

Resumo: Esta pesquisa explora a disponibilidade de dados abertos indígenas em portais governamentais sul-americanos. O objetivo é identificar, classificar e analisar dados abertos sobre povos indígenas de América do Sul. A metodologia caracteriza-se como exploratória, dividida em três etapas: coleta (seleção de conjuntos de dados relevantes), armazenamento (os dados são classificados com base em uma tabela elaborada a partir dos desafios enfrentados pelos povos indígenas) e análise dos dados. Foram encontrados 813 conjuntos de dados, mas apenas 122 atenderam a critérios específicos, sendo 23 considerados relevantes. A falta de padronização dificultou a comparação entre os conjuntos. Os resultados mostram que a maioria dos países oferece poucos dados, exceto Brasil e Colômbia, no entanto quando analisamos a relevância desses dados observamos uma queda drástica no número de conjuntos. O estudo sugere a criação de uma ferramenta para recompilação e difusão automática de informações sobre povos indígenas, facilitando o acesso para pesquisadores.

Palavras-chave: Dados Indígenas; Identidade dos Povos Indígenas; Portal de dados; Povos originários.

Abstract: This research explores the availability of indigenous open data on South American government portals. The objective is to identify, classify, and analyze open data on indigenous peoples in South America. The methodology is characterized as exploratory, divided into three stages: collection (selection of relevant datasets), storage (data is classified based on a framework developed from the challenges faced by indigenous peoples), and data analysis. A total of 813 datasets were found, but only 122 met specific criteria, with 23 considered relevant. The lack of standardization hindered comparisons between datasets. The results show that most countries provide little data, except for Brazil and Colombia; however, when analyzing the relevance of this data, we observe a drastic drop in the number of datasets. The study suggests the creation of a tool for the automatic compilation and dissemination of information on indigenous peoples, facilitating access for researchers.

Keywords: Indigenous Data; Identity of Indigenous Peoples; Data Portal; Native Peoples.

Introdução

A Ciência de Dados está em crescente evolução, a qual combina estatística, análise de dados e aprendizado de máquina para extrair visualizações de dados valiosas a partir de grandes volumes de dados. Atualmente, ela desempenha um papel crucial em diversas áreas, podendo ser integrada a Ciência da Informação com a finalidade de dar mais ênfase a organização e a representação de dados governamentais, facilitando a recuperação desses dados.

Outro ponto primordial para a recuperação de dados governamentais é a evolução da Internet e a criação de políticas públicas para a disponibilização de dados abertos pelos governos, o que estimulou a criação de portais de dados abertos governamentais. Segundo Bandeira, Alcantara, Sobrinho e Avila (2015) os dados abertos possibilitam que pessoas e organizações utilizem informações públicas de forma livre, para criar aplicativos e realizar

análises. Para serem considerados abertos, esses dados devem ser acessíveis, utilizáveis e redistribuíveis sem restrições, além de estarem indexados e serem legíveis por máquinas, sem impedimentos legais. Outra questão importante é que segundo Hawkins (2001) a Ciência da Informação contribui com os processos de geração, organização, representação, processamento, distribuição, comunicação e uso da informação, no contexto da busca por informações para atender necessidades informacionais.

A escolha do tema desta pesquisa recai sobre os dados indígenas, uma vez que as questões relacionadas a esses povos têm ganhado destaque crescente nos meios de comunicação, especialmente na Internet, que desempenha um papel crucial na coleta e disseminação de informações. Esse aumento pode ser atribuído a várias razões, incluindo a luta pelos direitos territoriais, a preservação cultural e ambiental, e a exploração de recursos naturais em terras indígenas.

Quando nos referimos aos dados indígenas, estes podem possuir várias características de acordo com os povos indígenas, como por exemplo: uma identidade cultural, linguística, histórica e territorial comum. Cada povo indígena representa um grupo maior e mais amplo que pode incluir várias comunidades: Guarani, Mapuche, Quechua, Aymara são exemplos de povos indígenas. Cada um desses povos pode ser encontrado em diferentes regiões e países da América Latina apresentando características diferentes.

Maimone e Matos (2019) enfatizam que a recompilação e difusão de material sobre o desenvolvimento dos povos originários são fundamentais para assegurar que a sociedade tenha acesso a informações indígenas relevantes e atualizadas. Essa prática não apenas enriquece o conhecimento coletivo, mas também fortalece a visibilidade das questões enfrentadas por essas comunidades. Além disso, a ênfase em compartilhar dados indígenas com um enfoque voltado para a defesa desses povos reflete um compromisso profundo com a aprendizagem coletiva e a troca de experiências. Essa abordagem é vital para promover o desenvolvimento dos povos indígenas, garantindo que suas vozes sejam ouvidas.

Dessa forma, a disponibilização de dados abertos sobre povos indígenas é essencial para promover a proteção de seus direitos, entender melhor suas necessidades e realidades gerando pesquisas relevantes. Contudo, surgem as seguintes questões a serem exploradas na investigação: Os países sul-americanos disponibilizam dados indígenas em seus portais governamentais? Podemos agrupá-los ou classificá-los para facilitar a indexação desses dados? Portanto, o objetivo deste trabalho é identificar e analisar dados abertos sobre povos indígenas da América do Sul. Para isso, foram realizados procedimentos de identificação e coleta de dados para armazenagem e classificação de modo a realizar a análise de dados.

Dados abertos governamentais

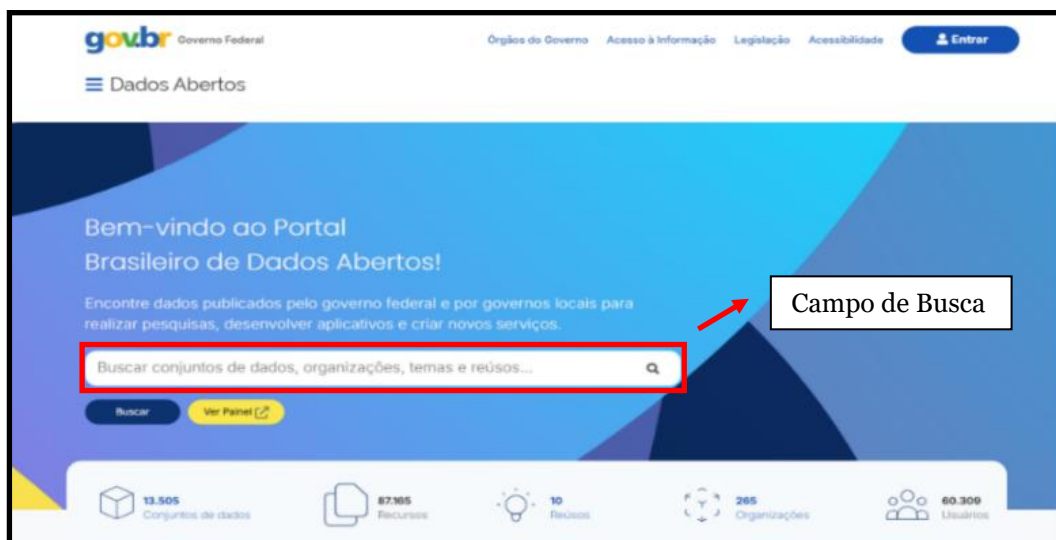
O crescimento da informação que circula na Internet trouxe novos desafios para a organização e recuperação da informação em plataformas digitais, pois nem sempre essas informações estão organizadas o que facilitaria sua recuperação. Lemos e Souza (2020) afirmam que pesquisas nas áreas de Ciência da Informação e Ciência da Computação têm avançado para abordar o excesso de informações e sua organização, visando melhorar a interoperabilidade e a integração de dados heterogêneos entre sistemas e bases disponíveis na *Web*.

Dessa forma o movimento de Dados Governamentais Abertos (DGA¹) surge com o intuito de fortalecer essa relação, disponibilizando dados abertos para que estimulem a participação da sociedade na administração pública. Contudo, este tópico é destinado ao esclarecimento dos termos referentes a tecnologia da informação como: Portais Governamentais, Dados Abertos e Dados estruturados ou não estruturados.

Os portais de dados abertos governamentais têm como objetivo disponibilizar o acesso a informação pública para a sociedade. Segundo Zuiderwijk, Janssen, Choenni, Meijer (2012a), há uma diversidade de portais, mas não há um "melhor" portal, uma vez que cada portal tem seus próprios objetivos, servindo para tornar os dados do governo pesquisáveis e encontráveis.

Porém segundo Chatfield e Reddick, (2018:124) “Os portais de dados abertos falham ao facilitar aos cidadãos comuns a reutilização de dados abertos sem o fornecimento de conjuntos de dados de alta qualidade e uma listagem completa dos campos de metadados.” A Fig. 1 demonstra um portal de dados e seu campo de busca.

Fig. 1 – Print da tela do portal de dados abertos do Brasil



Fonte: Dados abertos, Brasil, [s. d.].

Bernes-Lee (2009) define dados abertos como o esforço para a publicação e disseminação das informações do setor público na *Web*, permitindo a reutilização e a integração destes dados. Por isso, o uso dos dados abertos governamentais deve ser acompanhado de uma política capaz de dispor dados oficiais de fácil acesso e plenamente compreensíveis a pessoas não técnicas, em formato legível para homens e máquinas, com previsibilidade legal, segurança e por fim devidamente atualizados (DAVIES, 2010).

¹ Os DGA são dados abertos no contexto governamental, ou seja, referem-se ao setor público, tornando a informação livremente disponível em formatos abertos e permitindo vias de acesso públicas que facilitem a exploração (KALAMPOKIS, TAMBOURIS e TARABANIS, 2011).

Vaz, Ribeiro e Matheus (2011) relatam que, com o desenvolvimento das tecnologias da informação e comunicação (TIC), bases de dados brutos passaram a ser, livremente, manipuladas, sendo os dados filtrados ou cruzados, constituindo-se em novas aplicações e conhecimentos para a sociedade, caracterizando os Dados Governamentais Abertos (DGA).

Contudo, buscando definir dados abertos, Open Knowledge Brasil (2015) demonstra os oito princípios dos DGA, estabelecendo que os dados para que possam ser considerados abertos devem ser:

Quadro 1 – 8 princípios de dados abertos

1	Completos	Todos os dados públicos devem ser disponibilizados. Dado público é aquele que não está sujeito a restrições de privacidade, segurança ou outros privilégios.
2	Primários	São apresentados tal como colhidos da fonte, com o maior nível possível de granularidade, sem agregação ou modificação (por exemplo, um gráfico não é fornecido aberto, mas os dados utilizados para construir a planilha que deu origem a ele podem ser abertos).
3	Atuais	Devem ser publicados o mais rápido possível para preservar seu valor. Em geral, têm periodicidade: quanto mais recentes e atuais, mais úteis para seus usuários.
4	Acessíveis	São disponibilizados para a maior quantidade possível de pessoas, atendendo, assim, aos mais diferentes propósitos.
5	Compreensíveis por máquina	Devem estar estruturados de modo razoável, possibilitando que sejam processados automaticamente (por exemplo, uma tabela em PDF é muito bem compreendida por pessoas, mas para um computador é apenas uma imagem; uma tabela em formato estruturado, como CSV ou XML, é processada mais facilmente por softwares e sistemas).
6	Não discriminatórios	Devem estar disponíveis para qualquer pessoa, sem necessidade de cadastro ou qualquer outro procedimento que impeça o acesso.
7	Não proprietários	Nenhuma entidade ou organização deve ter controle exclusivo sobre os dados disponibilizados.
8	Livres de licenças	Não devem estar submetidos a copyrights, patentes, marcas registradas ou regulações de segredo industrial.

Fonte: Open Knowledge Brasil, 2015.

Outro ponto importante é o esclarecimento de um arquivo estruturado ou não estruturado. O primeiro exemplo refere-se a um arquivo não estruturado cujo formato é PDF. Como pudemos observar, embora os dados estejam organizados em colunas, não é possível selecionar dados específicos, ao tentarmos selecionar a coluna "Usuários ativos" não será possível, pois esse formato de arquivo permite a seleção de dados apenas por linhas, e não por colunas.

Fig. 2 – Exemplo de arquivo não estruturado extraído do portal do Paraguai

Navegador		↓ Usuarios activos	Usuarios nuevos	Sesiones con interacción	Porcentaje de interacciones	Sesiones con interacción por usuario activo
<input checked="" type="checkbox"/>	Total	223.865 100 % respecto al total	228.233 100 % respecto al total	223.428 100 % respecto al total	61,28 % Media 0 %	1,00 Media 0 %
<input checked="" type="checkbox"/>	1 Chrome	167.563	171.259	173.869	63,96 %	1,04
<input checked="" type="checkbox"/>	2 Safari	24.484	24.248	17.755	52,17 %	0,73
<input checked="" type="checkbox"/>	3 Samsung Internet	19.487	19.281	15.584	48,53 %	0,80
<input checked="" type="checkbox"/>	4 Edge	4.964	4.924	5.431	67,37 %	1,09
<input checked="" type="checkbox"/>	5 Firefox	4.142	4.123	4.937	69,68 %	1,19
<input type="checkbox"/>	6 Android Webview	2.865	2.805	2.510	69,09 %	0,88
<input type="checkbox"/>	7 Opera	1.212	1.212	1.365	65,44 %	1,13
<input type="checkbox"/>	8 Safari (in-app)	219	219	123	51,04 %	0,56
<input type="checkbox"/>	9 Amazon Silk	75	75	57	57,58 %	0,76
<input type="checkbox"/>	10 Internet Explorer	29	29	13	41,94 %	0,45

Fonte: Datos abiertos, Paraguai, [s. d.].

Já o segundo exemplo refere-se a um arquivo estruturado, pois neste caso podemos selecionar ou importar as colunas que desejamos em um sistema de informação e dessa forma obtemos as visualizações dos dados que desejamos.

Fig. 3 – Exemplo de arquivo estruturado extraído do portal da Colômbia

	A	B	C	D	E	F
1	NOMBRE DEL DEPARTAMENTO	NOMBRE DEL MUNICIPIO	RESGUARDO AL QUE PERTENECE	NOMBRE DE LA COMUNIDAD		
2	AMAZONAS	PUERTO SANTANDER	ANDOQUE DE ADUCHE	ANDOQUE DE ADUCHE		
3	AMAZONAS	LA PEDRERA-LETICIA	COMEYAFU	YAVIRA ANGOSTURA		
4	AMAZONAS	LA PEDRERA-LETICIA	COMEYAFU	TANIMUCA		
5	AMAZONAS	LA PEDRERA-LETICIA	COMEYAFU	YUCUNA		
6	AMAZONAS	LA PEDRERA-LETICIA	COMEYAFU	BACURI		
7	AMAZONAS	LA PEDRERA-LETICIA	COMEYAFU	RENACER TANIMUCA		
8	AMAZONAS	LA PEDRERA-LETICIA	CURARE-LOS INGLESES	CURARE		
9	AMAZONAS	LA PEDRERA-LETICIA	CURARE-LOS INGLESES	LOS INGLESES (BORICADA)		
10	AMAZONAS	LETICIA	KILOM. 6 Y 11 CARRETERA LETICIA TARAPACA	SAN JOSE KM 6		
11	AMAZONAS	LETICIA	KILOM. 6 Y 11 CARRETERA LETICIA TARAPACA	CIUDAD HUITOMA KM 7		
12	AMAZONAS	LETICIA	KILOM. 6 Y 11 CARRETERA LETICIA TARAPACA	JUSY MONILLA AMENA KM 9		
13	AMAZONAS	LETICIA	KILOM. 6 Y 11 CARRETERA LETICIA TARAPACA	NIMAIRA NAIMEKI IBIRI KM 11		
14	AMAZONAS	LETICIA	KILOM. 6 Y 11 CARRETERA LETICIA TARAPACA	ZIORA AMENA KM 7		
15	AMAZONAS	LETICIA	LA PLAYA	CANAAAN		

Fonte: Datos abiertos, Colômbia, [s. d.].

Enfim, após conhecermos as principais definições dos recursos utilizados nesta pesquisa partiremos para os procedimentos metodológicos, os quais irão ajudar a elucidar os termos citados acima.

Metodologia

A metodologia desta pesquisa caracteriza-se por ser de natureza exploratória para avaliação qualitativa e quantitativa, uma vez que se propõe a explorar o conteúdo dos arquivos disponibilizados nos portais de dados governamentais dos países sul-americanos e avaliar suas características de organização e representação para indexação e classificação. Os portais de dados governamentais analisados foram: *datos.gob.ar* (Argentina), *datos.gov.br* (Brasil), *datosabiertos.gob.pe* (Peru), *datos.gov.co* (Colômbia), *datos.gob.bo* (Bolívia), *datos.gob.cl* (Chile), *datosabiertos.gob.ec* (Equador), *catalogodatos.gub.uy* (Uruguai), *datos.gob.ve* (Venezuela).

Esta seção apresenta o desenvolvimento da pesquisa mediante execução de três fases:

- 1.** Coleta dos dados indígenas: identificar os países que disponibilizam dados abertos sobre povos indígenas e verificar a disponibilização dos dados abertos nos portais governamentais para identificar os tipos de dados disponibilizados nos portais.

Aborda a identificação dos endereços eletrônicos dos portais governamentais e o acesso a esses portais. Em seguida, utilizam-se as palavras-chave 'indígenas', 'indígenas' e 'pueblos/povos indígenas' no campo de busca, com o objetivo de filtrar e eliminar conjuntos de dados que não são relevantes para esta pesquisa.

- 2.** Armazenamento e processamento dos dados coletados em uma planilha eletrônica para organizar e classificar os dados indígenas extraídos dos conjuntos de dados, utilizando a classificação baseada nos desafios enfrentados por esses povos e desse modo estruturar os dados importados na planilha, articulando-os em colunas.

Essa classificação foi elaborada com base no acesso ao *site* do International Work Group for Indigenous Affairs (IWGIA), especificamente na seção dedicada aos países. Ao clicar em cada país, é possível visualizar os desafios enfrentados pelos povos indígenas. A partir da análise dessas informações geramos as categorias que nos permitirão agrupar diferentes conjunto de dados. Por exemplo, um aspecto comum em diversos países sul-americanos são os territórios e as terras. Por tanto, geramos uma categoria chamada 'Território e Terras', e assim sucessivamente.

Dessa forma, para que os dados sejam armazenados e indexados com eficiência, é necessária a criação de uma planilha de dados com o menor número de colunas vazias², logo os dados capturados nesta pesquisa serão referentes aos dados que são recorrentes nos portais de dados dos países sul-americanos evitando assim um número excessivo de colunas vazias.

² Nesta pesquisa, ao nos referirmos a coluna vazia, estamos destacando a ausência de dados indígenas nos países investigados. Por exemplo, na planilha, não foram importados dados sobre os nomes dos povos indígenas do Brasil, enquanto o Peru disponibilizou essas informações, que foram incluídas na planilha. Essa ausência resultou em colunas/linhas vazias na planilha.

Os dados capturados foram armazenados em uma planilha eletrônica, conforme ilustrado no protótipo representado na Fig. 4.

Fig. 4 – Protótipo da planilha eletrônica para armazenar dados indígenas

A	B	C	D	E	F	G	H	I
id	País	Pueblo/Povo	Comunidad/aldeia	Terra/Reserva	Provincia /Estado	Departamento	Município	etnia

Fonte: Elaborado pelos autores.

O protótipo acima foi elaborado exclusivamente com dados sobre a identificação dos povos indígenas, pois a coleta dessas informações é o primeiro passo no desenvolvimento da planilha. Esses dados são fundamentais para a identificação e compreensão das características desses povos.

3. Análise dos dados coletados

A terceira fase envolve a análise dos dados importados na planilha criada na etapa anterior, à qual inicialmente foram importados dados sobre a identificação dos povos indígenas considerados fundamentais para o desenvolvimento da planilha.

A análise envolve a estrutura organizada em colunas ao qual permite a realização de análises personalizadas, através da aplicação de filtros, possibilitando a extração de informações relevantes. Essa análise irá responder como os dados indígenas são disponibilizados nos portais de dados governamentais. Para isso, a análise realizada consistiu de três tipos: Análise Quantitativa, Análise Classificatória e Análise Descritiva.

Resultados

Esta seção apresenta três análises: a primeira, denominada de Análise Quantitativa, refere-se a quantidade de conjuntos de dados e arquivos extraídos; a segunda, denominada Análise Classificatória, trata da classificação dos dados indígenas coletados; a terceira, denominada Análise Descritiva, é referente a importação dos dados na planilha, ao qual irá responder se os dados estão estruturados e prontos para serem recuperados em um sistema de informação.

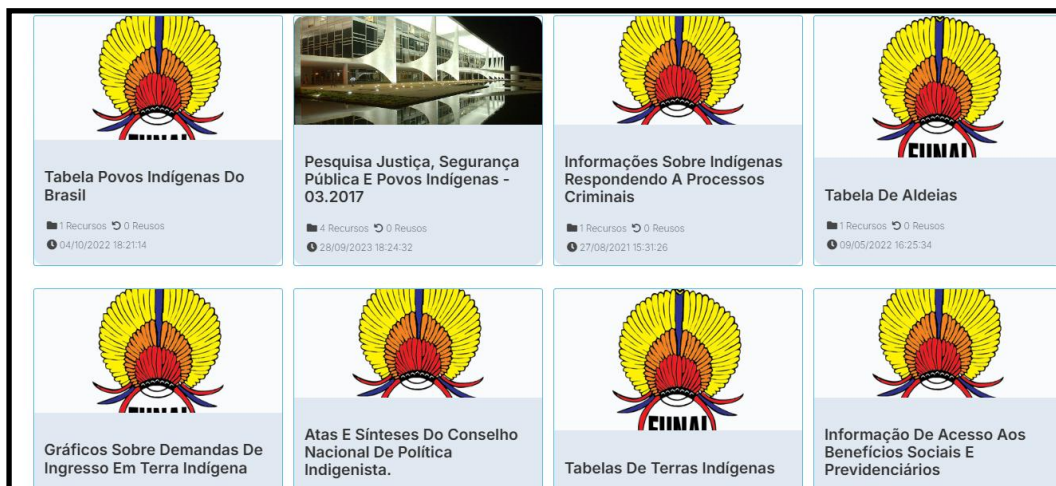
- 1.** A primeira análise, denominada de Análise Quantitativa, refere-se a quantidade de conjuntos de dados e arquivos extraídos.

Na primeira busca, mais genérica, com a palavra-chave "INDIGENA", foram identificados 813 conjuntos de dados. Na segunda busca, mais específica, com a palavras-chave "PUEBLOS/POVOS INDIGENAS" identificaram-se 122 conjuntos de dados utilizados para análise e classificação.

Desses 122 conjuntos, foram coletados 23 conjuntos de dados abertos referentes aos grupos (A, B, C, D e E) e importados os dados referentes aos grupos (A e B), pois são dados recorrentes nos países: Argentina, Brasil, Colômbia e Peru. Os outros países não

disponibilizaram dados referentes a classificação (A e B). A Fig. 5 mostra um exemplo da coleta de conjuntos de dados no portal brasileiro.

Fig. 5 – Exemplo de pesquisa em um portal de dados abertos



Fonte: Dados abertos Brasil, [s. d.].

2. A segunda análise, denominada Análise Classificatória, trata da classificação dos dados indígenas coletados

Dando ênfase à classificação criada de acordo com os desafios enfrentados pelos povos indígenas, a síntese *Os Povos indígenas na América Latina*, de autoria da Comissão Econômica para a América Latina e o Caribe, refere-se ao Convênio 169 da Organização Internacional do Trabalho (OIT) sobre Povos Indígenas e Tribais e estabelece que: “[...] é possível identificar ao menos quatro dimensões relacionadas aos povos indígenas: o reconhecimento da identidade, a origem comum, a territorialidade e a dimensão linguística e cultural, que devem ser levadas em conta ao se estabelecer critérios operacionais” (COMISSÃO..., 2015:38).

Assim, de acordo a análise dos desafios disponíveis no *website* IWGIA, foi criada a classificação dos dados indígenas, conforme demonstrado no Quadro 2.

Quadro 2 – Classificação dos dados indígenas

Grupo	Categoria	Definição	Desafio
A	Dados da Comunidade Indígena	É responsável pela caracterização da comunidade indígena disponibilizando: nome, município, estado da comunidade indígena	Falta de reconhecimento dos povos
B	Território e Terras	Este atributo pode conter informações sobre o território tradicional da comunidade indígena, incluindo tamanho, fronteiras, uso da terra, direitos territoriais e outras informações relacionadas à terra e recursos naturais.	Apropriação de terras indígenas

C	Dados Demográficos	Isso pode incluir informações sobre a população da comunidade indígena, como número de membros, faixa etária, sexo, densidade populacional, e outras características demográficas relevantes.	Censo Indígena
D	Aspectos Culturais e Tradicionais	Pode incluir informações sobre a cultura, tradições, línguas, práticas religiosas, costumes, arte, artesanato e outros aspectos culturais e tradicionais da comunidade indígena.	Discriminação da cultura indígena
E	Outros tipos de dados indígenas	Dados capturados nesta pesquisa são referentes a informações administrativas.	Falta de apoio do governo

Fonte: Elaborado pelos autores.

Foram recuperados 23 conjuntos de dados, cuja figura abaixo (Fig. 6) revela três características desses conjuntos: País, Grupo ou classificação e o nome do conjunto.

Fig. 6 – Conjuntos de dados coletados nesta pesquisa com sua classificação

	A	B	C
1	Pais	Grupo	Conjunto
2	Argentina	A	Listado de comunidades indígenas
3	Brasil	A	Tabela Povos Indígenas do Brasil
4	Brasil	E	Informações sobre indígenas respondendo a processos criminais
5	Brasil	A	Tabela de Aldeias
6	Brasil	A	Gráficos sobre demandas de Ingresso em Terra Indígena
7	Brasil	A	Tabelas de Terras Indígenas
8	Brasil	E	Saúde Indígena - Indicadores de enfrentamento e monitoramento à covid-19 dos Povos Indígenas
9	Brasil	A	Tabela das Bases de Proteção Etnoambiental
10	Brasil	E	Dados organizacionais sobre a unidade Museu do Índio
11	Brasil	E	Dados sobre contratos administrados pelo Museu do Índio
12	Brasil	E	Dados do acervo bibliográfico custodiado pelo Museu do Índio
13	Brasil	E	Solicitações de informações ao SIC
14	Brasil	E	Base de dados da Biblioteca Curt Nimuendajú
15	Brasil	E	Dados sobre o acervo museológico custodiado pelo Museu do Índio
16	Brasil	E	Cartografia de Ataques Contra Indígenas - CACI
17	Brasil	E	Sistema de Atenção à Saúde Indígena (SIASI) - Módulo de Vigilância Alimentar e Nutric...
18	Brasil	A	Comunidades Indígenas de Alagoas
19	Peru	A	Pueblos Indígenas en Aislamiento y Contacto Inicial
20	Peru	D	Lenguas Videntes
21	Colômbia	A	Resguardos Indígenas Formalizados
22	Colômbia	A	Accionar Institucional Territorios Indigenas E
23	Colômbia	A	Open Data Accionar Institucional Territorios Ind
24	Colômbia	A	Pueblos indígenas a nivel Nacional 2020

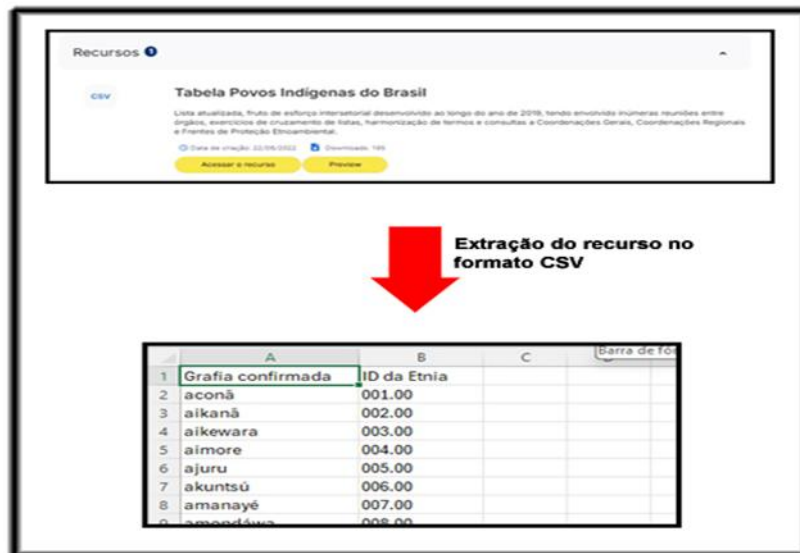
Fonte: Dados abertos Brasil, [s. d.].

Um fato interessante que ocorreu durante a pesquisa foi o acesso ao recurso Tabelas Povos Indígenas localizado no portal de dados do Brasil. Neste caso é disponibilizado um arquivo no formato CSV³, cujo teor, se observássemos o seu nome, deveria conter dados sobre os povos indígenas como nome do povo indígena, comunidade, aldeia, etc. No entanto,

³ CSV - Formato de arquivo separado em colunas por vírgulas.

quando abrimos o arquivo, nos deparamos com um arquivo totalmente incompleto, como mostra a figura abaixo.

Fig. 7 – Exemplo: Acesso de um recurso do conjunto de dados



Fonte: Dados abertos Brasil, [s. d.].

Também é importante demonstrarmos a diferença de um arquivo que é relevante, ou não, para esta pesquisa. Os arquivos relevantes para esta pesquisa são aqueles que possuem o seu conteúdo voltado para uma das classificações do Quadro 2; abaixo (Fig. 8) segue um exemplo de arquivo extraído do portal de dados brasileiro, relevante para esta pesquisa, porém contendo falhas.

Fig. 8 – Arquivo extraído do portal de dados brasileiro com falhas

terrai_nome	etnia_nome	municipio_nome	uf_sigla	fase_ti	modalid
Acapuri de Cima	Kokama	Fonte Boa	AM	Regularizada	Tradicic
Acimã	Apurinã	Lábrea	AM	Regularizada	Tradicic
-66.17767317027474 -7.7	-66.17762646872347 -7.79	-66.17730723542437 -7.7968	-66.17726	-66.177190994065	-66.177
-66.03993403036404 -7.7	-66.03988525233892 -7.78	-66.03981417761518 -7.7885	-66.03971	-66.0395005702914	-66.039
-66.03903690110812 -7.8	-66.03936485902626 -7.83	-66.03981486163639 -7.8376	-66.04042	-66.040605452942	-66.040
-66.13836539367621 -7.8	-66.13856096655063 -7.88	-66.13874078972947 -7.8869	-66.13892	-66.139042492857	-66.139
Aconã	Tingui-Botã	Traipu	AL	Regularizada	Reserva
Aldeia Escola Floresta	Maxakali	Teãfilo Otoni	MG	Regularizada	Reserva
Aldeia Katurama	"Pataxã	Pataxo Hãj-Hãj-Hãj"	Sãõ Joaq	346.124	Regular
Aldeia Kondã	Kaingang	Chapecã	SC	Encaminhada RI	Reserva
Aldeia Limão Verde	Guarani Kaiowã	Amambai	MS	Regularizada	Tradicic
Aldeia Velha	Pataxã	Porto Seguro	BA	Homologada	Tradicic
Alto Rio Guamã	"Timbira	Tembã	"Santa Luz Nova Esperanã	Garrafã	
-46.906308731973 -2.02	-46.906307989687 -2.0255	-46.906305055127 -2.025558	-46.90621	-46.905997009347	-46.905
-46.95519826 -1.767407	-46.95516671 -1.76735585	-46.95508927 -1.76735011	-46.95487	-46.95458159 -1.76	-46.954
-46.59290445 -2.820977	-46.59294234 -2.8208007	-46.59306409 -2.82044728	-46.59317	-46.59330557 -2.81	-46.593
Alto Rio Negro	"Arapãso	Karapanã	Barasãne Wanana"	"Japurã	
-67.347424576842 0.169	-67.347924833581 0.16884	-67.348195805982 0.168216	-67.34836	-67.347695549243	-67.347
-67.51855738 -0.400643	-67.51876336 -0.40042193	-67.51898989 -0.40023124	-67.51923	-67.51976711 -0.39	-67.519
-67.67667545 -0.467286	-67.67674738 -0.467758	-67.67689294 -0.46822068	-67.67712	-67.67743955 -0.46	-67.677

Fonte: Dados abertos Brasil, [s. d.].

Com relação ao arquivo acima, embora ele esteja com algumas falhas, foi possível fazer uma limpeza nos dados que estão confusos e importarmos na planilha criada nesta pesquisa.

Agora iremos mostrar um arquivo de dados que não é relevante para esta pesquisa cujo teor refere-se a uma planilha de veículos utilizados por órgãos indígenas.

Fig. 9 – Exemplo de arquivo sem relevância, extraído do portal de dados brasileiro

A	B	C	D	E	F
CÓDIGO DA UG UNIDADE GESTORA	NOME DA UG UNIDADE GESTORA	MODELO	TIPO DE VEÍCULO (INFORMAÇÃO DO DENATRAN)	PLACA	ESTADO
194035	Diretoria de Administração e Gestão - DAGES/ FUNAI SEDE	FIAT/PALIO WEEK TREKKING	AUTOMOVEL	NUA1673	DF
194035	Diretoria de Administração e Gestão - DAGES/ FUNAI SEDE	FIAT/PALIO WEEK TREKKING	AUTOMOVEL	NUB6773	DF
194035	Diretoria de Administração e Gestão - DAGES/ FUNAI SEDE	FIAT/PALIO WEEK TREKKING	AUTOMOVEL	NTZ7893	DF
194035	Diretoria de Administração e Gestão - DAGES/ FUNAI SEDE	FIAT/SIENA ESSENCE 1.6	AUTOMOVEL	LMH2142	DF
194035	Diretoria de Administração e Gestão - DAGES/ FUNAI SEDE	CITROEN/JUMP GREENCAR ES	VAN	OB01454	DF
194035	Diretoria de Administração e Gestão - DAGES/ FUNAI SEDE	RENAULT/MASTER MARIM PAS	VAN	RAK6042	DF
194035	Diretoria de Administração e Gestão - DAGES/ FUNAI SEDE	Mitsubishi L200	CAMIONETA	QCJ3801	DF
194035	Diretoria de Administração e Gestão - DAGES/ FUNAI SEDE	Mitsubishi L200	CAMIONETA	QCJ3741	DF
194035	Diretoria de Administração e Gestão - DAGES/ FUNAI SEDE	Mitsubishi L200	CAMIONETA	QCH7429	DF
190004	Coordenação Regional de Ji- Paraná - RO	MMC/L200 4X4 GL	CAMINHONETE	NDX9827	RO
190004	Coordenação Regional de Ji- Paraná - RO	MMC/L200 OUTDOOR	CAMINHONETE	NCV9485	RO
190004	Coordenação Regional de Ji- Paraná - RO	MMC/L200 4X4 GL	CAMINHONETE	NDZ9624	RO
190004	Coordenação Regional de Ji- Paraná - RO	MMC/L200 TRITON GL D	CAMINHONETE	QBU3530	RO
190004	Coordenação Regional de Ji- Paraná - RO	MMC/L200 TRITON GL D	CAMINHONETE	QBU3370	RO
190004	Coordenação Regional de Ji- Paraná - RO	MMC/L200 TRITON 3.2 D	CAMINHONETE	OBJ2801	RO
194003	Coordenação Regional de Cacoal - RO	MMC/L200 TRITON GL D	CAMINHONETE	QBL9561	RO
194003	Coordenação Regional de Cacoal - RO	IVECO/VERTIS 90V18	CAMINHÃO	NJT0693	RO
194003	Coordenação Regional de Cacoal - RO	MMC/L200 SPORT GL	CAMINHONETE	QTD-1H37	RO
194003	Coordenação Regional de Cacoal - RO	MMC/L200 SPORT GL	CAMINHONETE	QTF-1H15	RO
194003	Coordenação Regional de Cacoal - RO	MMC/L200 SPORT GL	CAMINHONETE	QTF-1H25	RO
194003	Coordenação Regional de Cacoal - RO	MMC/L200 4X4 GL	CAMINHONETE	QBL-9561	RO
194003	Coordenação Regional de Cacoal - RO	MMC/L200 SPORT GL	CAMINHONETE	QTC-1455	RO

Fonte: Dados abertos Brasil, [s. d.].

3. A Análise Descritiva é referente a importação dos dados na planilha, a qual irá responder se os dados estão estruturados e prontos para serem recuperados em um sistema de informação.

A análise entre os conjuntos de dados do Brasil e Colômbia em relação a relevância à temática indígena com a palavra “indígena” retornou 311 conjuntos de dados no portal brasileiro, porém apenas 32 referem-se a dados que podem ser relevantes para esta pesquisa. Desses 32 conjuntos foram identificados 16 arquivos de dados abertos que referem-se as categorias A, B, C, D e E.

Com relação ao portal de dados da Colômbia, foram identificados 445 conjuntos de dados, dos quais foram selecionados apenas 4 pois referem-se as categorias estipuladas no quadro 2, sendo extraídos 29 arquivos de dados abertos e importado na planilha apenas 1.

A Argentina disponibilizou apenas 14 conjuntos de dados, sendo extraído apenas um arquivo do qual foram importados 1.878 registros, muito mais que para Brasil e Colômbia, que disponibilizaram 756 de conjuntos de dados. No entanto, a falta de padronização nos atributos disponibilizados entre os países dificultou a comparação direta e a análise agregada dos dados coletados.

O Peru retornou 18 conjuntos de dados abertos sobre povos indígenas e 2 arquivos de dados abertos, sendo selecionado apenas 1 arquivo para importação de seus dados na planilha devido ao arquivo piaci.csv possuir dados necessários para identificação desses povos.

O portal de dados do Paraguai contém dados estatísticos relacionados à população indígena, incluindo informações demográficas sobre a população dopaís, sendo grande parte referente aos anos de 2016 a 2019.

O portal do Uruguai disponibiliza em seu portal gráficos sobre homicídios, violência doméstica, compras e contratações, dentre outras. Quando realizamos a busca com a palavra-chave ‘indígenas’ aparecem publicações e não dados sobre esses povos.

O portal da Bolívia detalha os projetos de desenvolvimento econômico territorial com inclusão em nível nacional, discriminados por departamentos e municípios da gestão de 2015. Venezuela, Chile, Equador, Guiana e Suriname não retornaram conjuntos de dados referentes às questões indígenas com as palavras indicadas na primeira fase do tópico 3.

Com relação as colunas vazias, este fato poderá ser ocasionado devido a falta de dados indígenas entre os países investigados, por isso esta pesquisa capturou dados relacionados com a classificação A e B, para que dessa forma o número de colunas vazias fosse o menor possível. Por exemplo: com relação as colunas povo/pueblo e comunidade/aldeia localizados na planilha criada nesta pesquisa, a Argentina disponibilizou ambos os dados, já a Colômbia disponibilizou apenas dados referente a comunidade e o Peru disponibilizou apenas dados referentes a povos indígenas que habitam em seu país. O Brasil não disponibilizou nenhuma das informações. Dessa forma foram geradas colunas vazias demonstradas na figura abaixo (Fig. 10):

Fig. 10 – Planilha de dados gerada nesta pesquisa

A	B	C	D	E	F	G	H	I
Id	País	Pueblo/Povo	Comunidad/aldeia	Terra/Reserva	Provincia/Estado	Departamento	Municipio	etnia
1	Argentina	Moqoit (Mocovi)	Comunidad So Icolaq		Santa Fe	La Capital		
2	Argentina	Guarani	Comunidad Guarani Nande Sy Yvy		Buenos Aires	Quilmes		
3	Argentina	Diaguita (Cacano)	Comunidad Aytlu Cacano San Isidro		Santiago del Este	Atamisqui		
4	Argentina	Moqoit (Mocovi)	Comunidad Nacoñiguit so Nashi Cacique Raimundo Villaba		Santa Fe	San Javier		
5	Brasil			Aldeia Velha	BA		Porto Seguro	Pataxó
6	Brasil			Alto Rio Guamã	PA		Santa Luzia do Pará, Timbira, Temb	
7	Brasil			Alto Rio Negro	AM		Japurá, São Gabriel d Arapáso, Kara	
8	Brasil			Alto Rio Purus	AC		Santa Rosa do Purus	Kaxinawá, Kul
9	Brasil			Alto Sepatini	AM		Lábrea	Apurinã
10	Brasil			Alto Taraucacá	AC		Jordão, Feijó	Isolados
11	Colômbia		ANDOQUE DE ADUCHE	ANDOQUE DE ADUCHE	AMAZONAS		PUERTO SANTANDER	
12	Colômbia		YAVIRA ANGOSTURA	COMEFYFU	AMAZONAS		LA PEDRERA-LETICIA	
13	Colômbia		TANIMUCA	COMEFYFU	AMAZONAS		LA PEDRERA-LETICIA	
14	Colômbia		YUCUNA	COMEFYFU	AMAZONAS		LA PEDRERA-LETICIA	
15	Peru	Machiguenga del alto Paquiria (Kirineri)		Kugapakori, Nahua, Nanti y Otros	La Convención	CUSCO		Arawak
16	Peru	Machiguenga del Alto Serjali (Kirineri)		Kugapakori, Nahua, Nanti y Otros	La Convención	CUSCO		Arawak
17	Peru	Machiguenga del Camisea (Nanti)		Kugapakori, Nahua, Nanti y Otros	La Convención	CUSCO		Arawak
18	Peru	Machiguenga del Cashiriani (Nanti)		Kugapakori, Nahua, Nanti y Otros	La Convención	CUSCO		Arawak
19	Peru	Machiguenga del Timpia (Nanti)		Kugapakori, Nahua, Nanti y Otros	La Convención	CUSCO		Arawak

Fonte: Elaborado pelos autores.

A síntese final dos resultados desta pesquisa abrange a classificação elaborada a partir dos desafios enfrentados pelos povos indígenas, conforme o *website* da IWGIA. Essa classificação foi organizada em cinco categorias, A, B, C, D e E, que contemplam temas como identificação dos povos, terras indígenas, dados demográficos, cultura e informações administrativas.

Após a conclusão da classificação, iniciamos a análise dos conjuntos de dados disponíveis nos portais dos países da América do Sul que disponibilizaram dados, sejam eles relevantes ou não, os quais totalizam 50% dos países deste continente. A outra metade não retornou dados ou não possui portais de dados abertos. Finalizando, os dados disponibilizados demonstram uma certa preocupação, porque se fossem importados todos os arquivos na planilha iriam gerar um número excessivo de colunas vazias dificultado a recuperação desses dados.

Considerações finais

Este estudo teve como objetivo investigar a disponibilidade dos dados abertos sobre povos indígenas nos portais governamentais da América do Sul, com ênfase na coleta, organização e agrupamento dessas informações, para que dessa forma esses dados possam ser recuperados e analisados de forma mais eficaz.

Atualmente, os portais de dados abertos dos países da América do Sul oferecem um número limitado de conjuntos de dados, com exceção notável do Brasil e da Colômbia, o que gerou falha na primeira e segunda fases da metodologia que tratam da coleta e armazenamento dos dados, uma vez que nem todos os países disponibilizaram dados e também esses dados não possuem a mesma classificação gerando colunas vazias na planilha. No entanto, quando passamos para a fase número 3, a qual trata da análise desses dados, o primeiro quesito observado é a relevância dos dados⁴ para a pesquisa em questão, tendo havido uma queda drástica no número de conjuntos disponíveis, ou seja, dados pertencentes aos outros grupos.

Uma das descobertas fundamentais foi a heterogeneidade dos dados dos portais governamentais, onde se observaram diferentes nomenclaturas, como departamento, província, estado e município. Essa diversidade reflete as variações nas subdivisões administrativas dos países, evidenciando as diferenças na organização territorial e na terminologia utilizada.

Ainda assim, para que não houvesse muitas colunas vazias devido a falta de dados de um país, foram armazenados na planilha apenas os dados dos grupos A e B, que se referem a identificação dos povos indígenas como: nome do povo, nome da comunidade, terra, localização, os quais são fundamentais para melhorar a indexação dos dados na planilha e também contribuir para diminuição do número de colunas vazias.

Em se tratando dos dados indígenas nos portais de dados sul-americanos, observamos que, entre os oito princípios dos dados abertos, o princípio número 3 - 'Atuais', é o que menos tem sido seguido pelos países desta região. Isso se deve ao fato de que os dados disponíveis não estão sendo atualizados regularmente, portanto não refletem informações recentes.

Contudo, ainda há muitos dados que podem ser disponibilizados pelos governos, porém seria um bom começo se todos os países disponibilizassem em seus portais dados dos grupos A e B, os quais tratam da identidade dos povos indígenas.

Para projetos futuros, importa analisar a possibilidade da criação de uma ferramenta ou sistema para a recompilação, processamento e difusão automática de dados abertos de América do Sul. A difusão automática e periódica dos dados sobre povos indígenas poderá ser dirigida a pessoas ou pesquisadores interessados nesta temática.

⁴ Dados referentes as categorias A e B do quadro 2, as quais são recorrentes nos países que disponibilizaram arquivos de dados abertos.

Referências bibliográficas

BANDEIRA, Jusdon; ALCANTARA, Williams; BARBOSA SOBRINHO, Armando

2015 Dados abertos conectados para a Educação. *Jornada de Atualização em Informática na Educação*. [Em linha]. 4:1 (2015) 47-69. [Consult 28 nov 2024]. Disponível em: https://www.researchgate.net/publication/283569633_Dados_Abertos_Conectados.

BERNERS-LEE, Tim

2009. Putting government data online. [Em linha]. 2009. [Consult. 30 nov. 2024]. Disponível em: <https://www.w3.org/DesignIssues/GovData.html>.

CHATFIELD, Akemi Takeoka; REDDICK, Christopher G.

2017 The Role of policy entrepreneurs in open government data policy innovation diffusion: An analysis of Australian Federal and State Governments. *Government Information Quarterly*. 35:1 (2018) 123-134.

COMISSÃO ECONÔMICA PARA A AMÉRICA LATINA

2015 Os Povos indígenas na América Latina: avanços na última década e desafios pendentes para a garantia de seus direitos. [Em linha]. Santiago do Chile: Nações Unidas; Brasília: CNPq, 1984. [Consult, 28 out 2024]. Disponível em: <https://repositorio.cepal.org/server/api/core/bitstreams/674dfaa2-fe61-484a-a61f-390330e9174a/content>.

DADOS ABERTOS BRASIL

[20--] *Dados abertos [Brasil]*. [Em linha]. [20--]. [Consult. 30 dez. 2024]. Disponível em: <https://dados.gov.br>.

DATOS ABIERTOS ARGENTINA

[20--] *Datos [abiertos] Argentina*. [Em linha]. [20--]. [Consult. 30 nov. 2024]. Disponível em: <https://datos.gob.ar>.

DATOS ABIERTOS BOLÍVIA

[20--] *Datos abiertos Bolívia*. [Em linha]. [20--]. [Consult. 30 nov. 2024]. Disponível em: <https://datos.gob.bo>.

DATOS ABIERTOS CHILE

[20--] *[Datos abiertos Chile]*. [Em linha]. [20--]. [Consult. 30 nov. 2024]. Disponível em: <https://datos.gob.cl>.

DATOS ABIERTOS COLÔMBIA

[20--] *Datos abiertos [Colômbia]*. [Em linha]. [20--]. [Consult. 30 nov. 2024]. Disponível em: <https://datos.gov.co>.

DATOS ABIERTOS ECUADOR

[20--] *Datos abiertos [Ecuador]*. [Em linha]. [20--]. [Consult. 30 nov. 2024]. Disponível em: <https://datosabiertos.gob.ec>.

DATOS ABIERTOS PARAGUAI

[20--] *[Datos abiertos Paraguai]*. [Em linha]. [20--]. [Consult. 30 out 2024]. Disponível em: <https://datos.gov.py>.

DATOS ABIERTOS PERU

[20--] *Datos abiertos [Peru]*. [Em linha]. [20--]. [Consult. 30 out 2024]. Disponível em: <https://datosabiertos.gob.pe>.

DATOS ABIERTOS URUGUAI

[20--] *Datos abiertos [Uruguai]*. [Em linha]. [20--]. [Consult. 28 out 2024]. Disponível em: <https://catalogodatos.gub.uy>.

DATOS ABIERTOS VENEZUELA

[20--] *Datos abiertos [Venezuela]*. [Em linha]. [20--]. [Consult. 28 out 2024]. Disponível em: <https://datos.gob.ve>.

DAVIES, T.

2010 Open data, democracy and public sector reform: A look at open government data use from data.gov.uk. [Em linha]. 2010. [Consult. 27 out 2024]. Disponível em: <http://practicalparticipation.co.uk/odi/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf>.

HAWKINS, Donald T.

2001 Information science abstracts: tracking the literature of information science. Part 1: definition and map. *Journal of the American Society for Information Science and Technology*. [Em linha]. 52:1 (2001) 44-54. [Consult. 28 out 2024]. Disponível em: [http://onlinelibrary.wiley.com/doi/10.1002/1532-2890\(2000\)52:1%3C44::AID-ASI1057%3E3.O.CO;2-6/epdf](http://onlinelibrary.wiley.com/doi/10.1002/1532-2890(2000)52:1%3C44::AID-ASI1057%3E3.O.CO;2-6/epdf).

LEMONS, Daniela Lucas da Silva; SOUZA, Renato Rocha

2020 Knowledge organization systems for the representation of multimedia resources on the web: A comparative analysis. *Knowledge Organization*. [Em linha]. 47:4 (2020) 300-319. [Consult. 28 out 2024]. Doi: <https://doi.org/10.5771/0943-7444-2020-4-300>.

MAIMONE, Geovana Deliberali; MATOS, Amanda Pedrosa de

2019 Cultura indígena sob a perspectiva da Ciência da Informação. *Brazilian Journal of Information Studies: Research Trends*. [Em linha]. 13:3 (2019) 46-55. [Consult. 28 out 2024]. Disponível em: <http://doi.org/10.36311/1981-1640.2019.v13n3.05.p46>.

OPEN KNOWLEDGE BRASIL

2015 Quando surgiram os 8 princípios dos dados abertos? [Em linha]. 2015. [Consult. 28 out 2024]. Disponível em: <https://ok.org.br/noticia/quando-surgiram-os-oito-principios-dos-dados-abertos>.

RAUTENBERG, Sandro; DO CARMO, Paulo Ricardo Viviurka

2019 Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. [Em linha]. *Brazilian Journal of Information Science*. 13:1 (2019) 56-67. [Consult. 30 nov 2024]. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=6983493>.

ROBREDO, J.

2003 *Da Ciência da Informação revisitada aos sistemas humanos de informação*. Brasília: Thesaurus, 2003.

SHERA, J. H.; EGAN, M. E.

1961 Exame atual da Biblioteconomia e da Documentação. In BRADFORD, S. C. - *Documentação*. Rio de Janeiro: Fundo de Cultura, 1961, p. 15-60.

ZUIDERWIJK, Anneke [et al.]

2012 Socio-technical Impediments of Open Data. *Electronic Journal of e-Government*. [Em linha]. 10:2 (2010) 156-172. [Consult. 30 nov. 2024]. Disponível em: <https://academic-publishing.org/index.php/ejeg/article/view/571/534>.

Ricardo Costa Rossi | ricardo.rossi@unesp.br

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Brasil

Mariângela Spotti Lopes Fujita | mariangela.fujita@unesp.br

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Brasil

Isidoro Gil Leiva | isgil@um.es

Universidad de Murcia, Espanha