

ORGANIZAÇÃO DOS DADOS DE PESQUISA EM CIÊNCIAS SOCIAIS E HUMANAS: UMA REFLEXÃO SOBRE ASSUNTOS E TIPOS DE CONTEÚDO NO *REGISTRY OF RESEARCH DATA REPOSITORIES*

ISABELLA HENRIQUE LIMA PEREIRA*

SUELLEN OLIVEIRA MILANI**

Resumo: Sistemas de organização do conhecimento desempenham um papel central na estruturação e recuperação da informação especialmente em ambientes digitais onde a diversidade de dados e metadados exigem abordagens complexas e inclusivas. Nesse contexto, o objetivo deste trabalho é analisar o re3data enquanto um sistema de organização do conhecimento em pesquisa, considerando os aspectos das Ciências Sociais e Humanas, a partir da análise das categorias do filtro «content type». Para tanto, foi realizada uma análise exploratória do re3data, na qual realizou-se um levantamento de dados utilizando os filtros «subject» e «content type». Os resultados demonstram que o re3data se configura como um instrumento global de organização, adotando uma postura generalista e universal em sua categorização de tipologias de dados. Conclui-se que os sistemas de organização do conhecimento são instrumentos que possuem dimensões epistemológicas e culturais e que não apenas estruturam dados, mas também moldam a forma como o conhecimento é produzido e acessado.

Palavras-chave: Representação do conhecimento; Dados de Pesquisa; Repositório de dados.

Abstract: Knowledge organization systems play a central role in the structuring and retrieval of information, especially in digital environments, where the diversity of data and metadata demands complex and inclusive approaches. In this context, the aim of this study is to analyze re3data as a knowledge organization system for research, considering the specificities of the Social Sciences and Humanities, based on the analysis of the categories under the «content type» filter. To this end, an exploratory analysis of re3data was conducted, in which data were collected using the «subject» and «content type» filters. The results show that re3data functions as a global organization tool, adopting a generalist and universal approach in its categorization of data typologies. It is concluded that knowledge organization systems are instruments that encompass epistemological and cultural dimensions, and that they not only structure data but also shape the ways in which knowledge is produced and accessed.

Keywords: knowledge organization; Research Data; Data Repositories.

* Universidade Federal Fluminense (UFF) – Brasil. Email: isabellahlpereira@gmail.com. ORCID: <https://orcid.org/0000-0002-8463-0629>.

** Universidade Federal Fluminense (UFF) – Brasil. Email: suellemilani@id.uff.br. ORCID: <https://orcid.org/0000-0001-7183-6030>.

INTRODUÇÃO

A representação do conhecimento em ambientes informacionais reflete escolhas e perspectivas inerentes aos contextos nos quais seus sistemas de organização do conhecimento foram concebidos. Nesse estudo, analisaremos o ambiente informacional denominado *Registry of Research Data Repositories (re3data)* que é um registro global de repositórios de dados de pesquisa, que funciona como uma ferramenta de descoberta, indexando e descrevendo repositórios que armazenam dados de pesquisa de todas as áreas do conhecimento.

Parte-se da premissa de que os aspectos sociais, que refletem a realidade de um tempo e espaço, podem ser percebidos nas interpretações conceituais, inclusive sobre o que são dados, nas políticas de coleta e organização. Com o intuito de refletir sobre as categorias que podem ser usadas para classificar documentos das Ciências Sociais e Humanas (CSH), pergunta-se: para representar conjuntos de dados de pesquisa referentes às áreas das CSH é necessário um conjunto específico de metadados?

Essa pesquisa parte de uma carência de literatura sobre a representação do conhecimento no âmbito de repositórios de dados de pesquisa das Ciências Sociais e Humanas para propor uma reflexão acerca das especificidades deste domínio do conhecimento e a perspectiva cultural inerente aos sistemas de organização do conhecimento utilizados para representá-los.

1. ORGANIZAÇÃO DO CONHECIMENTO, REPOSITÓRIOS DE DADOS E A PERSPECTIVA CULTURAL NAS CIÊNCIAS SOCIAIS E HUMANAS

Os sistemas de organização do conhecimento desempenham um papel central na área de organização do conhecimento. Desde padrões de metadados, ontologias e taxonomias até tesouros e sistemas de classificação, os sistemas de organização do conhecimento são utilizados para promover a representação e a recuperação da informação, bem como para viabilizar a interoperabilidade, a preservação do conhecimento e o acesso. Em ambientes digitais, possibilitam que dados e informações sejam sistematizados de forma a garantir o acesso eficiente e completo. No entanto, ao organizarem e classificarem o conhecimento, a eficiência e a completude do acesso dependem de questões que vão além da simples estruturação da organização, muitas vezes esses sistemas refletem escolhas epistemológicas, culturais e éticas que impactam diretamente sua neutralidade e eficácia (Guimarães 2019).

Os processos de organização do conhecimento não são neutros e livres de interferências, pois os critérios utilizados muitas vezes refletem as perspectivas culturais, sociais e políticas daqueles que os projetam (Hjørland 2002). No ambiente digital, essas questões tornam-se ainda mais evidentes devido à complexidade e à diversidade de dados e metadados que são organizados por sistemas globais, como o *re3data*.

Segundo Fernández-Molina e Guimarães (2002), os aspectos éticos no contexto da organização do conhecimento ganham destaque, sobretudo no que tange à inclusão, exclusão e priorização de determinadas informações. Os autores apontam que a definição e a representação do conhecimento em ambientes digitais devem considerar o impacto dessas escolhas sobre os usuários, especialmente em contextos culturais diversos.

Guimarães (2019) avança nesse debate ao propor a perspectiva da «organização enviesada do conhecimento», que questiona a ideia de neutralidade na representação e sugere que os sistemas de organização do conhecimento devem ser projetados de forma a reconhecer explicitamente seus vieses culturais e epistemológicos. Essa abordagem é particularmente relevante para as Ciências Sociais e Humanas, uma vez que as formas de coleta e organização dos dados estão profundamente imersas nos contextos culturais e sociais dos pesquisadores e das comunidades analisadas.

A literatura destaca que a concepção de sistemas como o *re3data* deve estar alinhada a princípios éticos que assegurem uma representação justa e inclusiva. Em síntese, os sistemas de organização do conhecimento não apenas estruturam dados e informações, mas também desempenham um papel epistemológico e cultural significativo, uma vez que contribuem ou não para a comunicação e disponibilização de dados de pesquisa.

No contexto da pesquisa, especialmente nas Ciências Sociais e Humanas, a análise desses sistemas deve considerar não apenas sua funcionalidade técnica, mas também os valores e pressupostos éticos que os sustentam, visto que o compartilhamento de dados ainda é uma questão em desenvolvimento na área, que ainda não se sente totalmente representada (Mozersky et al. 2020).

1.1. **Repositórios de dados de pesquisa: o papel do *re3data***

O *Registry of Research Data Repositories* (*re3data*) surgiu em 2012, como um projeto que, a partir de diversas parcerias entre universidades e instituições, passou em 2020 a integrar o que conhecemos hoje como o repositório (Re3data.org 2024b). Ele é um ‘repositório’¹ que funciona como um registro global de repositórios de dados de pesquisa. Esse portal funciona como uma ferramenta de busca que reúne em um só local repositórios de dados de todas as disciplinas, visando promover a cultura do compartilhamento, aumentar o acesso e promover a comunicação dos dados através do aumento da visibilidade dos diferentes tipos de dados e seus repositórios.

O *re3data* tem como objetivo facilitar a utilização e o encontro de dados por pesquisadores, agências de fomento, instituições de pesquisa e ensino e editores de

¹ Apesar de ser entendido e reconhecido como um repositório, o *re3data* não o é, é um diretório de pesquisa que reúne listas de repositórios de dados.

revistas, desempenhando um papel fundamental na promoção do acesso e da organização do conhecimento científico em múltiplas disciplinas. Sua estrutura baseia-se em princípios de interoperabilidade e padronização, sendo amplamente utilizado para localizar repositórios confiáveis e bem documentados.

O *re3data* atua como um sistema que descreve repositórios por meio de metadados padronizados. Esses metadados incluem informações como tipo de dados armazenados, escopo disciplinar, políticas de acesso, linguagem utilizada, conformidade com princípios FAIR (*Findable, Accessible, Interoperable, and Reusable*), etc.

Essa tentativa de padronização facilita a integração entre os repositórios e a recuperação eficiente de informações, mas também levanta questões sobre como a ferramenta pode funcionar num contexto global a partir da utilização de categorias e descriptores gerais para abranger diferentes contextos culturais e disciplinares.

A relevância do *re3data* para as Ciências Sociais e Humanas está diretamente ligada à diversidade de dados e métodos empregados nessas áreas. Dados em Ciências Sociais e Humanas são frequentemente qualitativos, contextuais e culturalmente situados, o que pode dificultar sua categorização e organização em repositórios amplos e multidisciplinares. Nesse sentido, os critérios utilizados pelo *re3data* para descrever e categorizar repositórios demandam uma análise crítica, considerando os potenciais vieses culturais que podem estar implícitos em seus esquemas de metadados.

Para atuar como um sistema efetivo de organização do conhecimento, o *re3data* precisa não apenas garantir a acessibilidade dos repositórios, mas também assegurar que as descrições sejam culturalmente apropriadas. Portanto, a análise do *re3data* no contexto das Ciências Sociais e Humanas não se limita a avaliar sua funcionalidade técnica.

É necessário explorar como o sistema reflete, ou não, as especificidades culturais e disciplinares dessas áreas, reconhecendo a importância de uma abordagem ética e inclusiva na organização do conhecimento. Esse exame é particularmente pertinente para compreender como sistemas de organização do conhecimento globais podem, simultaneamente, facilitar o acesso a dados e não perpetuar visões limitadas ou enviesadas de determinadas disciplinas e culturas.

1.2. A Perspectiva cultural nas Ciências Sociais e Humanas

Nas Ciências Sociais e Humanas, a cultura não é apenas objeto de estudo, mas também um elemento central na forma como o conhecimento é produzido, organizado e interpretado. Os dados gerados nessas áreas refletem valores, tradições e contextos locais que muitas vezes desafiam modelos universais de categorização. Esse cenário evidencia a necessidade de uma abordagem culturalmente apropriada na organização do conhecimento, que respeite as especificidades locais sem ignorar o potencial para a comunicação global (Beghtol 2003).

Segundo Beghtol (2003), os sistemas de organização do conhecimento enfrentam o desafio de equilibrar conceitos universais, ou seja, o reconhecimento de que os sistemas devem atender às necessidades e perspectivas das culturas específicas em que estão inseridos. Nesse contexto, Beghtol (2003) introduz o conceito de hospitalidade cultural, que enfatiza a importância de criar sistemas que não apenas respeitem as especificidades culturais, mas também promovam o diálogo transcultural. Essa abordagem é especialmente relevante para as Ciências Sociais e Humanas, cujos dados frequentemente transcendem categorias rígidas e demandam flexibilidade na sua representação.

A autora explica que garantia cultural significa que «qualquer tipo de sistema de organização e/ou representação do conhecimento pode ser sumamente apropriado e útil para os indivíduos de uma cultura somente se for baseado nas premissas, valores e predisposições daquela mesma cultura» (Beghtol 2002, p. 511). Por essa razão, os sistemas de organização do conhecimento são considerados artefatos culturais.

Olson (2002) argumenta que os sistemas de organização do conhecimento são inevitavelmente influenciados pelos valores e pressupostos culturais de seus criadores e mantenedores. No contexto das Ciências Sociais e Humanas, essa influência é ainda mais evidente, pois os dados frequentemente emergem de interações sociais e culturais específicas, que podem ser difíceis de padronizar ou categorizar sem perder nuances importantes.

Adicionalmente, García Gutiérrez (2002) contribui para o debate ao propor uma «ética transcultural de mediação» baseada na ideia de uma «cultura de fronteira». Para o autor, a organização do conhecimento deve reconhecer e integrar a diversidade cultural, funcionando como um espaço de mediação entre diferentes perspectivas e sistemas de valores.

No caso das Ciências Sociais e Humanas, essa mediação é essencial para evitar a imposição de categorias universalistas que possam obscurecer ou excluir as especificidades dos dados, os quais, diferentemente de áreas que lidam predominantemente com dados quantitativos e replicáveis, frequentemente são qualitativos, onde muitas vezes o resultado é a própria reflexão realizada a partir dos métodos propostos, situadas em contextos específicos e marcadas por significados culturais únicos. Os dados refletem valores, narrativas e contextos históricos específicos que devem ser pensados em plataformas como o *re3data*, para evitar a homogeneização do conhecimento e a restrição do que entendemos como dado.

Em síntese, é preciso refletir sobre a noção de neutralidade nos sistemas de organização do conhecimento, especialmente em plataformas globais que buscam padronizar a descrição de dados enquanto lidam com repositórios provenientes de diversos contextos culturais e domínios do conhecimento.

Ao refletir sobre os dilemas éticos que perpassam a organização do conhecimento, Milani (2024) estuda a garantia cultural enquanto um valor ético desta área, e explica que

[o] valor ético “garantia cultural”, tão caro às reflexões contemporâneas, possui os seguintes atributos: apresentação de diferentes tipos de informação e construção de diferentes caminhos para uma mesma informação. É importante para todas as comunidades de usuários, principalmente aquelas que não pertencem ao contexto dominante porque diferentes usuários têm diferentes identidades e necessidades informacionais específicas, as quais devem ser respeitadas (Milani 2024, p. 86-87).

Da negação, ou não observação, deste valor ético, tem-se a má-representação, ou incompletude, como afirma a autora, causando problemas na recuperação da informação. As Ciências Sociais e Humanas não pertencem ao contexto dominante das áreas científicas, sendo assim a análise da representação das suas especificidades em sistemas de organização do conhecimento trarão benefícios a todos os interessados nelas.

2. PROCEDIMENTOS METODOLÓGICOS

Para responder à pergunta de partida, foi realizada uma pesquisa descritiva que, a partir de uma abordagem qualitativa, realizou uma navegação e examinou a interface do *re3data* para familiarização com suas funcionalidades e estrutura. Após uma exploração inicial, foi feita uma análise utilizando os filtros «subject» e «content type» para selecionar os repositórios identificados como pertencentes às áreas de CSH e mapear as tipologias associadas a elas. Também foi realizada uma análise das métricas da plataforma para identificar as subcategorias associadas a cada categoria do filtro «content type». Por fim, foram definidos critérios para selecionar os repositórios para análise.

Para inclusão no *corpus* de análise, considerou-se os repositórios pertencentes às CSH e que apresentavam uma descrição detalhada dos tipos de dados armazenados no filtro «content type». Os critérios de exclusão levaram à remoção de repositórios com informações incompletas ou insuficientes sobre as tipologias de dados, garantindo a relevância dos dados analisados.

A análise dos resultados foi dividida em duas etapas, a saber: identificação e estabelecimento das tipologias de dados encontradas; e discussão dos resultados à luz das premissas que embasam os marcos teóricos sobre os aspectos sociais e culturais inerentes aos sistemas de organização do conhecimento.

A pesquisa é uma pesquisa exploratória e descritiva de abordagem qualitativa, e tem como objetivo analisar o *re3data* enquanto um sistema de organização

do conhecimento em pesquisa, considerando os aspectos culturais das Ciências Sociais e Humanas, a partir da análise das categorias presentes no diretório em relação a área.

Para a coleta de dados, foi realizada uma análise exploratória do repositório e decidimos utilizar os filtros disponíveis para a realização do recorte da pesquisa. Em se tratando da coleta de dados, utilizamos os filtros «subject» e «content type» para selecionar os repositórios que foram identificados como pertencentes às áreas de Ciências Sociais e Humanas e identificar as tipologias indicadas como sendo dessas áreas.

Além disso, utilizamos a aba referente às métricas da plataforma, associada ao filtro de conteúdo para identificar que subcategorias estavam associadas a cada categoria do filtro. A Figura 1 abaixo ilustra o processo de coleta das subcategorias.

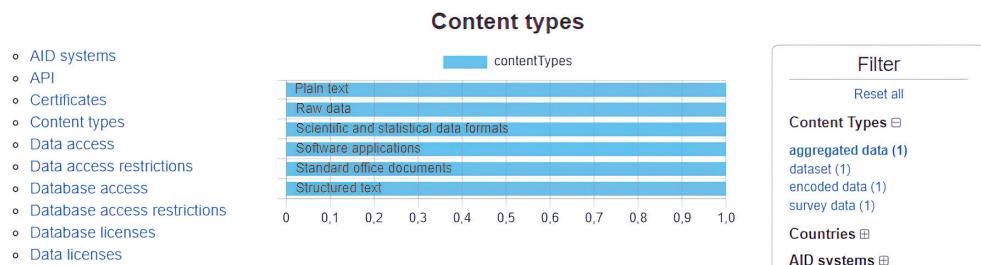


Fig. 1. Captura das subcategorias presentes na categoria «Dados Agregados»

Fonte: Re3data.org 2024c

Para garantir a rigidez da pesquisa foram adotados alguns critérios de inclusão e exclusão de dados. Na inclusão, olhamos somente os repositórios categorizados no *re3data* como pertencentes à área de CSH, e que realizaram a descrição detalhada dos tipos de dados armazenados, de modo que as tipologias estão presentes no filtro referente ao tipo de conteúdo (dados) armazenado pelos repositórios. Na exclusão, levando em consideração a utilização do filtro «content type» que diz respeito à tipologia de dados presente no repositório, optamos por excluir os repositórios com informações incompletas ou insuficientes sobre as tipologias de dados.

Na análise dos resultados, o trabalho foi dividido em duas etapas. Na primeira etapa, realizamos uma análise, visando identificar e estabelecer as tipologias de dados encontradas, identificando os padrões e as categorias utilizadas nas Ciências Sociais e Humanas. Na segunda etapa, foi realizada uma discussão dos resultados levando em consideração alguns princípios culturais da área. Nessa etapa, utilizamos os estudos de Olson (2002) e Hjørland (2002) para analisar como os repositórios e o *re3data* incorporaram ou deixaram de incorporar, os contextos e especificidades culturais das Ciências Sociais e Humanas, reconhecendo possíveis vieses nos critérios de organização

e descrição dos dados. Com as etapas da análise, pretendeu-se compreender não apenas a estrutura das tipologias no *re3data*, mas também as implicações culturais e epistemológicas subjacentes a ele.

Por fim, ressaltamos que apesar do rigor metodológico, a análise está pautada na consistência das informações disponíveis no repositório em dezembro de 2024 e que a interpretação dos aspectos culturais está sujeita aos vieses decorrentes das próprias definições e fundamentos teóricos adotados nesta pesquisa.

3. LEVANTAMENTO DE DADOS E DISCUSSÃO

Com a utilização do primeiro filtro «subject» foi possível identificar a primeira forma de organização do repositório que abarca quatro grandes áreas do conhecimento, conforme ilustrado na Tabela 1.

Tabela 1. Área do conhecimento presentes no Re3data.org

Áreas	Humanidades e Ciências Sociais	Ciências da Saúde	Ciências da Natureza	Ciência das Engenharias
N.º de Repositórios	1268	1888	1707	744

Fonte: Dados da pesquisa (Re3data.org 2024a)

É possível observar que as áreas de Ciências naturais e da Saúde são as que mais possuem repositórios de dados, isto é, a categoria *Humanities and Social Sciences*² contém um total de 1268 repositórios.

Seguindo a pesquisa, identificamos as tipologias de dados apresentados para cada área, mesmo já tendo definido a área de estudo, visto que queríamos estabelecer uma comparação inicial entre as tipologias presentes em cada área. Optamos por utilizar os dois filtros: o de área e o de conteúdo dos repositórios ao mesmo tempo para elaborar a Tabela 2.

A Tabela 2 foi elaborado como uma tentativa de fornecer uma visão geral de como os dados são classificados em cada área do conhecimento. É possível observar que a maioria das tipologias se repete, o que demonstra um caráter generalista do vocabulário adotado e evidencia que a plataforma priorizou a usabilidade e a aplicabilidade universal, desconsiderando em um primeiro momento as especificidades de cada área do conhecimento.

² No presente artigo iremos entender a área de *Humanities and Social Sciences* como Ciências Sociais e Humanas.

Tabela 2. Conteúdo de dados indicados pelos repositórios

Área	Categoria	Área	Categoria
Ciências Sociais e Humanas	29 Conjunto de dados	Ciências da Saúde	30 Conjunto de dados
	10 Dados de pesquisa		11 Imagem
	8 Texto		11 Texto
	7 Outros		9 Outros
	7 Imagem		6 Dados de pesquisa
	4 Dados observacionais		6 Dados genéticos
	3 Software		5 Dados de mensuração e teste
	3 Dados experimentais		4 Dados experimentais
	3 Dados de mensuração e teste		4 Dados observacionais
	2 Imagem estática		3 Software
	2 Dados geoespaciais		2 Dados de Gravação
	2 Dados de simulação		2 Dados de simulação
	1 Software de pesquisa		2 Imagem estática
	1 Dados codificados		1 Código-fonte
	1 Dados agregados		1 Dados geoespaciais
	1 Código-fonte		1 Software de pesquisa
Ciências da Natureza	18 Conjunto de dados	Ciência das Engenharias	15 Conjunto de dados
	9 Imagem		7 Imagem
	6 Texto		6 Outros
	5 Outros		6 Texto
	4 Dados de pesquisa		4 Dados experimentais
	4 Dados experimentais		3 Dados de pesquisa
	2 Dados de simulação		2 Dados de simulação
	2 Dados geoespaciais		2 Dados observacionais
	2 Dados observacionais		2 Software
	1 Software		1 Código-fonte
	1 Software de pesquisa		1 Dados de mensuração e teste
	1 Dados de Gravação		1 Dados geoespaciais
	1 Dados de mensuração e teste		1 Software de pesquisa
	1 Imagem estática		

Fonte: Dados da pesquisa, 2024

Dados os resultados presentes na Tabela 2, podemos perceber que as categorias de tipos de dados são gerais e levemente direcionadas para a área de pesquisas qualitativas. A área de CSH possui 16 categorias, que explicitamos na Tabela 3, abaixo, visando entender o que cada categoria significa.

Tabela 3. Categorias das CSH

Categoría	Definição
Dados agregados	Estatísticas que se referem a classes, grupos ou categorias amplas. Os dados são calculados como médias, totais ou derivados de dados em nível individual, tornando impossível distinguir as características individuais dentro dessas classes, grupos ou categorias.
Conjunto de dados	Uma coleção de fatos relacionados e codificados em uma estrutura definida.
Dados codificados	Dados qualitativos (textuais, de vídeo, áudio ou imagem fixa) originalmente produzidos para outros propósitos, transformados em dados quantitativos (expressos em matrizes de unidade por variável) usando técnicas de codificação de acordo com esquemas de categorização predefinidos.
Dados experimentais	Dados resultantes do método de pesquisa experimental envolvendo a manipulação de algumas ou todas as variáveis independentes incluídas nas hipóteses.
Dados geoespaciais	Dados de qualquer tipo com coordenadas espaciais que permitem que sejam mapeados na superfície da Terra. Podem representar objetos físicos, áreas discretas ou superfícies contínuas.
Imagen	Uma representação visual que não seja texto, incluindo todos os tipos de imagens em movimento e fixas.
Dados de mensuração e teste	Dados obtidos a partir da avaliação de propriedades específicas (ou características) de seres, coisas, fenômenos (e/ou processos); da aplicação de padrões preestabelecidos e/ou instrumentos especializados.
Dados observacionais	Dados resultantes de pesquisas observacionais, que envolvem coletar observações conforme ocorrem (por exemplo, comportamentos, eventos, desenvolvimento de condições ou doenças, etc.) sem tentar manipular qualquer variável independente.
Outros	Um tipo de recurso que não está incluído em termos existentes.
Software de pesquisa	Software usado para gerar, processar ou analisar resultados que você pretende incluir em uma publicação (seja um artigo de revista, apresentação de conferência, monografia, livro ou tese). O software de pesquisa pode variar de algumas linhas de código escritas por você a um pacote de software profissionalmente desenvolvido.
Dados de simulação	Dados resultantes da modelagem ou representação imitativa de processos, eventos ou sistemas do mundo real, geralmente usando programas de computador.
Software	Um programa de computador em código-fonte (texto) ou formato compilado.
Código-fonte	Qualquer coleção de código, com ou sem comentários, escrita em uma linguagem de programação legível por humanos, geralmente como texto simples.
Imagen estática	Uma representação visual registrada estática. Essa classe de imagens inclui diagramas, desenhos, gráficos, gráficos de <i>design</i> , plantas, fotografias, impressos e <i>prints</i> .
Dados de pesquisa	Dados resultantes de uma pesquisa, definida como uma investigação sobre as características de uma população, por meio da coleta de dados de uma amostra dessa população e da estimativa de suas características por metodologia estatística sistemática.
Texto	Um recurso composto principalmente por palavras para leitura. Exemplos incluem livros, cartas, dissertações, poemas, jornais, artigos e arquivos de listas de correspondência. Observe que <i>fac-símiles</i> ou imagens de textos ainda são do gênero «Texto».

Fonte: Strecker et al. 2023, pp. 21-22³

³ As informações presentes no quadro foram retiradas do texto *Metadata Schema for the Description of Research Data Repositories*, que afirma que o filtro tipo de conteúdo é baseado nos *Vocabulários Controlados COAR para Repositórios*.

Além das categorias, o repositório possui 15 subcategorias, a saber: formatos de dados científicos e estatísticos; documentos padrão de escritório; imagens; dados brutos; texto simples; outros; gráficos estruturados; texto estruturado; dados arquivados; bancos de dados; dados audiovisuais; aplicações de *software*; código fonte; dados baseados em rede; e dados de configuração⁴.

Essas subcategorias podem se associar a cada uma das categorias principais de acordo com a tipologia de dados dos repositórios. Elas, apesar de não possuírem uma definição estruturada no repositório, detalham os formatos e tipos específicos de conteúdo, adicionando um olhar mais prático e técnico às categorias gerais.

4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

A partir da análise do filtro «content type», podemos entender o mesmo como um sistema de organização do conhecimento, que baseado na estrutura de um vocabulário controlado sistematiza o tipo de dados presente nos repositórios, com o objetivo de classificar os repositórios ali preservados, estruturando uma forma de recuperação da informação.

A forma de organização do sistema, apesar de buscar a generalização, como apontado pela própria plataforma (Strecker et al. 2023), acaba refletindo uma abordagem amplamente influenciada pelas ciências exatas e naturais.

No contexto das Ciências Sociais e Humanas, as categorias presentes na Tabela 3 apresentam uma base útil, mas nem sempre abrangem completamente as especificidades da área. Dados qualitativos, como entrevistas e narrativas, são centrais nas Ciências Sociais e Humanas e frequentemente apresentam características híbridas que não se encaixam facilmente em definições rígidas. Um bom exemplo são as subcategorias: formatos de dados científicos e estatísticos e gráficos estruturados. Essas subcategorias sugerem um foco em formatos que priorizam dados quantitativos, estruturados e tecnicamente formatados.

Essa padronização pode ser vista como uma tentativa de universalizar categorias desenvolvidas para as ciências ‘duras’, impondo uma lógica que nem sempre é adequada às particularidades das Ciências Sociais e Humanas.

Uma outra subcategoria que merece destaque é a de «dados brutos», que nas Ciências Sociais e Humanas, não têm uma definição única e universal, como acontece em áreas quantitativas. Dados brutos podem ser transcrições de entrevistas,

A utilização do COAR reflete a decisão de adotar uma abordagem mais geral e acessível. Esse vocabulário, desenvolvido por um grupo internacional, foi projetado para atender a uma ampla gama de repositórios institucionais e temáticos, contemplando diversas disciplinas acadêmicas (Strecker et al. 2023).

⁴ No original: «scientific and statistical data formats»; «standard office documents»; «images»; «raw data»; «plain text»; «other»; «structured graphics»; «structured text»; «archived data»; «databases»; «audiovisual data»; «software applications»; «source code»; «network-based data»; «configuration data». Essas subcategorias foram retiradas da página métricas do *re3data* conforme ilustrado na Figura 1 (Re3data.org 2024c).

registros etnográficos, manuscritos históricos, vídeos não editados ou documentos originais coletados diretamente no campo de pesquisa. Esses dados costumam ser interpretados e analisados no contexto sociocultural em que foram gerados, o que desafia a ideia de ‘bruto’ como algo objetivo e neutro.

Além disso, questões culturais e fruto de interpretações, que são fundamentais nas Ciências Sociais e Humanas, podem não ser classificadas nas categorias «dados codificados» ou «dados observacionais», especialmente se dependem de interpretações.

Nas Ciências Sociais e Humanas, os dados frequentemente incluem *nuances* contextuais, culturais e interpretativas que não se enquadram facilmente em sistemas projetados para dados numéricos ou experimentais, de modo que as categorias e subcategorias da plataforma apesar de bem pensadas, podem ser aprimoradas.

Assim, embora essas categorias forneçam uma estrutura inicial para organização e classificação, sua adequação total às Ciências Sociais e Humanas requer uma análise crítica e, possivelmente, a criação de subcategorias mais específicas que refletem melhor as particularidades das metodologias e objetos de estudo dessa área, uma vez que as subcategorias existentes na plataforma são gerais e não possuem uma definição.

Os principais resultados da análise demonstram que o *re3data* se apresenta como um instrumento global de organização, adotando uma postura generalista e universal em sua categorização das tipologias de dados.

Em síntese, a análise do filtro «subject» revelou quatro grandes áreas do conhecimento presentes no *re3data*, sendo «Humanidades e Ciências Sociais» uma delas, com um número significativo de repositórios. A comparação das tipologias de dados entre as diferentes áreas mostrou que a maioria das tipologias se repete, evidenciando o caráter generalista do vocabulário adotado pela plataforma.

A análise detalhada das categorias de tipos de dados para a área de CSH (totalizando 16 categorias) e suas 15 subcategorias revelou que, apesar de fornecerem uma base útil, nem sempre abrangem suas especificidades. As subcategorias, embora tragam um detalhamento técnico dos formatos e tipos de conteúdo, reforçam uma ênfase em dados quantitativos e estruturados, mas comum às Ciências duras.

Embora o *re3data* seja um sistema funcional e valioso para generalizações na organização de repositórios de dados de pesquisa, ele carece de uma precisão epistemológica e terminológica para representar adequadamente as especificidades das Ciências Sociais e Humanas.

A sua estrutura organizacional é influenciada pela lógica das Ciências Exatas e Biológicas, com um foco em formatos padronizados e amplamente reconhecidos, o que pode levar a uma sub-representação das peculiaridades das Ciências Sociais e Humanas.

A ausência de categorias mais específicas para dados interpretativos, históricos, culturais, híbridos ou multimodais reforça a ideia de que o sistema não considera

a especificidade da área. A amplitude das categorias existentes pode gerar confusão entre pesquisadores ao definir seus dados, e a possibilidade de atribuição de múltiplas categorias para um mesmo repositório levanta questões sobre a identificação do tipo de dado predominante.

Assim, é fundamental incorporar princípios que assegurem uma representação apropriada entre as áreas do conhecimento, levando em consideração a diversidade de perspectivas dos usuários e das comunidades científicas. Sistemas de informação como o *re3data* são mais do que meras ferramentas, pois refletem estruturas epistemológicas e, muitas vezes, sociais e culturais que moldam a produção e o acesso ao conhecimento científico. Portanto, é necessário um esforço contínuo de pesquisa para auxiliar no desenvolvimento do sistema como uma ferramenta de recuperação de informação mais equitativa garantindo a recuperação de informações relevantes em todas as áreas do conhecimento.

CONSIDERAÇÕES FINAIS

A análise das categorias e subcategorias presentes no *re3data* evidencia uma estrutura organizacional amplamente influenciada pela lógica das Ciências Exatas e Biológicas. O foco em formatos padronizados e amplamente reconhecidos, reflete uma tentativa de universalizar práticas que, embora eficientes para algumas áreas, nem sempre conversam com a diversidade epistemológica e metodológica de outras áreas como as Ciências Sociais e Humanas.

O *re3data* adota uma lógica generalista e utilitária, que privilegia formatos amplamente usados. Enquanto categorias como texto simples e imagem podem ser usadas nas Ciências Sociais e Humanas, a falta de especificidade para dados interpretativos, históricos ou culturais aponta para uma sub-representação das peculiaridades dessas áreas.

Além disso, a ausência de categorias mais específicas pode influenciar tanto a representação dos dados de pesquisa pelos pesquisadores, quanto a recuperação da informação pelos usuários do *re3data*.

A categorização dos dados nos repositórios não é excludente, permitindo a atribuição de mais de uma categoria de tipologia de dados para os repositórios. Essa política precisa ser explicitada e o impacto dessa decisão de quem está representando o repositório deve ser explicado, pois os repositórios podem trabalhar com mais de um tipo de dados, mas também podem possuir um tipo de dado predominante. Isso deve ser considerado na estruturação do sistema de organização do conhecimento.

O sistema de organização do conhecimento utilizado pelo *re3data*, embora altamente funcional e útil para generalizações, carece de uma precisão epistemológica e terminológica para capturar plenamente as necessidades das Ciências Sociais e Humanas. O sistema pode integrar categorias que representem as especificidades

dos dados nessas áreas, ampliando sua funcionalidade para incluir formatos e descrições mais alinhados a elas.

Em síntese, incorporar definições mais detalhadas e termos que representem as especificidades de diferentes áreas pode não apenas tornar o sistema mais inclusivo, mas também promover uma compreensão mais ampla da diversidade de práticas de pesquisa na Ciência. Assim, uma abordagem que respeite as especificidades culturais é essencial para evitar uma dominação cultural e permitir um uso mais equitativo do sistema. Nesse sentido, as pesquisas precisam ser feitas para ajudar o sistema a se desenvolver como uma ferramenta de recuperação da informação.

REFERÊNCIAS

- ANDERSON, Barbara A., 2015. The Role of Data in Research and Policy. Em: R. A. SCOTT, S. M. KOSSLYN, e M. C. BUCHMANN, ed. *Emerging trends in the social and behavioral sciences* [Em linha]. [S.l.]: John Wiley & Sons [consult. 2025-07-01]. DOI: <https://doi.org/10.1002/9781118900772.etrds0350>.
- BARDIN, Laurence, 2016. *Análise de conteúdo*. São Paulo: Edições 70.
- BEGHTOL, Clare, 2003. Universal concepts, cultural warrant and cultural hospitality. *Bulletin of the American Society for Information Science and Technology*. **29**(2), 19-23.
- BEGHTOL, Clare, 2002. A proposed ethical warrant for global knowledge representation and organization systems. *Journal of Documentation*. **58**(5), 507-532.
- FERNÁNDEZ-MOLINA, Juan Carlos, e José Augusto C. GUIMARÃES, 2002. Ethical aspects of knowledge organization and representation in the digital environment: their articulation in professional codes of ethics. Em: Maria José LOPEZ-HUERTAS, ed. *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries*. Wurzburg: Ergon Verlag, pp. 487-492.
- GARCÍA GUTIÉRREZ, Antonio, 2002. Knowledge organization from a “culture of the border” towards a transcultural ethics of mediation. *Knowledge Organization*. **29**(2), 79-92.
- GUIMARÃES, José Augusto C., 2019. Slanted knowledge organization as a new ethical perspective. *Knowledge Organization*. **46**(8), 601-609.
- GUIMARÃES, José Augusto C., 2017. Slanted knowledge organization as a new ethical perspective. Em: Jack ANDERSEN, e Laura SKOUVIG, ed. *The organization of knowledge: caughts between global structures and local meaning*. Bingley: Emerald, cap. 6, pp. 87-102.
- HJØRLAND, Birger, 2002. Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Technology*. **53**(4), 257-270.
- HUDON, Michèle, 1997. Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts. *Knowledge Organization*. **24**(2), 84-91.
- KNOTTNERUS, J. André, 2016. Research data as a global public good. *Journal of Clinical Epidemiology* [Em linha]. 70, 270-271 [consult. 2025-10-20]. Disponível em: <https://www.jclinepi.com/action/showPdf?pii=S0895-4356%2815%2900330-3>.
- MILANI, Sueli Oliveira, 2024. Estudos críticos e sistemas de organização do conhecimento: aspectos conceituais e éticos. Em: C. C. de ALMEIDA, R. SAN SEGUNDO, e D. MARTÍNEZ-ÁVILA, org. *Estudos críticos em organização do conhecimento* [Em linha]. São Paulo: Cultura Acadêmica, pp. 73-104 [consult. 2025-06-20]. Disponível em: https://ebooks.marilia.unesp.br/index.php/lab_editorial/catalog/view/510/4461/8832.

- MOZERSKY, Jessica, et al., 2020. Are we ready to share qualitative research data? Knowledge and preparedness among qualitative researchers, IRB members, and data repository curators. *IASSIST Quarterly*. 43(4), 952.
- NASSI-CALÔ, Lilian, 2019. Promovendo e acelerando o compartilhamento de dados de pesquisa [Em linha]. *SciELO em Perspectiva*. 2019-06-13 [consult. 2024-10-19]. Disponível em: <https://blog.scielo.org/blog/2019/06/13/promovendo-e-acelerando-o-compartilhamento-de-dados-de-pesquisa/>.
- OLSON, Hope, 2002. *The power to name: locating the limits of subject representation in libraries*. Dordrecht: Springer.
- RE3DATA.ORG, 2024a. *Registry of Research Data Repositories* [Em linha] [consult. 2024-12-08]. DOI: <https://doi.org/10.17616/R3D>.
- RE3DATA.ORG, 2024b. *Registry of Research Data Repositories. About* [Em linha] [consult. 2024-11-02]. Disponível em: <https://www.re3data.org/about>.
- RE3DATA.ORG, 2024c. *Registry of Research Data Repositories. Metrics* [Em linha] [consult. 2024-12-08]. Disponível em: <https://www.re3data.org/metrics/contentTypes>.
- SAYÃO, Luis Fernando, e Luana Farias SALES, 2016. Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação* [Em linha]. 21(2), 90-115 [consult. 2024-10-18]. DOI: <https://doi.org/10.5433/1981-8920.2016v21n2p90>.
- STRECKER, Dorothea, et al., 2023. *Metadata Schema for the Description of Research Data Repositories: version 4.0* [Em linha]. [S.l.]: re3data, 33 pp. [consult. 2024-12-02]. DOI: <https://doi.org/10.48440/re3.014>.

