

# Diretrizes para a Compatibilização Semântica de Sistemas de Organização do Conhecimento em Biodiversidade

*Guidelines for Semantic Compatibility Knowledge Organization Systems in Biodiversity*

---

**Carla Beatriz Marques Felipe**

Universidade Federal do Rio de Janeiro, Brasil  
[felipecarla12@gmail.com](mailto:felipecarla12@gmail.com)

**Luana Farias Sales**

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Brasil  
[luanafsales@gmail.com](mailto:luanafsales@gmail.com)

## Resumo

A presente pesquisa versa sobre a compatibilização semântica de Sistemas de Organização do Conhecimento voltados para o domínio da Biodiversidade, tendo como objetivo geral propor diretrizes para a compatibilização terminológica entre diversos vocabulários usados na indexação de bases de dados em Biodiversidade. Para tal, os objetivos específicos desdobram-se em estudar os conceitos de compatibilização semântica propostos pela Ciência da Informação e Investigar as possibilidades de aplicação das técnicas de compatibilização semânticas na integração de bases de dados de Biodiversidade. Com a finalidade de alcançar os objetivos específicos foi realizado buscas em bases de dados da área da Ciência da Informação com fins de identificar autores basilares que estudam a temática e foi realizado um estudo empírico com termos presentes no projeto Pró Espécies, cujo objetivo é unificar informações acerca de espécies brasileiras. A análise dos termos e as investigações na literatura serviram de base para propor as diretrizes voltadas para a compatibilização. Conclui-se que realizar

## Abstract

*This research deals with the semantic compatibility of Knowledge Organization Systems focused on the Biodiversity domain, with the general objective of proposing guidelines for terminological compatibility between different vocabularies used in indexing Biodiversity databases. To this end, the specific objectives are to study the concepts of semantic compatibility proposed by Information Science and Investigate the possibilities of applying semantic compatibility techniques in the integration of Biodiversity databases. In order to achieve the specific objectives, searches were carried out in databases in the area of Information Science in order to identify key authors who study the subject and an empirical study was carried out with terms present in the Pró Espécies project, whose objective is to unify information about Brazilian species. The analysis of terms and investigations in the literature served as a basis for proposing guidelines aimed at compatibility. It is concluded that carrying out semantic compatibility with a view to semantic integration is not an*

a compatibilização semântica com vistas a integração semântica não é uma tarefa fácil, mas que as pesquisas que versam sobre compatibilização e compartilhamento de dados podem mostrar um caminho a ser seguido. *easy task, but that research that deals with compatibility and data sharing can show a path to be followed.*

**Palavras-chave:** Biodiversidade. Dados Científicos. Compatibilização Semântica. **Keywords:** *Biodiversity; Scientific data; Semantic compatibility.*

## 1. Introdução

O GEF<sup>1</sup> “Pró-Espécies: Estratégia Nacional de Conservação de Espécies Ameaçada”, instituído pelo Ministério do Meio Ambiente (MMA) do Brasil, em 2014 (Pró-Espécies, [2019]), tem como meta central integralizar a União e os estados com o objetivo de criar políticas públicas visando diminuir as ameaças e aprimorar o estado de conservação de espécies ameaçadas.

Entre as iniciativas desenvolvidas no âmbito do projeto está uma Proposta de Padrão de Dados e Metadados para Espécies Ameaçadas utilizados pelas instituições que participam do projeto, a saber, Jardim Botânico Rio de Janeiro, ICMBio e Ministério do Meio Ambiente. Embora haja vocabulários dentro dos sistemas de informação utilizados pelas instituições, eles são diversificados e isso dificulta a escolha do termo adequado para a representação do tema por parte de quem disponibiliza a informação, assim como sua recuperação e, conseqüentemente, a interoperabilidade semântica de sistemas.

A Proposta de Padrão de Dados e Metadados para Espécies Ameaçadas tem como objetivo unificar a forma como a representação da informação ocorre entre os sistemas de informação integrantes do projeto Pró-Espécies. No entanto, um outro problema que se coloca é que as instituições que já possuem seus próprios padrões de metadados, bem como seus vocabulários, nem sempre querem ceder trabalho já realizado para cumprir uma determinação governamental, ainda que saibam que essa determinação possa beneficiá-los a longo prazo. Fato verídico e comprovado é o quão a mudança da terminologia de um sistema passa também pela mudança cultural das instituições.

Neste sentido, os dados e a informação em Biodiversidade devem ser acessíveis e de fácil compreensão para quem deseja utilizá-los. Assim, os SOC podem auxiliar o pesquisador na representação semântica dos dados e, posteriormente, facilitar sua recuperação por assunto. Com o crescimento do número de bancos de dados, é preciso que estes estejam conectados para favorecer o uso por parte do usuário.

O objetivo geral da pesquisa é propor diretrizes para a compatibilização conceitual entre diversos vocabulários usados na indexação de bases de dados em Biodiversidade, promovendo a integração semântica neste domínio.

- a) Estudar os conceitos de compatibilização semântica propostos pela Ciência da Informação;

---

<sup>1</sup> A presente pesquisa tem como base a tese de doutoramento intitulada Compatibilização Semântica: uma proposta de melhoria para integração de dados científicos em Biodiversidade, apresentada no Programa de Pós-Graduação em Ciência da Informação (PPGCI), convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e a Universidade Federal do Rio de Janeiro/Escola de Comunicação (UFRJ/ECO).

- b) Investigar as possibilidades de aplicação das técnicas de compatibilização semântica na integração de bases de dados de biodiversidade.

Nesse contexto, acredita-se que as teorias desenvolvidas dentro da Organização do Conhecimento, como a Teoria do Conceito proposta por Dahlberg (1978a), somada a teorias que versam sobre compatibilização semântica, como os estudos de Neville (1970, 1972), Dahlberg (1981) e Sales (2022), podem contribuir para a recuperação da informação e a integração de vocabulários utilizados para descrever os dados e informações em Biodiversidade.

## 2. Dados em Biodiversidade

Dados de pesquisa são gerados de acordo com o domínio, com as tecnologias adotadas, com o objetivo da pesquisa, entre outros fatores. Na área de Biodiversidade esse fato não é diferente. Os dados das pesquisas em Biodiversidade são gerados de formas heterogêneas, isso porque a Biodiversidade engloba as mais variadas fontes de informação e de pesquisa. Além disso, segundo Kays, McShea e Wikelski (2020), devido a mudanças que estão ocorrendo rapidamente no planeta, as pesquisas em Biodiversidade se tornam cada vez mais importantes, ampliando conseqüentemente a quantidade e a diversidade de dados necessários para a compreensão dessas mudanças, suas conseqüências para a Terra e para a humanidade, entre outras questões que surgem neste contexto. Isso porque, esses dados trazem informações preciosas acerca dessas mudanças nos diversos biomas e espécies que habitam o globo terrestre.

Segundo Pereira e Peterson (2001), os dados em Biodiversidade se dividem em duas classificações: dados primários e dados secundários. “Os dados primários consistem em coletas de espécimes, observações e estudos diretos de espécies” (Pereira & Peterson, 2001, seção Fontes de Informação de Biodiversidade, para. 1). Estes dados são coletados pelos biólogos, ecólogos e outros profissionais que tratam da Biodiversidade e são armazenados nos museus e herbários. Por sua vez, ainda de acordo com Pereira e Peterson (2001, seção Fontes de Informação de Biodiversidade, para. 2), “os dados secundários consistem em resumos baseados nos dados primários que adquiriram a forma de mapas regionalizados, guias de campo e registros municipais”. Entende-se aqui que esses dados são fontes de informações geradas a partir das pesquisas e análises realizadas nos dados primários.

Os tipos de dados gerados em Biodiversidade dependerão do tipo de pesquisa realizado, bem como também da disciplina, do pesquisador e da instituição a qual o dado pertencerá. Uma busca realizada no Re3data – um diretório de repositórios de dados científicos – com o termo *Biodiversity*, mostra que os dados de Biodiversidade podem ser encontrados em repositórios ligados às áreas do conhecimento como Zoologia, Biologia, Ecologia, Oceanografia, Bioinformática, Ciências Naturais e seus desdobramentos. Estes mesmos dados são apresentados nos mais variados formatos, tais como imagens, gráficos estruturados, dados baseados em rede, textos, dados audiovisuais e aplicações de software.

Acerca de dados sobre espécies, aqui englobando fauna e flora, segundo Torres (2004), os sistemas de informação em Biodiversidade muitas vezes disponibilizam dados de ocorrência das espécies, isso implica em dizer que são dados geográficos (onde as espécies ocorrem) e dados temporais (quando as espécies são observadas). A união desses dois tipos de dados facilita, para quem for consultar a informação, compreender como a espécie se comporta. Ainda segundo Torres (2004), os dados sobre

espécies também podem ser taxonômicos, ou seja, apresentam a classificação e características das espécies.

Outro tipo de dado de pesquisa encontrado na área de Biodiversidade são os dados de alterações climáticas, que mostram as alterações pelas quais a terra vem passando ao longo dos anos. Esses dados são frutos de estudos multidisciplinares, que englobam pesquisas acerca das florestas, oceanos, lixo, agricultura, meteorológicos, indústria e outros. Para Brown et al. (2011), é por meio do acesso a esses dados que é possível debater políticas e gerenciamento de ações no combate às mudanças climáticas.

Partindo do trabalho dos professores Sayão e Sales (2020), que apresentam uma taxonomia de dados de pesquisa, verifica-se que os dados pertencentes ao domínio da Biodiversidade (Tabela 1), podem ser classificados como:

**Tabela 1 - Classificação dos dados em Biodiversidade**

Classe	Subclasse	Tipo	Exemplo
<b>Origem</b>	Dados de pesquisa Dados para pesquisa	Observacionais Governamentais Computacionais	Dados de ocorrência de espécies; Dados sobre mudança climática; Dados biogeográficos.
<b>Processamento</b>	Dados brutos Dados finais/processado/terciários	Limpos Anonimizados  Publicados	Dados topográficos; Dados de ocorrência de espécies ameaçadas; Dados disponíveis para consulta em sistemas de informação.
<b>Abordagem da pesquisa</b>	Qualitativo  Quantitativo	Multimídia  Número	Imagens; Fotografia; Vídeo; Dados sobre conservação de espécies.
<b>Materialidade</b>	Digital  Físico	Multimídia  Amostra	Dados disponibilizados em repositórios digitais; Coleções de espécies.

Fonte Elaborado pela autora com base em Sayão e Sales (2020).

A Tabela 1 tem como objetivo sintetizar as informações apresentadas anteriormente. Cabe frisar que os dados citados no quadro são apenas para exemplificação, e que em alguns casos, podem ser complementados com outros tipos de dados, como por exemplo os dados biogeográficos também podem ser governamentais e computacionais.

Apenas coletar o dado não significa que será possível a sua recuperação, por esse motivo, as instituições criam seus bancos de dados, para garantir a preservação e uso dos dados. Segundo Cavalcanti (2005, p. 200):

“a criação de bancos de dados por projetos que envolvem a coleta e o estudo de espécimes biológicos, as espécies a que pertencem e os locais em que ocorrem constitui-se num pré-requisito indispensável à conservação da biodiversidade e uso sustentável dos recursos naturais.”

Esses bancos de dados podem concentrar todas as informações relacionadas a um ecossistema inteiro, facilitando a disseminação da informação para pesquisadores e interessados. Um exemplo interessante é o Projeto Biotupé, situado na Desenvolvimento Sustentável do Tupé (RDS do Tupé), na Amazonia, que no início de sua jornada criou um banco de dados acerca de um lago dentro da reserva onde era possível armazenar os dados da Biodiversidade da área (Projeto Biotupé, 2019).

De acordo com Kays et al. (2020), existe uma classificação de dados em biodiversidade chamada “nascido digital”, são dados gerados por meio de sensores digitais e de outras tecnologias que facilitam a coleta dos dados. Segundo os autores, estes dados têm o mesmo valor dos dados guardados por museus em suas coleções e merecem o devido cuidado com relação à preservação dos mesmo por meio de atividades como a curadoria digital. A captura desses dados alcançou um ritmo acelerado nos últimos anos.

Esses dados digitais fizeram com que os sistemas de informação começassem a compartilhar dados de forma digital e online. Esse movimento fica cada vez mais perceptível com a criação de repositórios e sistemas que disponibilizam em larga escala os dados.

Sabe-se que esses dados, tanto os primários quanto os secundários, e os nascidos digitais, são fundamentais para o desenvolvimento das pesquisas em Biodiversidade. Assim, é importante que ocorra gestão desses dados em amplo espectro, que englobe a formulação de políticas e estratégias que considerem as peculiaridades próprias do domínio da Biodiversidade, além da representação da informação de forma correta. Nesse sentido, os estudos desenvolvidos dentro da Organização do Conhecimento, incluindo os estudos de compatibilização semântica podem contribuir para uma gestão de dados eficiente.

### **3. Compatibilização Semântica**

A Organização do Conhecimento enquanto subárea da Ciência da Informação serve como base para a gestão de dados, uma vez que se ocupa das questões de representação da informação para fins de recuperação, preservação, acesso, interoperabilidade e reuso.

A compatibilização semântica é uma dessas vertentes de estudo da Organização do Conhecimento que pode colaborar para uma gestão de dados efetiva, em particular para a controvérsia esta pesquisa se debruça em busca de uma possível proposta de solução.

No entanto, no âmbito da Organização do Conhecimento, os estudos referentes aos aspectos de compatibilização de linguagens ou dos seus sinônimos como interoperabilidade semântica, compatibilização de SOC e compatibilização de ontologias (Barbosa, 2021) não são relativamente novos, pois no âmbito da Ciência da Informação e da Terminologia eles já eram desenvolvidos há algumas décadas, mas se tornaram mais necessários à medida que os sistemas precisaram ser integrados para facilitar a recuperação da informação. Conforme Sales (2022, p. 81),

“Mesmo nos sistemas clássicos de recuperação da informação, a existência de linguagens distintas e a necessidade de compatibilização entre elas já se colocava há muito tempo como um desafio, fazendo com que a temática compatibilização de linguagens de indexação para compartilhamento e intercâmbio entre diferentes sistemas não seja nova.”

Segundo Svenonius (1983), as teorias da área da indexação, já na década de 1960, estavam preocupados com os aspectos inerentes à compatibilização entre sistemas. Ainda segundo a autora, no final dessa década e início dos anos 70, o entusiasmo quanto aos problemas sobre a compatibilidade de linguagens atenuou. A autora aponta que os estudos a respeito da compatibilização levaram a prática na década de 80, na Europa, com a criação de tesouros multilíngues. De todas as iniciativas criadas, o mais famoso foi o do UNIST, que “serviu como um mecanismo de integração em diferentes classificações e tesouros no processo de transferência de informação” (Svenonius, 1983, p. 3).

No Brasil, consta a tese de Batista (1986) sobre a orientação da professora Hagar Espanha Gomes, sendo um marco teórico sobre a temática. Há, também, Campos (2005, 2007, 2009) com as pesquisas sobre integração de ontologia. E, no contexto mais atual, Sales (2022) com o estudo voltado para a gestão de dados de pesquisa e compatibilização de linguagens na perspectiva da representação dos dados de pesquisa.

Para Glushkov et al. (citado por Sales, 2022, p. 81) a definição de compatibilidade é indicada como “a medida de similaridade entre duas linguagens, onde se introduz o conceito de graus de compatibilidade e estabelecem a distinção entre compatibilidade em plano semântico e no plano estrutural”. Nesse caso, a compatibilidade ocorre quando dois ou mais vocabulários apresentam semelhanças na temática e termos que os constituem.

Os sistemas ou as linguagens teriam a possibilidade de interagir um com o outro tocante à indexação e o mesmo assunto em um e outro. Para Batista (1986, p. 1), por sua vez: “Os estudos de compatibilidade e convertibilidade entre linguagens de indexação, visam, principalmente, a criação de instrumentos de conversão e/ou desenvolvimento de linguagens de indexação compatíveis, que viabilizem o acesso a múltiplas bases de dados que operem em bases cooperativas”.

Dessa forma, visam, além da integração de linguagens, a elaboração de um instrumento que permita a compatibilização de linguagens. Um dos mais importantes estudos já realizados sobre compatibilização de linguagens é o de Neville (1970, 1972) que teve seu estudo direcionado para os tesouros. A problemática encontrada pelo autor é que se deve construir uma única linguagem, ou uma nova, se os sistemas não forem do mesmo país.

Para evitar um retrabalho na construção de novos tesouros, o autor sugere a compatibilização mediante reconciliação de tesouro, e afirma que “O processo de reconciliação como um todo, resulta na criação de um conjunto de códigos numéricos. Cada palavra-chave em cada tesouro participante [da compatibilização] recebe um código, equiparando as palavras-chave” (Neville, 1970, p. 314).

Para melhor exemplificar, na Tabela 2 ilustra-se como funciona a proposta.

**Tabela 2 - Demonstração simples do método de Neville**

	<i>Source thesaurus</i>	<i>Thesaurus B</i>	<i>Thesaurus C</i>
<b>Original entries</b>	AIRFIELDS	AIRFIELD	FLUGPLÄTZE
<b>Reconciled entries</b>	AIRFIELDS (0101)	AIRFIELD (0101)	FLUGPLÄTZE (0101)
<b>Keys to code numbers</b>	0101=AIRFIELDS	0101=AIRFIELD	0101 = FLUGPLÄTZE

Fonte: Recuperado de “Feasibility study of a scheme for reconciling thesauri covering a common subject” de H. H. Neville, 1970, *Journal of Documentation*, 26. p. 316.

Na Tabela 2, percebe-se que mesmo com o emprego de termos expressos em idiomas diferentes, mas utilizados da mesma forma pelos tesouros, para que ocorra a reconciliação/compatibilização, faz-se preciso inserir o mesmo código em cada entrada. De acordo com Neville (1970, p. 316, tradução nossa), “formas singulares e plurais da mesma palavra podem ser consideradas como correspondendo exatas, e o uso de diferentes linguagens não tem importância, desde que os conceitos subjacentes sejam os mesmos”<sup>2</sup>. Sendo assim, o autor defende que os conceitos são indexados pelas palavras-chave que são a representação do conceito. Se este for indicado de modo

<sup>2</sup> Texto original: “Singular and plural forms of the same word may be taken as corresponding exactly, and the use of different languages is of no consequence as long as the underlying concepts are the same”.

diferente em dois ou em mais tesouros, a reconciliação ocorre, dispensando a adição de novos termos.

Por sua vez, Dahlberg (1981) apresenta uma proposta de registro do conceito para o desenvolvimento da compatibilização entre linguagens, porém antes de apresentar o que vem a ser esse registro do conceito é necessário lançar o que vem a ser o próprio conceito. Dahlberg(1978a) declara que mediante a linguagem natural o homem nomina objetos e comunica-se com seus semelhantes. Quanto ao objeto, este ser social elabora enunciados, com suas características, tornando-o, assim, singular. Assim, a reunião destes enunciados forma o conceito.

Ainda para Dahlberg (1978a, p. 102), “com ajuda da linguagem natural é possível formular enunciados a respeito tanto dos conceitos individuais como conceitos gerais. É com base em tais enunciados que se elaborou os conceitos relativos aos diversos objetos”. Através dos enunciados verdadeiros é concebível a identificação dos atributos que formam o conceito.

Concernente ao registro do conceito para a compatibilização de linguagens, Dahlberg (1981) propõe que todas as vezes em que as linguagens forem analisadas para uma possível compatibilização as sobreposições dos conceitos devem ser descritas no registro do conceito, a ser preenchido em uma ficha terminológica constituída por metadados que apresentam como o termo está estruturado e o grau de sobreposição entre as linguagens.

Desse modo, a ficha terminológica proposta pela autora deve ser formada pelos seguintes metadados:

- (1) nome do conceito
- (2) notação
- (3) conceito genérico mais próximo
- (4) categoria de assunto ou conceito em nível hierárquico mais alto
- (5) Indicação do nível hierárquico do conceito
- (6) número de subconceitos;
- (7) forma categorial do conceito (ex: objeto, processo, qualidade, relação, espaço, tempo, domínio etc.)
- (8) definição do conceito
- (9) outros nomes para o conceito ou classe
- (10) Fonte do conceito
- (11) Observações e comentários (Dahlberg, 1981, p. 87, tradução nossa <sup>3</sup>).

O uso dessas metainformações promove a análise e a comparação entre as linguagens de forma consistente, promovendo a descoberta em que grau ocorre a compatibilização das linguagens, sobretudo no tocante aos conceitos utilizados nestas.

---

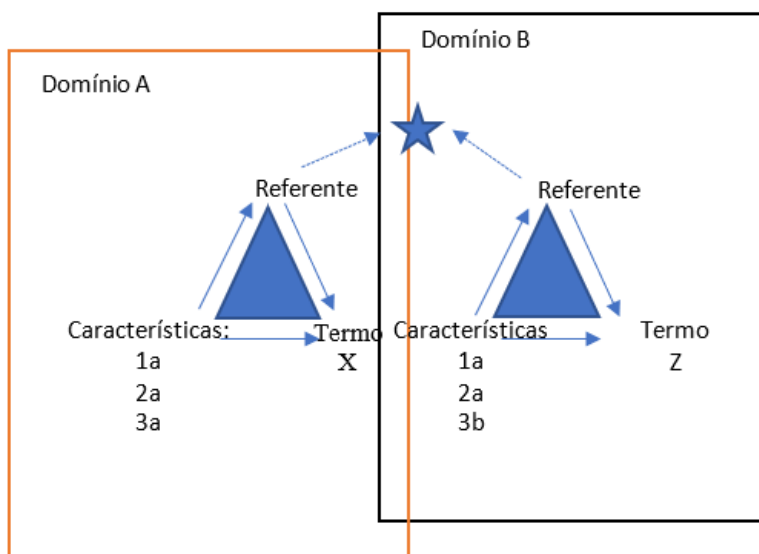
<sup>3</sup> Texto original: “(1) Name of concept or-class (2) Notation (3) Next broader concept (4) Highest concept in hierarchy/subject category (5) Indication of hierarchical level of concept (A) highest level (B) next lowest level (C) third level, etc. (6) Number of subconcepts, if comparison only on a certain level, in brackets for each level (7) Form category of concept (O) Object, entity (P) Process, activity, state (Q) Quantity, quality (R) Relation (S) Space-related concept (T) Time-related concept (W) Subject-field or discipline (8) Definition of concept (if necessary and possible) (9) Other names of concept or class (10) Source of concept abbreviation of IL accord. to (29) (11) Remarks”.

Além do registro do conceito, Dahlberg (1981) apresenta em sua pesquisa a proposta de estabelecer uma matriz de compatibilização, que deve ser capaz de descobrir como uma linguagem pode ser compatível com a outra semanticamente. O primeiro passo sugerido pela autora é o estabelecimento da elaboração de uma matriz de comparação alfabética. Os termos das linguagens devem ser ajustados uns ao lado dos outros para a identificação das relações. Posterior a isso, deve-se criar uma outra coluna apresentando os termos que têm coincidência verbal. A autora ainda afirma que é possível com o método da matriz estabelecer o quão uma linguagem é compatível com a outra.

No contexto atual, Sales (2022) propõe um o modelo conceitual como uma representação gráfica de compatibilização semântica voltada para os dados de pesquisa. A autora defende que mesmo com a existência da curadoria de dados, a representação temática não é evidenciada dentro dos ciclos de vida dos dados tradicionais, e isso causa uma falha na recuperação da informação, entre humanos e máquinas. Ainda, “A ausência de tratamento temático dos dados de pesquisa pode se revelar como uma barreira à sua recuperação, à sua compreensão e conseqüentemente ao seu reuso” (Sales, 2022, p. 85). Mesmo com o uso de metadados para descrever os dados isso não é o suficiente para a recuperação da informação nos repositórios, pois o pesquisador precisa compreender sobre o que aquele dado versa (assunto).

Sales apresenta um modelo conceitual (Figura 1) que representa graficamente uma proposta de compatibilização semântica para dados de pesquisa, baseado no triângulo conceitual de Dahlberg (1978b).

**Figura 1. Modelo Teórico de Compatibilização Semântica.**



Fonte: Compatibilização semântica entre dados de pesquisa: promovendo fairificação em domínios interdisciplinares, de L. F. Sales, 2022, p. 87.

A autora discorre que quando os pesquisadores compreendem o que é o referente (objeto principal), ainda que apresentem uma nomenclatura diferente, esse modelo terminológico pode ajudar os pesquisadores a identificarem que estão a falar do mesmo objeto. Segundo Sales (2022, p. 86):

“Empiricamente falando, uma proposta de matriz de compatibilização semântica, que se construa a partir dos sistemas ordenados existentes no domínio, pode melhorar a comunicação entre domínios interdisciplinares, fazendo a tradução da linguagem de um sistema para o outro e

consequentemente permitindo que dados sejam descobertos, acessados, interoperáveis e assim disseminados e reinterpretados.”

Assim, a matriz de compatibilização, se aplicada, seria capaz de permitir a comunicação entre os sistemas com domínios diferentes, visto que mesmo dentro de domínios do conhecimento distintos o referente trata do mesmo conceito para ambos. A autora sugere, também, o desenvolvimento de fichas terminológicas: “A elaboração de fichas terminológicas que reúnam metadados terminológicos sobre o conceito é um procedimento interessante que pode auxiliar na identificação de níveis de compatibilização semântica mais profundos, revelando, por exemplo, compatibilizações do tipo correlação conceitual” (Sales, 2022, p. 87).

Essas fichas seriam capazes de apresentar a distinção e a identificação entre os conceitos. Desse modo, considera-se que a proposta apresentada por Sales (2022) pode contribuir para a compatibilização semântica entre dados científicos em Biodiversidade.

#### **4. Procedimentos Metodológicos**

A pesquisa é caracterizada de acordo com sua natureza, objetivos e procedimento de coleta. A presente pesquisa se caracteriza como teórica e empírica, quanti-qualitativa, do tipo descritiva e exploratória, bibliográfica e documental.

Para atender os objetivos foi realizado um estudo empírico no contexto do Projeto GEF Pró-Espécies: Estratégia Nacional para a Conservação de Espécies Ameaçadas.

No âmbito do projeto supracitado, foi contratada uma consultoria cujo objetivo era propor uma melhoria na arquitetura e recuperação da informação nos sistemas de informação participantes. O consultor foi o responsável por levantar termos e padrões de metadados das instituições que compõem o projeto, com o objetivo de padronizar os termos. Este consultor possuía experiência na área de informática para biodiversidade e gestão de dados sobre biodiversidade.

O consultor nos forneceu uma planilha com os termos presentes nos sistemas de recuperação de informação presentes no projeto Pró-Espécies, do qual fazem parte as seguintes instituições, com seus respectivos sistemas de informação:

- a) Jardim Botânico Rio de Janeiro, com o sistema Flora e Funga do Brasil, CNC Flora e Catálogo da fauna;
- b) ICMBio, com o Salve;
- c) SiBB, com SiBB;
- d) Ministério do Meio ambiente.

As presentes instituições utilizam padrões de metadados, para descrever os dados tais como Darwin Core e GBIF Metadadata, mesmo assim, não existe uma padronização nos termos que representam esses dados. No momento da coleta dos dados, em outubro e novembro do ano de 2022, o consultor conseguiu identificar 458 termos presentes nos sistemas, disponíveis para a estratégia de compatibilização.

Cabe frisar que, a consultoria acabou posterior a esse momento, e o consultor englobou após esse período termos que pertencem ao Ministério do Meio Ambiente, logo aqui, no desenvolvimento da

presente tese não foram considerados estes termos que foram englobados posteriormente, assim ficando esses termos em escopo para pesquisas futuras. Assim sendo, a amostra inicial contava com os 458 termos identificados em outubro e novembro.

Após a primeira análise dos termos, foi perceptível notar que alguns termos não eram utilizados oficialmente pelas instituições que participavam do projeto, dessa forma, esses termos foram dispensados, uma vez que o objetivo da presente pesquisa era auxiliar as instituições pertencentes ao projeto. Posto isso, foram eliminados 60 termos, logo, os termos considerados válidos para análise totalizaram 398.

Após essa etapa, os conceitos foram separados por sistema de informação e alfabetados na ordem de A a Z. Dessa forma, foi possível identificar a quantidade de termos que cada sistema possui e quais os termos que poderiam apresentar semelhanças e estavam presentes na maioria dos sistemas que compõem o projeto no momento da presente coleta. Para a elaboração do recorte da pesquisa, foi aplicada a Lei de Pareto, que apresenta a relação 80/20. Para Koch (2015, p. 8):

“O princípio 80/20 nos diz que, em qualquer população, algumas coisas são muito mais importantes que outras. Uma boa referência ou hipótese é que 80% dos resultados ou dos produtos derivam de 20% das causas e, às vezes, até de uma proporção ainda menor.”

Assim, entende-se que 80% dos resultados são consequência de 20% das ações.

Logo, ao se fazer os cálculos, dos 398 termos disponíveis, o recorte selecionado foi de 81 termos, que convencionou-se chamá-los de conceitos, uma vez que foram analisadas suas características para a compatibilização. Nessa etapa, foram selecionados conceitos que se considerou possível fazer uma compatibilização semântica. Foi possível observar que alguns termos não apresentavam semelhanças, que outros termos eram específicos e pertenciam somente a um sistema de informação, não apresentando alguma característica passível de compatibilização. Logo esses critérios foram utilizados para a não inclusão desses termos na análise final.

Após a seleção dos 81 termos, ficou perceptível que alguns termos dentro da planilha não possuíam definições. Assim, foi necessário investigar seus significados para propor a compatibilização. Para tal, foi realizada uma consulta com os especialistas das instituições, para obter os significados dos termos. Além disso, foi perguntado a esses especialistas se esses termos eram importantes para os seus sistemas de informação. A resposta para essa questão foi sim, os termos são necessários para os sistemas de informação.

Mesmo com o contato com os especialistas, após a análise dos 81 termos, foi identificado que alguns termos não possuíam documentação, dessa forma foram selecionados os termos que possuíam proveniência de esquemas de metadados conhecidos. Logo, termos como autor e seu correspondente encontrado “author”, “bibliographiccitation\_how\_to\_cite”, “Data\_Sens”, “Endemic brasil”, “Brasil endemic”, “família”, “habit”, “habitats”, “identCert”, “identified\_by”, “nome”, “nome completo”, “scientificNameAuthorship” não foram considerados válidos para a análise<sup>4</sup>. A vista disso, permaneceram 67 termos válidos para a pesquisa.

---

<sup>4</sup> Como estes termos estão presentes nos sistemas, entende-se que são importantes e, portanto, podem no futuro compor pesquisas futuras dentro do projeto Pró-Espécies.

Dos 67 termos possíveis para a compatibilização, pôde-se perceber que alguns deles eram comuns aos sistemas de informação e não precisavam de compatibilização, pois eram escritos da mesma forma e possuíam o mesmo significado, sendo eles: “bibliographicCitation”, “Family” (termo presente em 4 sistemas de informação), “genus”, “identifie”, “kidgom”, “locality”, “name”, “status”, “title”, assim foram excluídos da análise final. Ainda foram excluídos termos que foram considerados complementares, tais como “acceptedNameUsage” e “acceptedNameUsageID”, Data “Resource ID” e “Data Resource Name”, “Institution” e “Institution ID”, “parentNameUsage” e “parentNameUsageID”. Assim, 38 termos foram os que permaneceram para a proposta final de compatibilização. Na Seção dos resultados, serão mostrados os termos selecionados para análise.

#### 4.1. Análise Dos Dados

A proposta de análise dos dados teve como base a literatura da área de Ciência da Informação, especificamente do campo da Organização do Conhecimento, que versa sobre a compatibilização semântica dos SOC. Assim, será realizada com base em Neville (1970), Dahlberg (1981) e Sales (2022)<sup>5</sup>.

Com os termos selecionados, a primeira análise foi a elaboração de fichas terminológicas, conforme propostas por Dahlberg (1981), uma vez que é feita uma avaliação dos termos para a sua compatibilização conceitual, assim sendo realizado o registro do novo conceito compatibilizado. Essa etapa foi baseada nos aspectos elencados na Tabela 3, sobre os quais Dahlberg (1981) sugere que seja elaborada:

**Tabela 3 - Orientações para o registro do conceito**

Aspecto a ser analisado	Significado
Nome do conceito	Nome que o conceito recebeu
Notação	Número que recebe em uma determinada organização
Conceito genérico mais próximo	Conceito geral mais próximo ao conceito
categoria de assunto ou conceito em nível hierárquico mais alto	Conceito mais específico próximo ao conceito analisado
Indicação do nível hierárquico do conceito	Se o conceito é um termo geral, termo específico, como pode ser usado dentro de um sistema de informação
Número de subconceitos	Identificação se o conceito tem termos abaixo dele, conceitos mais gerais
Forma categorial do conceito (ex: objeto, processo, qualidade, relação, espaço, tempo, domínio etc.);	Classe ao qual o conceito pertence
Definição do conceito	Significado do conceito (identificação de suas características)
Outros nomes para o conceito ou classe	Sinônimos do conceito
Fonte do conceito	Origem do termo (vocabulário o qual pertence)
Observações e comentários	Espaço destinado para explicações sobre o conceito que não cabe nos aspectos anteriores.

Fonte: Recuperado de “Towards the establishment of compatibility between indexing languages” de I. Dahlberg, 1981, International Classification, 8. p. 87.

Percebe-se na Tabela 3, de Dalhberg (1981), que os termos ali selecionados são metadados terminológicos, ou seja, dados que descrevem o termo. A fim de adaptá-lo ao contexto investigado, a

<sup>5</sup> Essas teorias serão explicadas na seção de compatibilização semântica.

Tabela 4 foi construído para o registro dos conceitos da área de Biodiversidade, a partir de uma adaptação da Tabela 3, por esta autora.

**Tabela 4 - Proposta de registro do conceito desenvolvida**

Aspecto analisado	Resultado
Nome do conceito	Apresentação do termo analisado
Identificador	Número que o termo recebeu dentro da tabela do consultor do projeto Pró-Espécie
Forma categorial do conceito: (O) Objeto, entidade	Apresentação da classe do termo, na presente pesquisa tratam se de Metadado.
Definição do conceito	Identificação do significado e características do metadado analisado
Outros nomes para o conceito ou classe	Termos selecionados para tentar elaborar a compatibilização
Fonte do conceito	Padrão de metadado do qual o termo se origina
Observações e comentários	Apresentação da identificação do sistema de informação ao qual o termo pertence

Fonte: Elaboração da autora com base em "Towards the establishment of compatibility between indexing languages" de I. Dahlberg, 1981, International Classification, 8.

A razão de analisar os termos por meio do registro do conceito de Dahlberg (1981) é identificar as características dos metadados, verificar de qual padrão se origina e conhecer qual sistema de informação dentro do projeto possui o termo, para assim facilitar a tentativa de compatibilização.

Posterior a isso, a análise seguiu a proposta de Sales (2022), que permitiu a identificação de características em comum entre os metadados, e objetivou apontar se é possível a compatibilização, uma vez que essa metodologia aborda a importância do referente na representação temática dos dados.

E por fim, a metodologia de Neville (1970) servirá para testar se realmente é possível compatibilizar os termos que são idênticos.

A análise dos conceitos que formam o corpus foi realizada da seguinte forma: como não se pode compatibilizar todos os termos de uma só vez, foram selecionados termos que possuem a grafia parecida – mesmo termos que são escritos em português e inglês, os termos foram divididos em grupo de cerca de 2, 3 a 4 metadados. Em seguida, os grupos foram analisados na seguinte ordem: primeiro, de acordo com Dahlberg (1981), em seguida com Sales (2022) e posteriormente conforme Neville (1970).

Isto posto, na próxima Seção serão apresentados os resultados da pesquisa.

## 5. Resultados

A proposta desta subseção é apresentar os resultados dos testes desenvolvidos para se chegar a uma compatibilização semântica entre os conceitos pertencentes ao projeto Pró-Espécies e, assim, consequentemente, atingir o objetivo geral.

A Tabela 5 contém a lista com os 41 termos que participaram da análise.

**Tabela 5 - Lista completa dos termos que participaram do experimento**

Termos	Termos
Basis Of Record	Longitude
Basis Of Record – original	Longitude – original
Basis of record badly formed	Occurrence Status
BasisOfRec	Occurrence status assumed to be present
CatalogNumber	Order
Catalogue Number	Order
Class	Ordername
Class	Order
classname	Scientific Name
collectID	scientificNameID
Collection	Scientific Name – original
Collection	ScientificName
Collection Code	scientificNameAuthorship
Collection code not recognized	speciesName
Collection ID	species
collectionCode	species
Country – parsed	vernacularName
Country inferred from coordinates	Vernacular name
countryCode	Vernacular name – original
countryCode	Longitude

Fonte: Dados da pesquisa (2023).

Para a apresentação dos resultados, foi selecionado um conceito que foi analisado de acordo com os parâmetros estabelecidos nos procedimentos metodológicos. A apresentação dos outros conjuntos pode ser consultada na tese final. Com base nos resultados serão apresentadas as diretrizes para a compatibilização semântica. A seguir, Tabela 6 a análise do conceito “Species”.

**Tabela 6 - Registro do conceito “Species”**

Aspecto analisado	Resultado
<b>Nome do conceito</b>	<i>Species</i>
<b>Identificador</b>	56
<b>Forma categorial do conceito: (O) Objeto, entidade</b>	Metadado
<b>Definição do conceito</b>	Nome da espécie.
<b>Outros nomes para o conceito ou classe</b>	species
<b>Fonte do conceito</b>	Padrão IUCN
<b>Observações e comentários</b>	Termo ou conceito pertence ao sistema de informação CNCFlora e SALVE

Fonte: Elaborado pela autora, com os dados da pesquisa (2023).

O conceito acima está voltado para as espécies e pertence aos sistemas CNCFlora e SALVE, tendo sua origem no padrão IUCN. A seguir, na Tabela 7, a análise do conceito do mesmo conjunto para identificação do seu significado e origem.

Tabela 7 - Registro do conceito “Species”

Aspecto analisado	Resultado
Nome do conceito	Species
Identificador	396
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Nome da espécie.
Outros nomes para o conceito ou classe	species
Fonte do conceito	Padrão SiBBr-ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Fonte: Elaborado pela autora, com os dados da pesquisa (2023).

Mesmo tendo a origem e pertencer a um sistema que não está ligado ao JBRJ, o termo acima possui a mesma definição que o conceito anterior e apresenta a mesma grafia. A seguir (Tabela 8), a análise do terceiro conceito presente no conjunto Species.

Tabela 8 - Registro do conceito “SpeciesName”

Aspecto analisado	Resultado
Nome do conceito	SpeciesName
Identificador	190
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Nome da espécie.
Outros nomes para o conceito ou classe	species
Fonte do conceito	Padrão IUCN
Observações e comentários	Termo ou conceito pertence ao sistema de informação CNC Flora e SALVE

Fonte: Elaborado pela autora, com os dados da pesquisa (2023).

Ao observar os três registros do conjunto de conceitos acima, percebeu-se que os conceitos possuem o mesmo referente, a espécie ao qual o dado pertence. Logo, para uma melhor análise de uma possível compatibilização, a análise conforme Sales (2022) será apresentada e as características do termo lado a lado (Tabela 9).

Tabela 9 - Análise do conjunto Species conforme Sales

	Termo 01 – Species	Termo 02 – Species	Termo 03 – SpeciesName
Referente	A espécie	A espécie	A espécie
Característica 01	Voltado para o nome das espécies.	Voltado para o nome das espécies.	Voltado para o nome das espécies.
Característica 02	Presente nos sistemas CNCFlora e SALVE	Presente no sistema SiBBr	Presente nos sistemas CNCFlora e SALVE

Fonte: Elaborado pela autora, com os dados da pesquisa (2023).

Pode-se observar que, ao aplicar a metodologia de Sales (2022), consegue-se identificar que os conceitos são equivalentes semanticamente, pois possuem o mesmo referente e uma mesma característica “são voltados para o nome das espécies”, possuem similaridade no significado e no objeto de representação, assim sendo, pode-se afirmar que esse conjunto de conceitos é passível de compatibilização semântica.

A seguir, na Tabela 10, a análise conforme Neville (1970) e um exemplo de aplicação de códigos para o conjunto de conceitos.

**Tabela 10 - Atribuição do Código 0404 ao conjunto Species conforme o Método de Neville**

	<b>Termo 01</b>	<b>Termo 02</b>	<b>Termo 03</b>
<b>Termo original</b>	<i>Species</i>	<i>Species</i>	<i>SpeciesName</i>
<b>Identificador</b>	<i>Species (0404)</i>	<i>Species (0404)</i>	<i>SpeciesName (0404)</i>
<b>Termos com os identificadores</b>	<i>0404=Species</i>	<i>0404=Species</i>	<i>0404=SpeciesName</i>

Fonte: Elaborado pela autora, com os dados da pesquisa (2023).

Assim, ao atribuir o mesmo código para os conceitos, mesmo com grafias diferentes, a compatibilização pode ocorrer sem nenhum prejuízo de perda de informação para o usuário.

Os resultados mostram que dentro dos próprios sistemas de informação existem metadados que são apresentados de forma diferente, porém foram criados para representar a mesma informação. Assim sendo, dentro do contexto investigado, sugere-se que as instituições gestoras dos sistemas de informação façam uma revisão dos conceitos e de suas funções, para assim evitar duplicação de informações nos sistemas. Além disso, percebe-se que os conceitos utilizados são voltados especificamente para representação de dados em Biodiversidade, pois englobam questões específicas, como os estudos ambientais e suas vertentes.

No projeto Pró-espécies, pode-se observar que nem todos os conceitos estão presentes nos cinco sistemas de informação, assim sugere-se que seja feita uma análise por parte dos especialistas do domínio, junto com o consultor do projeto, para observar se os conceitos podem ser incluídos em todos os cinco sistemas ou não, bem como identificar quais conceitos devem ser eliminados, uma vez que existem conceitos que não possuem documentação.

Também foi possível verificar que se pode estabelecer uma compatibilização semântica entre os sistemas presentes dentro do projeto, pelos seguintes aspectos:

- a) os sistemas possuem focos parecidos, mesmo quando voltados para objetos específicos, como flora e fauna. Os metadados utilizados por sistema mais abrangente, como o SiBBr, podem ser utilizados pelos mais específicos e vice-versa.
- b) os metadados que os sistemas utilizam, mesmo derivados de padrões distintos como Darwin Core, IUCN e SiBBr-Ala, possuem foco em Biodiversidade e foram construídos para atender uma demanda voltada para a representação das informações com esse foco.

Como o foco é contribuir para pesquisadores da Ciência da Informação, além de pares interessados em compatibilização semântica, como apresentado no objetivo geral, serão apresentadas, a seguir, diretrizes que visam auxiliar na implementação da compatibilização semântica.

### 5.1 Diretrizes para a Compatibilização Semântica

Com a análise dos resultados ficou evidente que apenas a aplicação das metodologias apresentadas na literatura não é o suficiente para que ocorra uma compatibilização semântica em sistemas de informação que disponibilizam dados e informações em Biodiversidade. O trabalho é amplo e demorado. Deste modo, sugere-se aqui diretrizes que visam auxiliar não só os pesquisadores do projeto Pró-Espécies, mas pesquisadores, em âmbito geral, que queiram promover a compatibilização semântica.

As diretrizes propostas têm o objetivo de auxiliar os pesquisadores da área da Ciência da Informação em estudos ou aplicações futuras. Mesmo que a presente pesquisa seja voltada para a área da Biodiversidade, as diretrizes a seguir não estarão voltadas para um domínio específico, acredita-se assim contribuir com as mais diversas áreas do conhecimento.

As diretrizes servem como indicações, instruções que podem ser seguidas, e acredita-se que facilitam a execução da compatibilização semântica. Segundo Ciavatta e Ramos (2012 p. 11), “diretrizes são orientações para o pensamento e a ação”, ou seja, servem para o planejamento e a execução do trabalho. Assim sendo, servem como guia para o estabelecimento de um trabalho futuro.

As diretrizes aqui propostas têm como ponto de partida o texto de Dalhberg (1981) sobre compatibilização semântica. Além de fornecer orientações de como proceder, a autora, em suas considerações finais, apresenta orientações complementares para a compatibilização semântica, a saber. Além disso, os trabalhos de Nevill (1970) e Sales (2022) também serviram de base para a elaboração das diretrizes a seguir, que estão sendo apresentadas em uma ordem que se considera lógica para a execução do trabalho.

#### DIRETRIZ 01: REGISTRO DOS CONCEITOS A SEREM AVALIADOS EM TABELA TERMINOLÓGICA INDIVIDUAL GUARDANDO SEUS METADADOS.

O primeiro passo a ser considerado em uma compatibilização semântica é o de Registro dos conceitos a serem avaliados em tabela terminológicos individual guardando seus metadados.

Esse registro se faz importante em duas etapas: na primeira se quer investigar a possibilidade de compatibilização entre os termos; e ao final do processo, quando se necessita realizar a revisão da compatibilização e chegar ao veredito. Dessa forma, o responsável pela compatibilização e possíveis futuros profissionais envolvidos, terão em mãos um importante instrumento contendo todas as informações necessárias acerca de cada termo que se pretende compatibilizar, além da possibilidade de identificação do referente do conceito.

A recomendação de registrar o conceito junto aos seus metadados, visa possibilitar uma consulta futura às informações referentes àquele conceito. Acredita-se que os metadados serão capazes de informar todas as características e informações relacionadas aos conceitos de ambos os SOC que se deseja compatibilizar a informação.

#### DIRETRIZ 02: REGISTRO DOS CONCEITOS EM TABELA TERMINOLOGIA COMPARATIVA

A segunda recomendação é da elaboração de fichas terminológicas para a análise comparativa entre conceitos. Ao se registrar os conceitos em tabelas terminológicas comparativas, os profissionais envolvidos serão capazes de estabelecer o primeiro passo para uma compatibilização efetiva.

Essa tabela deve conter todas as características dos termos que pretende fazer a compatibilização e, assim, descrever as características e enunciados que formam os conceitos selecionados. Seguindo as orientações de Dalhberg (1981), que sugere uma matriz de compatibilidade verbal, as tabelas devem

apresentar espaços para que seja realizada a comparação entre conceitos lado a lado, assim, o responsável não vai encontrar dificuldades na tentativa de estabelecer um grau de compatibilização semântica.

Como as características são identificadas a partir das definições, é importante se valer de documentos de referência das áreas, para que sejam coletadas definições corretas, adequadas e coerentes. Também é recomendável que essas fichas tenham suas definições validadas por especialistas.

#### DIRETRIZ 03: ANÁLISE COMPARATIVA DAS CARACTERÍSTICAS DO CONCEITO

A terceira diretriz sugerida é a análise comparativa entre as características do conceito e só pode ser realizada se a segunda diretriz for desenvolvida. Essa parte do desenvolvimento da compatibilização é necessária pois alguns termos podem informar o mesmo referente ou não e podem se apresentar como complementos para determinado aspecto da informação. Além disso, na estrutura dos SOC, é possível identificar conceitos que podem ser polissemias ou homônimos, termos com o mesmo significado e grafias e idiomas diferentes.

Nesse contexto, ao realizar a análise comparativa dos conceitos, o profissional envolvido já será capaz de obter um panorama geral do que se pode ou não estabelecer para a compatibilização semântica. Além do mais, já poderá antever problemas e pensar em possíveis soluções para a compatibilização, em caso que nesta etapa tenha identificado.

#### DIRETRIZ 04: IDENTIFICAÇÃO DE CONCEITOS COM GRAUS DE EQUIVALÊNCIA

Para a quarta diretriz, tem-se a identificação de conceitos com graus de equivalência. Recomenda-se seguir esse passo porque pode haver conceitos com um maior grau de compatibilidade disponíveis ou, ainda, conceitos que diferem em especificidade, mas que podem ser utilizados como complemento ou até mesmo substitutos de outros conceitos. Essa identificação só será possível após a comparação realizada entre os conceitos.

#### DIRETRIZ 05: ATRIBUIÇÃO DO MESMO CÓDIGO AOS CONCEITOS PARA EFEITOS DE COMPATIBILIZAÇÃO NO SISTEMA

Como quinta diretriz, estabelecer o mesmo código para conceitos aptos à compatibilização, se faz necessário. Com o intuito de utilizar os SOC já existentes nos sistemas de informação e não a elaboração de um terceiro SOC, ao se empregar o mesmo código para conceitos equivalentes o compatibilizador conseguirá identificar e estabelecer, entre sistemas, quais conceitos foram compatibilizados, sem perda de informação. Tal qual sugere Neville (1970), a aplicação dos códigos conseguirá informar que os conceitos são compatíveis mesmo que escritos em idiomas distintos.

Nesse caso, se for um profissional da informação bibliotecário que esteja responsável pela compatibilização, sugere-se que o mesmo entre em contato e peça auxílio para o profissional da informática, para que ocorra a compatibilização de forma precisa.

## 6. Considerações Finais

No atual contexto em que a ciência se desenvolve, conhecido como o quarto paradigma da ciência, onde os dados de pesquisa são desenvolvidos em larga escala e tornam-se insumos valiosos para o desenvolvimento da pesquisa, tem-se uma grande produção de informação. Assim sendo, os dados produzidos são por natureza heterogêneos e requerem uma gestão de dados concreta, para que ocorra o uso e reúso por parte de quem necessitar.

No âmbito da Ciência da Informação, tem-se a Organização do Conhecimento, disciplina responsável por desenvolver pesquisas e instrumentos com o propósito da recuperação da informação. E dentro desses estudos existem pesquisas acerca da Compatibilização Semântica.

No contexto da heterogeneidade dos dados de pesquisa, a compatibilização semântica pode se mostrar uma importante ferramenta para a padronização da informação com vistas a sua recuperação, uma vez que a compatibilização permite a uniformização das informações, independentemente do domínio no qual ela é gerada.

O projeto Pró-Espécies surge com o intuito de aprimorar e favorecer iniciativas de conservação de espécies e conta com a participação da União e dos estados, bem como conta com a participação de instituições brasileiras que disseminam dados e informações acerca da Biodiversidade. Assim, dentro do projeto surge a necessidade de integralizar a informação em instituições brasileiras que disseminam dados e informações acerca da Biodiversidade.

No contexto do projeto Pró-Espécies, dos sistemas de informação escolhidos para servir de corpus para a pesquisa, ainda que os termos utilizados para o desenvolvimento da presente pesquisa tenham sido metadados, percebe-se que, a compatibilização semântica não é uma tarefa fácil.

Assim, a presente pesquisa sugeriu a adoção de 5 diretrizes que podem auxiliar no desenvolvimento da compatibilização semântica. As diretrizes apresentadas foram embasadas, em todas as suas etapas, em autores seminais da área de Organização do Conhecimento, cujas teorias fundamentam a construção de SOC

Acredita-se que estas instruções são atuais e enquadram-se de maneira palpável para o funcionamento da compatibilização semântica no contexto de ambientes digitais de disseminação de dados de pesquisa, não somente no domínio da Biodiversidade. No entanto, seria interessante que essas diretrizes pudessem ser aplicadas também a outros domínios ou realidades para que possam ser melhor fundamentadas.

## 7. Referências Bibliográficas

---

- Barbosa, N. T. (2021). Para uma economia da informação semântica: a construção de ambientes semânticos para a recuperação inteligente da informação (Tese de Doutorado, Universidade Federal Fluminense, Niterói). Recuperado de: [https://www.academia.edu/103450551/Para\\_uma\\_economia\\_da\\_informa%C3%A7%C3%A3o\\_sem%C3%A2ntica\\_a\\_constru%C3%A7%C3%A3o\\_de\\_ambientes\\_sem%C3%A2nticos\\_para\\_a\\_recupera%C3%A7%C3%A3o\\_inteligente\\_da\\_informa%C3%A7%C3%A3o](https://www.academia.edu/103450551/Para_uma_economia_da_informa%C3%A7%C3%A3o_sem%C3%A2ntica_a_constru%C3%A7%C3%A3o_de_ambientes_sem%C3%A2nticos_para_a_recupera%C3%A7%C3%A3o_inteligente_da_informa%C3%A7%C3%A3o)
- Batista, G. H. R. (1986). Compatibilidade e Convertibilidade entre linguagens de indexação: um estudo de caso (Dissertação de Mestrado, Escola de Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro). Recuperado de: <https://pantheon.ufrj.br/handle/11422/8845>
- Brown, C. J., Schoeman, D. S., Sydeman, W. J., Brander, K., Buckley, L. B., Burrows, M., .... Richardson, A. J. Quantitative approaches in climate change ecology (2011). *Global Change Biology*, 17(12), 3697-3713. doi: <https://doi.org/10.1111%2Fj.1365-2486.2011.02531.x>
- Campos, M. L. de A. (2005). A problemática da compatibilização terminológica e a integração de ontologias: o papel das definições conceituais. In *Anais do VI Encontro Nacional de Pesquisa em Ciência da Informação*, Florianópolis, SC. Florianópolis, SC: UFSC.
- Campos, M. L. de A. (2007, janeiro-junho). Integração de Ontologias: o domínio da Bioinformática. *RECIIS – R. Eletr. de Com. Inf. Inov. Saúde*, 1(1), 117-121, jan./jun. 2007. doi: <https://doi.org/10.29397/reciis.v1i1.896>
- Campos, M. L. de A. (2009). Aspectos semânticos da compatibilização terminológica entre ontologias no campo da Bioinformática. In *Anais do X Encontro Nacional de Pesquisa em Ciência da Informação*, João Pessoa, PB. João Pessoa, PB: UFPB.
- Cavalcanti, M. J. (2005). Bancos de dados sobre biodiversidade na Amazônia: a experiência do Projeto Biotupé. In E. N. Santos-Silva, F. M. Aprile, V. V. Scudeller, & S. Melo (Orgs.). *Biotupé: Meio Físico, Diversidade Biológica e Sociocultural do Baixo Rio Negro, Amazônia Central* (pp 199-213). Manaus: INPA.
- Ciavatta, M.; Ramos, M. (2012, janeiro-abril). A “era das diretrizes”: a disputa pelo projeto de educação dos mais pobres. *Revista Brasileira de Educação*, 17(49), 11-37, 231-232. Recuperado de: <https://www.scielo.br/j/rbedu/a/nDS3v6XBFdJG3jQGLRk687m/?format=pdf&lang=pt>
- Dahlberg, I. (1978a). Teoria do conceito. *Ciência da Informação*, 7(2), 101-107. doi: <https://doi.org/10.18225/ci.inf.v7i2.115>
- Dahlberg, I. (1978b). A referent-oriented analytical concept theory of interconcept. *International Classification*, 5(3), 142-150. Recuperado de: <https://www.nomos-elibrary.de/10.5771/0943-7444-1978-3-142/a-referent-oriented-analytical-concept-theory-for-interconcept-volume-5-1978-issue-3?page=1>
- Dahlberg, I. (1981). Towards establishment of compatibility between indexing languages. *International Classification*, 8(2), 86-91. doi: [https://www.nomos-elibrary.de/10.5771/0943-7444-1981-2-86.pdf?download\\_full\\_pdf=1](https://www.nomos-elibrary.de/10.5771/0943-7444-1981-2-86.pdf?download_full_pdf=1)

- Kays, R., McShea, W. J., & Wikelski, M. (2020). Born-digital biodiversity data: Millions and billions. *Diversity and distributions*, 26(5), 644-648, 2020. doi: <https://doi.org/10.1111/ddi.12993>
- Koch, R. (2015). *O Poder 80/20: Os segredos para conseguir mais com menos nos negócios e na vida*. São Paulo: Gutenberg.
- Neville, H. H. (1970, December). Feasibility study of a scheme for reconciling thesauri covering a common subject. *Journal of Documentation*, 26(4), 313-336. doi: <https://doi.org/10.1108/eb026502>
- Neville, H. H. (1972, November). Thesaurus reconciliation. *Aslib Proceedings*, 24(11), 620-626. doi: <https://doi.org/10.1108/eb050379>
- Pereira, R. S., & Peterson, A. T. (2001, 10 junho). O uso de modelagem na definição de estratégias para a conservação da biodiversidade. Campinas, SP: SBPC/LABJOR. Recuperado de: <https://www.comciencia.br/dossies-1-72/reportagens/biodiversidade/bio18.htm>
- Pró-Espécies ([2019, 8 julho]). Pró-Espécies: Todos contra a extinção. Recuperado de: <https://prospecies.eco.br/projeto/>
- Projeto Biotupé (2019). Recuperado de: <http://biotupe.org/site/node/2>
- Sales, L. F. (2022). Compatibilização semântica entre dados de pesquisa: promovendo fairificação em domínios interdisciplinares. In G. Saldanha, P. C. Castro, & R. M. Pimenta (Orgs.). *Ciência da Informação: sociedade, crítica e inovação* (pp. 75-91). Rio de Janeiro: IBICT. Recuperado de: <https://ridi.ibict.br/bitstream/123456789/1227/1/saldanha-castro-pimenta.pdf>
- Sayão, L. F., & Sales, L. F. (2020, julho-dezembro). Afinal o que é dados de pesquisa? *Biblos: Revista do Instituto de Ciências Humanas e da Informação*, 34(2), 32-51. doi: <https://doi.org/10.14295/biblos.v34i2.11875>
- Svenonius, E. (1983). Compatibility of Retrieval Languages Introduction to a Forum. *International Classification*, 10(1), 2-4. Recuperado de: <https://www.nomos-elibrary.de/10.5771/0943-7444-1983-1-2.pdf>
- Torres, R. da S. (2004). *Ambiente de Gerenciamento de Imagens e Dados Espaciais para Desenvolvimento de Aplicações em Biodiversidade*. (Tese de Doutorado, Instituto de Computação, Universidade Estadual de Campinas, Campinas). Recuperado de: <https://repositorio.unicamp.br/acervo/detalhe/318094>