bases de dados clínicos.

A existência de base dados clínicos permite a aplicação de poderosas técnicas de mineração em análises epidemiológicas, testes clínicos e Genômica, permitindo estabelecer associações entre doenças e novos tratamentos para grupo de pacientes com padrões genéticos específicos. Experimentos em Genômica, Proteômica e Dinâmica Molecular estão constantemente produzindo novos dados, correlacionando-os, por meio de análise, aos dados já existentes, o que requer sofisticadas ferramentas de busca e *screening* (PITASSI; GONÇALVES; MORENO JÚNIOR, 2014, p. 258-259).

Os levantamentos bibliográficos realizados no domínio da saúde, geralmente, resultam em artigos científicos e estudos que trazem prestígio ao autor e a instituição a que pertence, contribuindo para a população como um todo, independente do suporte em que as informações estejam armazenadas. Isto porque as informações e dados coletados são reusados e transformados pelos pesquisadores em novos conhecimentos contribuindo para o desenvolvimento científico e prático. Fato que pode levar ao desencadeamento de tratamentos, evolução de procedimentos e até a extinção de doenças.

A revolução do pensamento científico permite que outras áreas do conhecimento abordem o campo da oncologia, trazendo um olhar multidisciplinar sobre o objeto de pesquisa câncer. Fato que amplia as possibilidades de investigação sobre o tema, não o abordando somente ao que se refere à saúde pública, mas a seu impacto social, assim como para o indivíduo e para a coletividade. A ideia de que a partir do conhecimento construído coletivamente haverá revolução científica no campo de investigação, é aclarada no seguinte trecho:

Assim, as investigações culminarão na substituição total ou parcial de um paradigma, resultando no desenvolvimento da ciência, entendendo-a como não resultante de acúmulo de descobertas isoladas, ou seja, desconstruindo a ideia de linearidade da construção do conhecimento em qualquer campo de investigação (RAMOS, 2013, 474).

Desta forma conclui-se que a pesquisa em saúde, assim como outras áreas do conhecimento, só é exequível por meio do acesso a informação e dados pertinentes. Através do uso e acesso a informação obtêm-se benefícios coletivos e individuais, para a sociedade e para quem se dedica a pesquisa. Isto porque, mesmo que não traga “resultados efetivos” ou “palpáveis” em forma de novos tratamentos, extinção de doenças ou criação de novos medicamentos, beneficia cientistas e pesquisadores com adição de saber e conhecimento.

2.2 Dados de pesquisa

É inegável que o incremento das TIC na sociedade transformou o mundo e a forma como as pessoas se comunicam. Adicionando velocidade às descobertas científicas, expandindo as formas de comunicação e gerenciamento de informação, as TIC modificaram também os processos inerentes ao fazer científico. Neste cenário de mudança, o surgimento da internet foi responsável pela potencialização da capacidade de inovação advinda das TIC. Contudo, há mais um elemento responsável pelos avanços inferidos: a informação.

Para aprimorar o que já se conhece é necessário conhecer o que já foi concebido. Neste contexto, a informação se destaca como protagonista na busca por novos conhecimentos, pois só é possível inovar ou gerar novos conhecimentos por meio do acesso à informação já produzida. Atualmente, em unidades de informação, não são raras as buscas por fontes de informação além das tradicionalmente conhecidas e disponíveis, como por exemplo: livros, teses, dissertações e periódicos. Pesquisadores e usuários têm solicitado dados de pesquisa, pois precisam não só dos resultados finais dos estudos, mas também do acesso aos dados que serviram para a estruturação do paradigma corrente.

Assim como ocorre com o conceito de informação, conceituar dados de pesquisa também é uma tarefa complexa, Sales e Sayão corroboram com a obscuridade da ação ao revelarem que dados de pesquisa podem ser coisas diferentes para diferentes pessoas e circunstâncias (2019, p. 35). Mas afinal, o que são dados de pesquisa?

Segundo relatório da Organização para Cooperação e Desenvolvimento Econômico (2007, p. 13 apud SAYÃO; SALES, 2019, p. 35), dados de pesquisa são “registros de fatos usados como fontes primárias na investigação científica e que geralmente são aceitos na comunidade científica como necessários para a validação dos resultados da pesquisa.”. Em uma perspectiva simplista, a Universidade Federal de São Carlos (2019) e a Fundação Oswaldo Cruz (2018, p. 1) definem dados científicos como registros realizados, coletados ou gerados durante uma pesquisa. Bertin, Visolli e Drucker (2017, p. 38), também de forma singela, definem os dados como “todo o tipo de registro produzido, compilado ou utilizado no decorrer da pesquisa”. Em uma visão mais ampla e robusta, que se assemelha a definição dada pela Organização para Cooperação e Desenvolvimento Econômico, Sales e Sayão (2019, p. 36) definem dados de pesquisa como “[...] todo e qualquer tipo de registro coletado, observado, gerado ou

usado pela pesquisa científica, tratado e aceito como necessário para validar os resultados da pesquisa pela comunidade científica”.

Dados de pesquisa podem ser gerados ou coletados por pesquisadores, um grupo ou núcleo de pesquisa. Uma de suas principais características é que são dependentes de interpretação, desse modo os envolvidos na sua criação e coleta atuam como interpretadores de dados científicos (SALES; SAYÃO, 2019, p. 35). Outra característica atribuída aos dados é que os mesmos possuem vida útil maior do que os projetos que os originam. Desta forma, os pesquisadores podem continuar a trabalhar com os dados também após o encerramento dos projetos e até mesmo iniciar um novo projeto a partir dos mesmos dados (BERTIN; VISOLLI; DRUCKER, 2017, p. 42-43). Para isso é essencial que os dados de pesquisa sejam curados e bem gerenciados, para manter seu potencial de fonte de informação e permitir que outros pesquisadores também o reutilizem em suas pesquisas. Sales e Sayão (2019, p. 34), apontam como necessário para viabilizar o reuso de dados de pesquisa estudar as propriedades dos dados em questão, identificando como eles se apresentam em cada domínio. A partir deste estudo podem ser construídos critérios de seleção que farão dos dados reutilizáveis (SALES; SAYÃO, 2019, p. 34).

O fato do conceito de dados de pesquisa se modificar para cada área do conhecimento não impede que dados produzidos em um domínio possam ser utilizados por outro. Como exemplificam Bertin, Visolli e Drucker (2017, p. 38), que dizem que “uma fotografia de uma construção municipal depositada em um arquivo histórico, [...], pode não representar muito para um agrônomo; enquanto que, para um historiador, aquela fotografia torna-se um dado de pesquisa”. Ou seja, o que para um pesquisador representa nenhuma informação ou até uma interferência, outro pesquisador pode interpretar como um vestígio. Esse potencial informativo dos dados de pesquisa só reforça a importância do seu gerenciamento, compartilhamento e preservação (BERTIN; VISOLLI; DRUCKER, 2017, p. 39). Ações essas que propiciam o reuso dos dados tanto para o pesquisador criador dos dados como para outros pesquisadores, que empreenderão novas pesquisas e pouparão esforços, e recursos, que seriam gastos na coleta de dados.

É sabido que dados de pesquisa são para serem reusados. O reuso dos dados é visto como oportunidade para promover a colaboração entre pesquisadores, o avanço no conhecimento e a elaboração de elucidações que atendam às demandas da sociedade (FUNDAÇÃO OSWALDO CRUZ, 2018). Vistos como base do conhecimento

científico, tecnológico e médico, dados de pesquisa não são simples de organizar, descrever e disponibilizar, para que sejam compreensíveis agora e no futuro (UNIVERSIDADE DE SÃO PAULO, 2016). Mediante tamanha complexidade, que tipos de materiais podem ser considerados dados de pesquisa?

Para Grogan (1970, p. 14), entre os tipos de dados de pesquisa estão: arquivos pessoais e de empresas, correspondências e relatórios de pesquisa internos, assim como cadernos, agendas e memorandos de laboratórios. Sales e Sayão (2019, p. 38, p. 39-40), complementam a lista adicionando números, imagens, textos, vídeos, áudio, software, algoritmos, equações, animações, simulações, fórmulas, filme, fotografia, códigos nucleares, tabelas, diagramas, modelos em 3D, desenhos, transcrição, certificados, cadernos de campo, espécimes, amostras, maquetes, protocolos e outros.

A Universidade De São Paulo (2016) acrescenta a lista de dados de pesquisa: fatos e estatísticas recolhidas para posterior referência ou análise, documentos (texto, Word), planilhas (Excel, etc.), cadernos de laboratório, questionários, sequências de proteínas ou genéticos, respostas de teste, slides, artefatos, coleção de objetos digitais adquiridos e gerados durante o processo de pesquisa, scripts, arquivos de log, software de simulação, metodologias e fluxos de trabalho. Além dos citados, cabe destacar que no domínio da saúde, prontuários e imagens histopatológicas também podem ser considerados dados de pesquisa, já que são elementos que podem ser usados para pesquisa, inclusive em domínios diferentes dos quais os dados foram originados.

De acordo com Sales e Sayão (2019, p. 45), é importante caracterizar os dados, tal caracterização permite identificar metadados específicos ao tratamento de cada tipologia de dados. Os autores também indicam que os dados podem ser sistematizados pela sua origem ou natureza em: observacionais, computacionais e experimentais. Onde:

Os dados observacionais são originados de observações de fenômenos e eventos únicos. São dados que não podem ser coletados uma segunda vez, e, portanto, devem ser arquivados para sempre, com a nível necessário de fidedignidade, autenticidade e integridade, em repositórios confiáveis que garantam sua preservação por longo prazo (SALES; SAYÃO, 2019, p. 42).

Os dados experimentais se diferem, pois são oriundos de ambientes científicos, de situações controladas em bancadas de laboratório.

Em tese, estes dados são provenientes de experimentos que podem ser precisamente reproduzidos e não precisam ser armazenados indefinidamente; entretanto, nem sempre é possível reproduzir precisamente todas as condições experimentais, e, na maioria das vezes o custo de gerenciar os dados é apenas uma fração do custo de refazer o experimento (SALES; SAYÃO, 2019, p. 42).

Por serem provenientes de ambientes ou empreendimentos científicos, Sales e Sayão (2019, p. 43) apontam que essas categorias de dados são propriamente dados de pesquisa. Por exemplo, dados governamentais, de registros médicos e arquivísticos são coletados para outros propósitos que são usualmente usados para pesquisa científica, ou seja, são dados usados para pesquisa, mas nem sempre de pesquisa. Fato que não ocorre com os dados experimentais.

Já os dados computacionais são resultados de empreitadas tecnológicas, aplicação de simulações ou modelos computacionais (SALES; SAYÃO, 2019, p. 42-43). Onde:

[...] devem ser submetidos a uma abordagem distinta, que pressupõe o arquivamento de um grande número de informações, expressos por um conjunto robusto de metadados, que incluem descrição de *hardware*, *software* e parâmetros de entrada (SALES; SAYÃO, 2019, p. 43).

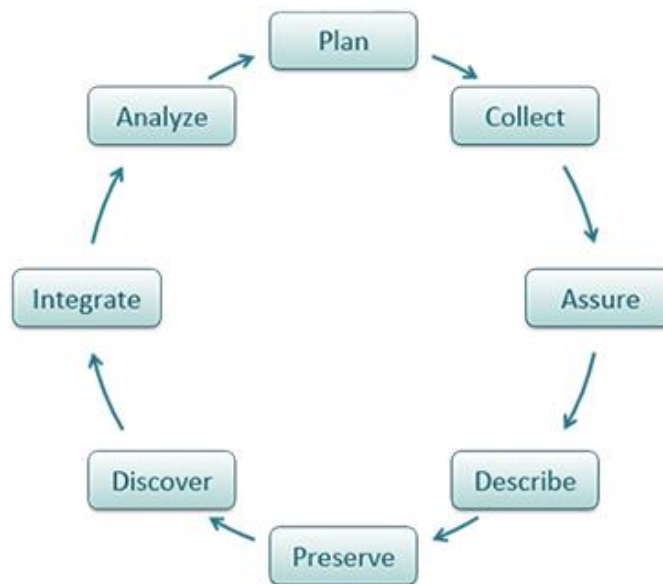
Os dados também podem ser caracterizados: pelo grau de processamento (referente ao grau de processamento que representa os fluxos de pesquisa de cada área); pela abordagem da pesquisa (referente ao tratamento distinto que os dados precisam adquirir conforme a metodologia usada no desdobramento da pesquisa); pela natureza dos dados (referente à variedade tipológica de dados que podem ser gerados no ambiente de pesquisa); pelo nível de sensibilidade (referente a graus de abertura durante o processo de gestão e curadoria, assim como pela complexidade de intervenção que deve ser imposta às coleções); pela materialidade (que possibilita o entendimento da necessidade de tratamento especial aos dados conforme a maneira que se materializam, digitais ou físicos); pela perenidade (referente ao lugar onde as coleções de dados são gerenciadas e preservadas conforme a sua relevância científica) e pela abertura (referente a diferentes níveis de abertura, atribuídos pelos criadores e pelos sistemas de gestão, a coleções de dados) (COSTA; 2017, p. 39; SALES; SAYÃO, 2019, p. 42-45). Segundo Sales e Sayão (2019, p. 45), os níveis de abertura estão relacionados "com o

nível de sensibilidade dos dados, propriedade intelectual, interesses comerciais, patentes, segurança nacional, interesse pessoal do pesquisador, etc.".

Há ainda os dados brutos ou de formato primário, que são classificados quanto ao nível de tratamento que recebem (COSTA, 2017). Esses dados são capazes de ser submetidos a inúmeros estágios de processamento e podem ser utilizados por diversas pesquisas, que não se relacionam a seu processo de produção ou coleta original (COSTA, 2017). Esta é a categoria de dados de maior interesse em ser compartilhada, pois pode facilmente ser reutilizada por outras pesquisas. Economizando recursos e o tempo dos pesquisadores que não se ocupam com a coleta, visto que utilizam dados já disponibilizados. Os dados de pesquisa em saúde e ambientais se encaixam aqui, pois têm grande potencial de reuso e são constantemente utilizados em pesquisas pertencentes a domínios diferentes dos quais foram gerados.

Corti et al (2014) propõe uma classificação dos dados relacionada ao estágio da pesquisa, onde os dados seriam preliminares ou finais. Dados preliminares seriam os dados resultantes dos processos preliminares das pesquisas, e os dados finais seriam aqueles que sustentam o argumento dos resultados da pesquisa e geralmente são selecionados para serem arquivados nas bases de dados (SAYÃO; SALES, 2015, p. 7).

Pela sua complexidade e importância no fazer científico, os dados de pesquisa precisam de um direcionamento diferenciado, quando comparado a outras fontes de informação. Isto para garantir seu principal objetivo: o reuso. O ciclo de vida dos dados de pesquisa é um instrumento auxiliar para gestão, que elucidada os processos referentes ao planejamento e a curadoria dos dados de pesquisa para reuso (COSTA, 2017; DATA OBSERVATION NETWORK FOR EARTH, 2019). Entre as possíveis visões do ciclo de vida dos dados de pesquisa, que possuem diferenças devido a variações de práticas entre domínios, destaco a do *Data Observation Network For Earth* (DataONE) (Figura 1). Cabe ressaltar que segundo o DataONE (2019), algumas atividades de pesquisa podem usar apenas parte das etapas do ciclo de vida dos dados proposto.

Figura 1- Ciclo de vida dos dados

Fonte: DataONE, 2019

Planejar, coletar, validar, descrever, preservar, descobrir, integrar e analisar. No ciclo de vida dos dados os processos podem ocorrer simultaneamente ou se repetirem. Quanto às etapas elucidadas: o planejamento se refere ao mapeamento das ações para o todo o ciclo de vida dos dados; a coleta é a definição da maneira mais adequada para a coleta de dados e seu armazenamento; a validação assegura a qualidade dos dados; a descrição tem como objetivo documentar de forma completa os dados para assegurar seu reuso e compartilhamento; a preservação se ocupa em proteger os dados em curto e longo prazo; a descoberta se refere à localização e recuperação dos dados por meio de metadados que propiciam o reuso; a integração combina dados de formas distintas; e a análise trata de analisar os dados com objetivos distintos e diferentes pesquisadores (BERTIN; VISOLLI; DRUCKER, 2017, p. 43-44). Conciliando com o papel da Ciência da Informação que objetiva chegar à compreensão das propriedades, comportamentos e circulação da informação, os processos aqui destacados são facilmente integráveis a pesquisa em saúde (SHERA, 1980, p. 98).

Conclui-se ressaltando a ideia que ciência produzida com recursos públicos deve atender a toda a sociedade. Nesta perspectiva atuam os dados de pesquisa, que são instrumentos que permitem a possibilidade de reflexão e ação frente às desigualdades inerentes ao fazer científico, por meio do compartilhamento de informação científica.

Os dados possibilitam o remanejamento das formas de concentração de conhecimento e recursos científicos, assim como a interpretação de dados em contextos diversos. Desta forma, as comunidades científicas reconhecem as coleções de dados de pesquisa como uma parte do patrimônio da ciência universal e um pilar imprescindível para seu avanço (SAYÃO; SALES, 2015, p. 181). Trazendo benefícios também para a saúde pública, a quem os dados de pesquisa contribuem na qualidade de vida da população, fortalecendo-a como direito e, conseqüentemente, corroborando para o SUS (FUNDAÇÃO OSWALDO CRUZ, 2018).

2.3 Gestão de dados de pesquisa

Atividades de pesquisa, que se diferem quanto ao domínio ou tipologia dos dados, são inúmeras e ocorrem em todo o mundo. Contudo, percebe-se que as atividades de pesquisa acontecem em ambientes fragmentados e fragmentadores, fato que dificulta a gestão dos dados gerados no decorrer da pesquisa. Não só isso, mas o fato de estarem dispersos, não ou mal documentados ou até inacessíveis, dificulta o reuso, impossibilitando que em novas análises novas descobertas sejam empreendidas (BERTIN; VISOLLI; DRUCKER, 2017, p. 36).

Por serem objetos complexos, os dados de pesquisa demandam diferentes estratégias de gestão. Sales e Sayão (2019, p.36) explicam que “essa heterogeneidade intrínseca aos dados de pesquisa implica na necessidade de formular estratégias de gestão de amplo espectro que englobem os vários tipos de dados”. Deste modo, surge a “gestão de dados de pesquisa”, que cobre todo o ciclo de vida dos dados, incluindo desde a sua coleta até a preservação de longo prazo dos dados (MEDEIROS, 2018).

Gestão de dados de pesquisa é uma expressão que se refere tanto aos processos de planejamento, aquisição, organização, estruturação, definição de fluxos analíticos e ferramenta computacional apropriada para o armazenamento de dados, quanto aos quesitos inerentes à preservação, à organização, ao compartilhamento, à proteção e à confidencialidade dos dados para a instituição que possui o direito sobre os mesmos, bem como o acesso e disponibilização para a sociedade (COSTA, 2017, p. 41; COX; PINFIELD, 2013; WHYTE; TEDDS, 2011). Considerada como um dos itens necessários às boas práticas de pesquisa, a gestão de dados também perpassa por processamentos de limpeza, curadoria, anotação, indexação e transformação (MEDEIROS, 2018).

A gestão de dados de pesquisa tem como elemento essencial o planejamento, que surge na gestão como “plano de gestão de dados de pesquisa”. É no planejamento que se identificam as características dos dados e com isso, são definidos os cuidados que eles requerem (COSTA, 2017). O plano de gestão de dados integra a documentação de pesquisa, revelando-se como um documento formal que detalha os dados processados, gerados ou coletados no decorrer de um projeto de pesquisa (UNIVERSIDADE DE SÃO PAULO, 2018; UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2019). De forma simplificada, é o documento que integra a documentação de pesquisa, onde o autor expõe como pensou em criar, armazenar, compartilhar e

preservar seus dados. A propósito, o plano de gestão de dados referente ao presente estudo está anexado no apêndice B.

No plano de gestão de dados são descritos métodos de gerenciamento de dados que serão implementados durante e após o desenvolvimento do projeto (UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2019). Além disso, o plano de gestão de dados responde indagações como quais as metodologias e padrões serão empregues no decorrer dos processos; se, onde, como e sob que condições os dados serão compartilhados para a comunidade científica, assim como de que forma os dados serão curados e preservados (SAYÃO; SALES, 2015, p. 15). Sayão e Sales (2015, p. 15) revelam que, apesar das especificações, o plano de gestão de dados não deve ser encarado como um procedimento meramente burocrático, mas sim como uma carta de intenções que deve explicitar o que é necessário para a preservação, compartilhamento e reuso dos dados.

Caracterizado por ser um documento que direciona as ações referentes à responsabilidade, custos de preservação, segurança, armazenamento, representatividade legal e disseminação dos dados, o plano de gestão de dados é um termo de compromisso do autor que se compromete a corroborar para o avanço científico, por meio da conduta responsável da pesquisa nos ambientes científicos atuais (SAYÃO; SALES, 2015, p. 16). Apesar da sua importância, não há um padrão para elaboração do documento. Ao elaborá-lo, o autor deverá observar se a instituição ou financiadora do projeto sugere um modelo de plano de gestão de dados de pesquisa. Caso o modelo não exista, o autor poderá criar a mão livre o documento, desde que siga as informações recomendadas para compor o seu plano de gestão. Quanto às informações recomendadas que auxiliarão na elaboração do plano de gestão de dados, Sayão e Sales (2015, p. 26) apresentam de o seguinte quadro:

Figura 2- Composição do plano de gestão de dados: recomendação

RESUMO DAS INFORMAÇÕES RECOMENDADAS PARA COMPOR O PLANO DE GESTÃO DE DADOS	
DESCRIÇÃO DOS DADOS	<ul style="list-style-type: none"> • Tipo de dados produzidos pela pesquisa • Quantidade de dados que será coletada • Como os dados serão coletados • Como os dados serão processados • Formatos de arquivo que serão usados • Como os arquivos serão nomeados • Medidas para garantir a qualidade dos dados • Coleções de dados disponíveis • Dados existentes que serão usados • Preservação de curto prazo • Responsáveis pela gestão de curto prazo
METADADOS	<ul style="list-style-type: none"> • Metadados necessários • Como os metadados serão criados • Esquema que será usado
POLÍTICA DE ACESSO, COMPARTILHAMENTO E REUSO	<ul style="list-style-type: none"> • Obrigações de compartilhamento • Como os dados serão compartilhados • Questões éticas e de privacidade • Propriedade intelectual e copyright • Usos futuros e usuários potenciais • Citação dos dados
GESTÃO DO ARQUIVAMENTO DE LONGO PRAZO: PRESERVAÇÃO DIGITAL DOS DADOS DE PESQUISA	<ul style="list-style-type: none"> • Que dados serão preservados • Onde os dados serão arquivados • Necessidade de formatação dos dados • Responsável pelo contato com o centro de dados
ORÇAMENTO: CUSTOS ENVOLVIDOS NA GESTÃO DE DADOS	<ul style="list-style-type: none"> • Custos previstos • Como os custos serão cobertos

Fonte: Sayão e Sales, 2015

No mais, já existem ferramentas online que disponibilizam modelos de plano de gestão de dados de pesquisa (FUNDAÇÃO DE AMPARO A PESQUISA DO ESTADO DE SÃO PAULO, 2019). Sites como o DMPTool (<https://dmptool.org/>) e o DMPonline (<https://dmponline.dcc.ac.uk/>), auxiliam pesquisadores a gerar um plano, os orientando na elaboração do documento por meio de perguntas, que facilitam o processo de criação do documento.

No que se refere a seus objetivos, a gestão de dados visa garantir que os dados permaneçam em alta qualidade, sejam sustentáveis, acessíveis, reutilizáveis e estejam bem organizados, assim como documentados e preservados (COSTA, 2017, p. 41). Um programa institucional de gestão de dados de pesquisa deve ter como objetivo:

oferecer, à comunidade de pesquisadores, as ferramentas, o treinamento, o apoio e a orientação que são necessários para o apropriado gerenciamento dos dados de pesquisa em todo o seu ciclo, assegurando assim o uso responsável desse valioso recurso para a PD&I (BERTIN; VISOLLI; DRUCKER, 2017, p. 45).

Sendo assim, profissionais da informação e suas equipes, assim como os profissionais integrantes de setores de busca e informação em instituições científicas de pesquisa e ensino, devem estar preparados para prestar apoio e orientação aos pesquisadores que precisam de auxílio. O apoio esperado se refere tanto ao auxílio na elaboração do plano de gestão de dados, quanto na orientação quanto à escolha da plataforma de gestão de dados de pesquisa, as quais os dados dos pesquisadores serão depositados.

Enfatizamos aqui, que é necessário e importante não somente explicitar o escopo e objetivo da pesquisa. Instituições e financiadores de pesquisa vêm requisitando um plano de gestão de dados para os projetos que financiam. Nesse sentido, pesquisadores precisam atender aos critérios dispostos em editais, os quais são estabelecidos pelas próprias agências (VEIGA, 2017). A justificativa para a exigência é a ideia que “os dados que sustentam publicações científicas, ou quaisquer dados com valor de longo prazo, sejam adequadamente organizados, armazenados e disponibilizados” (UNIVERSIDADE DE SÃO PAULO, 2018). A exigência da gestão de dados garante a disponibilidade dos dados pelo maior tempo possível, assegurando a possibilidade de reuso e, conseqüentemente, mantendo a integridade e a reprodutibilidade da pesquisa (FUNDAÇÃO DE AMPARO A PESQUISA DO ESTADO DE SÃO PAULO, 2019; MEDEIROS, 2018). Segundo o editorial *Data's Shameful Neglect* (apud BERTIN; VISOLLI; DRUCKER, 2017, p. 36), publicado pela Nature, a gestão de dados de pesquisa é um dos principais alicerces da produção científica e por isso deveria integrar currículos acadêmicos em todas as áreas do conhecimento. Em vista disso, quem deve gerenciar os dados?

Para a Universidade de São Paulo (2018), o gerenciamento dos dados de pesquisa é responsabilidade dos pesquisadores e das instituições, que devem compartilhar seus dados e outros materiais de apoio elaborados ou coletados no decorrer do trabalho de pesquisa. Já para Medeiros (2018), os projetos de pesquisa devem se ocupar do início e meio do ciclo, em etapas que correspondem ao planejamento de coleta dos dados que serão coletados, tudo isso à medida que limpam os dados e os armazenam de forma adequada. Complementando a lista de atores responsáveis pelo gerenciamento, destaca-se que:

“toda Universidade deve possuir funcionários de suporte para auxiliar a criação de um plano de gerenciamento de dados, incluindo funcionários de apoio à pesquisa do Departamento, a equipe de TI da faculdade, bibliotecários e financiadores” (UNIVERSIDADE DE SÃO PAULO, 2018).

Mesmo assim o principal responsável pela efetiva gestão dos dados de pesquisa é o pesquisador, que além de se organizar deverá organizar seus dados de pesquisa (UNIVERSIDADE DE SÃO PAULO, 2018).

Entre as principais razões para elaborar um plano de gestão de dados, destacam-se os seguintes tópicos:

- a) Ajustar o seu projeto de pesquisa às políticas mandatórias da sua instituição e/ou dos órgãos de fomento à pesquisa;
- b) Assegurar a integridade da pesquisa e o seu potencial de replicação;
- c) Assegurar que os dados e demais registros de pesquisa sejam acurados, completos, autênticos e confiáveis;
- d) Aumentar a sua eficiência como pesquisador – um plano que organize os dados e seu armazenamento permite que você foque na sua pesquisa. Você estará mais capacitado a localizar e usar os seus dados e compartilhá-los com os seus colaboradores;
- e) Permitir que os seus dados sejam compreensíveis agora e no futuro – se os dados são bem documentados antes e durante a formação da coleção de dados, eles serão mais facilmente entendidos e reutilizados;
- f) Economizar tempo e recursos a longo prazo;
- g) Aumentar a segurança dos dados e minimizar os riscos de perda;
- h) Evitar a duplicação de esforços na coleta ou regeneração dos dados, possibilitando que outros pesquisadores se beneficiem dos seus dados e os interprete em outros contextos e com novas visões;
- i) Aumentar a visibilidade da pesquisa – se os seus dados foram planejados para estarem organizados e corretamente arquivados, eles poderão ser identificados, recuperados e citados, aumentando a visibilidade da sua pesquisa e o seu prestígio como pesquisador;
- j) Tornar mais fácil a preservação e o arquivamento – ajustando antecipadamente a geração de dados tomando como referência as práticas, formatos e demais padrões mais adequados ao arquivamento e à preservação de longo prazo, torna a gestão de dados mais fácil e menos custosa; além do mais, tornam os dados mais aderentes aos requisitos dos repositórios e centros de dados (COMISSÃO NACIONAL DE ENERGIA NUCLEAR, 2019).

Para proteger os dados de pesquisa da obsolescência tecnológica e da fragilidade a que estão submetidos quando armazenados em computadores pessoais e institucionais, ferramentas de armazenamento online ou em qualquer outro tipo de mídia digital, estão às infraestruturas de gestão de dados de pesquisa. As infraestruturas de gestão de dados

de pesquisa permitem que as coleções de dados sejam localizadas, reutilizadas, compartilhadas e citadas. Oferecendo um serviço, geralmente gratuito, de preservação e gerenciamento de dados de pesquisa.

A Universidade de São Paulo (2018) aponta que para aumentar a visibilidade e a descoberta dos dados, coleções de dados podem ser cadastradas em: catálogos institucionais e nacionais; repositórios institucionais e temáticos; catálogos e bibliotecas virtuais de financiadores e agências; em perfis profissionais; websites de institutos, faculdades e grupos de pesquisa ou registrados no ORCID ou redes de pesquisa como o ResearchGate. Apesar dessa variedade de instrumentos de armazenamento e divulgação de coleções de dados, na atualidade, somente os Repositórios garantem a longo prazo que coleções de dados sejam reusadas (SAYÃO; SALES, 2015). Justificada sua relevância, os repositórios de dados são a temática da próxima seção terciária.

2.3.1 REPOSITÓRIOS

Segundo Marques e Maio (2007 apud PAVÃO ET AL., 2018, p. 980), “os repositórios surgiram na década de 1990, quando as revistas científicas eram o principal meio de divulgação da produção científica de pesquisadores e universidades”. Apesar das inovações trazidas pelo advento de aparatos tecnológicos estarem inseridas no fazer técnico e científico de instituições de pesquisa, tecnológicas e educacionais ao redor do mundo, em um processo que teve seu apogeu no século XX com o chamado “*boom*” da informação, há de se constatar demora quanto à priorização de salvaguarda de objetos informacionais. Ou seja, houve atraso quanto ao surgimento de zelo e preservação de objetos informacionais que quando acessíveis e compartilhados são dotados de capacidade de transformar o mundo, dada a sua complexidade e importância.

Caso os objetos informacionais fossem submetidos a um processo formado por etapas de organização, armazenamento e preservação para então serem disseminados, adicionariam equidade ao fazer científico, equidade que propiciaria inovações sem precedentes. Isto, pois, principalmente no passado, o acesso à informação disponibilizada em periódicos era elitizada. Poucos tinham acesso devido ao custo e a impossibilidade de acesso ao território restrito por onde circulavam os itens empregues na comunicação científica.

A inviabilidade de acesso irrestrito a informação disponível em acesso aberto é uma realidade experimentada até hoje, apesar de muitos fantasiarem um mundo onde

todos têm as mesmas oportunidades, justificando o contrário baseado em um discurso ligado a meritocracia e não a oportunidades e privilégios. Para perceber este fenômeno basta um olhar mais atento quanto ao desenvolvimento científico global. Países subdesenvolvidos não enfrentam somente problemas econômicos, sociais e de saúde, enfrentam também dificuldades científicas. Fato que reflete seu posicionamento científico quase sempre a sombra de países desenvolvidos que contam com um maior investimento, incentivo e conseqüentemente infraestrutura científica.

Atualmente, ainda que as discrepâncias mencionadas ainda existam mesmo que de “forma velada”, o acesso às publicações não se dá de forma tão analógica e restrita como antes. O advento da Internet permite que publicações científicas estejam acessíveis publicamente e sem custos, sujeitas somente a restrições referentes à amplitude dos acessos (Via Verde e Via Dourada). Apesar de um cenário melhor quanto ao acesso à informação, hoje em dia, pesquisadores não se contentam somente em acessar os documentos tradicionalmente encontrados em bibliotecas e centros de informação. Na contemporaneidade, cientistas e pesquisadores querem acesso aos dados que dão origem aos achados científicos. Desta forma, é essencial a preocupação em geri-los e preservá-los, independentemente da linguagem, formato ou sistema a qual esteja inserido. Para isso é necessário à adoção de uma infraestrutura para a gestão e armazenamento desses dados e é justamente nesse cenário que estão os repositórios que propiciam acesso, reuso e compartilhamento dos dados de pesquisa, quando apoiados em incumbências institucionais sustentáveis e de longo prazo (SAYÃO; SALES, 2013, p. 3-4).

“Atualmente um grande número de instituições de pesquisas, entre elas as universidades, vem apostando no depósito de suas investigações científicas dentro de seus repositórios eletrônicos” (BARRUECO, 2010, p. 2 apud NASCIMENTO, 2014, p. 48). Para Viana, Márdero Arellano e Shintaku (2005, p. 3) um repositório digital é “uma forma de armazenamento de objetos digitais que tem a capacidade de manter e gerenciar material por longos períodos de tempo e prover o acesso apropriado.”. Lynch (2003) acrescenta que os repositórios são utilizados pelas universidades também como ferramenta de apoio a comunicação dos resultados de suas pesquisas científicas. Cumprindo, dessa forma a função de validar e estimular a publicação dos estudos empreendidos no âmbito acadêmico. Viana, Márdero Arellano e Shintaku (2005, p. 4) corroboram com a característica evidenciada quando afirmam que se trata de um ambiente onde a produção científica de uma instituição pode ser depositada, permitindo

a discussão entre pares e, conseqüentemente, viabilizando a troca de experiências e ideias por meio do acesso e revisão constantes a um documento.

Consumando o entendimento de que estamos em uma nova era científica, Lynch (2003) atesta ao declarar que os repositórios surgiram como uma nova estratégia para que as universidades alavanquem em meio às mudanças que estão acontecendo na comunicação e na pesquisa. Em suma, não há como negar que os repositórios são uma maneira de preservar, gerenciar e dar visibilidade ao conhecimento científico produzido individualmente ou por grupos de pesquisa no âmbito institucional ou temático. Conhecimento este que caso não fosse gerido, disponibilizado e compartilhado, estaria à mercê da obsolescência tecnológica e invisível perante a comunidade científica e acadêmica já que circularia somente entre seus pares. Uzwyszyn (2016, p. 1 apud SAYÃO; SALES, 2016, p. 95) arremata ao acrescentar que “um repositório permite exame, prova, revisão, transparência de resultados de pesquisa por outros especialistas que vão além da revisão por pares do artigo acadêmico publicado.

Mas afinal o que são repositórios? Repositórios digitais são bases de dados desenvolvidas que possibilitam a recuperação de registros e possibilitam o gerenciamento de informação científica funcionando como via alternativa de comunicação científica (LEITE et al, 2012). Por serem ferramentas complexas e de grande importância, pesquisadores classificam os repositórios em: institucionais, temáticos ou disciplinares e de dados.

Para Rodrigues et al. (2010, p. 22), o “termo “repositório” designa um sistema informático em que existe uma plataforma de armazenamento de objectos representados em ficheiros, capaz de incorporar novos objectos à medida que são produzidos ou submetidos”.

Repositórios digitais são coleções de informação digital que podem ser colaborativos e com um controle suave dos conteúdos e da autoridade dos documentos [...] Mas podem, também, ter um alto nível de controle e ser concebidos para públicos específicos (BARRETO, 2010).

Segundo Leite (2009, p. 20), os repositórios institucionais são dirigidos à produção intelectual de uma instituição, em especial universidades e institutos de pesquisa. São veículos multiplicadores na disseminação de resultados de pesquisa, caracterizados por tornar os conteúdos disponíveis e acessíveis amplamente (LEITE et al., 2012). Crow (2002, p. 16) complementa a ideia ao afirmar que os repositórios

institucionais podem ser definidos como “arquivo digital da produção criada pela universidade, equipe de pesquisa e estudantes de uma instituição e que esteja acessível aos usuários finais dentro e fora da instituição”. Arrebatando a importância dos repositórios institucionais na atualidade, Kuramoto (2009) afirma que eles se destacam como uma das principais iniciativas para a implantação do acesso livre no mundo.

Para Leite (2009, p. 20), os repositórios temáticos ou disciplinares são dirigidos a comunidades científicas específicas, ou seja, são voltados para áreas do conhecimento específicas. Para Kuramoto (2006, p. 83), os repositórios temáticos correspondem a 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”.

Há ainda os repositórios de dados, que segundo Costa (2017, p. 46), “são bases de dados digitais onde são armazenados, disseminados e preservados os dados de pesquisa em formato digital”. Em suma, repositórios são sistemas que oferecem um serviço de armazenamento, preservação, recuperação, disseminação e gerenciamento, que promovem o acesso livre a conteúdos como produtos de pesquisa, entre outros objetos digitais (MONTEIRO, 2017, p. 34).

O campo empírico deste estudo é o Registry of Research Data Repositories (Re3data), que é um projeto conjunto do DataCite, do Instituto de Tecnologia Karlsruhe (KIT), da Escola de Biblioteconomia e Ciência da Informação de Berlim e do Escritório de Ciências Abertas Helmholtz no Centro de Pesquisa Alemão para Geociências (GFZ), financiado pela Fundação Alemã de Pesquisa (DFG) desde janeiro de 2020 (COUSIJN; FENNER, 2020). Trata-se de uma ferramenta de informação sobre repositórios de dados de pesquisa

O Re3data disponibiliza informações sobre mais de 2.450 repositórios de dados de pesquisa, em uma dinâmica onde universidades e centros de pesquisa registram seus repositórios de dados institucionais, disciplinares e interdisciplinares (COUSIJN; FENNER, 2020). Neste processo que tem como finalidade permitir, por meio da ferramenta, que pesquisadores selecionem repositórios apropriados para armazenamento e localização de dados de pesquisa, a visibilidade dos repositórios de dados de pesquisa disponibilizados é ampliada. Visto que frente à expansão geográfica mundial e a velocidade do surgimento de novos repositórios, é tarefa quase impossível que pesquisadores acompanhem em tempo real o surgimento de novos repositórios viáveis a seu interesse científico.

Tais características trazem à tona a capacidade versatilidade de serviços oferecidos por repositórios que são dirigidos a quem deposita, a quem pesquisa e aos administradores do sistema (RODRIGUES et al., 2010, p. 22-23). Porém, apesar de oferecerem serviços relevantes no atual contexto científico, os repositórios têm capacidade para ser algo muito maior e melhor. Ao oferecerem novos serviços de troca de experiências, como por exemplo: fóruns, os repositórios se transformarão também em espaço para trocas de experiências. Para isso, é necessário maior esforço na divulgação dos serviços oferecidos para que se obtenha uma maior conscientização sobre sua importância. Tal esforço também servirá para que o sistema se torne atrativo e amplamente conhecido pela comunidade científica.

Racionalizando recursos, validando resultados e propiciando troca de informação entre áreas do conhecimento distintas, da mesma forma que promove novas pesquisas corroborando para o avanço da Ciência, a gestão de dados e os repositórios tornam-se elementos imprescindíveis no fazer científico. Nessa perspectiva a gestão de dados iniciada pela elaboração de um plano de gestão de dados para o então depósito dos mesmos e futura recuperação nos repositórios, são aliados. Entre seus benefícios, criam condições para que os dados relevantes relacionados aos resultados e às publicações geradas pelas pesquisas estejam organizados, arquivados, gerenciados e acessíveis, permitindo que os resultados possam ser verificados, melhorados ou mesmo reutilizados no futuro (UNIVERSIDADE DE SÃO PAULO, 2018; COSTA, 2017). Fato que facilita também o treinamento de novas gerações de pesquisadores e a exploração de tópicos não previstos no projeto original (FUNDAÇÃO DE AMPARO A PESQUISA DO ESTADO DE SÃO PAULO, 2019). Desta maneira, se consolidam trazendo benefícios não só para a Ciência, mas também ao autor dos dados, que tem visibilidade científica ampliada, redução do tempo gasto na coleta de dados, possibilidade de novas parcerias e obtenção de prêmios por meio dos trabalhos desenvolvidos.

3 METODOLOGIA

O presente trabalho trata-se de um estudo exploratório, descritivo e empírico, onde foram feitas pesquisas bibliográficas e documentais a respeito dos assuntos: dados de pesquisa em saúde, câncer, repositórios de dados e informação em saúde. Trata-se também de um estudo de caso onde o objeto são os repositórios de dados de pesquisa em câncer compreendidos no Re3data, um diretório de registro de repositórios de dados.

Estudos exploratórios são comumente empreendidos quando se torna difícil formular hipóteses precisas sobre o tema de pesquisa escolhido, devido ao tema ser pouco explorado cientificamente. Segundo Gil (2008, p. 27), as pesquisas exploratórias “têm como principal finalidade desenvolver, esclarecer e modificar conceitos e ideias, tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores”. Assim sendo, como produto final do processo de pesquisa, tem-se um problema mais esclarecido e suscetível de investigação por meio de procedimentos mais estruturados (GIL, 2008, p. 27.).

O estudo também se caracteriza como descritivo, pois é feito um levantamento, que tem como objetivo a descrição das características de determinada população (GIL, 2008, p. 28). Neste caso o objetivo mapear os repositórios de dados de pesquisa em câncer registrados no Re3data.

Quanto à natureza é um estudo empírico pelo fato de confrontar a visão teórica com os dados da realidade da amostra estudada baseando-se na observação dos fatos (GIL, 2002, p. 43). Quanto a seu delineamento, este estudo inclui pesquisas bibliográficas e documentais com o objetivo de permitir a autora um conjunto de fenômenos mais amplos do que o que se poderia pesquisar diretamente (GIL, 2002, p. 45). No processo de pesquisa bibliográfica foram utilizadas contribuições de diversos autores, identificados a partir de fontes capazes de fornecer respostas adequadas à solução do problema apontado (GIL, 2002, p. 64). Entre os benefícios da pesquisa bibliográfica como delineamento da pesquisa, está a possibilidade de conhecer o passado, já que “os dados documentais, por terem sido elaborados no período que se pretende estudar, são capazes de oferecer um conhecimento mais objetivo da realidade” (GIL, 2008, p. 153). Além deste, outro benefício, principalmente em tempos de dificuldades para obtenção de financiamento, é a obtenção de dados com menor custo. Gil (2008, p. 153), corrobora ao confirmar que as pesquisas elaboradas a partir de dados

já existentes, requerem quantidade menor de recursos (humanos, materiais e financeiros) tornando-se mais viáveis.

Sendo assim, a partir de metodologia indicada por Gil (2002, p. 59), a pesquisa bibliográfica se delineou da seguinte forma: escolha dos termos de busca (já mencionados anteriormente); levantamento bibliográfico preliminar (busca das fontes); leitura do material; fichamento; organização lógica do assunto e redação do texto. As bases de dados escolhidas para as pesquisas bibliográficas e documentais foram: *Web of Science, Library and Information Science Abstracts (LISA)* e *Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS)*.

A *Web of Science* é uma base multidisciplinar e foi selecionada por ter como principal característica o fato de indexar somente os periódicos mais citados em suas respectivas áreas (COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR, 2019).

A LISA foi selecionada para o estudo, pois é uma base destinada aos profissionais de bibliotecas, ciência da informação e áreas correlatas (COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR, 2019).

Já a LILACS foi selecionada para o estudo a fim de localizar estudos que relacionassem as temáticas: informação e saúde. Isto, pois contém informação bibliográfica Latino-Americana em ciências da saúde, destacando-se como o mais importante e maior índice da literatura científica e técnica da América Latina e Caribe (WIKIPEDIA, 2019; BVS SALUD, 2019).

Além das bases de dados já citadas, também foram utilizados no estudo exploratório o repositório institucional Arca, da Fiocruz, e a Biblioteca Virtual em Saúde, que contribuíram para a pesquisa ao recuperar estudos relevantes sobre informação, saúde e câncer.

Na coleta de dados foi realizado um levantamento dos repositórios de dados de pesquisa em câncer no diretório Re3data. Dessa forma foi possível a identificação dos repositórios de dados de pesquisa em câncer que compõem a amostra e a sistematização dos dados levantados.

O campo empírico do trabalho é o Re3data, que como já foi citado, disponibiliza informações sobre mais de 2.450 repositórios de dados de pesquisa. Apesar do conhecimento de que o campo empírico escolhido não compreende todo o universo de diretórios de dados de pesquisa existentes na área Oncológica, a opção pelo Re3data foi

a mais viável por ser considerado fonte de referência mais abrangente para infraestruturas de dados de pesquisa em todo o mundo o diretório foi selecionado como universo de pesquisa (WITT, 2018).

Mesmo existindo outras opções de ferramentas que localizam infraestruturas de gestão de dados, outra razão pela qual o Re3data foi escolhido como campo empírico, foi devido aos critérios mínimos exigidos para inclusão de repositórios no diretório. Segundo o Re3data.org (2020), os requisitos mínimos são: preenchimento de formulário de solicitação; avaliações realizadas pela Equipe do Re3data; a administração do repositório requerente por uma entidade legal, como por exemplo: instituição de pesquisa, unidade de informação, universidade e etc.; declaração clara de termos de uso e condições de acesso aos dados e repositórios. Além disso, é necessário que o repositório tenha uma interface em inglês e que tenha como foco os dados de pesquisa.

Para o levantamento de repositórios de dados de pesquisa em câncer no Re3data, realizado em 26 de abril de 2020, foi utilizada a seguinte estratégia: **Keyword(s):(neoplasms OR cancer OR tumor* OR tumour* OR carcinoma* OR onco*)**. Tal levantamento possibilitou a identificação de 72 repositórios de dados de pesquisa em câncer apresentados no apêndice B desta pesquisa.

Como estratégia de recorte adotou-se como critério para condição de amostra de pesquisa a presença do termo “câncer” no título dos repositórios identificados. A escolha pelo verbete se justifica pela percepção, no dia-a-dia na pesquisadora no ofício de bibliotecária em unidade hospitalar, do fato do verbete se caracterizar como o mais usado pelo corpo clínico ao referir-se ao tema neoplasias. Além disso, o termo “câncer” foi selecionado, porque é a palavra que aparece com mais frequência em relação aos outros verbetes indexados no DeCS (Descritores em Ciências da Saúde) e que também representam a doença, inclusive quando a busca é expandida para os idiomas inglês e espanhol (esquema explicitado nas figuras a seguir).

Figura 3 - Ocorrência do termo de busca “câncer” em português

1 / 1 DeCS	
Descritor Inglês:	Neoplasms
Descritor Espanhol:	Neoplasias
Descritor Português:	Neoplasias
Sinônimos Português:	<p>Câncer</p> <p>Neoplasia</p> <p>Neoplasia Benigna</p> <p>Neoplasia Maligna</p> <p>Neoplasias Malignas</p> <p>Neoplasmas</p> <p>Tumor</p> <p>Tumor Maligno</p> <p>Tumores</p> <p>Tumores Malignos</p>
Categoria:	<p>C04</p> <p>SP4.001.012.098</p> <p>SP4.046.452.698.879.165</p>
Definição Português:	Crescimento novo anormal de tecido . As neoplasias malignas apresentam um maior grau de anaplasia e têm propriedades de invasão e de metástase quando comparadas às neoplasias benignas.
Nota de Indexação Português:	geral; prefira específicos; familiar: veja também SÍNDROMES NEOPLÁSTICAS HEREDITÁRIAS ; câncer metastático de origem desconhecida: indexe sob METÁSTASE NEOPLÁSTICA
Veja também termos em Português:	CÂNCER, CARCINO-, ONCO-, e TUMOR

Fonte: Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, 2019

Figura 4 - Ocorrência do termo de busca “câncer” em inglês:

1 / 1 DeCS	
Descritor Inglês:	Neoplasms
Descritor Espanhol:	Neoplasias
Descritor Português:	Neoplasias
Sinônimos Inglês:	<p>Benign Neoplasm</p> <p>Benign Neoplasms</p> <p>Cancer</p> <p>Cancers</p> <p>Malignancies</p> <p>Malignancy</p> <p>Malignant Neoplasm</p> <p>Malignant Neoplasms</p> <p>Neoplasia</p> <p>Neoplasias</p> <p>Neoplasm</p> <p>Neoplasm, Benign</p> <p>Neoplasm, Malignant</p> <p>Neoplasms, Benign</p> <p>Neoplasms, Malignant</p> <p>Tumor</p> <p>Tumors</p>
Categoria:	<p>C04</p> <p>SP4.001.012.098</p> <p>SP4.046.452.698.879.165</p>
Definição Inglês:	New abnormal growth of tissue . Malignant neoplasms show a greater degree of anaplasia and have the properties of invasion and metastasis, compared to benign neoplasms .
Nota de Indexação Inglês:	general; prefer specifics; familial: consider also NEOPLASTIC SYNDROMES, HEREDITARY; metastatic cancer of unknown origin: index NEOPLASM METASTASIS
Veja também termos em Inglês:	CANCER, CARCINO-, ONCO-, and TUMOR

Fonte: Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, 2019

Figura 5 - Ocorrência do termo de busca “câncer” em espanhol:

1 / 1 DeCS	
Descritor Inglês:	Neoplasms
Descritor Espanhol:	Neoplasias
Descritor Português:	Neoplasias
Sinônimos Espanhol:	<p>Câncer</p> <p>Neoplasia</p> <p>Neoplasia Benigna</p> <p>Neoplasia Maligna</p> <p>Neoplasias Malignas</p> <p>Neoplasma</p> <p>Neoplasmas</p> <p>Tumor</p> <p>Tumor Maligno</p> <p>Tumores</p> <p>Tumores Malignos</p>
Categoria:	<p>C04</p> <p>SP4.001.012.098</p> <p>SP4.046.452.698.879.165</p>
Definição Espanhol:	Crecimiento anormal y nuevo de tejido. Las neoplasias malignas muestran un mayor grado de anaplasia y tienen la propiedad de invasión y metástasis , comparados con las neoplasias benignas.
Nota de Indexação Espanhol:	general; prefiera específicos; familiar: vea también SÍNDROMES NEOPLÁSICOS HEREDITARIOS ; cáncer metastático de origen desconocido: índice bajo METÁSTASIS DE LA NEOPLASIA
Veja também termos em Espanhol:	CANCER, CARCINO-, ONCO-, y TUMOR

Fonte: Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, 2019

Isto posto, foram selecionados para a presente investigação os 13 repositórios indicados no quadro a seguir.

Quadro 1 – Amostra de pesquisa

REPOSITÓRIOS DE DADOS DE PESQUISA EM CÂNCER SELECIONADOS	
1.	<i>German Centre for Cancer Registry Data</i>
2.	<i>Australian Breast Cancer Tissue Bank</i>
3.	<i>Oral Cancer Gene Database</i>
4.	<i>CancerData.org</i>
5.	<i>Cancer GenomeAtlas Data Portal</i>
6.	<i>Cancer Genomics Hub</i>
7.	<i>Breast Cancer Surveillance Consortium</i>
8.	<i>The Cancer Imaging Archive</i>
9.	<i>Broad-Novartis Cancer Cell Line Encyclopedia</i>
10.	<i>National Cancer Data Base</i>
11.	<i>The Cancer Immunome Atlas</i>
12.	<i>Cancer Genome Anatomy Project</i>
13.	<i>Cancer in Young People in Canada</i>

Fonte: Elaboração própria

A extração dos dados coletados para análise foi feita por meio do software *Elastic*, que inquiriu os elementos apresentados no quadro 1 e esquadrinhados na próxima seção. O uso do software *Elastic* para coleta dos dados aqui estudados, só foi possível devido ao apoio da Rede GO FAIR Brasil Saúde, coordenada pela Fiocruz e participação do INCA. A iniciativa GO FAIR tem sede internacional na Holanda e escritório no Brasil, coordenado pelo IBICT. A análise de dados possuiu caráter qualitativo, onde os resultados obtidos foram utilizados com o objetivo de apresentar um panorama dos repositórios de dados de pesquisa utilizados para pesquisa em câncer.

Quadro 2 – Dimensões de análise da amostra e elementos de caracterização

CAMPO	ELEMENTOS DE CARACTERIZAÇÃO
<i>Repository types</i> (Tipos de repositórios)	Disciplinar Institucional Outro
<i>Subjects</i> (Assuntos)	Ciências Humanas e Sociais Ciências da Vida Ciências Naturais Ciências da Engenharia
<i>Countries</i> (Países)	Países responsáveis pelos repositórios de dados
<i>Software</i> (Sistema de gerenciamento)	Software utilizado pelo repositório de dados
<i>Data access restrictions</i> (Restrições de acesso a dados)	Restrições de acesso ao conjunto de dados
<i>Database access</i> (acesso ao banco de dados)	Condições de acesso ao repositório
<i>Database access restrictions</i> (restrições de acesso ao banco de dados)	Restrições quanto ao acesso ao repositório
<i>Enhanced publication</i> (publicação ampliada)	Publicações ampliadas
<i>Content Types</i> (Tipos de conteúdo)	Tipos de conteúdo que formam a coleção do repositório de dados

Fonte: Elaboração própria

4 REPOSITÓRIOS DE DADOS DE PESQUISA EM CÂNCER NO R3DATA

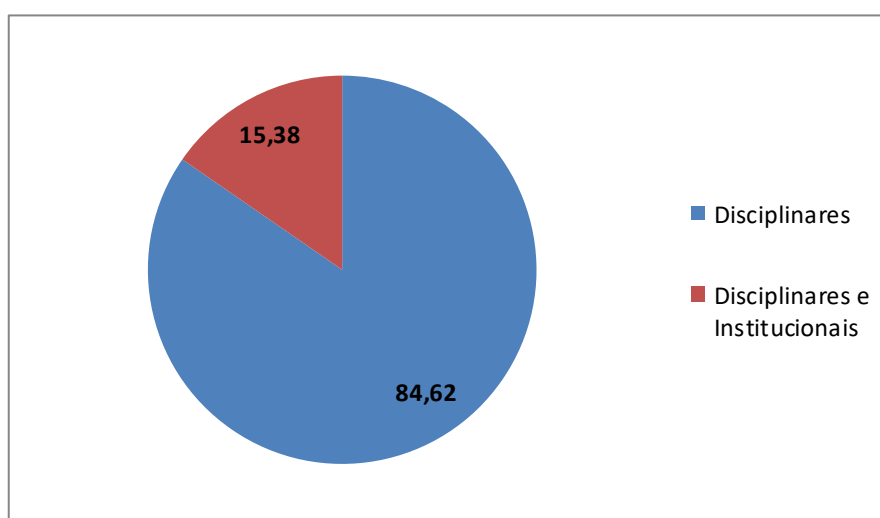
A partir dos dados coletados no Re3data em metodologia descrita na seção anterior, esta seção apresenta os dados descobertos e discute seu significado a luz do que os dados coletados representam. Para tal ação, a presente seção é composta por seções secundárias onde são analisadas as seguintes características dos dados coletados na amostra: tipos de repositório de dados de pesquisa em câncer; classificação de assunto; abrangência geográfica; tecnologias e conteúdos disponibilizados.

4.1 Tipos de repositórios de dados de pesquisa em câncer

Quanto aos tipos de repositórios de dados de pesquisa registrados no diretório Re3data constatou-se que são classificados nos seguintes tipos: disciplinares, institucionais e outros. Dos 2893 repositórios compreendidos no diretório: 1992 (68,85%) são disciplinares, 617 (21,33%) são institucionais e 284 (9,82%) estão registrados como outros (RE3DATA, 2020).

Em relação à amostra estudada, repositórios de dados de pesquisa em câncer, nenhum repositório foi classificado como do tipo outro. Sendo assim, as classificações da amostra estudada variam somente entre disciplinares e institucionais, onde dos 13 repositórios estudados: 11 (84,62%) são do tipo disciplinar e 2 (15,38%) são disciplinares e institucionais.

Gráfico 1- Tipos de repositórios de dados de pesquisa em câncer



Fonte: Elaboração própria

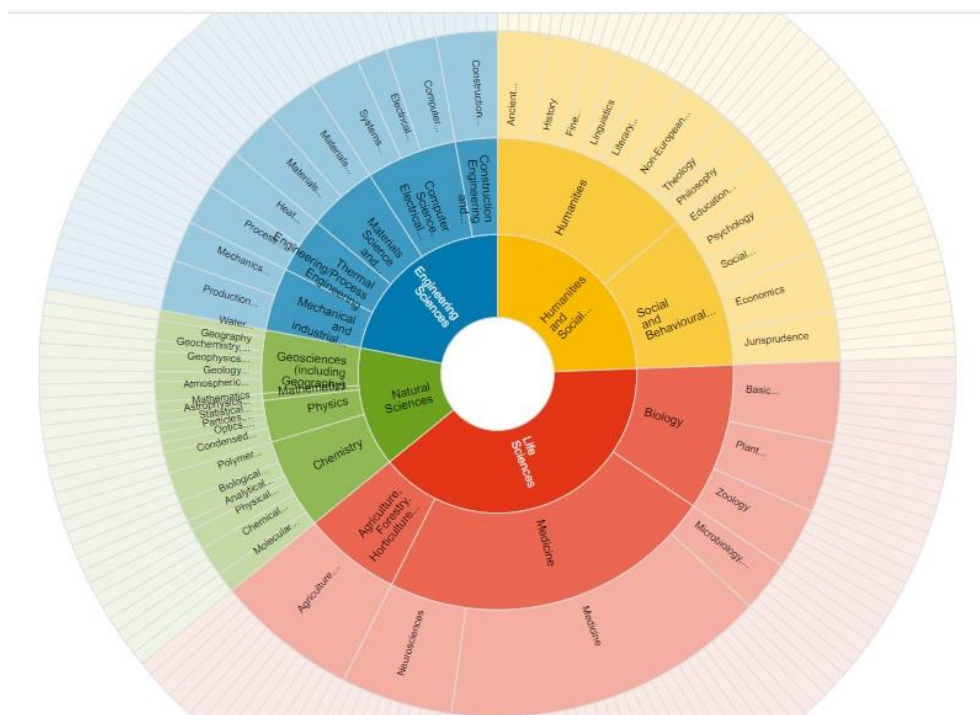
4.2 Classificação de assunto

Destacando-se como um dos elementos mais importantes no processo de comunicação científica, a classificação é um processo mental que reflete a forma a qual indivíduos enxergam a realidade. Classificar é o ato de dividir em categorias, é o processo pelo qual se reúne elementos por semelhança e se separa de acordo com as diferenças. Tudo isso em um processo mental no qual é possível, por meio de elementos comuns, reconhecer similaridades e relações.

Existem diferentes tipos de sistemas de classificação, que são escolhidos e incorporados de acordo com as necessidades de acervos de uma infinidade de ambientes, desde lares até bibliotecas, arquivos, escritórios, centros de informação e pesquisa e etc. Mediante essa infinidade de ambientes onde esquemas de classificação são incorporados, não surpreende o fato de um diretório de repositórios de dados de pesquisa como o Re3data contar com um “sistema” que organiza os repositórios indexados no Diretório por meio de um esquema de ordenação das Ciências.

No Re3data, os repositórios indexados são classificados por hierarquias e Classes, pré-estabelecidas. Como apresentado na figura a seguir:

Figura 6 – Classificações hierárquicas no Re3data



Fonte: Re3data, 2020

Nessa dinâmica, o representante do repositório interessado em registrar sua plataforma no Diretório, tem a opção de acrescentar dados complementares para a localização de seu respectivo repositório na plataforma.

No formulário de registro, destinado à sugestão de incorporação de novos repositórios ao Re3data, nos campos “palavras-chave” e “identificador de repositório”, o solicitante fica livre para propor as palavras-chave que achar adequada, pois não há indicação de vocabulário controlado.

Figura 7 - Formulário de indicação de repositórios

The image shows a web form from re3data.org. The form is titled 'Formulário de indicação de repositórios'. It contains several sections, each with a list of items and a corresponding 'Adicionar' button. The sections are: 'Nomes adicionais', 'assuntos', 'Contatos do repositório', 'Tipos de conteúdo', and 'Certificados'. Below these are two text input fields: 'Palavras-chave' and 'Identificadores de repositório'. Both of these fields are circled in red. To the right of each text field is a 'Retirar' button. At the bottom of the form, there is an 'Adicionar repository identifiers' button. The top of the page shows the re3data.org logo and navigation links: Procurar, Squeaky toy, Sugerir, Recursos, and Contato. The DataCite logo is also visible in the top right corner.

Fonte: Re3data, 2020

A adoção de campos como “palavras-chave” e “identificadores de repositório”, se assemelha a condutas adotadas por grandes bases de dados científicas e sistemas de gerenciamento de acervo presentes em unidades de informação. Essas fontes de informação utilizam os dados inseridos nesses campos como alternativa para a recuperação de informações contidas em seus bancos de dados. Sendo assim, a livre inclusão de termos nesses respectivos campos é uma forma de facilitar tanto pesquisadores experientes como usuários inexperientes no domínio científico. Atuando, portanto, como facilitador no processo de localização de repositórios que atendam as peculiaridades dos conjuntos de dados de pesquisa pertencentes ao pesquisador que tem interesse em salvaguardar e disponibilizar seus dados a comunidade científica quanto para pesquisadores que acessam o Re3data em busca de repositórios que contenham

conjuntos de dados de pesquisa que possam contribuir com o andamento de suas pesquisas e projetos.

Como mostrado na figura 7, não há espaço para orientação de como o solicitante deve proceder, em relação ao uso de palavras-chave. Caso o responsável pelo preenchimento do requerimento de inclusão não seja um bibliotecário ou um profissional conhecedor sobre o propósito de seu repositório de dados e a que os conjuntos de dados contidos nele se referem, enfrentará dificuldades no preenchimento do campo. Por ser um elemento importante, que dá visibilidade ao repositório é imprescindível que essas informações sejam preenchidas de forma adequada. Isto posto, sugiro a inclusão de orientações, por parte do Re3data, quanto ao preenchimento do campo. Utilizando como justificativa para tal empreitada a importância do campo para a eficiente recuperação de informação.

Quanto à proximidade a Sistemas de classificação amplamente conhecidos no Campo da Ciência da Informação, incluindo o fato de classificar numericamente as Ciências, a classificação utilizada no Re3data muito se assemelha a Classificação Decimal de Dewey (CDD). A CDD é um sistema de classificação internacional criado por Melvil Dewey em 1876, que organiza hierarquicamente os conhecimentos em classes decimais. No Re3data as Ciências são ordenadas em quatro classes, são elas: *Engineering Sciences*; *Humanities and Social Sciences*; *Life Sciences* e *Natural Sciences*, onde cada uma das quatro classes principais é dividida em no mínimo duas categorias, que por sua vez originam cadeias, as quais também se desmembram em subdivisões.

Em relação às áreas temáticas de abrangência, verificou-se que todos os repositórios da amostra estudada pertencem à classe “Life Sciences”, em uma relação onde foi atribuído, pelo menos, dois assuntos para cada repositório (quadro 3).

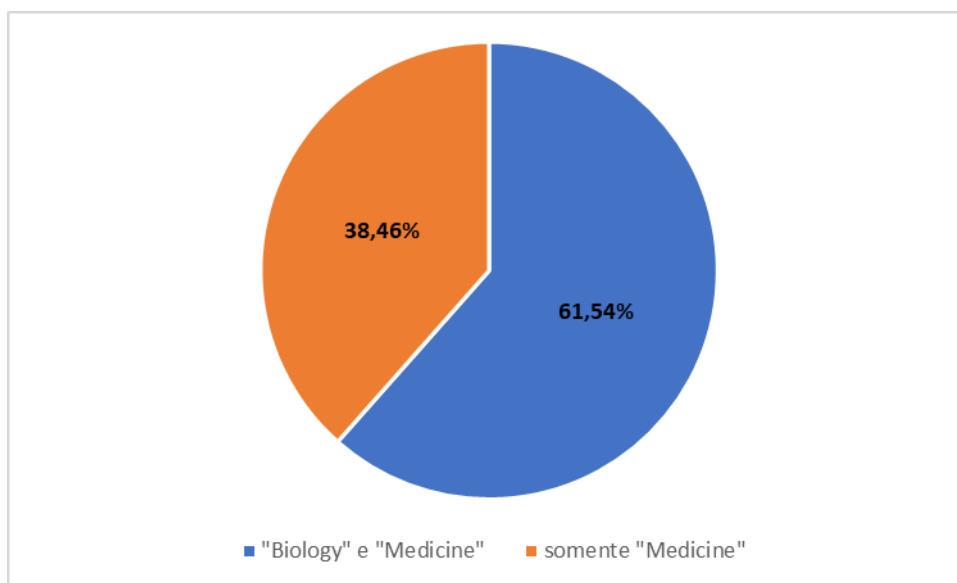
Quadro 3 – Áreas temáticas de abrangência

REPOSITÓRIO	ASSUNTOS
German Centre for Cancer Registry Data	Medicine, Epidemiology, Medical Biometry, Medical Informatics, Hematology, Oncology, Transfusion Medicine, Life Sciences, Medicine
Australian Breast Cancer Tissue Bank	Medicine, Life Sciences
Oral Cancer Gene Database	Basic Biological and Medical Research, Microbiology, Virology and Immunology, Medicine, Neurosciences, Biology, Life Sciences, Medicine
CancerData.org	Basic Biological and Medical Research, Medicine, Biology, Life Sciences
Cancer GenomeAtlas Data Portal	Basic Biological and Medical Research, Human Genetics, Biology, Life Sciences, Medicine, Medicine
Cancer Genomics Hub	Basic Biological and Medical Research, Human Genetics, Biology, Life Sciences, Medicine, Medicine
Breast Cancer Surveillance Consortium	Gynaecology and Obstetrics, Medicine, Medicine, Life Sciences
The Cancer Imaging Archive	Radiation Oncology and Radiobiology, Biology, Life Sciences, Medicine, Medicine, Radiology and Nuclear Medicine
Broad-Novartis Cancer Cell Line Encyclopedia	General Genetics, Basic Biological and Medical Research, Biology, Life Sciences, Human Genetics, Pharmacy, Medicine, Medicine
National Cancer Data Base	Hematology, Oncology, Transfusion Medicine, Medicine, Medicine, Life Sciences
The Cancer Immunome Atlas	Medicine, Microbiology, Virology and Immunology, Human Genetics, Hematology, Oncology, Transfusion Medicine, Public Health, Health Services Research, Social Medicine, Medicine, Immunology, Basic Biological and Medical Research, Biology, Life Sciences
Cancer Genome Anatomy Project	General Genetics, Animal Genetics, Cell and Developmental Biology, Human Genetics, Basic Biological and Medical Research, Biology, Life Sciences, Zoology, Medicine, Medicine
Cancer in Young People in Canada	Dermatology, Traumatology and Orthopaedics, Medicine, Medicine, Life Sciences

Fonte: Elaboração própria

De acordo com a análise dos dados encontrados, nenhum repositório de dados de pesquisa em câncer foi atribuído à categoria “*Agriculture, Forestry, Horticulture and Veterinary Medicine*”. Sendo assim, 8 (61,54%) dos repositórios de dados de pesquisa em câncer estudados estão imputados às categorias “Biology” e “Medicine”, ao mesmo tempo, e 5 (38,46%) somente a categoria “Medicine”.

Gráfico 2 - Categorias de repositórios de dados de pesquisa em câncer



Fonte: Elaboração própria

A categoria “*Medicine*” se desmembra em três cadeias, são elas: “*Neurosciences*”, “*Medicine*” e “*Microbiology, Virology and Immunology*”. Onde 1 (um) repositório foi classificado em *Neurosciences*, 11 (onze) em *Medicine* e 2 (dois) em “*Microbiology, Virology and Immunology*”, como esquematizado no quadro a seguir:

Quadro 4 - Cadeias de repositórios de dados de pesquisa

REPOSITÓRIOS	NEUROSCIENCES	MEDICINE	MICROBIOLOGY, VIROLOGY AND IMMUNOLOGY
German Centre for Cancer Registry Data		X	
Oral Cancer Gene Database	X	X	X
CancerData.org			
Cancer GenomeAtlas Data Portal		X	
Cancer Genomics Hub		X	
Breast Cancer Surveillance Consortium		X	
The Cancer Imaging Archive		X	
Broad-Novartis Cancer Cell Line Encyclopedia		X	
National Cancer Data Base		X	
The Cancer Immunome Atlas		X	X
Cancer Genome Anatomy Project		X	
Cancer in Young People in Canada		X	

Fonte: Elaboração própria

O repositório “*Australian Breast Cancer Tissue Bank*” não aparece no quadro, pois somente foi classificado em níveis de Classe e Categoria. Este fato não o inviabiliza perante uma busca mais sensível por assunto e também não impede que as relações entre áreas no conhecimento, existentes em relação ao conteúdo do repositório, sejam demonstradas. Isso pois, ao investigar um pouco mais sobre os detalhes do repositório, foi possível verificar a atribuição das palavras-chave: “*health*”, “*breast*”,

“*cancer*”, “*tissue bank*” e “*oncology*”. Todas possuem relação a área do conhecimento a qual o repositório pertence e poderia ser subclassificado.

Tendo em vista que as áreas temáticas de abrangência, além de servirem para organizar o Diretório, são um instrumento que permite que os repositórios sejam localizados por assunto, recomenda-se maior atenção quanto à representação temática dos repositórios localizados no Diretório. Isso por meio de um melhor aproveitamento de sua estrutura de organização, definindo os repositórios indexados não só nos níveis mais rasos de Categoria e Cadeia, mas explorando melhor as Disciplinas e Subdisciplinas as quais os repositórios podem estar relacionados.

Desse modo, uma alternativa para o aprimoramento da representação temática dos repositórios no Re3data, é analisar as palavras-chave atribuídas pelos requerentes de modo que no futuro, essas palavras possam ser incorporadas as já existentes em vocabulários controlados e validados por diversas disciplinas.

Aconselha-se também uma análise geral do repositório, considerando o título e propósito para o qual foi criado, ou seja, uma investigação mais profunda. Essas respostas, aliadas aos dados dos depositantes de dados no repositório permitiriam um melhor conhecimento quanto aos elementos que compõem a Ciência a qual esses dados estão debruçados. Essa readequação trará maior visibilidade ao repositório em questão, maior confiabilidade quanto à procedência dos dados localizados no Re3data e poupará o tempo do leitor na seleção desses repositórios, seja para downloads de conjuntos de dados ou para o upload de seus conjuntos de dados ou conhecer o estado da arte quanto ao compartilhamento de dados em determinada Área do conhecimento. A adoção de tais medidas evidencia a importância da classificação não só para organização do conhecimento, mas para a recuperação da informação e quanto ao alinhamento e adesão ao Movimento de Acesso Aberto a Informação em distintas áreas do conhecimento.

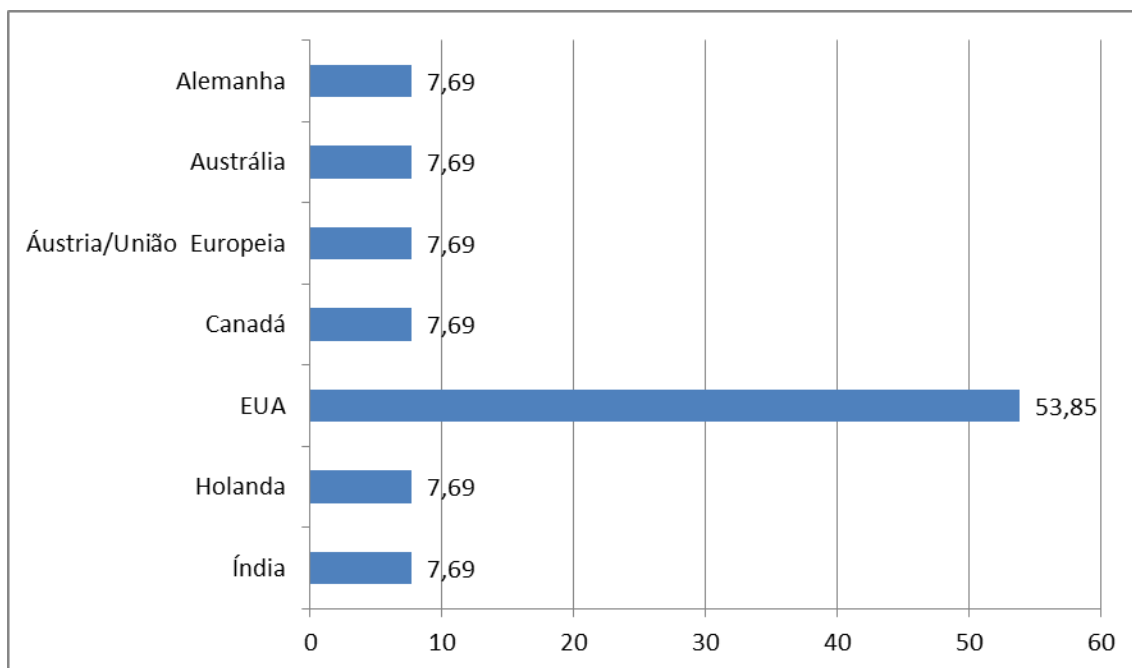
4.3 Abrangência geográfica

Em relação aos países pelos quais os repositórios estão sendo registrados, mantidos, geridos ou financiados, em sua maioria, a abrangência geográfica retrata os países ou regiões mantenedoras dos repositórios.

Referente à amostra estudada, o país com maior quantidade de repositórios de dados de pesquisa em câncer, sete (7) repositórios, são os Estados Unidos da América (EUA) de onde se originam 53,85% dos repositórios de dados de pesquisa em câncer.

Com o mesmo número de infraestruturas, totalizando um (1) repositório, está: a Alemanha (7,69%), a Austrália (7,69%), a Áustria e a União Europeia (7,69%), o Canadá (7,69%), a Holanda e a Índia (7,69%).

Gráfico 3 – Abrangência geográfica dos repositórios de dados de pesquisa



Fonte: Elaboração própria

O fato dos EUA se destacarem como o país com maior quantidade de repositórios de dados de pesquisa em câncer não surpreende. No site do Re3data é possível fazer uma busca por repositórios usando como critério o país de registro, essa mesma página aponta que dos EUA advêm a maioria dos repositórios de dados de pesquisa do mundo, 1072 repositórios.

No Re3data, existe registrado nenhum repositório de dados de pesquisa em câncer brasileiro. A título de curiosidade, no Diretório, foi consultada a seção busca de repositórios por países. Nessa consulta conferimos que o Brasil mantém, gera ou financia nove (9) repositórios de dados de pesquisa. Se comparado a outros países da América do Sul, o Brasil é o país do Continente que possui o maior número de repositórios registrados no Re3data. A Colômbia aparece em segundo lugar com seis (6) repositórios, a Argentina em terceiro lugar com cinco (5) repositórios, em seguida o Peru com três (3) repositórios e o Chile com dois (2) repositórios. Os outros países e territórios que compõem a América do Sul não possuem repositórios registrados no Re3data.

4.4 Tecnologias de repositórios de dados de pesquisa em câncer

No que se refere a tecnologia, a presente seção secundária informa sobre os softwares utilizados, assim como traz informações referentes a restrições de acesso aos dados e conjuntos de dados dos repositórios estudados. Por fim, analisa a existência de publicações ampliadas nos referidos repositórios de dados de pesquisa em câncer.

A busca por informações referentes ao Software foi realizada por meio de consultas a página de cada repositório estudado e ao acesso ao perfil dos mesmos na página do Re3data. No processo em que as páginas oficiais dos repositórios foram verificadas, as suas políticas também foram examinadas, pois poderiam conter informações referentes ao item apurado. Sendo assim, após conferência, dos 13 repositórios que compõem a amostra estudada, 12 não disponibilizam informações referentes ao software utilizado, nem sequer em suas páginas oficiais ou em suas políticas. O único repositório que apresentou informações referentes ao software utilizado foi o *National Cancer Data Base*, que utiliza o software desenvolvido pela *Open Knowledge Foundation: o Comprehensive Knowledge Archive Network (CKAN)*. O CKAN é um software de código aberto que vem sendo utilizado por governos e instituições de pesquisa, que têm como característica o fato de coletarem grandes volumes de dados, objetivando principalmente, o desenvolvimento de sites de dados abertos, gerenciando e promovendo a criação de coleções de dados (CKAN, 2021).

Os repositórios de dados são sistemas de informação que tem como base para seu funcionamento a existência de um software que atenda às necessidades de armazenamento para as quais foram criados, disseminando os objetos digitais neles alocados. Destacando-se como um dos principais elementos para o pleno funcionamento de um repositório, sua origem não deveria ser desconhecida, mas sim receber destaque. Isto porque essa informação justifica e permite aos pesquisadores e usuários entender quais formatos de dados podem estar contidos no repositório, aumentando a visibilidade dos dados disponibilizados e conseqüentemente propiciando melhor interação com o usuário.

Com relação a restrições de acesso dos repositórios que compõem a amostra, foram analisados três aspectos, são eles: acesso ao banco de dados, acesso aos conjuntos de dados de pesquisa e upload de conjuntos de dados. O primeiro aspecto, acesso ao banco de dados, faz referência a possibilidade de acesso ao banco de dados em geral por

parte dos usuários. Ou seja, diz se os usuários podem acessar o banco de dados em geral. O segundo aspecto, acesso aos conjuntos de dados de pesquisa, se relaciona ao acesso aos conjuntos de dados de pesquisa no repositório de dados de pesquisa em particular. Já o terceiro aspecto, upload de conjuntos de dados, faz menção a condições para fazer upload de conjuntos de dados de pesquisa nos repositórios de dados de pesquisa. Os três aspectos mencionados anteriormente se desmembram em: aberto, restrito e/ou fechado. Onde: aberto indica que não existem barreiras de acesso, restrito quer dizer que existem certos limites de acesso e fechado significa que não existe possibilidade de acesso a usuários externos.

Em relação ao tópico “acesso ao banco de dados”, segundo registros no Re3data apenas o repositório Broad-Novartis Cancer Cell Line Encyclopedia teria acesso restrito. Porém, ao acessar o repositório verificou-se que ele está em acesso aberto. Desta forma, todos os repositórios que compõem a amostra são abertos, fato que possibilita acesso livre por parte do usuário aos dados repositórios do corpus da pesquisa.

Quanto ao tópico “acesso aos conjuntos de dados de pesquisa”, verificou-se que os repositórios de dados de pesquisa em câncer estudados não se limitam a somente um valor de entrada. Nessa perspectiva, 4 (30,77%) dos repositórios analisados concedem acesso do tipo restrito; 1 (7,70%) concede acesso aberto; 2 (15,38%) oferecem acesso restrito, aberto ou fechado e 6 (46,15%) repositórios de dados de pesquisa em câncer têm “restrito ou fechado” como níveis de acesso a seus conjuntos de dados de pesquisa. Não foram encontradas informações suficientes para descrever exatamente quais tipos de dados de pesquisa em câncer se encaixam em cada nível de acesso, inclusive, essas informações também não estão presentes nas políticas dos repositórios que integram a amostra.

Se comparado ao tópico “acesso aos conjuntos de dados de pesquisa”, o tópico “upload de conjuntos de dados” apresenta valores mais homogêneos. Isto porque só os valores restrito ou fechado se manifestaram. O tópico “upload de dados”, faz referência ao envio de dados por parte de pesquisadores, instituições e grupos de pesquisa para o Re3data. De acordo com a pesquisa realizada, 8 repositórios têm permissão de upload restrita e em 5 repositórios não há permissão de upload. Isso quer dizer que em 61,54% são aplicados limites em referência ao envio de conjuntos de dados, para usuários cadastrados, e em 38,46% dos repositórios não há possibilidade de upload de conjunto de dados a usuários externos e/ou que não pertencem a Instituição ou grupo de pesquisa

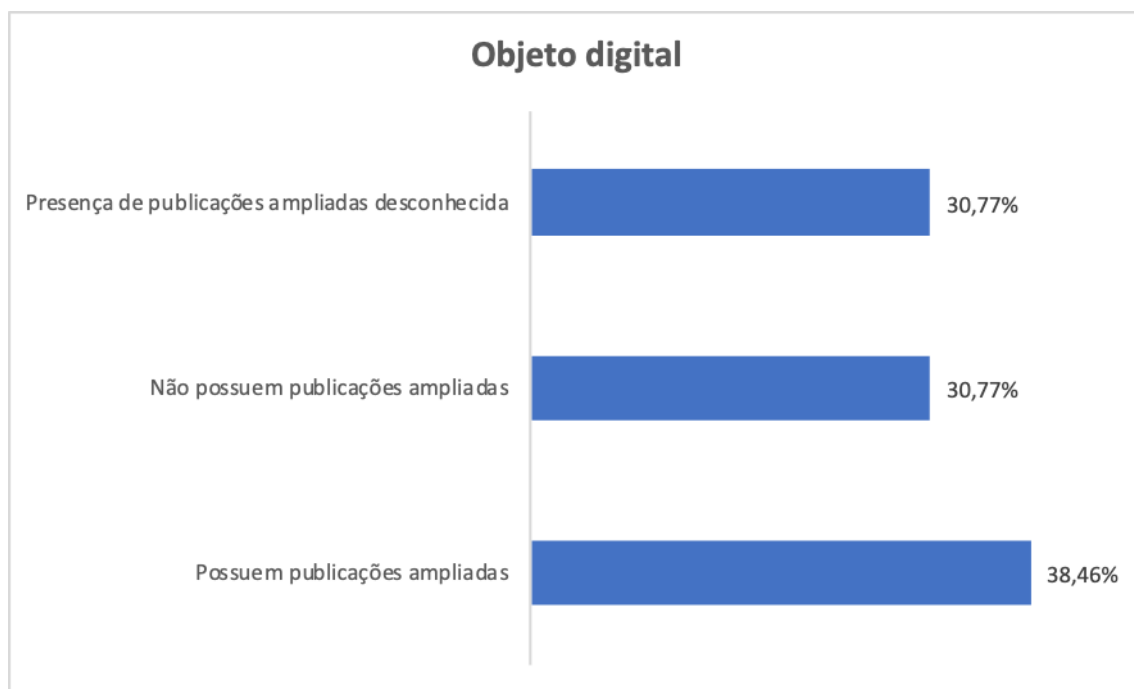
responsável pelo repositório de dados em questão. Como esquematizado no quadro a seguir:

Quadro 5 - Upload de conjuntos de dados

UPLOAD DE CONJUNTOS DE DADOS		
Repositório	Tipo de acesso	
	Restrito	Fechado
German Centre for Cancer Registry Data		X
Oral Cancer Gene Database		X
CancerData.org	X	
Cancer GenomeAtlas Data Portal	X	
Cancer Genomics Hub		X
Breast Cancer Surveillance Consortium	X	
The Cancer Imaging Archive	X	
Broad-Novartis Cancer Cell Line Encyclopedia	X	
National Cancer Data Base	X	
The Cancer Immunome Atlas		X
Cancer Genome Anatomy Project		X
Cancer in Young People in Canada	X	
Australian Breast Cancer Tissue Bank	X	

Fonte: Elaboração própria

Outro fator relevante que se destaca com relação a tecnologia disponibilizada nos repositórios que compõem a amostra é em relação a disponibilização de publicações ampliadas. Dos 13 repositórios de dados de pesquisa em câncer analisados 5 repositórios disponibilizam publicações ampliadas, 4 não disponibilizam em seu acervo publicações ampliadas e 4 não divulgaram informações sobre a presença ou não desse tipo de material em seus acervos.

Gráfico 4 – Publicações ampliadas em repositórios de dados de pesquisa em câncer

Fonte: Elaboração própria

4.5 Tipologia dos dados de pesquisa em câncer

O quadro abaixo apresenta a variedade tipológica de dados de pesquisa em câncer presentes nos repositórios estudados.

Quadro 6 - Tipos de conteúdo de coleções de repositório de dados de pesquisa em câncer

NATUREZA	DADOS IDENTIFICADOS
Imagens	Imagens Clínicas e/ou diagnósticas
	Imagens de tecido
	Imagens radiológicas
	Imagens de rastreamento <ul style="list-style-type: none"> • Mamografia com filme • Mamografia digital • Tomossíntese • Ressonância magnética • Ultrassom • Tomografia computadorizada • Histopatologia digital
Dados pessoais	Sexo (Masculino / feminino) Mês do nascimento Ano de nascimento Idade no diagnóstico CEP Estado Plano de saúde História pessoal de câncer História pessoal de biópsia, cirurgia, radiação Presença de sintomas Motivo principal da visita
	Fatores de estilo de vida <ul style="list-style-type: none"> • Uso de tabaco • Uso de álcool • Altura • Peso • Área de Superfície Corporal (BSA) • Índice de Massa Corporal (IMC)
Dados clínicos	Morfologia: Código de histologia Topografia: Topografia de neoformação Dignidade: Característica do tumor Classificação: Grau de diferenciação do tecido canceroso Tipo de célula: Tipo de tecido canceroso T - Local do tumor primário N - Metástases de linfonodos regionais M - Metástases distantes Classificação clínica / patológica do estágio

	<p>Tipo de câncer anterior Idade no diagnóstico Ano do diagnóstico Tratamento para o câncer Tipo de procedimento Data do procedimento Resultado da avaliação Fatores hormonais Histórico de parto</p>
	<p>Informação de Acompanhamento (Data de acompanhamento e estado da doença)</p>
Dados sobre terapia primária	<p>Cirurgia Radioterapia Quimioterápico Terapia hormonal Imunoterapia Transplante de medula óssea Outras terapias</p>
Dados científicos e estatísticos	<p>Espécimes de doadores não identificáveis</p> <ul style="list-style-type: none"> • Sangue <p>Método de preservação de tecido Tecido fresco e imagens</p>
	<p>Dados genômicos Dados de linhas de células Sequências do genoma do câncer Alinhamentos e informações sobre mutações</p>
	<p>Metadados de analito molecular Dados de caracterização molecular Expressão gênica Composição celular de infiltrados imunológicos Antígenos de linha germinativa de câncer Tipos de HLA Heterogeneidade do tumor Análises imunogenéticas Tags de sequência expressa Padrões de expressão gênica Polimorfismos de nucleotídeo único Citogenética Genes RNAi Cromossomos Índice de gene supressor de tumor</p>
	<p>Metilação Intabilidade de microssatélites (IMS) miRNAs Expressão de mRNA Expressão de proteína</p>

	caracterização molecular
Dados sobre mortalidade	Mês de morte Ano de morte Causa da morte

Fonte: Elaboração própria

5 CONSIDERAÇÕES FINAIS

Países subdesenvolvidos e de alto nível de desenvolvimento econômico e social têm enfrentado o que é considerada a maior crise sanitária mundial da atualidade, a COVID-19. Tanto países desenvolvidos, que possuem excelente acesso a saúde e contam com condições sanitárias apropriadas, como países subdesenvolvidos que não têm recursos considerados básicos para conter e prevenir o avanço do vírus, têm visto os números de mortos e contaminados se multiplicarem dia após dia. A gravidade do cenário atual trouxe à tona o trabalho de pesquisadores, equipes de pesquisa e institutos que se dedicam a pesquisa em saúde. O atual protagonismo desses atores se relaciona a necessidade e urgência em descobrir uma preparação biológica que forneça ao corpo imunidade para a doença.

Na perspectiva de tentar enxergar algo bom em situações ruins, podemos refletir sobre o surgimento da Pandemia como elemento de condução para a valorização da pesquisa em saúde e, conseqüentemente, da Ciência como elemento fundamental a manutenção da vida. A saúde é um direito e para garantir o mesmo a toda população é essencial o contínuo investimento em pesquisa em saúde.

Em um passado não tão distante, era comum se deparar com pessoas que entendiam a saúde como algo individual que não repercutia diretamente nos demais. Mas atualmente, o cenário é outro. A pandemia fez ressurgir a preocupação com a saúde coletiva, isso porque através de evidências científicas foi provado que para termos sucesso no enfrentamento a COVID-19, é necessário mais do que ações individuais. Para ficarmos longe do vírus é preciso um movimento coletivo, é necessário que o outro se cuide, tanto no que diz respeito a medidas de prevenção quanto para medidas de redução da transmissão.

Hoje, apesar da pandemia não ter acabado, felizmente vivemos em um cenário de esperança. Já existem vacinas eficazes sendo aplicadas por todo o mundo e em breve a COVID-19 terá sido somente um episódio ruim na história da humanidade. As vacinas produzidas, que estimulam uma resposta imunológica do organismo, são frutos de pesquisa. Para chegar à fórmula de sucesso em um tempo tão curto, de menos de um ano, foi necessário uma série de testes. Testes esses oriundos de pesquisa que foram realizadas em um ambiente de colaboração que envolveu pesquisadores, laboratórios e voluntários de todo o mundo. Nesse processo de pesquisa o número de dados produzidos e coletados deve ter sido gigantesco, compreendendo não só aspectos de

saúde. Mas também reunindo dados econômicos e sociais referentes ao período. São esses dados sensíveis e complexos que deveriam ser preservados, criando memória em relação a Pandemia enfrentada, ações realizadas e referentes ao impacto de cada decisão.

A pandemia de COVID-19 não é temática deste trabalho, mas nos faz refletir sobre a importância dos dados de pesquisa e conseqüentemente relacioná-los a pandemia do câncer, que são as temáticas estudadas. Diferentemente do COVID-19, onde há perspectiva de extinção do vírus, infelizmente não se espera o mesmo em relação ao câncer. Por ser uma doença complexa, não transmissível e proveniente do crescimento incontrolável de células cancerosas, pesquisadores e profissionais de saúde ainda não descobriram a cura para a doença. Por atingir milhares de pessoas no mundo e ser objeto de inúmeras pesquisas, há de se imaginar que dados de pesquisa em câncer são produzidos o tempo todo e em escalas sem precedentes. Fato que dificulta o acesso e gerenciamento dos conjuntos de dados produzidos e coletados. Nesse sentido o Brasil conta com uma “vantagem” em relação ao registro desses dados, por contar com o SUS, que atende mais de 190 milhões de brasileiros e está presente em todo o território nacional, incluindo institutos e serviços de atendimento especializados em câncer.

Essa vantagem diz respeito a possibilidade de reunir em um só sistema dados de saúde provenientes de pesquisas e pacientes oncológicos. Esses dados ao serem sistematizados e contextualizados serviriam para analisar o perfil epidemiológico da população acometida pela doença, coletar mutações, dados do tecido acometido, dados referentes a terapia aplicada e etc. Permitindo que novas estratégias de promoção da saúde e prevenção do câncer fossem elaboradas, assim como novas tecnologias e procedimentos fossem incorporados a partir de dados válidos, confiáveis, contextualizados e atualizados.

Para que tais medidas sejam aplicadas e/ou alcançadas, além de reconhecer os dados gerados e coletados como objetos informacionais, é preciso considerável esforço na atribuição de metadados aos conjuntos de dados, tal ação dará sentido aos dados disponibilizados. Destacando os dados como potenciais fios condutores da pesquisa no campo oncológico e, em uma visão macro, contribuindo para a consolidação dos dados como principais ativos da sociedade da informação.

Para chegar a fase em que metadados são atribuídos aos dados, é necessário percorrer algumas etapas. Uma das primeiras etapas é justamente o objetivo desse trabalho: conhecer os dados provenientes e usados para pesquisa oncológica. Toda a

discussão proveniente deste estudo teve como finalidade investigar mais profundamente as propriedades dos dados oncológicos.

Deste modo, por meio da pesquisa aqui realizada, foi possível verificar que em relação aos tipos de repositórios de dados de pesquisa em câncer, a maior parte dos mesmos são do tipo disciplinar.

Com relação aos repositórios indexados no Re3data, apesar da plataforma possuir um sistema de classificação que se assemelha a Classificação Decimal de Dewey, a mesma não possui um vocabulário controlado ou informações específicas referentes a essa vertente, fato que contribui para que o processo de indexação dos repositórios incorporados não seja feito de forma profunda. Em relação a área temática de abrangência, todos os dados da amostra pertencem à classe “Ciências da vida”, sendo a maioria atribuído a categoria “Biologia e Medicina”.

Quanto a abrangência geográfica, há nenhum repositório de dados de pesquisa em câncer brasileiro registrado no Re3data, enquanto os EUA se destacam por possuírem a maior quantidade de repositórios de dados de pesquisa em câncer.

A cerca dos softwares utilizados, somente 1 repositório pertencente a amostra estudada indicou informações sobre o software adotado. Em referência as restrições de acesso aos dados, todos os repositórios que integram a amostra são abertos. Já os conjuntos de dados dos repositórios analisados se diferenciam por oferecerem diferentes tipos de acesso: restrito, fechado e/ou aberto, a depender do nível de permissão de acesso do interessado em visualizar os dados ali contidos. Com relação ao acesso ao conjunto de dados de pesquisa, os dados de pesquisa em câncer se caracterizam por ter acesso restrito ou não permitido a seu conjunto de dados. Fato que deve ser justificado por questões éticas, legais/ou e particulares referentes a origem dos dados coletados. Outro fato que foi possível perceber a partir da pesquisa e análise foi a presença ainda tímida de publicações ampliadas, publicações essas que vem ganhando espaço em diversas áreas do conhecimento.

Quanto a tipologia dos dados de pesquisa que compõem a amostra estudada, há grande variedade. Os dados de pesquisa em câncer podem ser imagéticos; pessoais: provenientes de anamnese; dados clínicos; terapêuticos; assim como científicos e estatísticos.

No percurso para elaboração dessa pesquisa a pluralidade dos tipos de dados de pesquisa em câncer foi algo surpreendente. Além dos amplamente conhecidos e citados tipos de dados de pesquisa produzidos e coletados no campo oncológico e de outras

áreas do conhecimento, foram mapeados dados que se diferem dos conhecidos por sua natureza e particularidade. Nessa perspectiva, o trabalho aqui produzido apresenta-se como primeiro passo para que outras pesquisas que explorem, especificamente, dados de pesquisa oncológicos de determinada natureza, sejam desenvolvidas.

Outra questão que chamou a atenção, foram as restrições quanto ao acesso aos bancos de dados e, conseqüentemente, as coleções de dados ali abrigados. Fato que evidencia o cuidado dos repositórios com a segurança dos dados contidos e a necessidade de saber quem e porquê está acessando esses dados.

A partir do estudo desenvolvido, questões como: a padronização de metadados para a recuperação da informação, preservação, custos referentes ao tratamento e tempo de salvaguarda dos dados podem ser discutidos e abordados em estudos futuros de forma mais profunda considerando as características dos dados de pesquisa em câncer. Trazendo para áreas do conhecimento como a Ciência da Informação, aprofundamento de discussões mais complexas sobre o que se destaca atualmente como o principal ativo da sociedade: os dados. Dessa forma, a Ciência da Informação poderá contribuir sugerindo ações para tornar o conhecimento disponibilizado através dos dados acessível a toda a sociedade, aclarando o caráter interdisciplinar da Área.

No ponto de vista econômico e social o estudo aprofundado de dados de pesquisa em câncer e plataformas digitais de gestão, almeja contribuir como elemento no processo de redução de incidência de câncer. Poupano gastos com a coleta dos dados preservados, aplicando novas descobertas tecnológicas e terapêuticas por meio de tratamentos de saúde aprimorados e menos invasivos.

A discussão aqui iniciada incentiva também a criação de repositórios de dados de pesquisa em câncer brasileiros, que espelhem nossas características regionais, servindo como vitrine para pesquisadores e institutos de pesquisa e saúde brasileiros. Refletindo não só em benefícios profissionais aos profissionais dedicados a pesquisa em câncer, mas em benefícios de saúde para a população brasileira. Servindo como canal de divulgação sobre as pesquisas que vêm sendo desenvolvidas sobre a temática em território nacional.

REFERÊNCIAS

ACCART, J. **Serviço de referência: do presencial ao virtual**. Tradução Antonio Agenor Briquet de Lemos. Brasília, DF: Briquet de Lemos, 2012.

BARRETO, A. Bases de dados, repositórios de informação, bibliotecas digitais e virtuais. In: **ALDOBARRETO'S BLOG**. [S. l.], 2010. Disponível em: <<https://aldobarreto.wordpress.com/2010/04/21/bases-de-dados-e-repositorios-de-informacao/>>. Acesso em 13 ago. 2019.

BERTIN, P. R. B.; VISOLI, M. C.; DRUCKER, D. P. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de P&D. **PontodeAcesso**, Salvador, v. 11, n. 2, ago. 2017.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**: promulgada em 5 de outubro de 1988, art. 196. Disponível em: <https://www2.senado.leg.br/bdsf/bitstream/handle/id/518231/CF88_Livro_EC91_2016.pdf>. Acesso em: 04 ago. 2019

_____. Governo do Estado do Rio de Janeiro. **Rio sem fumo**. Rio de Janeiro, 2019. Disponível em: <<http://riocomsaude.rj.gov.br/riosemfumo/site/conteudo/a-lei.php>>. Acesso em: 22 ago. 2019.

_____. **Lei nº 5517, de 17 de agosto de 2009**. Proíbe o consumo de cigarros, cigarrilhas, charutos, cachimbos ou de qualquer outro produto fumígeno, derivado ou não do tabaco, na forma que especifica, e cria ambientes de uso coletivo livres de tabaco. Disponível em: <<https://gov-rj.jusbrasil.com.br/legislacao/818781/lei-5517-09>>. Acesso em: 05 abr. 2018.

BORKO, H. Information Science: what is it? **American Documentation**, v. 19, n. 1, p. 3- 5, Jan. 1968. Disponível em: <<http://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerneda/k---artigo-01.pdf>>. Acesso em: 15 jul. 2019.

BUCKLAND, M. Information as thing. **JASIST**, v. 42, n. 5, jun. 1991. Disponível em: <<http://ppggoc.eci.ufmg.br/downloads/bibliografia/Buckland1991.pdf>>. Acesso em: 12 jun. 2019.

BVS SALUD. **LILACS**. São Paulo: 2019. Disponível em: <<http://lilacs.bvsalud.org/>>. Acesso em: 12 jun. 2019.

CENTRO LATINO-AMERICANO E DO CARIBE DE INFORMAÇÃO EM CIÊNCIAS DA SAÚDE. **DeCS: Descritores em Ciências da Saúde**. São Paulo, 2019. Disponível em: <<http://decs.bvs.br/cgi-bin/wxis1660.exe/decsserver/>>. Acesso em 13 ago. 2019.

COMISSÃO NACIONAL DE ENERGIA NUCLEAR. **10 razões para você criar um plano de Gestão de Dados de Pesquisa**. Rio de Janeiro: CNEN, 2019. Disponível em: <<http://www.cnem.gov.br/component/content/article/78-cin/dicas-academicas/232-10-raozes-para-voce-criar-um-plano-de-gestao-de-dados-de-pesquisa>>. Acesso em: 13 ago. 2019>.

COMPREHENSIVE KNOWLEDGE ARCHIVE NETWORK. **About CKAN**. [S.l.]:CKAN, 2021. Disponível em: <<https://ckan.org/about/>>. Acesso em: 13 fev. 2021.

COUSIJN, H.; FENNER, M. German Research Foundation to fund new services of re3data. In: DATACITE. **DataCite blog**. [S. l.]: DataCite, 2020. Disponível em: <<https://blog.datacite.org/german-research-foundation-to-fund-new-services-of-re3data/>>. Acesso em: 11 mar. 2020.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR. **Buscar base**. Brasília, DF, 2019. Disponível em: <<http://www-periodicos-capes.gov.br.ez29.capes.proxy.ufrj.br/index.php?#base>>. Acesso em: 29 maio 2019.

COSTA, M. P. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. 2017. 269 f. Tese (Doutorado em Ciência da Informação)— Universidade de Brasília, Brasília, DF, 2017.

DATA OBSERVATION NETWORK FOR EARTH. **Data life cycle**. New Mexico: DataONE, 2019. Disponível em: <<https://www.dataone.org/data-life-cycle>>. Acesso em: 10 ago. 2019.

FUNDAÇÃO DE AMPARO A PESQUISA DO ESTADO DE SÃO PAULO. **Plano de gestão de dados**. São Paulo: FAPESP, 2019. Disponível em: <<http://www.fapesp.br/gestaodedados/>>. Acesso em: 13 ago. 2019.

FUNDAÇÃO DO CÂNCER. **Fatores de risco**. Rio de Janeiro, 2019. Disponível em: <<https://www.cancer.org.br/sobre-o-cancer/prevencao/>>. Acesso em 13 ago. 2019.

FUNDAÇÃO OSWALDO CRUZ. Grupo de Trabalho em Ciência Aberta. **Termo de referência: gestão e abertura de dados para pesquisa na Fiocruz**. Rio de Janeiro: FIOCRUZ, 2018.

GIL, A. C. **Como elaborar projetos de pesquisa**. Atlas: São Paulo, 2002.

_____. **Métodos e técnicas de pesquisa social**. Atlas: São Paulo, 2008.

GROGAN, Denis. **A prática do serviço de referência**. Tradução Antonio Agenor Briquet de Lemos. Brasília, DF: Briquet de Lemos, 1995.

FERREIRA, A. B. H. **Miniaurélio: o minidicionário da língua portuguesa**. Curitiba: Posigraf, 2004.

FIGUEIREDO, Nice. Bibliotecas universitárias e especializadas: paralelos e contrastes. **R. Bibliotec. Brasília**, Brasília, DF, v.7, n. 1, p. 9-12, jan./jun. 1979.

FINAMOR, M. S.; LIMA, C. R. M. Bibliotecários em hospitais: práticas informacionais. **P2P e INOVAÇÃO**, v. 4, n. 1, p. 109-129, 2018. DOI: 10.21721/p2p.2017v4n1.p109-129. Acesso em: 22 ago. 2019.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **ABC do câncer**: abordagens básicas para o controle do câncer. Rio de Janeiro: INCA, 2011.

_____. **Causas e prevenção**: prevenção e fatores de risco. Rio de Janeiro, 2018a. Disponível em: <<https://www.inca.gov.br/causas-e-prevencao/prevencao-e-fatores-de-risco>>. Acesso em 13 ago. 2019.

_____. **Estimativa 2018**: incidência de câncer no Brasil. INCA: Rio de Janeiro, 2018b.

_____. **Notícias**: Bebidas muito quentes podem causar câncer de esôfago. Rio de Janeiro, 2016. Disponível em: <<https://www.inca.gov.br/noticias/bebidas-muito-quentes-podem-causar-cancer-esofago>>. Acesso em: 11 jun. 2019.

_____. **Notícias**: Em 16 anos, restrição de fumar em ambientes públicos evita 15 mil mortes de crianças no Brasil, revela estudo. Rio de Janeiro, 2019a. Disponível em: <<https://www.inca.gov.br/noticias/em-16-anos-restricao-de-fumar-em-ambientes-publicos-evita-15-mil-mortes-de-criancas-no>>. Acesso em: 11 jun. 2019.

_____. **Notícias**: INCA promove debate sobre câncer de mama e tira dúvidas da população em tempo real. Rio de Janeiro, 2017. Disponível em: <<https://www.inca.gov.br/noticias/inca-promove-debate-sobre-cancer-de-mama-e-tira-duvidas-da-populacao-em-tempo-real>>. Acesso em: 14 jun. 2019.

_____. **O que é câncer?**. Rio de Janeiro, 2019b.. Disponível em: <<https://www.inca.gov.br/o-que-e-cancer>>. Acesso em 27 maio 2019.

_____. **Tipos de câncer**. Rio de Janeiro, 2019c. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>>. Acesso em 13 ago. 2019.

_____. **Tipos de câncer**: Câncer de pele não melanoma. Rio de Janeiro, 2018c. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-pele-nao-melanoma>>. Acesso em 13 ago. 2019.

_____. **Sobre o INCA**: cooperação com a União Internacional para Controle do Câncer (UICC). Rio de Janeiro, 2018d. Disponível em: <<https://www.inca.gov.br/atuacao-internacional/cooperacao-com-uniao-internacional-para-controle-cancer-uicc>>. Acesso em: 24 jul. 2019.

LE COADIC, Y. F. **A Ciência da Informação**. Tradução de Maria Yeda F. S. de Figueiredo Gomes. Brasília, DF: Briquet de Lemos, 1996.

LIMA, E. Bibliotecas em hospitais. **R. Esc. Bibliotecono. UFMG**, Belo Horizonte, v. 2, n. 2, p. 141-159, set. 1973.

LOFEGO, J.; PINHEIRO, R. Comunicação e informação no controle do câncer de colo uterino no Brasil: uma análise sob perspectiva da integralidade em saúde. **RECIIS**, v. 6, n. 4, fev. 2013.

MEDEIROS, C.B. **Gestão de dados científicos: da coleta à preservação**. SCIELO em Perspectiva, [S. l.], 2018. Disponível em: <<https://blog.scielo.org/blog/2018/06/22/gestao-de-dados-cientificos-da-coleta-a-preservacao/>>. Acesso em: 13 ago. 2019.

MIRANDA, A. C. C. Formação e desenvolvimento de coleções em bibliotecas especializadas. **Inf. & Soc.: Est.**, João Pessoa, v. 17, n. 1, jan./abr. 2007.

MOLINO, C. G. R. C.; MELO, D. O.; RIBEIRO, E. A criação de sítio para disseminar informações sobre medicamentos no SUS: um relato de experiência. **RECIIS**, v. 11, n. 1, 2017.

MONTEIRO, E. C. S. A. **Direitos autorais nos repositórios de dados científicos: análise sobre os planos de gerenciamento dos dados**. 2017. 115 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017.

PITASSI, C.; GONÇALVES, A. A.; MORENO JÚNIOR, V. A. Fatores que influenciam a adoção de ferramentas de TIC nos experimentos de bioinformática de organizações biofarmacêuticas. **Ciênc. saúde coletiva**, v. 19, n. 1, jan. 2014.

ORGANIZAÇÃO PAN AMERICANA DA SAÚDE. **Folha informativa: câncer**. Brasília, DF: OPAS Brasil, 2018. Disponível em: <https://www.paho.org/bra/index.php?option=com_content&view=article&id=5588:folha-informativa-cancer&Itemid=1094>. Acesso em : 13 ago. 2019.

RAMOS, R. S. Ciência e paradigma na pesquisa em câncer: a necessidade do estabelecimento de um debate. **Cad. saúde colet.** v. 21, n. 4, 2013.

REGISTRY OF RESEARCH DATA REPOSITORIES. **About**. Bonn: Re3data, 2020a. Disponível em: <<https://www.re3data.org/about>>. Acesso em: 13 set. 2020.

REGISTRY OF RESEARCH DATA REPOSITORIES. **Browse by subject**. Bonn: Re3data, 2020. Disponível em: <<https://www.re3data.org/browse/by-subject/>>. Acesso em: 13 set. 2020.

REGISTRY OF RESEARCH DATA REPOSITORIES. **FAQ**. Bonn: Re3data, 2020c. Disponível em: <<http://re3data.org/faq>>. Acesso em: 13 set. 2020.

RODRIGUES, C.; BLATTMANN, U. Gestão da informação e a importância do uso de fontes de informação para geração de conhecimento. **Perspect. ciênc. inf.**, v. 19, n. 3, 2014.

RODRIGUES, E. et al. Dados científicos e repositórios de dados. GRUPO DE TRABALHO CONJUNTO DA UNIVERSIDADE DO MINHO. UNIVERSIDADE DO PORTO. **Os repositórios de dados científicos: estado da arte**. RCAAP, jul. 2010.

SALES, L. F.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. **Conhecimento em Ação**, Rio de Janeiro, v. 4, n. 1, jan./jun. 2019.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015.

TARGINO, M. B. Comunicação científica: uma revisão de seus elementos básicos. **Inf. soc.**, João Pessoa, v. 10, n. 2, 2000.

UNION FOR INTERNATIONAL CANCER CONTROL'S. **About us**. Geneva: UICC, 2019. Disponível em: <<https://www.uicc.org/who-we-are/about-us>>. Acesso em: 24 jul. 2019.

UNIVERSIDADE DE SÃO PAULO. SIBI USP. **Acontece: gestão de dados de pesquisa: o que precisamos saber hoje!**. São Paulo: SIBI USP, 2018. Disponível em: <<http://www.sibi.usp.br/?p=17574>>. Acesso em: 13 ago. 2019.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. **Gestão de dados**. São Paulo: UFSCAR, 2019. Disponível em: <http://www.propq.ufscar.br/pesquisador/gestao-de-dados-1>. Acesso em: 13 ago. 2019.

SHERA, J. H. Sobre biblioteconomia, documentação e ciência da informação. FOSKETT, D. J. et al. **Ciência da informação ou informática?**. Organização e tradução Hagar Espanha Gomes. Rio de Janeiro: Calunga, 1980.

VEIGA, V. S. O. **A 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, Rio de Janeiro, 2017.

VITORINO, E. V.; PIANTOLA, D. Dimensões da competência informacional. **Ci. Inf.**, Brasília, DF, v. 40, n. 1, jan.-abr., 2011.

WHYTE, A.; TEDDS, J. **Making the Case for Research Data Management**. Digital Curation Centre, 2011. Disponível em: <<http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm>>. Acesso em: 13 ago. 2019.

WIKIPEDIA. **Hipócrates**. [San Francisco, CA: Wikimedia Foundation, 2019a]. Disponível em: <<https://pt.wikipedia.org/wiki/Hip%C3%B3crates>>. Acesso em: 12 jun. 2019.

WIKIPEDIA. **LILACS**. [San Francisco, CA: Wikimedia Foundation, 2019b]. Disponível em: <<https://pt.wikipedia.org/wiki/Lilacs>>. Acesso em: 12 jun. 2019.

WITT, M. 2,000 Data Repositories and Science Europe's Framework for Discipline-specific Research Data Management. In: DATAcite. **DataCite blog**. [S. l.]: DataCite, 2018. Disponível em: <<https://blog.datacite.org/re3data-science-europe/>>. Acesso em: 11 mar. 2020.

WORLD HEALTH ORGANIZATION. **Global Action plan for the prevention and control of noncommunicable diseases 2013-2020**. Geneva: WHO, 2013.

APÊNDICE A – LEVANTAMENTO BIBLIOGRÁFICO PRELIMINAR

O levantamento bibliográfico preliminar foi realizado no dia 23 de julho de 2019 para dar início ao processo de construção deste estudo. O acesso as bases de dados Web os Science e Medline via PubMed foram realizados via Portal de Periódicos CAPES, já o acesso a BVS Prevenção e Controle de Câncer e LILACS ocorreu via acesso direto aos endereços eletrônicos das fontes de informação em questão.

A seguir estão os quadros que apresentam as estratégias de busca utilizadas para busca em cada fonte de informação. No quadro também é apresentado os anos de publicação e a tipologia dos documentos localizados na busca são descritas na linha “resultado”.

Quadro 7 – Estratégia de busca na BVS Prevenção e Controle de Câncer

BVS PREVENÇÃO E CONTROLE DE CÂNCER
(tw:"Research data" OR tw:"scientific data" OR tw:"dados de pesquisa" OR ti:"datos de investigación") AND (tw:Brasil* OR tw:Brazil*)
Filtros: Não se aplica
Resultado: 5 artigos
Ano de publicação: 2004-2015

Fonte: Elaboração própria

Quadro 8 – Estratégia de busca na LILACS

LILACS
(tw:"Research data" OR tw:"scientific data" OR tw:"dados de pesquisa" OR tw:"datos de investigación") AND (tw:neoplasm\$ OR tw:cancer OR tw:tumor\$ OR tw:tumour\$ OR tw:onco\$ OR tw:carcinoma\$ OR mh:"Medical Oncology" OR tw:cancro OR tw:neoplasia\$) AND (tw:Brasil* OR tw:Brazil*)
Filtros: Não se aplica
Ano de publicação: 2015-2018
Resultado: 39 documentos (37 artigos e 2 teses)

Fonte: Elaboração própria

Quadro 9 – Estratégia de busca no Medline via PubMed

MEDLINE VIA PUBMED
(neoplasms[mh] OR neoplasm*[tiab] OR cancer[tiab] OR tumor*[tiab] OR tumour*[tiab] OR carcinoma*[tiab] OR onco*[tiab]) AND ("Research data"[tiab] OR "scientific data"[tiab]) AND (Brasil*[tiab] OR Brazil*[tiab])
Filtros: Não se aplica
Ano de publicação: 2015-2018
Resultado: 4 artigos

Fonte: Elaboração própria

Quadro 10 – Estratégia de busca no Web of Science

WEB OF SCIENCE
TS=(neoplasms OR neoplasm* OR cancer OR tumor* OR tumour* OR carcinoma* OR onco*) AND TS=("Research data" OR "scientific data") AND TS=(brasil* OR Brazil*)
Filtros: Não se aplica
Ano de publicação: 1999-2019
Resultado: 8 documentos (3 artigos, 3 revisões e 2 anais de eventos)

Fonte: Elaboração própria

Dos documentos recuperados nas buscas, apenas um apresenta proposta semelhante ao problema proposto neste estudo. Trata-se do artigo publicado no ano de 2015, encontrado na Web of Science, “*Simple Management of High Assurance Data in Long-Lived Interdisciplinary Healthcare Research: a Proposal*”. O citado artigo fala de forma superficial sobre a importância do gerenciamento de dados de pesquisa em um Grupo de pesquisa sobre câncer, tendo como foco a obtenção de informações para a elaboração de um sistema de informação a fim aprimorar o gerenciamento de dados do grupo de pesquisa, formado por cientistas, médicos e outros especialistas.

Apesar das temáticas abordadas no estudo selecionado se aproximarem das temáticas que norteiam o presente estudo, os mesmos são heterogêneos devido aos objetivos. No artigo “*Simple Management of High Assurance Data in Long-Lived*

Interdisciplinary Healthcare Research: a Proposal”, os autores objetivam criar um sistema de gestão de dados que melhor atenda às necessidades do grupo de pesquisa. Já neste estudo, a intenção é criar um mapa temático dos dados de pesquisa em câncer, apresentando a tipologia dos dados encontrados e indicando ações mundiais de gestão e compartilhamento de dados de pesquisa em câncer.

APÊNDICE B – REPOSITÓRIOS LOCALIZADOS NO RE3DATA

Quadro 11 – Repositórios com o termo “câncer” no resumo

REPOSITÓRIOS COM O TERMO “CÂNCER” NO RESUMO	
Nome do repositório	Assuntos
National Cancer Data Base	["Hematology, Oncology, Transfusion Medicine", "Medicine", "Medicine", "Life Sciences"]
Surveillance Epidemiology and End Results	["Statistics and Econometrics", "Economic and Social Policy", "Epidemiology, Medical Biometry, Medical Informatics", "Public Health, Health Services Research, Social Medicine", "Hematology, Oncology, Transfusion Medicine", "Economics", "Social and Behavioural Sciences", "Humanities and Social Sciences", "Medicine", "Medicine", "Life Sciences"]
Cancer in Young People in Canada	["Dermatology", "Traumatology and Orthopaedics", "Medicine", "Medicine", "Life Sciences"]
Australian Breast Cancer Tissue Bank	["Medicine", "Life Sciences"]
Carcinogenic Potency Database	["Toxicology and Occupational Medicine", "Medicine", "Medicine", "Life Sciences"]
Cancer Genomics Hub	["Basic Biological and Medical Research", "Human Genetics", "Biology", "Life Sciences", "Medicine", "Medicine"]
German Centre for Cancer Registry Data	["Medicine", "Epidemiology, Medical Biometry, Medical Informatics", "Hematology, Oncology, Transfusion Medicine", "Life Sciences", "Medicine"]
The Cancer Immunome Atlas	["Medicine", "Microbiology, Virology and Immunology", "Human Genetics", "Hematology, Oncology, Transfusion Medicine", "Public Health, Health Services Research, Social Medicine", "Medicine", "Immunology", "Basic Biological and Medical Research", "Biology", "Life Sciences"]
Progenetix	["Medicine", "Medicine", "Life Sciences", "Biology", "Hematology, Oncology, Transfusion Medicine", "Epidemiology, Medical Biometry,

	Medical Informatics", "Human Genetics", "Humanities and Social Sciences", "Bioinformatics and Theoretical Biology", "Basic Biological and Medical Research"]
Cancer Genome Anatomy Project	["General Genetics", "Animal Genetics, Cell and Developmental Biology", "Human Genetics", "Basic Biological and Medical Research", "Biology", "Life Sciences", "Zoology", "Medicine", "Medicine"]
CanGEM	["Life Sciences", "Biology", "Basic Biological and Medical Research", "General Genetics", "Medicine", "Human Genetics", "Medicine"]
Project Achilles	["Bioinformatics and Theoretical Biology", "Human Genetics", "Basic Biological and Medical Research", "Biology", "Life Sciences", "Medicine", "Medicine"]
Broad-Novartis Cancer Cell Line Encyclopedia	["General Genetics", "Basic Biological and Medical Research", "Biology", "Life Sciences", "Human Genetics", "Pharmacy", "Medicine", "Medicine"]
PROFILES Registry	["Microbiology, Virology and Immunology", "Biology", "Medicine", "Life Sciences"]
Breast Cancer Surveillance Consortium	["Gynaecology and Obstetrics", "Medicine", "Medicine", "Life Sciences"]
arrayMap	["Biology", "Basic Biological and Medical Research", "Life Sciences", "Medicine", "Medicine", "Human Genetics", "General Genetics", "Hematology, Oncology, Transfusion Medicine"]
Genomic Data Commons	["Public Health, Health Services Research, Social Medicine", "Life Sciences", "Medicine", "Medicine", "Human Genetics"]
Integrated Risk Information System	["Toxicology and Occupational Medicine", "Medicine", "Medicine", "Life Sciences"]
The Cancer Imaging Archive	["Radiation Oncology and Radiobiology", "Biology", "Life Sciences", "Medicine", "Medicine", "Radiology and Nuclear Medicine"]
CaPSURE	["Hematology, Oncology, Transfusion Medicine", "Urology", "Medicine", "Medicine", "Life Sciences"]
Oral Cancer	["Basic Biological and Medical Research", "Microbiology, Virology

Gene Database	and Immunology", "Medicine", "Neurosciences", "Biology", "Life Sciences", "Medicine"]
Mouse Tumor Biology Database	["Basic Biological and Medical Research", "Animal Genetics, Cell and Developmental Biology", "Biology", "Life Sciences", "Zoology", "Bioinformatics and Theoretical Biology"]
EPIC study	["Empirical Social Research", "Epidemiology, Medical Biometry, Medical Informatics", "Social Sciences", "Social and Behavioural Sciences", "Humanities and Social Sciences", "Medicine", "Medicine", "Life Sciences", "Public Health, Health Services Research, Social Medicine"]
Cancer Models Database	["Basic Biological and Medical Research", "Hematology, Oncology, Transfusion Medicine", "Bioinformatics and Theoretical Biology", "Clinical Chemistry and Pathobiochemistry", "Biology", "Life Sciences", "Medicine", "Medicine"]
Canada's Michael Smith Genome Sciences Centre	["Basic Biological and Medical Research", "Microbiology, Virology and Immunology", "Medicine", "Epidemiology, Medical Biometry, Medical Informatics", "Biology", "Life Sciences", "Medicine"]
HIstome	["Cell Biology", "General Genetics", "Basic Biological and Medical Research", "Biology", "Life Sciences"]
Cancer GenomeAtlas Data Portal	["Basic Biological and Medical Research", "Human Genetics", "Biology", "Life Sciences", "Medicine", "Medicine"]
Growing Up Today Study	["Public Health, Health Services Research, Social Medicine", "Medicine", "Medicine", "Life Sciences"]
PlasmID	["Cell Biology", "General Genetics", "Basic Biological and Medical Research", "Biology", "Life Sciences", "Animal Genetics, Cell and Developmental Biology", "Human Genetics", "Zoology", "Medicine", "Medicine"]
caArray	["Life Sciences", "Biology", "General Genetics", "Cell Biology", "Bioinformatics and Theoretical Biology", "Medicine", "Human Genetics", "Hematology, Oncology, Transfusion Medicine", "Basic Biological and Medical Research", "Medicine"]
Project Data	["Radiation Oncology and Radiobiology", "Medicine", "Public Health, Health Services Research, Social Medicine", "Hematology, Oncology,

Sphere	Transfusion Medicine", "Medicine", "Life Sciences"]
Nationale Kohorte	["Medicine", "Medicine", "Life Sciences", "Biology", "Public Health, Health Services Research, Social Medicine", "Basic Biological and Medical Research"]
RADAM database portal	["Life Sciences", "Biology", "Basic Biological and Medical Research"]
Exposome-Explorer	["Public Health, Health Services Research, Social Medicine", "Gastroenterology, Metabolism", "Medicine", "Medicine", "Life Sciences"]
CancerData.org	["Basic Biological and Medical Research", "Medicine", "Biology", "Life Sciences"]
caNanoLab	["Basic Biological and Medical Research", "Cell Biology", "Microbiology, Virology and Immunology", "Medical Microbiology, Molecular Infection Biology", "Biology", "Life Sciences", "Medicine"]
iReceptor	["Basic Biological and Medical Research", "Life Sciences", "Medicine", "Biology", "Microbiology, Virology and Immunology", "General Genetics", "Cell Biology", "Immunology"]
MolTable	["Medicine", "Molecular Chemistry", "Pharmacy", "Pharmacology", "Chemical Solid State and Surface Research", "Chemistry", "Medicine", "Life Sciences", "Natural Sciences"]
DTU Bioinformatics	["Bioinformatics and Theoretical Biology", "Basic Biological and Medical Research", "Microbiology, Virology and Immunology", "Human Genetics", "Biology", "Life Sciences", "Medicine", "Medicine"]
IMEx	["Cell Biology", "Microbiology, Virology and Immunology", "Basic Biological and Medical Research", "Biology", "Life Sciences", "Medicine"]
IntAct	["General Genetics", "Microbiology, Virology and Immunology", "Basic Biological and Medical Research", "Biology", "Life Sciences", "Medicine"]
TOXNET	["Biology", "Medicine", "Chemistry", "Basic Biological and Medical Research", "Molecular Chemistry", "Cell Biology", "Life Sciences", "Natural Sciences"]
Biogrid	["Biology", "Basic Biological and Medical Research", "General

Australia	Genetics", "Hematology, Oncology, Transfusion Medicine", "Life Sciences", "Medicine", "Medicine", "Public Health, Health Services Research, Social Medicine"]
West of Scotland Twenty-07 Study	["Social Sciences", "Public Health, Health Services Research, Social Medicine", "Social and Behavioural Sciences", "Humanities and Social Sciences", "Medicine", "Medicine", "Life Sciences"]
Centers for Disease Control and Prevention, Data & Statistics	["Medicine", "Epidemiology, Medical Biometry, Medical Informatics", "Public Health, Health Services Research, Social Medicine", "Social Sciences", "Psychology", "Basic Biological and Medical Research", "Medicine", "Life Sciences", "Social and Behavioural Sciences", "Humanities and Social Sciences", "Biology"]
International Toxicity Estimates for Risk	["Toxicology and Occupational Medicine", "Public Health, Health Services Research, Social Medicine", "Medicine", "Medicine", "Life Sciences"]
Apollo	["Humanities", "Social and Behavioural Sciences", "Biology", "Medicine", "Agriculture, Forestry, Horticulture and Veterinary Medicine", "Chemistry", "Physics", "Mathematics", "Geosciences (including Geography)", "Engineering Sciences", "Humanities and Social Sciences", "Life Sciences", "Natural Sciences"]
The Human Protein Atlas	["Structural Biology", "Cell Biology", "Structural Biology", "Basic Biological and Medical Research", "Biology", "Human Genetics", "Life Sciences", "Medicine", "Medicine"]
CDC WONDER	["Basic Biological and Medical Research", "Medicine", "Medicine", "Epidemiology, Medical Biometry, Medical Informatics", "Biology", "Public Health, Health Services Research, Social Medicine", "Life Sciences"]
ExAC Browser	["Cell Biology", "Structural Biology", "General Genetics", "Microbiology, Virology and Immunology", "Medicine", "Basic Biological and Medical Research", "Biology", "Life Sciences", "Medicine"]
HAGR	["Biology", "General Genetics", "Human Genetics", "Life Sciences", "Basic Biological and Medical Research", "Medicine", "Medicine", "Animal Genetics, Cell and Developmental Biology", "Zoology"]

Gazel	["Epidemiology, Medical Biometry, Medical Informatics", "Medicine", "Medicine", "Life Sciences", "Public Health, Health Services Research, Social Medicine"]
Canadian Epigenetics, Environment and Health Research Consortium Network	["Immunology", "Bioinformatics and Theoretical Biology", "Microbiology, Virology and Immunology", "Medicine", "Biology", "Life Sciences", "Basic Biological and Medical Research"]
HumanCyc	["Endocrinology, Diabetology", "Human Genetics", "Gastroenterology, Metabolism", "Medicine", "Medicine", "Life Sciences"]
Chemical Carcinogenesis Research Information System	["Basic Biological and Medical Research", "Molecular Chemistry", "Cell Biology", "Structural Biology", "Biology", "Life Sciences", "Chemistry", "Natural Sciences"]
45 and Up Study	["Public Health, Health Services Research, Social Medicine", "Social and Behavioural Sciences", "Humanities and Social Sciences", "Medicine", "Medicine", "Life Sciences", "Social Sciences"]
Mouse Atlas of Gene Expression	["Microbiology, Virology and Immunology", "Animal Genetics, Cell and Developmental Biology", "Biomedical Technology and Medical Physics", "Medicine", "Biology", "Life Sciences", "Zoology", "Medicine"]
PDX Finder	["Life Sciences", "Medicine", "Medicine", "Pharmacology", "Pharmacy", "Public Health, Health Services Research, Social Medicine"]
Mouse Phenome Database	["Basic Biological and Medical Research", "Cell Biology", "Medicine", "Human Genetics", "Biology", "Life Sciences", "Medicine"]
XNAT CENTRAL	["Systemic Neuroscience, Computational Neuroscience, Behaviour", "Cognitive Neuroscience and Neuroimaging", "Radiation Oncology and Radiobiology", "Biomedical Technology and Medical Physics", "Neurosciences", "Medicine", "Life Sciences", "Medicine"]
Androgen Receptor Mutations	["Life Sciences", "Biology", "General Genetics", "Medicine", "Public Health, Health Services Research, Social

Database	Medicine", "Medicine", "Basic Biological and Medical Research"]
Wiki-Pi	["Biology", "Basic Biological and Medical Research", "General Genetics", "Life Sciences"]
pSILAC	["Bioinformatics and Theoretical Biology", "Medicine", "Neurosciences", "Structural Biology", "Basic Biological and Medical Research", "Biology", "Life Sciences"]
ALEXA	["Medicine", "Microbiology, Virology and Immunology", "General Genetics", "Epidemiology, Medical Biometry, Medical Informatics", "Life Sciences", "Biology", "Basic Biological and Medical Research", "Medicine"]
ABCD Data Repository	["Life Sciences", "Medicine", "Medicine", "Public Health, Health Services Research, Social Medicine"]
Complete Genomics	["General Genetics", "Human Genetics", "Basic Biological and Medical Research", "Biology", "Life Sciences", "Medicine", "Medicine"]
LIPID MAPS Lipidomics Gateway	["Medicine", "Medicine", "Endocrinology, Diabetology", "Cell Biology", "Basic Biological and Medical Research", "Biology", "Life Sciences"]
Comprehensive Epidemiological Data Resource	["Medicine", "Electrical Energy Generation, Distribution, Application", "Medicine", "Life Sciences", "Electrical Engineering", "Computer Science, Electrical and System Engineering", "Engineering Sciences"]

Fonte: Elaboração própria

Quadro 12 – Universo de pesquisa

REPOSITÓRIOS DE DADOS DE PESQUISA EM CÂNCER	
1.	MolTable
2.	Karlsruhe Astrophysical Database of Nucleosynthesis in Stars
3.	Centers for Disease Control and Prevention, Data & Statistics
4.	The whole brain Atlas
5.	Melanoma Molecular Map Project
6.	Forschungsdatenzentrum am Robert Koch Institut
7.	Canada's Michael Smith Genome Sciences Centre
8.	Mouse Phenome Database
9.	CanGEM
10.	caArray
11.	caNanoLab
12.	ALEXA
13.	DTU Bioinformatics
14.	Apollo
15.	Mouse Atlas of Gene Expression
16.	IMEx
17.	IntAct
18.	German Centre for Cancer Registry Data
19.	Canadian Epigenetics, Environment and Health Research Consortium Network
20.	XNAT CENTRAL
21.	Surveillance Epidemiology and End Results
22.	Australian Breast Cancer Tissue Bank
23.	The Human Protein Atlas

24.	HIstome
25.	Oral Cancer Gene Database
26.	CancerData.org
27.	Cancer GenomeAtlas Data Portal
28.	Cancer Genomics Hub
29.	Chemical Carcinogenesis Research Information System
30.	Breast Cancer Surveillance Consortium
31.	HumanCyc
32.	Zentrum für Klinische Studien der Universität zu Köln
33.	TOXNET
34.	pSILAC
35.	Integrated Risk Information System
36.	International Toxicity Estimates for Risk
37.	Carcinogenic Potency Database
38.	The Cancer Imaging Archive
39.	Sicas Medical Image Repository
40.	Complete Genomics
41.	45 and Up Study
42.	Broad-Novartis Cancer Cell Line Encyclopedia
43.	CaPSURE
44.	Gazel
45.	Project Achilles
46.	West of Scotland Twenty-07 Study
47.	National Cancer Data Base
48.	HAGR
49.	Synapse

50.	Mouse Tumor Biology Database
51.	EPIC study
52.	CDC WONDER
53.	Androgen Receptor Mutations Database
54.	The Cancer Immunome Atlas
55.	Genomic Data Commons
56.	Clinical Proteomic Tumor Analysis Consortium Data Portal
57.	ExAC Browser
58.	Exposome-Explorer
59.	PROFILES Registry
60.	Cancer Genome Anatomy Project
61.	Nationale Kohorte
62.	Biogrid Australia
63.	PlasmID
64.	arrayMap
65.	National Biomedical Imaging Archive
66.	Network Data Exchange
67.	iReceptor
68.	Wiki-Pi
69.	Progenetix
70.	Cancer in Young People in Canada
71.	PDX Finder
72.	Project Data Sphere

Fonte: Elaboração própria

APÊNDICE C – PLANO DE GESTÃO DE DADOS DE PESQUISA

O presente documento trata-se de um plano de gestão de dados de pesquisa para o estudo “dados de pesquisa em câncer: um mapeamento das ações de gestão e compartilhamento”, que tem como autora: Jéssica Fernanda dos Santos Lima Ramos e será submetido ao Programa de Pós-Graduação em Ciência da Informação, convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro/Escola de Comunicação, como requisito parcial à obtenção do grau de Mestre em Ciência da Informação.

No decorrer da pesquisa os seguintes dados serão coletados e/ou criados: proposta de projeto; documentos no formato RTF, ODT e WORD; planilhas; cadernos de anotações; estratégias de busca; slides; metodologias; listas de referências; modelos e arquivos multimídia no formato JPEG. Em relação à quantidade de dados gerados pela pesquisa, ainda não há estimativa de volume de dados.

Os dados que formarão a coleção de dados serão coletados em fontes de informação primárias e secundárias, localizadas por meio de pesquisa exploratória.

No que se refere à nomeação dos arquivos, o processo se delinear-se-á de forma padronizada, sem abreviaturas e caracteres especiais. Sendo assim ocorrerá da seguinte forma: nome completo do arquivo (assunto) + título do projeto + data da última modificação.

Quanto às medidas de garantia, segurança e controle de qualidade, os dados em formatos digitais serão arquivados na nuvem, em serviços de armazenamento online, mais precisamente no *DropBox* e *Google Drive*. O acesso aos dados será feito mediante *login* e senha de conhecimento, somente, da autora dos dados.

Em relação ao licenciamento dos dados, optaremos por licenças menos restritivas, já que as coleções de dados não contêm dados pessoais, sensíveis ou projetos patenteáveis. Além disso, os dados coletados no decorrer da pesquisa não dependem de permissão do detentor do *copyright*.

A autora do estudo será responsável pelo gerenciamento, coleta e backup da coleção de dados de pesquisa elaborada. Quanto à gestão em longo prazo, os dados gerados na pesquisa serão arquivados e processados em repositório ainda não definido. No momento, verifica-se a possibilidade de armazenamento da coleção de dados gerada

no Repositório Institucional do Instituto Brasileiro de Informação em Ciência e Tecnologia (RIDI).

Não há custos de armazenamento e preservação em relação à gestão a longo ou curto prazo. Os direitos sobre os dados gerados são reservados à autora da pesquisa, sendo liberado seu uso para fins acadêmicos e científicos.

A fim de *linkar* a coleção de dados a publicações acadêmicas que serão originadas a partir dele, oferecendo maior visibilidade aos dados e contribuindo para a “ecologia de dados e informações”, pretende-se criar um identificador persistente pessoal (ORCID) assim como um identificador persistente (DOI) a coleção de dados.