

Uma proposta de taxonomia para dados de pesquisa

A taxonomy proposal for research data

Luana Farias Sales

ORCID: <http://orcid.org/0000-0002-3614-2356>

Doutora em Ciência da Informação pelo Programa de Pós-Graduação do IBICT/UFRJ

Docente do Programa de Pós-Graduação em Ciência da Informação do IBICT/UFRJ

luanafsales@gmail.com

Luís Fernando Sayão

ORCID: <http://orcid.org/0000-0002-6970-0553>

Doutor em Ciência da Informação pelo Programa de Pós-Graduação do IBICT/UFRJ

Docente do Programa de Pós-Graduação em Biblioteconomia da UNIRIO e do Programa de Pós-Graduação em Memória e Acervos da Fundação Casa de Rui Barbosa

lsayao@cnen.gov.br

RESUMO: No contexto da ciência contemporânea, os dados de pesquisa deixam de ser meros subprodutos das atividades de pesquisa e ressurgem como protagonistas na busca por novos conhecimentos. Esse fenômeno é impulsionado pelas tecnologias digitais que criam as condições para o surgimento de um genuíno big data, científico e engendram processos de pesquisa baseados na coleta, geração, processamento e análise de massivas quantidades de dados estruturados em bases de dados. Os pesquisadores, instituições acadêmicas, formuladores de políticas científicas e agências de fomento começam a compreender que os dados de pesquisa se bem gerenciados se tornam recursos informacionais imprescindíveis que podem ser compartilhados e reusados como input para novas pesquisas. Entretanto, os dados, diferentes das publicações, são heterogêneos, diversificados, gerados para diferentes propósitos, por diferentes tecnologias e em domínios disciplinares específicos. Observa-se que há uma lacuna terminológica que dificulta a gestão desses ativos informacionais. Partindo desse ponto, o presente trabalho, aceita o desafio de propor uma taxonomia para a classificação de tipos de dados de pesquisa, ancorado na abordagem teórico-metodológico da Organização do Conhecimento, que apoie a modelagem de sistemas de gestão de dados de pesquisa mais efetivos.

PALAVRAS-CHAVE: Dados de pesquisa. Taxonomia. Tipos de dados.

ABSTRACT: In the context of contemporary science, research data are not mere subproducts of research activities and resurfaces as protagonists in the search for new knowledge. This phenomenon is driven by digital technologies that create the conditions for the emergence of a genuine big data, scientific and engender research processes based on the collection, generation, processing and analysis of massive amounts of data structured in databases. Researchers, academic institutions, policy makers, and development agencies are beginning to understand that well-managed research data becomes an essential information resource that can be shared and reused as input for further research. However, the data, different from the publications, are heterogeneous, diversified, generated for different purposes, by different technologies and in specific disciplinary domains. It is observed that there is a terminological gap that makes it difficult to manage these informational assets. From this point, the present paper accepts the challenge of proposing a taxonomy for the classification of types of research data, anchored in the theoretical-methodological approach of the Knowledge Organization, which supports the modelling of more effective research data management systems.

KEYWORDS: Research data. Taxonomy. Data types

1 Introdução

Cientistas de todo o mundo têm abordado a necessidade de aumentar o acesso global aos dados de pesquisa que são produzidos em quantidade cada vez maior. Isto acontece essencialmente devido à tecnologia digital que se torna cada vez mais um elemento onipresente nos processos da construção do conhecimento científico e permite também que esse conhecimento seja compartilhado e construído de forma cooperativa. Os pesquisadores, as instituições acadêmicas e as agências de fomento à pesquisa começam a entender que estes dados, se preservados e bem gerenciados, constituem uma fonte imprescindível de recursos informacionais que podem ser compartilhados e reutilizados como insumo para novas pesquisas.

A Declaração de Berlim sobre o Acesso Aberto ao Conhecimento em Ciências e Humanidades, publicada em 2003, amplia o escopo do que se entende por acesso

livre ao definir que as “contribuições de acesso livre incluem resultados de pesquisas científicas originais, dados não processados e metadados, fontes originais, representações digitais de materiais pictóricos e gráficos e materiais acadêmicos multimídia” (BERLIN, 2003).

Como um ponto de inflexão histórica da ascensão dos dados, a *D-Lib Magazine*¹ – um periódico importante no que envolve as pes-

¹ Disponível em: <<http://www.dlib.org/dlib/january11/01contents.html>>. Acesso em: 20 maio 2013. O periódico encerrou suas atividades em julho de 2017.

quisas em bibliotecas digitais – publicou no início de 2011, um número especial sobre dados de pesquisa onde estão apresentadas questões como acesso livre, curadoria digital, aquisição e gestão, qualidade e confiabilidade e as possíveis conexões entre dados de pesquisa e as publicações acadêmicas tradicionais. Em 2014, essa mesma revista volta a publicar outro número sobre dados de pesquisa, mas dessa vez, enfatizando a criação do *Research Data Alliance*², também conhecido pela sua sigla, RDA – uma aliança

² <<https://rd-alliance.org/>>

internacional entre os diversos atores interessados na preservação e no tratamento de dados de pesquisa.

Nesse contexto científico de grandes novidades, é imprescindível considerar a relevância da temática “dados de pesquisa” para a ciência contemporânea, bem como todos os, seus desdobramentos, como Gestão de dados, Curadoria Digital, Ciência Aberta, e-Science e Big Data Científico entre tantos outros; é preciso e considerar ainda a interdisciplinaridade e a natureza díspar e heterogênea intrínseca do objeto “dado de pesquisa”, que pode se originar em diferentes áreas de conhecimento, com metodologias, jargões, práticas e fluxos específicos de cada domínio. Partindo dessa diversidade, o presente trabalho identifica uma lacuna na caracterização dos dados de pesquisa. Esta lacuna pode ser explicitada pela ausência de um instrumento terminológico e classificatório que fundamente a gestão e a curadoria de dados pesquisa, que seja capaz de apoiar uma interlocução entre os diversos atores que pesquisam esta temática e que estabelecem políticas, normas, padrões e práticas para propiciar o uso e o reuso de dados em seu potencial máximo e transversalmente, por diversas disciplinas.

Essa preocupação se justifica pelo fato do potencial de reuso constituir no principal parâmetro de avaliação dos dados. A real veracidade do mantra “dados de pesquisa são para serem reusados” implica em níveis crescentes de infraestruturas de gestão e de tecnologia, além de sustentabilidade temporal, política e econômica. O reuso de dados de pesquisa só se torna possível se os dados forem recuperados, acessados, compreensíveis e sejam estáveis ao longo do tempo. Para isso, esses dados precisam ser curados e bem gerenciados. Uma boa curadoria e gestão efetiva de dados se dará se for feita considerando as especificidades do domínio e também as características dos dados gerados nesse domínio. Isso significa dizer que estudar as propriedades do dado e como ele se manifesta em cada domínio é condição necessária para a construção de critérios de seleção que tornarão o dado reusado.

Neste mesmo sentido, plataformas de gestão de dados de pesquisa, bem como políticas de gestão necessitam de uma categorização precisa dos tipos de dados que serão gerenciados e curados. Assim, sob a abordagem teórico-metodológica da Organização do Conhecimento, o presente trabalho vem propor uma taxonomia para classificação de tipos dados de pesquisa na intenção de auxiliar os atores que lidam com dados de pesquisa a gerenciarem e curarem dados de pesquisa de forma efetiva.

2 Dados de pesquisa: uma tentativa de definição

Inspirado na Teoria do Conceito de Dahlberg (1978) e na Teoria de Eugene Wüster (1981), o presente estudo, parte de uma tentativa de compreensão do conceito de dado de pesquisa. Em uma abordagem exploratória, encontramos algumas definições que mereceram ser destacadas aqui.

O *National Research Council* dos EUA, em seu relatório sobre direitos privados e interesse público em bases de dados técnico-científicas, define dados como “fatos, números, letras, símbolos que descrevem um objeto, uma condição, uma situação ou outro fator” (NATIONAL RESEARCH COUNCIL, 1999, p.15). Esta definição pode variar consideravelmente entre colaboradores e de acordo com a área em que são

utilizados.

A Organização para Cooperação e Desenvolvimento Econômico (OCDE, 2007), em seu guia para acesso aos dados de pesquisas financiadas por recursos públicos, define como dados de pesquisa “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.

A Secretaria de Gestão e Orçamento dos EUA define dados de pesquisa como aqueles “coletados, observados ou criados para fins de análise para produzir resultados originais de pesquisa”. A Universidade de *Edmburg* os define como “[...] material factual registrado comumente aceito na comunidade científica como necessário para validar os resultados da pesquisa ...” Já o *National Endowment for the Humanities* define os dados como “materiais gerados ou coletados durante a realização de pesquisas [...]”. Como pode-se perceber são inúmeras as definições de dados de pesquisa, Borgman (2011) explica o porquê da dificuldade em definir dados de pesquisa. Para ela “Informação é um conceito complexo com centenas de definições [...]. Dado [por sua vez] é um conceito simples com poucas definições, porém sujeito a muitas e diferentes interpretações”. Ou seja, o que dificulta atribuir uma definição consensual ao dado de pesquisa é o fato idiossincrático que ele pode ser muitas coisas diferentes para pessoas e circunstâncias diferentes, ou conforme Buckland (1991) “os dados existem apenas aos olhos do observador”. Isto acontece porque dado de pesquisa é dependente de interpretação. O pesquisador é um interpretador de dados científicos.

Neste sentido, uma sequência de bits proveniente de um sensor sísmico é dado de pesquisa para os sismólogos; amostras de rochas são dados de pesquisa para um geomorfologista; conversas gravadas são dados de pesquisa para sociólogos; e inscrições em cuneiformes são dados de pesquisa para quem estuda linguagens do Oriente. “Porém, os cuneiformes podem ser também dados para o arqueólogo ou para o ambientalista que buscam padrões climáticos históricos; de forma similar, os dados sísmicos podem ser úteis para biólogos que estudam comportamento animal”. (BORG-

MAN, 2007, p.119)

Nesta pesquisa propomos a seguinte definição:

Dado de pesquisa é 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.

Essa definição é suficientemente ampla para abarcar todas as possibilidades de dados de pesquisa. No entanto, é importante destacar que para que o registro se configure como dado de pesquisa, ele precisa ser tratado e aceito pela comunidade. Isso coloca em voga a obrigação de uma gestão mínima, com atribuição de metadados que tornem o registro compreensível à comunidade científica.

De acordo com Borgman (2010, p.3), alguns tipos de dados têm tanto valor imediato e duradouro, alguns ganham valor ao longo do tempo, outros têm valor transiente, alguns dados são capturados num momento específico e não se repetirão jamais, enquanto outros são passíveis de serem reproduzidos. Tipos de dados podem incluir, por exemplo, números, imagens, textos, vídeos, áudio, *software*, algoritmos, equações, animações, modelos, simulações. Além do mais, a noção de dados pode variar consideravelmente entre pesquisadores e, ainda mais, entre áreas do conhecimento. A constatação de que os dados são gerados para diferentes propósitos, por diferentes comunidades acadêmicas e científicas e por meio de diferentes processos intensifica ainda mais essa percepção de diversidade e torna imperativo o estabelecimento de uma taxonomia de tipos de dados de pesquisa.

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. O reconhecimento dessas diferenças torna-se crucial para diversas ações no escopo da gestão de dados de pesquisa e do ciclo de vida da curadoria, como por exemplo, para as fases de desenvolvimento de coleções de dados, fase de preservação, fase de versionamento e até mesmo para o reuso. Assim, o presente trabalho vem apresentar uma proposta de taxonomia para dados de pesquisa que tem por finalidade auxiliar o gestor dos dados, seja o pesquisador, ou o bibliotecário no seu papel de

curador a desempenhar suas tarefas de forma mais efetiva.

3 Uma classificação para dados de pesquisa

A pesquisa realizada se iniciou a partir de uma abordagem exploratória e de cunho qualitativo que se configurou como um levantamento na literatura dos conceitos e tipos de dados de pesquisa. Em seguida, a pesquisa partiu para uma abordagem empírica que se configurou no levantamento dos tipos de dados existentes em uma área do conhecimento multidisciplinar: a área de ciências nucleares. Essa abordagem empírica teve por finalidade verificar a ausência de algum tipo de dado. O levantamento foi realizado a partir de entrevista com os líderes de pesquisa de uma instituição brasileira voltada a pesquisa nessa área: o Instituto de Engenharia Nuclear da Comissão Nacional de Energia Nuclear.

No levantamento realizado algumas tentativas de classificação dos dados foram encontradas, como a da *National Science Foundation* (NSF) (2007), a de Borgman (2010), de Harvey (2010), de Lyon (2007) e da OCDE (2007) e, em uma tentativa de sistematização, chegou-se ao seguinte quadro:

Quadro 1 - Tentativa de sistematização a partir da literatura.

CARACTERÍSTICA DE DIVISÃO	NSF	BORGMAN	HARVEY	LYON
Quanto à natureza	Número Imagem Software			
Quanto à origem	Observacionais Computacionais Experimentais	Registro		
Quanto ao nível de processamento	Intermediário Finais			
Quanto ao estágio de geração		Brutos Derivados		Crus Primários Derivados
Quanto à formação de coleções			De pesquisa De comunidade De referência	
Quanto à mutabilidade				Canônicos Episódicos

Fonte: Os autores.

No entanto, essas classificações eram limitadas a uma visão única sobre os dados. Juntar essas múltiplas visões parecia ser uma saída, mas elas se sobrepunham em alguns conceitos e pareciam divergir-se em outros. Além disso, com a abordagem empírica da pesquisa, que se deu a partir da entrevista realizada com pesquisadores da área de ciências nucleares, percebeu-se a ausência de alguns tipos de dados, como era pressuposto. O quadro a seguir revela os tipos de dados e uma tentativa de definição das tipologias encontradas realizada a partir de estudo conceitual.

Quadro 2 - Definição dos tipos de dados e encontrados na área de Ciências Nucleares

TIPOS DE DADOS		DEFINIÇÃO
Número	Medidas	Quantidade fixada por um padrão para determinar as dimensões ou o valor de uma grandeza da mesma espécie
	Resultados de levantamentos	Resultado de pesquisa prévia e mais ou menos aprofundada de um fenômeno, antes de se fazer um projeto, um programa, uma pesquisa científica etc. (coleta)
	Resultados de experimentos	Resultado de trabalho científico que se destina a verificar um fenômeno.
	Fórmulas	Expressão concisa e rigorosa, constituída em geral de símbolos, que resumem certo número de dados
	Equações	Redução de uma questão, um problema intrincado, a pontos simples e claros, para facilitar a obtenção de uma solução
	Algoritmos	Seqüência finita de regras, raciocínios ou operações que, aplicada a um número finito de dados, permite solucionar classes semelhantes de problemas
Multimídia	Imagens	Representação da forma ou do aspecto de ser ou objeto por meios artísticos
	Vídeo	Técnica de reprodução eletrônica de imagens em movimento
	Áudio	Sinal sonoro; som
	Animações	Ato ou efeito de imprimir movimento ou aceleração
	Filme	Seqüência de imagens registradas em filme cinematográfico ou videoteipe, para exibição em movimento ou não;
	Fotografia	Imagem obtida por arte ou processo de reprodução sobre uma superfície fotossensível (como um filme), pela ação de energia radiante, esp. a luz
Software	Bases de dados	Conjunto de dados inter-relacionados sobre determinado assunto, armazenados em sistemas de processamento de dados segundo critérios preestabelecidos (reúne)
	Simulações	Teste, experiência ou ensaio em que se empregam modelos para simular o ser humano, em especial em casos de grande perigo de vida
	Códigos nucleares	Programa de computador que representam as simulações matemáticas do núcleo do reator.

Visualização	Tabelas	Quadro sistemático de consulta de dados
	Gráficos	Curva num sistema de coordenadas, que representa uma função [A curva pode ser substituída por uma superfície, uma série de colunas etc.]
	Diagramas	Representação gráfica, por meio de figuras geométricas (pontos, linhas, áreas etc.), de fatos, fenômenos, grandezas, ou das relações entre eles
	Modelos em 3D	Modelo em formato tridimensional, que inclui a idéia de profundidade
	Modelos reduzidos	Esquema que possibilita a representação de um fenômeno ou conjunto de fenômenos físicos e eventualmente a previsão de novos fenômenos ou propriedades, tomando como base um certo número de leis físicas, em geral obtidas ou testadas experimentalmente
	Desenhos	Representação de seres, objetos, idéias, sensações, feita sobre uma superfície, por meios gráficos, com instrumentos apropriados
Textuais	Metadados	Dados que registram e preservam dados
	Questionários	Sequência de perguntas feitas para servir de guia a uma investigação
	Entrevistas	Coleta de declarações tomadas para divulgação
	Anotações	Indicação escrita breve
	Normas	Aquilo que regula procedimentos ou atos;
	Padrões	Base de comparação, algo que o consenso geral ou um determinado órgão oficial consagrou como um modelo aprovado. objeto que serve de modelo para outro
	Certificados	Documento no qual se atesta a existência de certo fato e dele se dá ciência
	Caderno de laboratório	Ferramenta usada por pesquisadores de várias áreas para fazer anotações sobre a pesquisa quando executada em laboratórios.
	Transcrição	Passar para o papel ou equivalente (algo) que está sendo ouvido (p.ex., um texto de discurso, uma música etc.)
	Correspondências	Intercâmbio de mensagens, cartas etc. entre pessoas, promovido através de serviço próprio
Artefatos	Diário	Escrito em que se registram os acontecimentos de cada dia
	Caderno de campo	Ferramenta usada por pesquisadores de várias áreas para fazer anotações quando executam trabalhos de campo. É um exemplo clássico de fonte primária.
	Espécimes	Exemplo, amostra, modelo
Artefatos	Amostras	Pequena porção de alguma coisa dada para ver, provar ou analisar, a fim de que a qualidade do todo possa ser avaliada ou julgada
	Maquete	Representação em escala reduzida de uma obra de arquitetura ou engenharia a ser executada. Cenário em miniatura destinado a filmagens de estúdio, quando a obtenção de certas imagens, em ambientes ou paisagens reais, se torna muito difícil ou impraticável; reprodução em miniatura de edifícios, meios de transporte, paisagens etc., us. na simulação de peripécias impossíveis de filmar (p.ex., cenas de catástrofes)
Processos	Procedimentos operacionais padronizados	Procedimento que busca fazer com que um processo, independente da área, possa ser realizado sempre de uma mesma forma, permitindo a verificação de cada uma de suas etapas. Ele deve ser escrito de forma detalhada para a obtenção de uniformidade de uma rotina operacional, seja ela na produção ou na prestação de serviços.

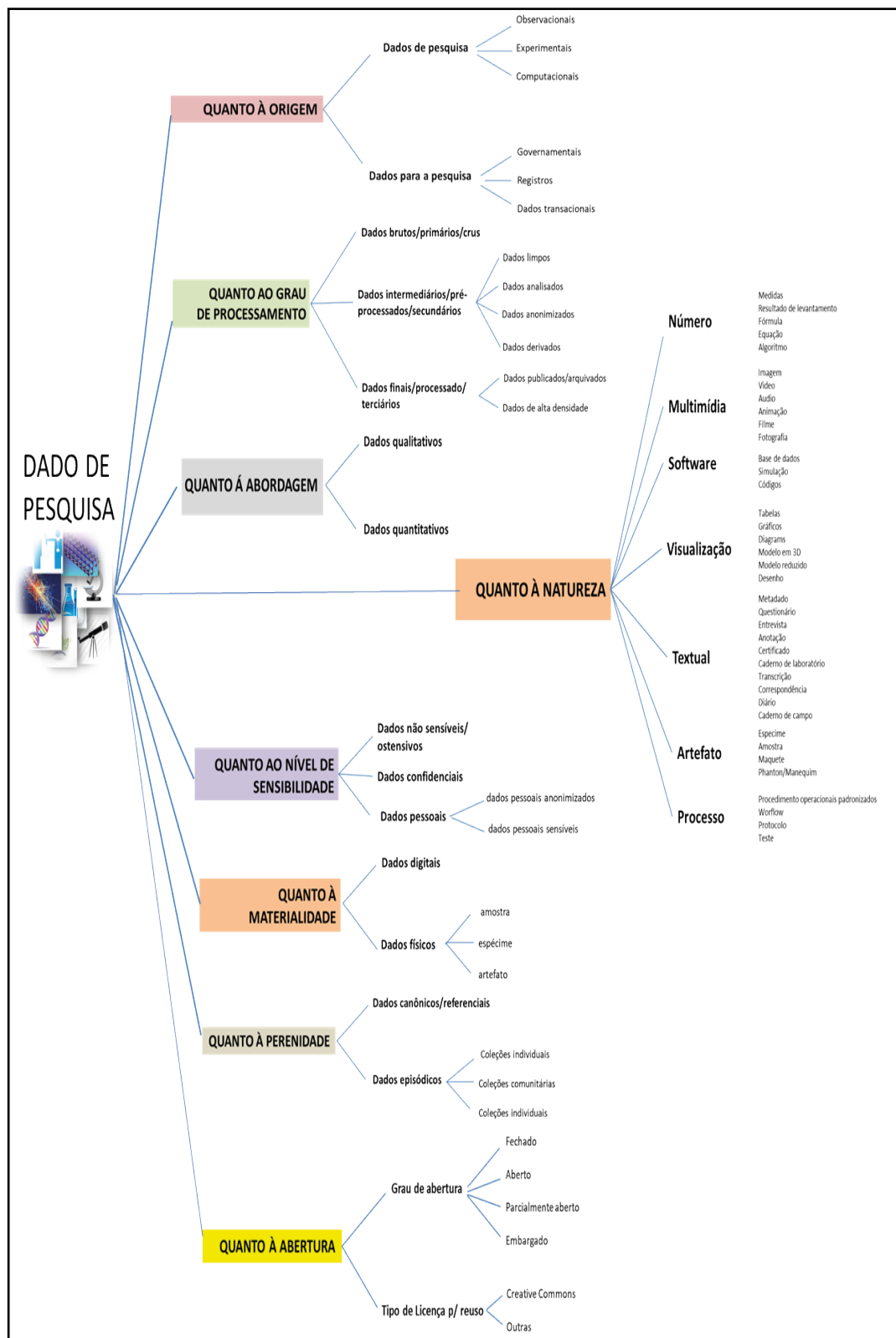
	Workflows	Sequência de passos necessários para que se possa atingir a automação de processos de negócio, de acordo com um conjunto de regras definidas, envolvendo a noção de processos, permitindo que estes possam ser transmitidos de uma pessoa para outra de acordo com algumas regras.
	Protocolos	Planejamento que visa responder uma pergunta ou problema em evidência, definindo a estrutura da pesquisa, selecionando o tipo e o número de variáveis a serem estudadas, e analisando os resultados encontrados
	Teste	Exame crítico ou prova das qualidades de uma pessoa ou coisa
Outros	Phanton ou Manequim	UP Simulador de Tecido Material que possui as mesmas características que o tecido humano com relação à absorção e espalhamento da radiação ionizante.

Fonte: Os autores.

Posteriormente, após o estudo conceitual das definições dos tipos de dados de pesquisa, chegou-se a uma proposta de sistematização que tentou incluir todas as visões encontradas na literatura estudada e ainda acrescentando os tipos indicados pelos pesquisadores durante as entrevistas. A sistematização realizada se pautou no método analítico-sintético de Ranganathan (1967), tentando, sempre que possível se valer de seus cânones e princípios para ordenação dos conceitos.

A sistematização parecia pronta, mas ao propô-la, refletindo sobre a orientação do instrumento para a gestão e curadoria, percebeu-se a necessidade de incluir outras abordagens que atendessem a outras formas de olhar os dados de pesquisa. Neste momento, foram incluídos conceitos classificados nas categorias “quanto à materialidade”, quanto a perenidade”, quanto ao nível de sensibilidade” e “quanto à abertura”. Assim, chegou-se a sistematização representada na Figura 1.

Figura 1 - Taxonomia de dados de pesquisa.



Fonte: Os autores.

A taxonomia adotada categorizou os tipos de dados de pesquisa de acordo com oito características de divisão (quanto à origem, quanto ao grau de processamento, quanto à natureza, quanto à abordagem da pesquisa, quanto ao nível de sensibilidade, quanto à materialidade, quanto à perenidade e quanto à abertura). Essas características adotadas revelam abordagens que precisam ser consideradas na curadoria de dados de pesquisa. Assim, acredita-se que essa taxonomia possa ser útil na modelagem mais precisa de plataformas dedicados à curadoria de dados pesquisas e seus múltiplos propósitos.

- **Quanto à origem** – este princípio de divisão é, provavelmente, o mais relevante para o tratamento dos dados, pois estabelece a profundidade e a abrangência da gestão e curadoria que se deve aplicar para cada categoria. Isto porque a diferença de criticidade dada pela possibilidade – ou impossibilidade - de replicação desses dados implicarão em tratamentos distintos. Os dados observacionais são originados de observações de fenômenos e eventos únicos, como a erupção de um vulcão num dado momento, que não se repetirão jamais. 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. Por outro lado, os dados experimentais são provenientes de situações controladas em bancadas de laboratórios. 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. Os dados computacionais - resultados da execução de modelos computacionais ou de simulações - embora mais fáceis de serem reproduzidos, 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.

Essas categorias de dados são geradas ou coletadas em ambientes ou empreendimentos científicos como, laboratórios, observatórios, expedições, ambientes virtuais, etc. São dados de pesquisa propriamente. Entretanto os dados governamentais, registros médicos e arquivísticos e dados provenientes das redes sociais, por exemplo, são dados coletados para outros propósitos que são frequentemente utilizados para a pesquisa científica. São dados usados para a pesquisa.

- **Quanto ao grau de processamento** – raramente os dados são utilizados diretamente quando são coletados, por exemplo, quando saem de um instrumento. Eles precisam de graus distintos de processamento que refletem os fluxos de pesquisa de cada área. Este conhecimento é fundamental para a precisa gestão e curadoria dos dados, e permite fazer a ligação entre as cadeias de dados que se transformam ao longo da pesquisa, facilitando assim a reprodutibilidade dos experimentos.

- **Quanto à abordagem da pesquisa**, sua relevância está alinhada com o tratamento distinto que os dados precisam receber de acordo com a metodologia adotada no desenvolvimento da pesquisa. Uma parcela relevante dos dados é proveniente de estudos qualitativos originados nas áreas de ciências sociais, humanidades, arte e cultura, e mesmo das áreas das ciências exatas que exigem infraestruturas tecnológicas e práticas distintas dos dados qualitativos.

- **Quanto à natureza dos dados** – retrata a grande diversidade e heterogeneidade de tipos de dados que podem ser originados no am-

biente de pesquisa em termos de formatos, mídias, suportes, expressões, arcabouço tecnológico, etc.

- **Quanto ao nível de sensibilidade** - sua importância está na possibilidade de definir os níveis de abertura e compartilhamento durante o processo de gestão e curadoria, e a profundidade de intervenção que deve ser imposta às coleções, especialmente no caso das pesquisas que envolvem pessoas e informações pessoais sensíveis e informações confidenciais. Isto porque, nem todos os dados devem ser compartilhados sem tratamentos adequados, por exemplo, a submissão dos dados a processos de anonimização e armazenamento em sistemas off-line. Existem dados sensíveis, de grande utilidade para a ciência, que só poderão ser acessados por meios de condições controladas (SAYÃO, SALES, 2015).

- **Quanto à materialidade** - permite vislumbrar a necessidade de tratamento especial aos dados de acordo com a forma que se materializam – digitais ou físicos. O tratamento fornecido aos dados nascidos digitais, ou que passaram por processos de digitalização, tem características específicas determinado pela fragilidade intrínseca dos objetos digitais e pela obsolescência tecnológica que ameaçam a sua capacidade de renderização, de autenticidade e de integridade. Por sua vez, as amostras físicas representam uma ampla variedade de dados de pesquisa. Elas são elementos básicos para referência, estudos e experimentação na pesquisa científica. Testes e análises são realizados diretamente baseados em amostras, tais como espécimes biológica, rochas e minerais, solos, sedimentos, plantas e sementes, artefatos arqueológicos, amostras de tecidos humanos e DNA; outros objetos físicos como mapas e fotografias impressas são objetos diretos de estudos (RDA, 2018). Cada uma desses tipos de dados necessitam de processos de curadoria e cuidados completamente dife-

rentes, por exemplo, um herbário necessita de adubos, irrigação e níveis de temperaturas, humidade e insolação controlados. Porém, a sua forma de representação para fins de recuperação e continua se valendo de metadados. Neste caso, em bases de dados referenciais.

- **Quanto a perenidade** – as coleções de dados são gerenciadas e preservadas de acordo com a sua relevância para a ciência. No mais alto grau estão as coleções internacionais que são referências importantes que precisam de estruturas estáveis que as mantenham para sempre, como o GenBank³ e o Protein Data Bank⁴; em um nível intermediário estão as coleções nacionais e regionais; por fim, as coleções institucionais, comunitárias e individuais que têm valores mais transitório que, porém, podem migrar para níveis mais perenes .

³ GenBank: <https://www.ncbi.nlm.nih.gov/genbank/>.

⁴ Protein Data Bank: <http://www.rcsb.org/>.

- **Quanto à abertura** – às coleções de dados são atribuídos pelos criadores e pelos sistemas de gestão diferentes níveis de abertura que estão relacionados com o nível de sensibilidade dos dados, propriedade intelectual, interesses comerciais, patentes, segurança nacional, interesse pessoal do pesquisador, etc. Às coleções de dados também são atribuídas licenças que determinam o grau de reuso que terceiros podem fazer com essas coleções.

Caracterizar os dados quanto a essas categorias é importante pois permite identificar metadados peculiares ao tratamento de cada tipo de dado. Temos defendido uma abordagem de tratamento dos dados em que a gestão vem antes da tecnologia e em que o tratamento do dado é anterior à sua abertura ou compartilhamento. Faz

parte do processo de gestão, mapear os tipos de dados que serão curados, bem como os padrões de metadados e os arcabouços tecnológicos necessários para a curadoria adequada e a recuperação precisa.

4 À guisa de conclusão

De acordo com duas premissas fundamentais apresentadas por Souza (2012, p.4), no que diz respeito às classificações “os mesmos objetos e ideias podem ser organizados e representados de formas diferentes e toda classificação está relacionada a um propósito definido de construção e uso de informação”. Esta afirmação se harmoniza com as reflexões de Borgman (2010) sobre a gênese dos dados de pesquisa. A autora ressalta que pesquisadores coletam dados para diversos fins, usando vários métodos, podendo tanto a finalidade quanto os métodos influenciar no que consideram como “dados”, e em que condições estes pesquisadores estão dispostos a compartilhar seus dados com os pares. Isso significa dizer que o conceito de dados de pesquisa pode variar não apenas de acordo com o domínio disciplinar, mas também de acordo com o propósito e até com a metodologia empregada na pesquisa.

Dados de pesquisa podem e devem ser visto sob diversas perspectivas. Essas diversas perspectivas precisam estar claras para gestores e desenvolvedores de sistemas e políticas de gestão de dados de pesquisa. Assim, a taxonomia aqui apresentada pode auxiliar esses atores no estudo, mapeamento e caracterização dos dados de pesquisa proporcionando uma modelagem de sistema de dados mais eficaz e uma definição precisa dos dados no âmbito da construção de políticas de gestão de dados.

Todas as classificações refletem o propósito para o qual foram construídas e um contexto específico. A presente taxonomia foi proposta para esboçar um retrato dos dados de pesquisa com nitidez e granularidade suficiente para assistir o planejamento de plataformas e de serviços voltados para a gestão e curadoria de dados. Isto porque as infraestruturas de dados só se tornam efetivas se considerarem as idiossincrasias de cada área em relação aos seus fluxos de geração/coleta de dados e de compartilhamento.

Ainda assim, mesmo considerando a ampla multidisciplinaridade das ciências e tecnologias nucleares, que geram, como consequência, um largo espectro de dados de pesquisa, é provável que a taxonomia aqui apresentada ainda deixe de fora algum tipo de dado, principalmente porque o levantamento empírico realizado abrangeu apenas este domínio específico. Neste caso, seria importante em algum momento, realizar pesquisa semelhante em outros domínios para validar a generalidade da presente taxonomia.

Por fim, o presente trabalho pode ser encarado também como uma tentativa de ajudar pesquisadores na classificação dos dados gerados por suas pesquisas, bem como na gestão e curadoria desses dados por bibliotecários e demais atores envolvidos no processo de gestão.

Referências:

BERLIN. **Declaration on open access to knowledge in the sciences and humanities**. Berlin, 2003. Disponível em: <http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf>. Acesso em: 20 dez. 2011.

BORGMAN, C. L. Research data: who will share what, with whom, when and why. **Rratswd working paper**, v.161, n.10, 2010. Disponível em: http://sydney.edu.au/research/data_policy/resources/ands_borgman_2010_research_data.pdf. Acesso em: 19 maio 2013.

BUCKLAND, Michael K. Information as thing. **Journal of the american society for information science**, v.42, n.5, p.351-360, 1991.

DAHLBERG, I A referent-oriented analytical concept theory of interconcept. **International classification**, Frankfurt, v. 5, n.3, p.142-150, 1978.

D-lib magazine. 2011. Disponível em: <http://www.dlib.org/dlib/january11/01contents.html>. Acesso em: 02 fev. 2019.

D-lib magazine 2014. Disponível em: <http://www.dlib.org/dlib/january14/01editorial.html>. Acesso em: 02 fev. 2019

HARVEY, Douglas Ross. **Digital curation: a how-to-do-it manual**. London: Facet, 2010.

LYON, Liz. Dealing with data; role, rights, responsibilities and relationships. **Consultancy report**, p.1-65, jan. 2007. Disponível em: <http://opus.bath.ac.uk/412/1/dealing_with_data_report-final.pdf>. Acesso em: 19 maio 2013.

NATIONAL RESEARCH COUNCIL – NRC. **A question of balance: private rights and the public interest in scientific and technical databases**. Washington, dc: National Academy Press. 1999. Disponível em: <http://www.nap.edu>. Acesso em: 19 maio 2013.

NATIONAL SCIENCE FOUNDATION – NSF (2007). **Cyberinfrastructure vision for 21st century discovery**.

Disponível em: http://escience.caltech.edu/workshop/ci_vision_march07.pdf. Acesso em: 19 maio 2013.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO – OCDE. Principles and guidelines for access to research data from public data. 2007. Disponível em: <http://www.oecd.org/dataoecd/9/61/38500813.pdf>. Acesso em: 17 fev. 2012.

RDA. **IG physical samples and collections research data ecosystem**. RDA eleventh Plenary Meeting. Berlin: Germany, 2018. Disponível em: <https://rd-alliance.org/ig-physical-samples-and-collections-research-data-ecosystem-rda-11th-plenary-meeting>. Acesso em: 14 jun. 2019.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de Gestão de Dados para Bibliotecários e Pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. 90p. Disponível em: <https://bit.ly/30SFzK5>. Acesso em: 01 jun. 2019.

SOUZA, Rosali Fernandez de. Universo de ciência e tecnologia: organização e representação em classificações do conhecimento. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 13, 2012, Rio de Janeiro. Anais [...]. Rio de Janeiro, 2012. Disponível em: <http://www.eventosecongressos.com.br/metodo/enancib2012/arearestrita/pdfs/19371.pdf>. Acesso em: 20 maio 2013.

WÜSTER, E. L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses. In: RONDEAU, G.; FELBER, E. (org.). **Textes choisis de terminologie**. Québec: Girserm, 1981, p. 57-114. (fondements théoriques de la terminologie, v. I).