

# ARQUITETURA DA INFORMAÇÃO EM REPOSITÓRIOS DE DADOS CIENTÍFICOS DISCIPLINARES: A INFLUÊNCIA DOS SISTEMAS DE NAVEGAÇÃO E BUSCA NA ENCONTRABILIDADE DOS DADOS

## INFORMATION ARCHITECTURE IN DISCIPLINARY RESEARCH DATA REPOSITORIES: THE INFLUENCE OF SEARCH AND NAVIGATION SYSTEMS ON DATA FINDABILITY

Lucas Henrique Alves da Silva<sup>a</sup>  
Marckson Roberto Ferreira de Sousa<sup>b</sup>  
Izabel França de Lima<sup>c</sup>

### RESUMO

**Objetivo:** Avaliar a estruturação dos sistemas de navegação e busca de repositórios de dados científicos disciplinares sob o ponto de vista da Arquitetura da Informação, analisando em que medida os sistemas de navegação e busca podem influenciar a encontrabilidade da informação nesses ambientes digitais. **Metodologia:** Caracteriza-se como qualitativa quanto à sua abordagem e descritiva quanto aos seus objetivos. Selecionaram-se quatro repositórios de dados de diferentes ramos do conhecimento a partir do diretório re3data.org, compondo-se uma amostra heterogênea pelo critério da intencionalidade. Adotou-se um checklist para nortear a avaliação, listando-se elementos observáveis com ênfase nos atributos da encontrabilidade da informação. **Resultados:** Os repositórios PANGAEA e Biodiversity.aq demonstraram maior aderência aos requisitos da encontrabilidade, enquanto o OLAC e o TalkBank revelaram maior fragilidade quanto aos seus sistemas. **Conclusões:** Observou-se discrepância entre os princípios da Arquitetura da Informação e os sistemas de navegação e busca dos repositórios de dados analisados. Por fim, constatou-se que muitos repositórios se assemelham mais a bancos de dados do que a repositórios propriamente ditos.

**Descritores:** Arquitetura de Informação. Repositórios de dados científicos. Repositórios

---

<sup>a</sup> Mestre em Ciência da Informação pela Universidade Federal da Paraíba (UFPB). E-mail: lucashenalsi@gmail.com

<sup>b</sup> Doutor em Engenharia Elétrica pela Universidade Federal da Paraíba (UFPB). Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba (UFPB). E-mail: marckson.dci.ufpb@gmail.com

<sup>c</sup> Doutora em Ciência da Informação pela Universidade Federal de Minas Gerais (UFMG). Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba (UFPB). E-mail: belbib@gmail.com

temáticos. Encontrabilidade da informação.

## 1 INTRODUÇÃO

A disponibilidade de tecnologias cada vez mais avançadas de armazenamento, transmissão e análise de dados, aliada às mudanças paradigmáticas do próprio fazer científico ao longo do tempo, tornaram a ciência intensivamente conduzida por dados. Em decorrência disso, observou-se um fortalecimento da cultura de preservar, compartilhar e divulgar os conjuntos de dados produzidos nas pesquisas dos mais diversos ramos do conhecimento.

A socialização do conhecimento científico tem passado por profundas transformações ao longo da última década. Entre as diversas razões para essas mudanças, pode-se apontar uma tendência de valorização dos dados científicos como um produto da atividade investigativa que também deve ser objeto de preservação e compartilhamento entre a comunidade científica.

A disponibilização dos dados de pesquisa visa aumentar a visibilidade e democratizar o acesso aos conjuntos de dados, tanto para que possam ser reutilizados por outros pesquisadores quanto para proporcionar uma validação mais rigorosa dos resultados da pesquisa científica, tornando a ciência mais célere e transparente, além da economia de tempo, dinheiro e esforços humanos (CAVALCANTI; SALES, 2017).

A comunicação dos dados de pesquisa enfrenta um desafio que diz respeito à gestão desses recursos, a qual perpassa os processos de coleta, documentação, representação, armazenamento, compartilhamento, reuso, entre outros que constituem o ciclo de vida dos dados. A gestão de dados científicos inicia-se com o próprio pesquisador, a partir do momento em que este define que tipos de dados serão coletados ou gerados, a forma de coleta e o modo como os conjuntos de dados serão documentados, descrições essenciais para possibilitar a reinterpretação e o reuso dos dados por outros pesquisadores.

No que concerne às atividades referentes à gestão de dados de pesquisa, os repositórios digitais figuram como uma das ferramentas aparelhadas com os recursos necessários para atender ao conjunto de demandas envolvidas nessa

agenda. De acordo com Guandalini, Furnival e Arakaki (2019), os repositórios digitais contribuem para alicerçar uma cultura de boas práticas científicas, que incluem a divulgação e disseminação da produção científica tanto para a comunidade acadêmica quanto para a sociedade como um todo.

Os dados de pesquisa, em virtude de sua estrutura e natureza distintas em relação aos materiais bibliográficos tradicionais, requerem um tratamento especializado para que sejam apropriadamente preservados e reutilizados, o que suscitou a criação dos repositórios de dados. Para Monteiro e Sant'Ana (2016), os repositórios de dados científicos desempenham papel essencial na gestão de dados, na medida em que reúnem conjuntos de dados e possibilitam seu armazenamento, compartilhamento, acesso, recuperação e reuso, amparando a cultura de colaboração científica.

O compartilhamento de dados de pesquisa por meio de repositórios digitais permite que pesquisadores trabalhem colaborativamente em escala internacional, propiciando a reutilização dos dados com hipóteses distintas e ideias originais, além da disponibilidade dos conjuntos de dados para acesso a longo prazo (MEDEIROS; CAREGNATO, 2012; SAYÃO; SALES, 2014).

Os repositórios digitais têm como uma de suas funções principais disponibilizar estoques de dados ou recursos informacionais em meio eletrônico, fazendo com que a aplicação de princípios da Arquitetura da Informação seja uma ação fundamental para tornar esse conteúdo acessível e recuperável.

Abordar a Arquitetura da Informação em repositórios de dados científicos justifica-se pelo fato de serem, de acordo com Sanchez e Vechiato (2019, p. 143), “incipientes na literatura científica estudos que abordem critérios para o projeto, a implementação e a avaliação desse tipo específico de repositório digital”. Os estudos em Arquitetura da Informação abrangem os mais diversos tipos de ambientes informacionais, seja contribuindo para a concepção de novos ambientes ou aperfeiçoando aqueles já existentes.

A Arquitetura da Informação é composta pelos sistemas de organização, rotulação, navegação e busca, embora alguns autores considerem as estruturas de representação da informação como um quinto sistema. Contudo, para os propósitos desta pesquisa optou-se por esmiuçar apenas os aspectos relativos

aos sistemas de navegação e busca, visto que uma abordagem holística demandaria um estudo mais extenso. Portanto, delimitou-se a abordagem em torno da encontrabilidade da informação em repositórios de dados sob o ponto de vista da navegação e da busca, as quais são tidas como os principais meios pelos quais os usuários recuperam a informação em ambientes digitais.

Em vista disso, este artigo tem como objetivo geral avaliar os sistemas de navegação e busca de repositórios de dados científicos disciplinares sob o ponto de vista da Arquitetura da Informação, analisando em que medida esses sistemas podem influenciar a encontrabilidade dos dados nesses repositórios. Para isso, buscou-se estabelecer a relação entre Arquitetura da Informação e encontrabilidade da informação em repositórios de dados científicos disciplinares e averiguar em que medida a Arquitetura da Informação influencia a recuperação da informação em repositórios de dados científicos.

## **2 DADOS CIENTÍFICOS**

Os dados gerados ou coletados no âmbito da pesquisa acadêmica constituem o ponto de partida para a geração do conhecimento científico desde os primórdios da ciência. Hipóteses, análises, inferências e interpretações tornam-se possíveis a partir da apreensão de dados extraídos da realidade, os quais permitem compreender os fenômenos naturais, humanos e sociais.

Dados científicos podem ser classificados em pelo menos três tipos, conforme sua origem: dados observacionais, coletados em contextos temporais e espaciais específicos, a partir da observação da realidade imediata; dados computacionais, gerados na execução de modelos ou em simulações de computador; e dados experimentais, coletados em ambientes controlados de laboratório (NATIONAL SCIENCE BOARD, 2005).

De acordo com Sayão e Sales (2014), os dados de pesquisa passaram a ter seu potencial reconhecido a partir do momento em que deixaram de ser vistos como meros subprodutos da investigação científica, armazenados em mídias pessoais sujeitas à obsolescência tecnológica e de formatos, além de vulneráveis ao esquecimento, ao descarte ou à destruição com o tempo.

Entretanto, o compartilhamento de dados entre pesquisadores não é uma prática recente. A pesquisa colaborativa possui raízes antes mesmo do surgimento das tecnologias contemporâneas, as quais forneceram ferramentas mais aptas para arquivar e disseminar dos dados de pesquisa. Nesse sentido, os repositórios assumem uma importante função no que diz respeito tanto à custódia quanto à democratização do acesso a esses valiosos recursos digitais.

### **3 REPOSITÓRIOS DE DADOS CIENTÍFICOS**

Ao longo da história da humanidade e do conhecimento por ela construído, bibliotecas, arquivos e museus custodiaram coleções físicas de documentos e objetos que representavam a memória intelectual e cultural da humanidade. Com a evolução das tecnologias da informação e comunicação, surgiram novos meios de preservar e disseminar a produção intelectual humana, e dentre essas ferramentas destacam-se os repositórios digitais.

Existem diferentes tipos de repositórios digitais, que atendem aos objetivos próprios de sua natureza e, portanto, veiculam diferentes tipos de conteúdo. Os repositórios de dados, um dos tipos de repositório digital, são conceitualizados por Sanchez, Vechiato e Vidotti (2019, p. 52), como “ambientes informacionais digitais que buscam armazenar, organizar, representar, prover acesso, disseminar e preservar dados oriundos de pesquisas científicas”.

Os repositórios de dados científicos podem ser classificados da seguinte forma: 1) Repositórios de dados institucionais: criados com o fim de armazenar e disponibilizar conjuntos de dados produzidos por pesquisadores de uma instituição acadêmico-científica; 2) Repositórios de dados disciplinares: voltados para um domínio específico do conhecimento, tendem a oferecer estruturas mais especializadas para gerenciar os conjuntos de dados; e 3) Repositórios de dados multidisciplinares: reúnem dados provenientes de pesquisas de diversos ramos do conhecimento (RIBEIRO; VIDOTTI, 2009; SAYÃO; SALES, 2016).

Todavia, essa classificação não é inflexível nem mutuamente excludente, uma vez que repositórios institucionais costumam ser também multidisciplinares e alguns repositórios disciplinares podem também estar vinculados a uma determinada instituição de ensino ou pesquisa. Sejam quais forem os objetivos

que se propõem a cumprir, os repositórios de dados possibilitam a gestão, a representação, a disseminação e o acesso aos conjuntos de dados que incorporam em suas coleções (MONTEIRO, 2017).

Os repositórios de dados científicos fazem parte da ciberinfraestrutura de gestão de dados, que pode ser definida como a infraestrutura tecnológica subjacente à preservação, curadoria e disseminação do grande volume de dados de pesquisa que circula no ciberespaço. Assim, os repositórios de dados fornecem uma base tecnológica para a disseminação, preservação e contextualização de dados científicos, provendo serviços de valor agregado e mecanismos de busca e navegação que propiciam o acesso e uso desses recursos (SAYÃO; SALES, 2016; PAVÃO; ROCHA; GABRIEL JUNIOR, 2018).

Oferecer metadados que possibilitam a recuperação e a contextualização dos conjuntos de dados arquivados, atribuir identificadores persistentes que contribuem para assegurar o acesso unívoco aos conjuntos de dados a longo prazo, e proporcionar mecanismos de busca e estruturas de organização da informação que conduzem à descoberta dos conjuntos de dados no contexto digital, são tarefas indispensáveis aos repositórios de dados de qualidade. Além disso, é fundamental que esses repositórios sigam princípios de Arquitetura da Informação para garantir que a comunidade possa acessar os dados de pesquisa com o mínimo de dificuldades relativas à busca e à descoberta de dados.

## **4 ARQUITETURA DA INFORMAÇÃO**

A Arquitetura da Informação (AI) consiste em um campo teórico e prático que encontra na Ciência da Informação um alicerce epistemológico que lhe proporciona pertencimento a uma disciplina no contexto da pesquisa acadêmica, ainda que receba contribuições advindas de outros ramos do conhecimento, como a Ciência da Computação, o Design e até mesmo da Arquitetura.

### **4.1 DEFINIÇÕES E FUNÇÕES DA ARQUITETURA DA INFORMAÇÃO**

A expressão “Arquitetura da Informação” foi oficialmente cunhada em 1976 por Richard Saul Wurman, arquiteto de formação e entusiasta do

tratamento da informação sob uma perspectiva híbrida entre ciência e arte (ALBUQUERQUE; LIMA-MARQUES, 2011). Todavia, o termo “*architecture*” já era empregado por volta de 1959 pela empresa IBM para se referir à arquitetura de computadores, além de que em 1970 a empresa Xerox também já utilizava a expressão “*architecture of information*” para se referir aos seus estudos focados na interface gráfica e na interação humano-computador (RONDA LEÓN, 2008).

A AI pode ser definida em quatro concepções básicas:

1. O design estrutural de ambientes de informações compartilhadas.
2. A combinação de sistemas de organização, rotulação, navegação e busca em sites e intranets.
3. A arte e a ciência de moldar produtos e experiências de informação para dar suporte à usabilidade e à encontrabilidade.
4. Uma disciplina emergente e comunidade de prática focada em trazer princípios de design e arquitetura para o cenário digital (ROSENFELD; MORVILLE; ARANGO, 2007, p. 24, tradução nossa).

Resmini e Rosati (2012), por sua vez, compreendem a AI como uma arte aplicada, embora a definam também como um campo de estudos cujo foco reside na solução de problemas referentes ao acesso e uso do grande volume de informação corrente no meio digital.

Na visão de Dillon e Turnbull (2009), a AI abarca quatro tarefas principais: moldar a organização e o acesso à informação; construir associações abstratas entre unidades de conteúdo; desenvolver mecanismos de navegação e busca; e projetar interfaces e técnicas de interação que permitam aos usuários de um ambiente informacional acessar um conteúdo de interesse. Macedo (2005, p. 136) define “ambiente informacional” como um espaço de integração entre três elementos: um **contexto**, constituído por um fluxo de **conteúdos**, que por sua vez são direcionados a uma **comunidade** de usuários, por essa razão constituindo objeto de estudo da AI. Em síntese, os fenômenos de interesse da AI mostram-se atrelados ao desenho de ambientes informacionais.

Assim como o presente estudo, Ribeiro e Vidotti (2009) também discutem a aplicabilidade da Arquitetura da Informação em repositórios digitais, revelando que a estrutura dos repositórios costuma ser diferente do design dos *websites* comuns, pois os repositórios normalmente são desenhados a partir de um software que oferece um arranjo pré-definido, embora costumem permitir algum grau de customização. De qualquer maneira, os repositórios digitais devem ser

estruturados de modo a proporcionar uma navegação intuitiva e uma recuperação da informação facilitada ao conteúdo que disponibilizam.

## **4.2 OS COMPONENTES DA ARQUITETURA DA INFORMAÇÃO**

Considera-se que a Arquitetura da Informação é constituída por quatro componentes interdependentes que, juntos, permitem o pleno funcionamento de espaços informacionais digitais. Esses componentes são denominados de sistemas de organização, rotulação, navegação e busca. Além destes, há também os esquemas de representação da informação subjacentes (ex.: metadados, vocabulários controlados, tesauros).

O sistema de organização remete aos critérios de como as informações são categorizadas, se por assunto, ordem alfabética ou cronológica. O sistema de rotulação concerne à forma como a informação é representada, que pode ser por meio de rótulos textuais ou icônicos. O sistema de navegação fornece ao usuário os meios de se locomover ao longo do ambiente digital. O sistema de busca, por sua vez, permite a pesquisa direta a partir da aplicação de estratégias de busca e diz respeito também ao modo como a informação é indexada e como ela pode ser pesquisada (ROSENFELD; MORVILLE; ARANGO, 2015).

De acordo com Barker (2005), a navegação intuitiva não acontece por acaso. Isso significa que é necessário ter em vista todo um conjunto de critérios de natureza empírica e teórica que sejam centrados nos usuários potenciais e reais do ambiente informacional e nos objetivos que este pretende alcançar.

Krug (2008) considera que a navegação possui os seguintes objetivos: auxiliar a encontrar as informações desejadas; informar onde o usuário está localizado no ambiente informacional; evitar que o usuário se perca; revelar o conteúdo existente; instruir implicitamente como usar o sítio; e conferir credibilidade à equipe responsável pela construção e manutenção do ambiente digital, aumentando-se a probabilidade de que os usuários retornem a ele.

O sistema de navegação é composto por três subsistemas principais: a navegação global, geralmente situada horizontalmente em uma área de destaque e fácil visualização (como no topo das páginas, na forma de menu), de modo a oferecer uma noção geral de todo o conteúdo disponível no ambiente; a

navegação local, frequentemente disposta verticalmente nas páginas dos *websites*, provendo acesso às suas subseções; e a navegação contextual, composta por uma teia de hiperlinks que provê acesso ao conteúdo de outras páginas. Muitos autores consideram ainda a existência de um quarto subsistema, a navegação suplementar, que inclui recursos como mapa do site, índice e outros guias (AGNER, 2009; ROSENFELD; MORVILLE; ARANGO, 2015).

Para Krug (2008), os usuários não leem integralmente as páginas pelas quais navegam, apenas passam rapidamente os olhos por seu conteúdo, pois desejam economizar tempo em suas tarefas e apenas uma parcela do conteúdo informacional disponível é efetivamente relevante para seus interesses.

O sistema de busca, por sua vez, tem como objetivo central oferecer um acesso mais direto às informações, principalmente para os usuários que dispõem de pouco tempo ou paciência para explorar os recursos de navegação. Segundo Rosenfeld, Morville e Arango (2015), o sistema de busca oferece o mecanismo de pesquisa como um meio para recuperação da informação, além de englobar aspectos referentes à interface e à apresentação dos resultados da pesquisa. A presença de um sistema de busca implica a indexação do conteúdo que se pretende que seja recuperável. Contudo, as formas de buscar a informação podem ser muito variadas e são fortemente influenciadas tanto pelas intenções quanto pelas condições sociocognitivas e emocionais do usuário.

Para Agner (2009), sistemas de busca são aplicações de *software* que permitem o cruzamento entre as estratégias de busca expressas pelos usuários e as informações efetivamente indexadas, o que pode incluir documentos completos ou apenas metadados básicos, como autor, título e assunto. As estratégias de busca podem ocorrer a partir de linguagem natural, linguagem controlada e com o emprego de operadores *booleanos*. Quanto à apresentação dos resultados, Nonato *et al.* (2008) apontam que as informações recuperadas podem ser ordenadas por diferentes critérios, como relevância, ordem alfabética e data de publicação. Em todo caso, ambos os sistemas têm como função em comum garantir a encontrabilidade dos objetos digitais de interesse do usuário.

### 4.3 A INTERSECÇÃO ENTRE ARQUITETURA DA INFORMAÇÃO E O CONCEITO DE ENCONTRABILIDADE

O emprego do termo “encontrabilidade” é uma tradução literal da palavra inglesa “*findability*” e tem sido empregada para designar a capacidade de um objeto digital ser localizável em ambientes informacionais. Morville (2005) define encontrabilidade como a qualidade de ser localizável ou recuperável e a facilidade com que um objeto é descoberto ou localizado. Vechiato e Vidotti (2014, p. 110) acrescentam que a encontrabilidade “ocorre a partir da busca prévia de informação por meio da navegação ou de estratégias de pesquisa em um mecanismo de busca”.

De acordo com Sanchez, Vechiato e Vidotti (2019), a noção de encontrabilidade está diretamente associada à AI, principalmente no que tange à recuperação da informação e à navegação em ambientes digitais. Nesse contexto, Alvarez *et al.* (2016) pontuam nem sempre os mecanismos de busca e esquemas de navegação são os principais determinantes da encontrabilidade, a qual também irá depender do estado psicossocial do usuário no momento da navegação ou busca. Assim, sob uma perspectiva cognitiva, o indivíduo nem sempre tem consciência de todas as suas necessidades informacionais, o que pode representar uma barreira à localização de dados e documentos.

Em uma abordagem voltada para repositórios de dados científicos, Sanchez e Vechiato (2019) classificam os atributos da encontrabilidade da informação de acordo com cada um dos quatro sistemas da AI. Para os propósitos deste trabalho, serão abordados apenas os atributos referentes aos sistemas de navegação e busca, os quais são expostos no Quadro 1.

**Quadro 1 – Aplicação dos atributos da Encontrabilidade da Informação aos repositórios de dados científicos**

ATRIBUTOS E DEFINIÇÕES	INSTRUMENTOS/TECNOLOGIAS/NORMAS
<b>SISTEMAS DE NAVEGAÇÃO</b>	
<b><i>Affordances</i></b> Elemento que compõe a interface de um ambiente, fornecendo pistas que incentivam as possíveis ações naquele espaço. Orienta, localiza, oferece acesso	1. Mudança do cursor do mouse de “seta” para mão, quando indicar link; 2. Links em cores diferentes no texto, pistas metafóricas, como o uso de imagens do cotidiano para remeter a ação no repositório; 3. Mudança de cor quando o cursor do mouse passa pelo menu, indicando

e propicia encontrabilidade e descobertas de dados.	menu <i>dropdown</i> ; <b>4.</b> Utilizar cores fortes que preencham uma caixa de conteúdo ou botão de ação, para que chame a atenção do usuário.
<b><u>Wayfindings</u></b> Compõe a interface auxiliando na orientação espacial do ambiente durante a navegação.	<b>1.</b> A orientação espacial dentro do ambiente deve ser clara e simples, ou seja, compreender que ao clicar na logo da instituição ou nome do repositório ele retornará a <i>home</i> do ambiente; <b>2.</b> Inserir a descrição da trilha de navegação na parte superior da interface como links; <b>3.</b> Mapas do site à disposição, sinalizações com ícones para completar ações e/ou navegar.
<b><u>Mobilidade, convergência e ubiquidade</u></b> Direcionado ao acesso e uso dos ambientes digitais sem barreiras, ou seja, acesso por diferentes dispositivos. Conceito associado aos espaços externos e não apenas no contexto digital.	<b>1.</b> Utilizar de recursos que permitam que o repositório seja responsivo para acesso em dispositivos móveis. Criar aplicações destinadas ao acesso exclusivo por dispositivos móveis, ampliando as funcionalidades e visibilidade do repositório.
<b>SISTEMAS DE BUSCA</b>	
<b><u>Descoberta de dados</u></b> Facilidade da interface em propiciar a encontrabilidade (tanto das necessidades explícitas quanto implícitas do sujeito informacional) através da navegação ou mecanismo de busca.	<b>1.</b> Inserção de recursos que potencializem o mecanismo de busca, como <i>autosuggest</i> , <i>autocomplete</i> e corretor ortográfico; <b>2.</b> Possibilitar a descoberta de dados por meio da busca por palavras-chaves; <b>3.</b> Enriquecer os mecanismos com: linguagem natural, lógica <i>booleana</i> e adição de filtros que possibilitem o afinilamento da pesquisa.

**Fonte:** Adaptado de Sanchez e Vechiato (2019).

Conforme exposto no Quadro 1, pode-se considerar pelo menos quatro atributos da encontrabilidade em repositórios de dados, sendo três referentes aos sistemas de navegação e um relativo aos sistemas de busca.

## 5 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa caracteriza-se metodologicamente como qualitativa quanto à sua abordagem e descritiva quanto aos seus objetivos. A pesquisa qualitativa tem como objetivo entender e atribuir sentido a casos particulares colhidos da realidade, sendo caracterizada pela reflexividade por parte do pesquisador na compreensão desses casos. A pesquisa descritiva, por sua vez, busca fazer observações minuciosas e descrições detalhadas sobre as características de um fenômeno de interesse (SANDELOWSKI, 2004).

Para a seleção da amostra de repositórios de dados, utilizou-se o *Registry of Research Data Repositories*, (Re3data), um diretório que elenca mais de 2.500

repositórios de dados institucionais e disciplinares de todas as áreas do saber e de todas as regiões do mundo. Desse modo, utilizaram-se filtros como “*Database access: open*”, “*Data access: open*” e “*Repository types: disciplinary*” para selecionar repositórios disciplinares de acesso livre aos dados.

A partir de uma breve análise de um conjunto de repositórios, selecionou-se a amostra pelo critério da intencionalidade, de modo a ressaltar características da AI nos repositórios de dados que apresentassem o maior número de elementos para análise. Conforme aponta Sandelowski (2004), a pesquisa qualitativa é uma forma de investigação empírica que frequentemente envolve uma amostragem intencional de casos ricos em informações.

Assim, foram selecionados quatro repositórios de dados de diferentes áreas do conhecimento, realizando-se, portanto, uma amostragem heterogênea. Foram selecionados para estudo os seguintes repositórios de dados:

- PANGAEA<sup>4</sup> – propõe arquivar e disseminar dados georreferenciados de pesquisas nas áreas de ciências ambientais e da terra;
- *Open Language Archives Community* (OLAC)<sup>5</sup> – tem como missão o arquivamento digital de dados linguísticos de mais de 7.000 línguas vivas e mortas de todos os continentes;
- *Biodiversity.aq*<sup>6</sup>. – fornece acesso a mais de 600 conjuntos de dados de pesquisa relacionados à biodiversidade da Antártida;
- *TalkBank*<sup>7</sup> – tem como finalidade contribuir com os estudos da comunicação humana, com ênfase na comunicação oral, reunindo e disponibilizando dados sonoros e textuais de dezenas de idiomas.

Para a definição dos critérios a serem aplicados na análise dos repositórios, adaptou-se um *checklist* elaborado por Vechiato, Oliveira e Vidotti (2016) para a avaliação de ambientes informacionais híbridos. Por este estudo focalizar os sistemas de navegação e busca de ambientes digitais, o *checklist* foi simplificado, acrescentando-se ao original outros itens com base na literatura.

---

<sup>4</sup> Página inicial do PANGAEA: <https://www.pangaea.de/>

<sup>5</sup> Página inicial do OLAC: <http://www.language-archives.org/>

<sup>6</sup> Página inicial do Biodiversity.aq: <https://data.biodiversity.aq/>

<sup>7</sup> Página inicial do TalkBank: <https://www.talkbank.org/>

Os *checklists*, de acordo com Downey e Banerjee (2010), são um mecanismo para evidenciar questões ou tópicos específicos, sendo uma lista de itens a serem verificados em um estudo, os quais contribuem para a discussão dos resultados. O Quadro 2 representa o instrumento aplicado neste estudo para avaliar os sistemas de navegação e busca dos repositórios de dados.

**Quadro 2 – Checklist para avaliação dos repositórios de dados científicos**

ATRIBUTO	ITENS DE VERIFICAÇÃO
SISTEMAS DE NAVEGAÇÃO	
<b>Navegação global</b>	Presente em todas as páginas do repositório?
	A barra de navegação global fornece link para a página inicial?
<b>Navegação local</b>	Permite aos usuários explorar a área imediata de forma eficaz?
<b>Navegação contextual</b>	Evita a desorientação cognitiva do usuário?
	Facilita a encontrabilidade de dados relevantes?
<b>Navegação suplementar</b>	O repositório apresenta recursos de navegação suplementares, como mapa do site, índice ou guia?
<b>Affordances</b>	Facilitam o entendimento por diferentes tipos de usuários?
<b>Wayfindings</b>	Há marcos e/ou metáforas que dão pistas ao sujeito para orientá-lo no espaço digital?
<b>Responsividade</b>	Possui interface responsiva?
	Permite a continuidade das ações dos sujeitos informacionais entre os diferentes dispositivos?
SISTEMAS DE BUSCA	
<b>Descoberta de informações</b>	O repositório possui uma barra de pesquisa visível e fácil de localizar?
	Há opção de pesquisa avançada com possibilidade de refinamento por filtros?
	O mecanismo de busca utiliza recurso <i>autocomplete</i> ou autossugestão?
	Os resultados de busca fornecem facetas para o refinamento da pesquisa?
	Os resultados de busca apresentam os documentos com base na estratégia de busca inicial, ordenando-os por critérios lógicos?

Fonte: Adaptado de Vechiato, Oliveira e Vidotti (2016).

A primeira coluna do Quadro 2 refere-se aos atributos da encontrabilidade da informação, ao passo que a segunda coluna relaciona os itens a serem verificados nos repositórios de dados examinados.

## 6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

O Quadro 3 exhibe o resultado da avaliação dos repositórios, cujo sistema de verificação aplicado considerou três tipos de resposta: atende plenamente (✓✓); atende parcialmente (✓); e não atende (X).

**Quadro 3 – Resultado da avaliação dos repositórios de dados científicos**

ITENS DE VERIFICAÇÃO	REPOSITÓRIO			
	PANGAEA	OLAC	BIODIV.	TALKBANK
<b>SISTEMA DE NAVEGAÇÃO</b>				
Presente em todas as páginas do repositório?	✓✓	X	✓✓	X
A barra de navegação global fornece link para a página inicial?	✓✓	✓✓	✓✓	X
Permite aos usuários explorar a área imediata de forma eficaz?	✓✓	X	X	✓
Evita a desorientação cognitiva do usuário?	✓✓	✓	✓✓	X
Facilita a encontrabilidade de dados relevantes?	✓✓	✓	✓✓	✓
O repositório apresenta recursos de navegação suplementares, como mapa do site, índice ou guia?	X	X	X	X
Facilitam o entendimento por diferentes tipos de usuários?	✓✓	✓	✓	✓
Há marcos e/ou metáforas que dão pistas ao sujeito para orientá-lo no espaço digital?	✓	X	✓	X
Possui interface responsiva?	✓✓	✓	✓✓	✓
Permite a continuidade das ações dos sujeitos informacionais entre os diferentes dispositivos?	✓✓	X	✓✓	✓
<b>SISTEMA DE BUSCA</b>				
O repositório possui uma barra de pesquisa visível e fácil de localizar?	✓✓	✓✓	✓✓	X
Há opção de pesquisa avançada com possibilidade de refinamento por filtros?	✓✓	✓	✓	✓
O mecanismo de busca utiliza recurso <i>autocomplete</i> ou autossugestão?	✓✓	X	✓	X
Os resultados de busca fornecem facetas para o refinamento da pesquisa?	✓✓	✓✓	✓✓	X
Os resultados de busca apresentam os documentos com base na estratégia de busca inicial, ordenando-os por critérios lógicos?	✓	X	X	✓

**Fonte:** Dados da pesquisa.

Conforme exposto no Quadro 3, os repositórios PANGAEA e Biodiversity.aq demonstraram um elevado grau de aderência aos atributos da encontrabilidade em relação aos seus sistemas de navegação e busca, ao passo que nos repositórios OLAC e TalkBank não foram identificados elementos correspondentes à maior parte dos atributos em questão.

## 6.1 PANGAEA

Dentre os repositórios avaliados, o PANGAEA destacou-se por atender a quase todos os requisitos de forma absoluta, não interpondo dificuldades de navegação ou busca. Esse achado converge com o fato de este repositório ser

certificado por instâncias como o *CoreTrustSeal*, organização que confere confiabilidade e um elevado padrão de qualidade aos repositórios digitais submetidos ao seu processo de certificação. Além dos sistemas de navegação global, local e contextual bem estruturados, o repositório apresenta *affordances* que favorecem uma navegação fluida e intuitiva, mudando o cursor de “seta” para “mão” para indicar links e evidenciando os botões clicáveis (Figura 1).

Figura 1 – Sistemas de navegação e busca do repositório PANGAEA



Fonte: Repositório PANGAEA.

Observou-se que os resultados da busca são ordenados exclusivamente por “score”, uma pontuação indicativa de relevância, não permitindo ordenação por outros critérios que poderiam ser úteis, como data de publicação. Em contrapartida, existe um mecanismo de refinamento que permite a aplicação de filtro por diversos critérios, como autor, data, assunto, local, método de coleta de dados, entre outros. Além disso, o PANGAEA atendeu plenamente ao quesito responsividade, visto que adequa sua interface aos diferentes tamanhos de tela e possui um design específico para dispositivos móveis.

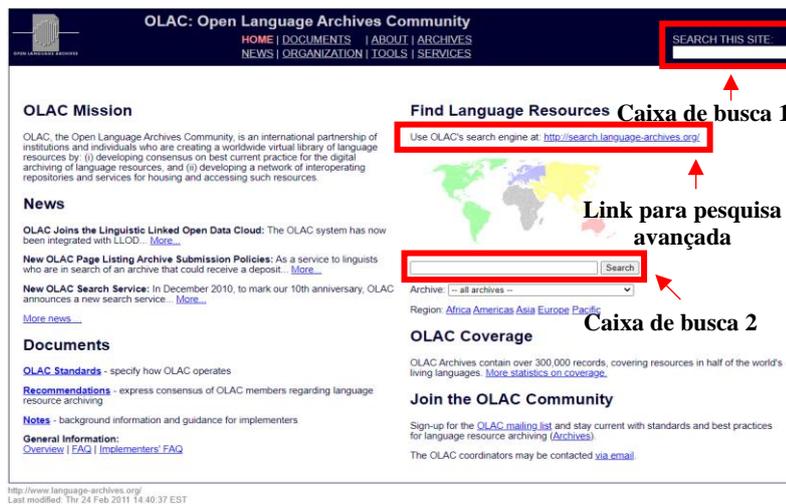
## 6.2 OLAC

O repositório OLAC abarca uma coleção de registros de metadados relativos a recursos linguísticos de centenas de idiomas. No entanto, apesar de sua rica coleção, a plataforma enfrenta numerosos fatores que dificultam a encontrabilidade dos dados. O repositório não exibe um sistema de navegação global em todas as suas páginas, sendo um recurso restrito à sua página inicial. Além disso, o sistema de busca não proporciona uma encontrabilidade efetiva, conduzindo a páginas que não revelam o que o usuário esperaria encontrar.

A respeito das *affordances*, o OLAC possui qualidades similares às aquelas apresentadas pelo PANGAEA, no sentido de indicar com clareza os botões de ação e os links clicáveis. Por outro lado, falha quanto aos *wayfindings*, não exibindo indicações da trajetória percorrida pelo usuário ao longo do repositório, o que força a utilização excessiva da opção “voltar” do *browser* para qualquer ação que se deseje realizar, ou clicar no logo do repositório que remete diretamente à página inicial. Essas condicionantes podem tornar a navegação cansativa e impedir que o usuário explore o repositório de forma plena, como ocorreria em um ambiente estruturado de forma mais centrada no usuário.

No que tange à responsividade, o repositório permite ajuste proporcional ao tamanho da tela. Todavia, não apresenta interface customizada para dispositivos móveis. Na página inicial (Figura 2), há duas caixas de busca: a caixa 1 remete a resultados no Google e a caixa 2 conduz a uma página que não apresenta resultados, mas sim a uma espécie de navegação secundária em que se deve escolher a região do mundo cujas línguas se deseja explorar.

Figura 2 – Página inicial do repositório OLAC



Fonte: Repositório OLAC.

Selecionando-se a opção “Américas”, por exemplo, o usuário é conduzido a uma página que lista os países do continente e, entre parênteses, a quantidade de registros que serão encontrados a respeito de cada país. Constatou-se que o OLAC propicia uma encontrabilidade muito mais centrada na navegação do que na busca, posto que aquilo deveria ser um procedimento de busca acaba se assemelhando mais a uma navegação contextual. Inusitadamente, o OLAC apresenta duas interfaces diferentes, no intuito de oferecer um sistema de busca mais eficaz. A Figura 3 ilustra essa segunda interface, que possibilita a pesquisa avançada, cujos resultados não parecem ser ordenados de uma maneira lógica, mas oferecem várias opções de filtros.

Figura 3 – Pesquisa avançada do OLAC



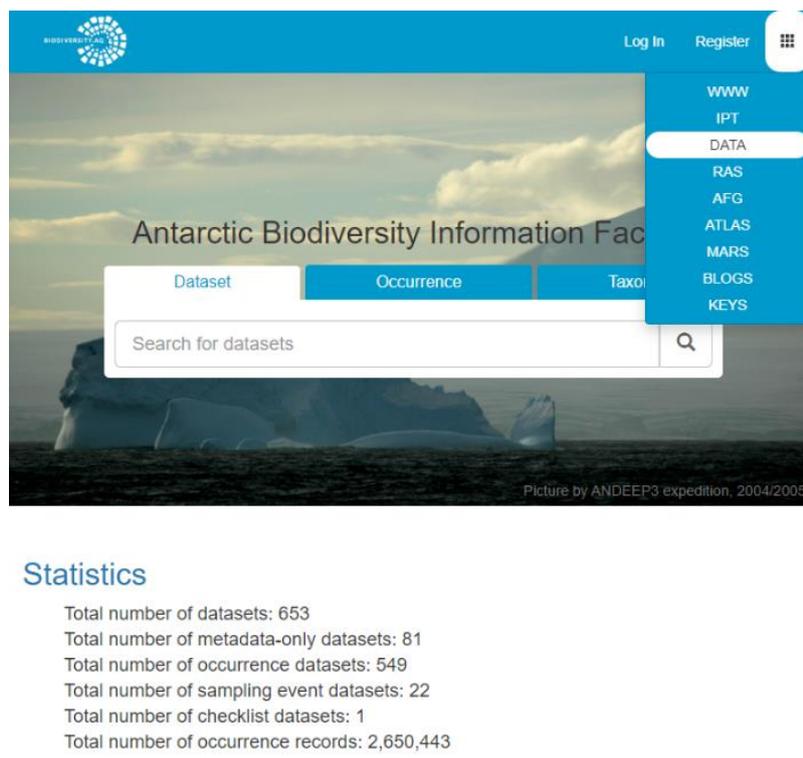
Fonte: Repositório OLAC.

Em última análise, o OLAC parece exibir mais características de um arquivo de registros de metadados do que de um repositório que visa à disseminação e ao reuso dos dados.

## 6.2 BIODIVERSITY.AQ

O repositório Biodiversity.aq apresenta uma interface minimalista e um sistema de busca evidente ao usuário logo em sua *homepage*. O sistema de navegação global não segue os moldes tradicionais, resumindo-se apenas ao logotipo do repositório, às opções de *login* e registro e a um menu *dropdown*, como mostra a Figura 4.

**Figura 4 – Página inicial do repositório Biodiversity.aq**



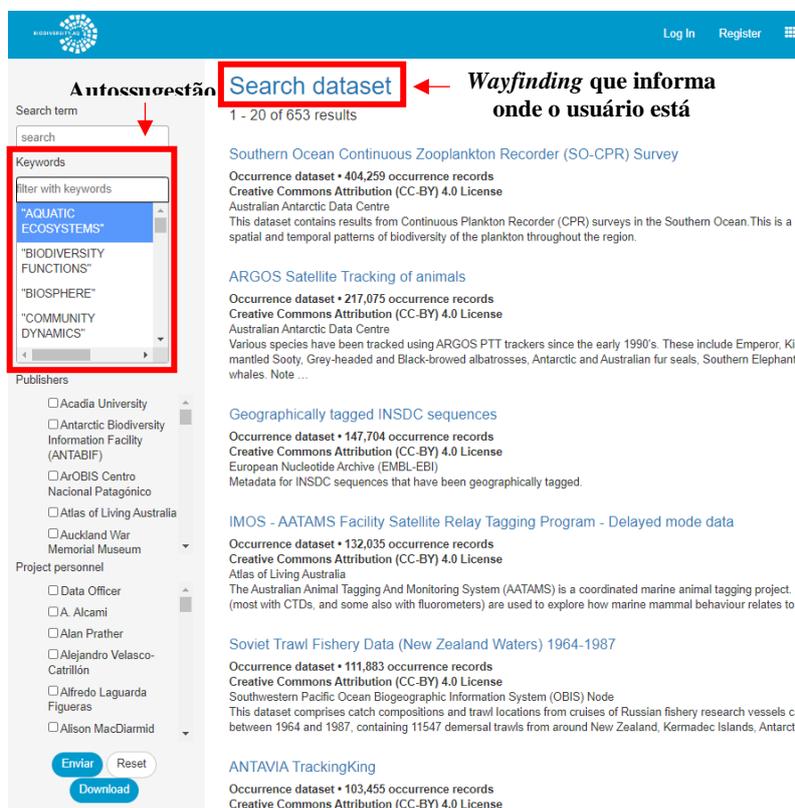
**Fonte:** Repositório Biodiversity.aq.

Como é possível observar na Figura 4, o sistema de busca permite a pesquisa por conjuntos de dados, por ocorrência (coordenadas geográficas) e por taxonomia. O repositório apresenta *affordances* e *wayfindings* que proporcionam uma orientação cognitiva ao longo da navegação e que deixam

inequívoco o que é link, o que é botão e o que é apenas texto. Assim como o repositório PANGAEA, o Biodiversity.aq possui uma interface responsiva que assegura a mobilidade entre dispositivos, visto que disponibiliza uma interface personalizada para dispositivos móveis.

Em relação ao sistema de busca, os mecanismos de refinamento da pesquisa e de autossugestão de palavras-chave cumprem a função de facilitar a encontrabilidade dos conjuntos de dados que alguns usuários procuram e não conseguem obter usando estratégias de busca com linguagem natural. Contudo, não foi possível perceber com clareza o critério estabelecido pelo repositório para ordenar os resultados de busca. A Figura 5 ilustra essas características.

**Figura 5 – Sistema de busca do Biodiversity.aq**



Fonte: Repositório Biodiversity.aq.

Em consonância com a avaliação relatada no Quadro 3, o Biodiversity.aq demonstrou possuir sistemas de navegação e busca com forte presença dos atributos da encontrabilidade referente aos sistemas de navegação e busca. Em vista disso, pode-se inferir que a busca por conjuntos de dados nesse repositório tende a ser bem-sucedida, do ponto de vista do usuário.

## 6.4 TALKBANK

O TalkBank compreende uma vasta coleção de gravações de áudio e vídeo de mais de cinquenta idiomas, vinculadas a transcrições detalhadamente anotadas. O repositório possui um sistema de navegação pouco convencional, de modo que se torna difícil distinguir navegação global, local e contextual. A plataforma reflete uma arquitetura da informação idiossincrásica, apresentando fragilidades não apenas quanto à navegação e busca, mas também quanto ao sistema de organização da informação (Figura 6).

Figura 6 – Página inicial do repositório TalkBank

The screenshot shows the TalkBank homepage layout. At the top, there is a 'TalkBank' logo, a building icon with 'TALK' above it, and 'The TalkBank System' text. Below this is a paragraph of introductory text. The main content is organized into a grid of categories, each with a list of links. The categories are: System, Programs, Manuals, Conversation Banks, Child Language Banks, Multilingualism Banks, Clinical Banks, Other, Recording, Resources, and Symposia. A red box highlights the link 'TalkBank DB - Search the Databases' in the 'Other' category, with a red arrow pointing to it from the text below.

System	Programs	Manuals
<a href="#">**Ground Rules**</a>	<a href="#">CLAN</a>	<a href="#">CHAT - CLAN - MOR</a>
<a href="#">**Hints on Downloading**</a>	<a href="#">MOR grammars</a>	<a href="#">Tutorial Screencasts</a>
<a href="#">Contributing</a>	<a href="#">XML creator</a> and <a href="#">XML Schema</a>	<a href="#">SLP's Guide to CLAN</a> and <a href="#">P&amp;X</a>
<a href="#">IRB Principles</a>	<a href="#">Other Software</a>	

Conversation Banks	Child Language Banks	Multilingualism Banks
<a href="#">CABank</a>	<a href="#">CHILDES</a>	<a href="#">Second Language Tutors</a>
<a href="#">SamtaleBank</a>	<a href="#">PhonBank</a>	<a href="#">BilingBank</a>
<a href="#">ClassBank</a>	<a href="#">HomeBank</a>	<a href="#">SLABank</a>

Clinical Banks	Clinical Banks	Other
<a href="#">DementiaBank</a>	<a href="#">AphasiaBank</a>	<a href="#">Database Versioning</a>
<a href="#">RHDBank</a>	<a href="#">ASDBank</a>	<a href="#">TalkBank DB - Search the Databases</a>
<a href="#">TIBank</a>	<a href="#">FluencyBank</a>	

Recording	Resources	Symposia
<a href="#">Digital Video</a>	<a href="#">CLARIN Knowledge Center</a>	<a href="#">CMU June 5-8</a>
<a href="#">Digital Audio</a>	<a href="#">GoogleGroups</a>	
	<a href="#">Picture Stimuli</a>	

**Link que direciona para uma interface de busca**

Fonte: Repositório TalkBank.

Como destaca a Figura 6, o TalkBank disponibiliza um link que remete a uma página em que é possível realizar uma busca, embora não da maneira convencional. A pesquisa não permite a busca livre por palavras em linguagem natural, mas sim uma lista de opções que podem ser selecionadas. Essa é realizada por facetas, selecionando-se os aspectos de interesse e executando-se a busca. Os resultados são ordenados de forma demasiadamente técnica,

não sendo possível ter de imediato uma ideia clara sobre o que se trata cada conjunto de dados, ao contrário dos repositórios PANGAEA e Biodiversity.aq. Quanto ao quesito responsividade, o TalkBank atende de forma razoável, visto que possibilita ajuste automático de tela no computador, mas não adequa sua interface aos dispositivos móveis, com exceção da interface de busca.

De acordo com Krug (2008), o usuário médio deseja economizar tempo e geralmente não está disposto a descobrir como as coisas funcionam de forma detalhada, mas sim encontrar o que deseja em tempo hábil. O repositório do TalkBank, apesar da riqueza de dados linguísticos que apresenta, não aparenta ter entre seus objetivos o foco no usuário. Apesar da riqueza de dados, o tempo e a energia despendidos para encontrá-los podem frustrar a experiência do usuário. Dentre os quatro repositórios estudados, apresentou o maior número de fragilidades em relação à navegabilidade e à encontrabilidade dos dados.

## **7 CONSIDERAÇÕES FINAIS**

Nesta pesquisa buscou-se adquirir percepções acerca da estruturação de repositórios de dados científicos, considerando aspectos da Arquitetura da Informação que podem influenciar de forma significativa a encontrabilidade da informação em repositórios de dados. A partir de um olhar sobre os sistemas de navegação e de busca dos repositórios avaliados, constatou-se que muitos elementos ainda podem ser aperfeiçoados nesse tipo de plataforma.

Se os repositórios de dados têm como missão não apenas armazenar, mas também promover o acesso e o reuso dos dados, precisam planejar mais criteriosamente a construção de sistemas de navegação, busca e organização da informação adequados à experiência do usuário. Por outro lado, se já existem milhares de repositórios digitais de dados de pesquisa, significa que a cultura de gestão de dados científicos vem sendo cada vez mais consolidada. O crescimento do número de repositórios de dados sinaliza um aumento da preocupação com a preservação e a disseminação de conjuntos de dados produzidos nas mais diversas áreas do conhecimento, ainda que a estrutura dessas plataformas em muitos casos não corresponda à fisionomia e funcionalidade tidas como ideais aos olhos da comunidade.

Ao contrário dos repositórios de dados institucionais, que possuem maior tendência em apresentar uma estrutura mais padronizada devido às escolhas de *software* que as instituições de ensino e pesquisa geralmente partilham em comum, alguns repositórios de dados disciplinares parecem necessitar ainda de um redesenho estrutural que facilite a descoberta de dados pela comunidade-alvo. Observou-se, a partir da análise realizada neste trabalho, um certo grau de discrepância entre os princípios da Arquitetura da Informação difundidos na literatura e os sistemas de navegação e busca efetivamente observados nos repositórios de dados disciplinares examinados.

Os repositórios selecionados como amostra refletem tanto os maiores acertos quanto os maiores erros praticados pelos repositórios de dados científicos que têm sido criados na Web. Outra constatação proporcionada por esta pesquisa foi a de que muitos desses repositórios assemelham-se mais a bancos de dados do que a repositórios propriamente ditos, apesar de serem registrados como “repositórios” nos diretórios especializados.

Conclui-se, por fim, que a encontrabilidade da informação pode ser significativamente prejudicada caso os repositórios de dados não sigam padrões e princípios da Arquitetura da Informação comprovadamente convergentes com os esquemas mentais dos usuários de repositórios de dados e demais ambientes informacionais, os quais esperam localizar dados e informações de interesse no menor tempo possível e, preferencialmente, sem sobrecarga cognitiva.

## REFERÊNCIAS

AGNER, L. **Ergodesign e arquitetura de informação**: trabalhando com o usuário. 2. ed. Rio de Janeiro: Quartet, 2009.

ALBUQUERQUE, A. R. R.; LIMA-MARQUES, M. Sobre os fundamentos da Arquitetura da Informação. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 1, número especial, p. 60-72, out. 2011. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/pgc/article/view/10827>. Acesso em: 15 jul. 2020.

ALVAREZ, E. B.; SIRIANI, A. L. R.; VIDOTTI, S. A. B. G.; CARVALHO, A. M. G. Os Sistemas de Recomendação, Arquitetura da Informação e a

Encontrabilidade da Informação. **Transinformação**, Campinas, v. 28, n. 3, p. 275-286, set./dez. 2016.

BARKER, I. What is information architecture? **Step Two Designs**, 2005.  
Disponível em: [https://www.steptwo.com.au/files/kmc\\_whatinfoarch.pdf](https://www.steptwo.com.au/files/kmc_whatinfoarch.pdf).  
Acesso em: 11 jul. 2020.

CAVALCANTI, M. T.; SALES, L. F. Gestão de dados de pesquisa: um panorama da atuação da União Europeia. **Biblos**, Rio Grande, v. 31, n. 1, p. 73-98, jan./jun. 2017.

DILLON, A.; TURNBULL, D. Information Architecture. *In*: BATES, M. J.; MAACK, M. N. (ed.). **Encyclopedia of Library and Information Sciences**. 3. ed. Boca Raton: Taylor & Francis, 2009.

DOWNEY, L.; BANERJEE, S. Building an Information Architecture checklist: encouraging and enabling IA from infrastructure to the user interface architecture. **Journal of Information Architecture**, [S. l.], v. 2, n. 2, 25-42, 2010.

GUANDALINI, C. A.; FURNIVAL, A. C. M.; ARAKAKI, A. C. S. Boas práticas científicas na elaboração de planos de gestão de dados. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 17, p. 1-20, 2019.

KRUG, S. **Não me faça pensar!** Uma abordagem de bom senso à usabilidade na Web. 2. ed. Rio de Janeiro: Alta Books, 2008.

MACEDO, F. L. O. **Arquitetura da Informação**: aspectos epistemológicos, científicos e práticos. 2005. 190 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2005.

MEDEIROS, J. S.; CAREGNATO, S. E. Compartilhamento de dados e e-*Science*: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, p. 311-322, set. 2012.

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

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. Repositórios de dados científicos nas universidades brasileiras e portuguesas. *In*: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 6., 2016, Londrina. **Anais** [...] Londrina: UEL, 2016. p. 652-664.  
Disponível em:  
<http://www.uel.br/eventos/cinf/index.php/secin2016/secin2016/paper/viewFile/338/166>. Acesso em: 5 jul. 2020.

MORVILLE, P. **Ambient findability**. Sebastopol: O'Reilly, 2005.

NATIONAL SCIENCE BOARD. The elements of the digital data collections universe. *In*: NATIONAL SCIENCE BOARD. **Long-lived digital data collections: enabling research and education in the 21st century**. National Science Foundation, 2005. Disponível em: <https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>. Acesso em: 6 jul. 2020.

NONATO, R. S.; BORGES, G. S. B.; MACULAN, B. C.; LIMA, G. A. B. O. Arquitetura da Informação em bibliotecas digitais: uma abordagem da Ciência da Informação e da Biblioteconomia. **Informação & Informação**, Londrina, v. 13, n. 2, p. 125-141, jul./dez. 2008.

PAVÃO, C. G.; ROCHA, R. P.; GABRIEL JUNIOR, R. F. Proposta de criação de uma rede de dados abertos da pesquisa brasileira. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 16, n. 2, p. 329-343, maio./ago. 2018.

RESMINI, A.; ROSATI, L. A brief history of Information Architecture. **Journal of Information Architecture**, [S. l.], v. 3, n. 2, p. 33-45, 2012.

RIBEIRO, O. B.; VIDOTTI, S. A. B. G. Otimização do acesso à informação científica: discussão sobre a aplicação de elementos da Arquitetura da Informação em repositórios digitais. **Biblos**, Rio Grande, v. 23, n. 2, p. 105-116, jul./dez. 2009. Disponível em: <https://periodicos.furg.br/biblos/article/view/1309>. Acesso em: 15 jun. 2020.

RONDA LEÓN, R. Arquitectura de Información: análisis histórico-conceptual. **No Solo Usabilidad**, n. 7, 2008. Disponível em: [http://www.nosolousabilidad.com/articulos/historia\\_arquitectura\\_informacion.htm](http://www.nosolousabilidad.com/articulos/historia_arquitectura_informacion.htm). Acesso em: 15 jul. 2020.

ROSENFELD, L.; MORVILLE, P.; ARANGO, J. **Information Architecture: for the Web and beyond**. 4. ed. Sebastopol: O'Reilly, 2007.

SANCHEZ, F. A.; VECHIATO, F. L. Avaliação de repositórios de dados de pesquisa segundo critérios da Encontrabilidade da Informação. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 142-160, nov. 2019.

SANCHEZ, F. A.; VECHIATO, F. L.; VIDOTTI, S. A. B. G. Encontrabilidade da informação em repositórios de dados: uma análise do DataONE. **Informação & Informação**, Londrina, v. 23, n. 1, p. 51-79, jan./abr. 2019.

SANDELOWSKI, M. Qualitative research. *In*: LEWIS-BECK, M. S.; BRYMAN, A.; LIAO, T. F. (ed.). **The SAGE encyclopedia of social science research methods**. Thousand Oaks, California: SAGE, 2004. v. 1, p. 893-894.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90-115, maio/ago. 2016.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, Rio de Janeiro, v. 8, n. 2, p. 76-92, abr./jun. 2014.

VECHIATO, F. L.; OLIVEIRA, H. P. C.; VIDOTTI, S. A. B. G. Arquitetura da Informação Pervasiva e Encontrabilidade da Informação: instrumentos para avaliação de ambientes informacionais híbridos. **Informação & Tecnologia**, Marília, v. 3, n. 1, p. 47-65, jan./jun. 2016. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/itec/article/view/38365>. Acesso em: 5 jul. 2020.

VECHIATO, F. L.; VIDOTTI, S. A. B. G. **Encontrabilidade da informação**. São Paulo: Cultura Acadêmica, 2014.

## INFORMATION ARCHITECTURE IN DISCIPLINARY RESEARCH DATA REPOSITORIES: THE INFLUENCE OF SEARCH AND NAVIGATION SYSTEMS ON DATA FINDABILITY

### ABSTRACT

**Objective:** Aims to evaluate the navigation and search systems of disciplinary scientific data repositories from the point of view of Information Architecture, analyzing to what extent these systems can influence the information findability in data repositories. **Methodology:** Qualitative in terms of its approach and descriptive in terms of its objectives. Four data repositories from different fields of knowledge were selected from the re3data.org directory, composing a heterogeneous sample by the criterion of intentionality. A checklist was adopted to guide the assessment, listing observable elements with an emphasis on the attributes of information findability. **Results:** The PANGAEA and Biodiversity.aq repositories showed greater adherence to the findability requirements, while OLAC and TalkBank showed greater fragility regarding their systems. **Conclusions:** There was a discrepancy between the principles of Information Architecture and the navigation and search systems of the analyzed data repositories. It was found that many repositories are more like databases than repositories themselves.

**Descriptors:** Information Architecture. Research data repositories. Disciplinary repositories. Information findability.

## ARQUITECTURA DE INFORMACIÓN EN REPOSITARIOS TEMÁTICOS DE DATOS DE INVESTIGACIÓN: LA INFLUENCIA DE LOS SISTEMAS DE NAVEGACIÓN Y

## BÚSQUEDA EN LA ENCONTRABILIDAD DE DATOS

### RESUMEN

**Objetivo:** Evaluar los sistemas de navegación y búsqueda de repositorios de datos científicos disciplinarios desde el punto de vista de la Arquitectura de Información, analizando en qué medida los sistemas de navegación y búsqueda pueden influir en la encontrabilidad en estos entornos. **Metodología:** Cualitativa en cuanto a su enfoque y descriptiva en cuanto a sus objetivos. Se seleccionaron cuatro repositorios de datos de diferentes áreas de conocimiento del directorio re3data.org, componiendo una muestra heterogénea por el criterio de intencionalidad. Se adoptó una lista de verificación para guiar la evaluación, en la que se enumeran los elementos observables con énfasis en los atributos de la encontrabilidad. **Resultados:** Los repositorios PANGEA y Biodiversity.aq mostraron mayor adherencia a los requisitos de búsqueda, mientras que OLAC y TalkBank mostraron mayor fragilidad. **Conclusiones:** Hubo una discrepancia entre los principios de Arquitectura de Información y los sistemas de navegación y búsqueda de los repositorios de datos. Se descubrió que muchos repositorios son más similares a las bases de datos que a los repositorios reales.

**Descriptores:** Arquitectura de Información. Repositorios de datos. Repositorios temáticos. Encontrabilidad de la información.

**Recebido em:** 26.08.2021

**Aceito em:** 30.08.2022