CARLA BEATRIZ MARQUES FELIPE

Compatibilização Semântica: uma proposta de melhoria para integração de dados científicos em Biodiversidade

Tese de doutorado

Junho de 2023

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO – UFRJ
ESCOLA DE COMUNICAÇÃO – ECO
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBICT
PROGRAMA DE PÓSOGRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – PPGCI

CARLA BEATRIZ MARQUES FELIPE

Compatibilização Semântica: uma proposta de melhoria para integração de dados científicos em Biodiversidade

Rio de Janeiro
2023

CARLA BEATRIZ MARQUES FELIPE

Compatibilização Semântica: uma proposta de melhoria para integração de dados científicos em Biodiversidade

Tese de Doutorado apresentada ao 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), como requisito parcial à obtenção do título de Doutor em Ciência da Informação.

Área de concentração: Informação e Mediações Sociais e Tecnológicas para o Conhecimento

Rio de Janeiro, 16 de junho de 2023.

_____

Profa. Dra. Luana Farias Sales Marques
Convênio MCT/IBICT-UFRJ/ECO (Orientadora)

_____

Profa. Dra. Rosali Fernandez de Souza, PhD
Convênio MCT/IBICT-UFRJ/ECO (Membro Interno)

_____

Prof. Dra. Naira Christofoletti Silveira
Convênio MCT/IBICT-UFRJ/ECO (Membro Interno)

_____

Dra. Debora Pignatari Drucker
EMBRAPA (Membro Externo)

_____

Dr. Cleverson Rannieri Meira dos Santos
Museu Paraense Emílio Goeldi (Membro Externo)

_____

Prof. Dra. Vania Lisboa da Silveira Guedes
Convênio MCT/IBICT-UFRJ/ECO (Suplente Interno)

_____

Dra. Alegria Celia Benchimol
Museu Paraense Emílio Goeldi (Suplente Externo)

CIP - Catalogação na Publicação

F315c

Felipe, Carla Beatriz Marques
Compatibilização Semântica: uma proposta de melhoria para integração de dados científicos em Biodiversidade / Carla Beatriz Marques Felipe. -- Rio de Janeiro, 2023.
142 f.

Orientadora: Luana Farias Sales Marques.
Tese (doutorado) - Universidade Federal do Rio de Janeiro, Escola da Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-Graduação em Ciência da Informação, 2023.

1. Biodiversidade. 2. Dados Científicos. 3. Compatibilização Semântica. I. Marques, Luana Farias Sales , orient. II. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

Ao meu tio José de Arimatéia (*in memória*) que sempre me disse: estuda menina!

A Deus por não ter me deixado enlouquecer.

Aos meus pais, por terem proporcionado a educação necessária para eu conseguir vencer os obstáculos da vida e chegar até aqui e por fazerem o almoço quando que precisava estudar. A minha irmã também por fazer o almoço. A Murilo por proporcionar os periféricos que precisei para ajudar na digitação da tese.

Aos outros membros de minha família que de certo modo contribuíram para eu chegar até aqui.

Ao meu amor Jesué, por todo apoio que me deu durante esse processo e me ajudar em figuras, quadros e outras coisas mais rsrs.

Bibliotecária Érica por encontrar textos que ninguém jamais encontrou.

A Luana minha orientadora, pelo auxílio, inspiração e o presente do tema, no qual pude juntar duas paixões antigas com uma nova paixão que é a gestão de dados.

Ao pessoal do departamento, mas em especial a Danilo e Gustavo que puderam proporcionar mais tempo para eu me dedicar ao doutorado.

A todos os meus amigos que escutaram meus desabafos: Fernanda, Delana, Fátima, Bruna, Kelly, Eliane, Patrícia, Malkene, Ana Cláudia.

Aos meus antigos amigos (aqui o pessoal de Natal) e novos amigos (aqui o pessoal) do *Green Special*.

A mim mesma.

Os dados no século XXI são como o petróleo no século XVIII: um bem valioso e inexplorado. Assim como o petróleo, para aqueles que entendem o valor essencial dos dados e aprendem a extraí-los e utilizá-los, haverá grandes recompensas.

(Joris Tonders, Yonogo)

RESUMO

A presente tese 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 identificar, na literatura, o panorama acerca dos SOC voltados para o domínio da Biodiversidade; compreender os principais conceitos estudados na área da Biodiversidade; 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. Para alcançar o primeiro objetivo que é identificar, na literatura, o panorama acerca dos SOC voltados para o domínio da Biodiversidade, foi realizado uma busca bibliográfica em bases de dados da área da Ciência da Informação, além de bases mais gerais como a Web of Science e a BioOne. Com o intuito de alcançar o segundo objetivo que versa sobre compreender os principais conceitos estudados na área da Biodiversidade, foi desenvolvido uma análise de domínio conforme propostas por Hjørland (2022), com base em três abordagens sugeridas pelo autor, a saber: abordagem de estudos epistemológicos e críticos; abordagem de estudo bibliométrico; e abordagem de estudo terminológico. Com a finalidade de conseguir o terceiro objetivo que é estudar os conceitos de compatibilização semântica propostos pela Ciência da Informação, foi realizado buscas em bases de dados da área da Ciência da Informação com fins de identificar autores basilares que estudam a temática. Para obter o quarto objetivo que versa sobre investigar as possibilidades de aplicação das técnicas de compatibilização semânticas na integração de bases de dados de biodiversidade, 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 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.

Palavras-chave: biodiversidade; dados científicos; compatibilização semântica.

ABSTRACT

This thesis deals with the semantic compatibility of Knowledge Organization Systems focused on the field of Biodiversity, with the general objective of proposing guidelines for terminological compatibility between different vocabularies used in the indexing of databases on Biodiversity. To this end, the specific objectives unfold in identifying, in the literature, the panorama about the knowledge organization systems focused on the field of Biodiversity; understand the main concepts studied in Biodiversity; to study the concepts of semantic compatibility proposed by Information Science and to investigate the possibilities of applying semantic compatibility techniques in the integration of Biodiversity databases. To reach the first objective, which is to identify, in the literature, the panorama about knowledge organization systems focused on the field of Biodiversity, a bibliographic search was carried out in databases in Information Science, in addition to more general bases such as the Web of Science and BioOne. To reach the second objective, which deals with understanding the main concepts studied in the field of Biodiversity, a domain analysis was developed as proposed by Hjørland (2022), based on three approaches suggested by the author, namely: studies approach epistemological and critical; bibliometric study approach; and terminological study approach. To achieve the third objective, which is to study the concepts of semantic compatibility proposed by Information Science, searches were carried out in databases in Information Science in order to identify key authors who study the subject. To obtain the fourth objective, which deals with investigating the possibilities of applying semantic compatibility techniques in the integration of biodiversity databases, 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 performing semantic compatibility with a view to semantic integration is not an easy task, but that research that deals with compatibility and data sharing can show a path to be followed.

Keywords: biodiversity; scientific data; semantic compatibility.

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE QUADROS

LISTA DE ABREVIATURAS E SIGLAS

ABCD	*Access to Biological Collections Data*
BRAPCI	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CENBAM	Centro de Estudos Integrados da Biodiversidade Amazônica
CI	Ciência da Informação
DwC	Darwin Core
EML	*Ecological Metadata Language*
ETS	*Ecological Trait-data Standard*
FLOPO	*Flora Phenotype Ontology*
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
JBRJ	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro
LISA	*Library & Information Science Abstracts*
LISTA	*Library, Information Science & Technology Abstracts*
MMA	Ministério do Meio Ambiente
OC	Organização do Conhecimento
PPBio	Programa de Pesquisa em Biodiversidade
SiBBr	Sistema de Informação sobre a Biodiversidade Brasileira
SOC	Sistemas de organização do Conhecimento
SWI	*Semantic Web Interactive Gazetteer*
TO	*Plant Trait Ontology*
TOP	*Thesaurus of Plant characteristics*

SUMÁRIO

1 INTRODUÇÃO

A Ciência da Informação (CI) tem como foco o estudo da informação, desde a sua origem até o seu uso pelo usuário, passando pelos aspectos de coleta, organização, armazenamento e recuperação da informação (Borko, 1968). Entende-se que outras disciplinas fazem uso da informação, porém poucas se apropriam dela como objeto de pesquisa. Nesse sentido, a Ciência da Informação, enquanto metaciência, visa traçar formas e mecanismos que facilitem o acesso à informação.

No âmbito da CI, para que ocorra a recuperação da informação, é imprescindível organizá-la, de maneira que permita ao usuário o acesso em sua forma mais completa. Essa ação está totalmente ligada à geração do conhecimento e sua organização. E neste contexto, a Organização do Conhecimento (OC) surge como a disciplina da Ciência da Informação que objetiva estudar e desenvolver sistemas voltados para a organização de unidades de conhecimento (conceitos) e para a recuperação da informação, quando representada por estes conceitos.

Nesta disciplina, os estudos são voltados para o tratamento temático e descritivo da informação. O primeiro ocupa-se do assunto pelo qual um objeto é tratado, enquanto o segundo versa sobre as características do objeto. Para funcionar, a disciplina aludida gera instrumentos que auxiliam na representação do conhecimento, os quais são denominados Sistemas de Organização do Conhecimento (SOC) e visam auxiliar o profissional da informação na representação da informação.

Na contemporaneidade, emerge, no contexto científico, a *e-Science*. Conforme Costa (2017, p. 22), a *e-Science* "se refere a uma nova forma de fazer ciência, cuja principal característica é a produção em grandes volumes de dados que precisa estar *on-line* para facilitar a colaboração entre pesquisadores", orientada por máquina.

Em razão da *e-Science* e do Movimento da Ciência Aberta, movimento esse que versa sobre a abertura das informações para todos, o compartilhamento de dados de pesquisa ganhou notoriedade no contexto acadêmico. Neste caso, os dados são considerados produtos gerados no decorrer das pesquisas científicas, nos mais variados formatos (textos, imagens, sons, planilhas, vídeos etc.) e nas múltiplas áreas do conhecimento.

Neste sentido, a presente pesquisa se debruça sobre informações geradas no âmbito de uma área específica, a da Biodiversidade, mais especificamente no intuito de propor melhorias na recuperação da informação para este domínio. De acordo com a Convenção da Biodiversidade (CBD), a Biodiversidade ou diversidade Biológica

> [...] significa a variabilidade entre os organismos vivos de todas as origens, incluindo, entre outros, ecossistemas terrestres, marinhos e outros ecossistemas aquáticos e os complexos ecológicos dos quais fazem parte; isso inclui a diversidade dentro de espécies, entre espécies e de ecossistemas (Convention on Biological Diversity, 2016, documento não paginado).

Em vista disso, pode-se afirmar que o conceito de Biodiversidade engloba toda a diversidade biológica presente na terra. Por conseguinte, sabe-se que os cientistas desta área fazem uso e compartilham constantemente dados visando o progresso científico. Posto isto, esta pesquisa discorre sobre a compatibilização semântica entre os SOC, no contexto da Biodiversidade, visto a diversidade de bases de dados existentes nesse domínio, e o uso de diferentes vocabulários para a representação desses dados.

Sendo a Ciência da Informação uma ciência social aplicada, a presente pesquisa pode ser classificada como multidisciplinar, interdisciplinar e transdisciplinar, uma vez que versa sobre questões da Ciência da Informação, aplicadas ao domínio da Biodiversidade.

A seguir, serão apresentados a questão da pesquisa, os objetivos e a justificativa.

1.1 QUESTÃO DA PESQUISA

No domínio da Biodiversidade existem diversas iniciativas de compartilhamento de dados de pesquisa, sejam em entidades governamentais, sejam em instituições de pesquisa. No Brasil, dentre as ações de compartilhamento de dados podem ser citados: o Portal da Biodiversidade[1], vinculado ao Ministério do Meio Ambiente; o repositório de dados[2] em Estudos Ecológicos, do Programa de Pesquisa em Biodiversidade (PPBio), ligado ao programa de pesquisa em Biodiversidade, do Centro de Estudos Integrados da Biodiversidade Amazônica; o Portal de Dados da Diretoria de Pesquisas[3], do Jardim Botânico do Rio de Janeiro; e o Sistema de Informação sobre a Biodiversidade Brasileira – SiBBr[4].

Esses dados podem ser apresentados sob múltiplas formas, como texto, imagens, planilhas, *software* etc. Corroborando com esta perspectiva, Daltio e Medeiros (2007), dissertam que esses dados muitas vezes são heterogêneos, por serem coletados pelos mais variados grupos de pesquisadores que utilizam metodologias e vocabulários distintos em suas pesquisas. Albuquerque *et al*. (2010, p. 1, tradução nossa) validam o discurso de Daltio e Medeiros (2007) ao mostrarem que “os estudos de biodiversidade lidam com uma grande

[1] Disponível em: https://portaldabiodiversidade.icmbio.gov.br/portal/. Acesso em: 30 mar. 2021.
[2] Disponível em: https://ppbio.inpa.gov.br/Sobre. Acesso em: 30 mar. 2021.
[3] Disponível em: http://dados.jbrj.gov.br/sobreoportal.php. Acesso em: 30 mar. 2021.
[4] Disponível em: https://www.sibbr.gov.br/. Acesso em: 30 mar. 2021.

variedade de dados, incluindo registros de espécies, geográficos, ecológica, socioeconômica e outro" [5]. Deste modo, a busca por informações de forma integrada na *web* é uma tarefa difícil. Neste sentido, as ontologias podem auxiliar na integração dos dados.

O GEF[6] "Pró-Espécies: Estratégia Nacional de Conservação de Espécies Ameaçada", instituído pelo Ministério do Meio Ambiente (MMA), 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, consequentemente, 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.

Assim se coloca a problemática da presente pesquisa no contexto de um domínio em que já existem muitos sistemas de dados e informações e algumas tentativas de organização, algumas isoladas, outras mais amplas, lideradas pelo governo ou por grupos de pesquisadores, mas que de uma forma ou de outra, fato é que esses dados e informações serão melhores reaproveitados se eles puderem ser localizáveis, acessíveis, interoperáveis e reutilizáveis, em inglês, FAIR (*findable, acessible, interopearable, reusable*).

Em pleno ano de 2023, não há mais possibilidade de se pensar em gestão, organização, representação de dados e informação sem se pensar que esses elementos precisam se tornar FAIR. Embora não seja este o objetivo direto dessa pesquisa, a questão que aqui se levanta tem

---

[5] Texto original: "*Biodiversity studies handle a wide variety of data, including records of species, geographical, ecological, socioeconomics and others*".

[6] O GEF é o financiador de um projeto para o Programa de espécies ameaçadas (Brasil, [2019]).

total relação com a fairificação de dados e informações, pois ela nasce na percepção da necessidade de melhoria na integração semântica entre sistemas de biodiversidade.

Desta forma, o problema que esta pesquisa apresenta baseia-se no seguinte questionamento: **Como melhorar a integração semântica de dados e informações em Biodiversidade?**

A **hipótese** que essa pesquisa defende é a de que a aplicação das teorias que embasam a construção de Sistemas de Organização do Conhecimento (SOC) pode melhorar o tratamento semântico de dados e informações em Biodiversidade, possibilitando a busca integrada e a interoperabilidade semântica entre os sistemas. Acredita-se que a Teoria do Conceito proposta por Dahlberg (1978a), em conjunto com os estudos de Neville (1970, 1972), Dahlberg (1981) e Sales (2022), que versam sobre a integração de SOC, podem contribuir para a integração de dados e informações em Biodiversidade. Essas abordagens citadas, serão desenvolvidas no capítulo que trata sobre Compatibilização Semântica.

1.2 OBJETIVOS

Para a resolução da questão da pesquisa e tentar provar a hipótese da pesquisa, os objetivos definidos foram:

Objetivo geral: Propor diretrizes para a compatibilização conceitual entre diversos vocabulários usados na indexação de bases de dados em Biodiversidade, promovendo a integração semântica neste domínio.

Os objetivos específicos se desdobram em:

a) Identificar, na literatura, o panorama acerca dos SOC voltados para o domínio da Biodiversidade;
b) Compreender os principais conceitos estudados na área da Biodiversidade;
c) Estudar os conceitos de compatibilização semântica propostos pela Ciência da Informação;
d) Investigar as possibilidades de aplicação das técnicas de compatibilização semântica na integração de bases de dados de biodiversidade.

1.3 JUSTIFICATIVA

A presente pesquisa parte da trajetória da pesquisadora, que em sua vida acadêmica dedicou-se aos estudos relacionados à Organização do Conhecimento e suas teorias, começando em seu Trabalho de Conclusão de Curso, passando por grupos de pesquisa, desenvolvimento da Dissertação de mestrado e, dando continuidade, no curso de doutorado. Além disso, surge também do interesse em contribuir diretamente para os estudos de Biodiversidade, considerados pela autora ações fundamentais para o desenvolvimento e permanência da vida em todo o globo terrestre.

Os estudos sobre Biodiversidade são importantes, atuais e impactam diretamente a sociedade, pois envolvem questões acerca da mudança climática, do desmatamento das florestas, dos combustíveis e até mesmo da alimentação (Walls *et al.*, 2014), bem como a geração de energia, alimentos e questões relacionadas à saúde.

Nesse contexto, as informações são importantes para o desenvolvimento de toda a sociedade. Assim, os pesquisadores da área da Ciência da Informação estão se ocupado em disseminar informações sobre a Biodiversidade. Pode-se citar a iniciativa **Saberes do Cerrado**, um projeto desenvolvido pelo Jardim Botânico de Brasília, Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e Universidade Federal de São Carlos (Jardim Botânico de Brasília, 2022). Um dos objetivos centrais do projeto é organizar e disseminar informações sobre os saberes que surgem por meio do Cerrado.

Pode-se citar, ainda, trabalhos desenvolvidos no âmbito do PPGCI IBICT/UFRJ[7], que versam sobre a Biodiversidade e a informação. Como pesquisas doutorais tem-se: "Sistemas agroecológicos: regime de informação de um campus universitário como um Living lab" defendida em 2022, desenvolvido por Lilliana Vallejo, e "Produção de energia elétrica e licenciamento ambiental: cidadania no Brasil em tempo de crise ecológica" defendida em 2017, por Marcia Bettencourt. Tem-se também as dissertações: "As coleções de plantas em herbários: a organização e representação da informação sob aspectos históricos e parâmetros metodológicos" por Thaís Pinheiro, defendida em 2017; "Geociências como área do conhecimento no Brasil" defendida em 2018, por Jéssica Gonsalves; e "A Amazônia nas publicações científicas: mapeando temáticas e atores" defendida em 2018, de Cleiton de Souza.

[7] Programa de Pós-Graduação em Ciência da Informação, do Instituto Brasileiro de Informação em Ciência e Tecnologia, em convênio com a Universidade Federal do Rio de Janeiro.

Isso demonstra o PPGCI IBICT/UFRJ preocupado e atento para questões em que a informação pode contribuir para o desenvolvimento da Biodiversidade.

A presente pesquisa debruça-se sobre os sistemas de informação voltados para o domínio da Biodiversidade. Considerando a existência de vários tipos de sistemas de informação e repositórios de dados no âmbito da Bioinformática, Walls *et al.* (2014) afirmam que:

> Os meios para descrever e inter-relacionar essas diferentes fontes e tipos de dados são essenciais para que esses recursos cumpram seu potencial de flexibilidade de uso e reúso em uma ampla variedade de métodos de monitoramento científicos e aplicações orientadas a políticas (Walls *et al.*, 2014, p. 2, tradução nossa[8]).

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.

Este trabalho pode contribuir com as pesquisas em CI, ao aprofundar a investigação sobre os mecanismos desenvolvidos por ela, como os SOC. Mas, também contribuir com a Biodiversidade, ao mostrar uma forma eficaz de integralização da informação que facilita a recuperação da informação e a interoperabilidade entre sistemas, posto que a Biodiversidade é constituída de informações heterogêneas, e a padronização da informação é essencial para a sua recuperação e uso.

Por conseguinte, o estudo relaciona-se à CI, porque esta pode ser compreendida como a ciência que se dedica aos aspectos gerais da informação, bem como com questões de organização, disseminação e uso da informação. Para Borko (1968, p. 3):

> A Ciência da Informação está preocupada com o corpo de conhecimentos relacionados à origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação, e utilização da informação".

Assim sendo, tem como propósito o estudo acerca da informação, desde a origem até seu uso por parte do usuário. Conforme Saracevic (1996, p. 47),

> A Ciência da Informação é um campo dedicado às questões científicas e à prática profissional voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, no contexto social,

[8] Texto original: "*The means to properly describe and interrelate these different data sources and types is essential if such resources are to fulfill their potential for flexible use and re-use in a wide variety of monitoring, scientific, and policy-oriented applications*".

> institucional ou individual do uso e das necessidades de informação. No tratamento destas questões são consideradas de particular interesse as vantagens das modernas tecnologias informacionais.

Assim, as teorias e as metodologias desenvolvidas por esta ciência podem auxiliar na disseminação e acesso à informação, no contexto da Biodiversidade, considerando que para que ocorra a recuperação da informação é imprescindível organizá-la, de modo a permitir ao usuário o seu acesso na forma completa. Essa ação está totalmente ligada à representação do conhecimento. Deste modo, o uso dos SOC é fundamental para a recuperação da informação, além das teorias englobadas na OC.

Na visão de Smiraglia (2014, p. 3, tradução nossa), a "Organização do Conhecimento é fundamental para o bom funcionamento da Ciência da Informação. Sem o que é aprendido na Organização do Conhecimento, a recuperação da informação não pode funcionar[9]". Isto porque é a disciplina voltada para a elaboração de mecanismos e teorias com vistas à recuperação da informação. Segundo Souza (2007, p. 103), "A Organização do Conhecimento é uma área central de ensino e pesquisa em Ciência da Informação e Biblioteconomia", visto que para a informação ser usada é preciso haver sua recuperação e estar organizada.

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.

Além disso, pensando nos objetivos da presente pesquisa e na questão da pesquisa, que versam sobre integração de vocabulários em sistemas de informação que disseminam dados de pesquisa, cabe citar os princípios FAIR[10], aspectos ligados à gestão de dados, algo que vem sendo bastante discutido na Ciência da Informação e contribui nas mais diversas áreas do conhecimento acerca da gestão de dados de pesquisa.

Na atualidade, não se pode falar sobre gestão de dados, representação dos dados e informação sem mencionar os princípios FAIR. O acrônimo para *Findable* (encontrável), *Acessible* (acessível), *Interoperable* (interoperável) e *Reusable* (reutilizável) versa sobre orientações para gestores e usuários de como aprimorar a gestão de dados, e para tal, apresentam aspetos ligados a metadados, identificadores persistentes, semântica, vocabulários,

[9]Texto original: "*Knowledge organization is critical for the proper functioning of the science of information. Without that which is learned in KO, information retrieval cannot work*".

[10] Os desdobramentos que versam sobre o assunto serão abordados no capítulo sobre metadados.

proveniência e outros, que colaboram com a recuperação da informação não só por parte do humano, mas também por parte da máquina. Os princípios FAIR podem contribuir não só na representação descritiva da informação como na representação temática, uma vez que o objetivo final dos princípios FAIR é o reúso dos dados, passando pela recuperação da informação acessível.

Nesse sentindo, ao se pensar em integração semântica em sistemas voltados para dados de pesquisa, tais como os investigados na presente pesquisa, considera-se a farificação dos dados, uma vez que as diretrizes apresentadas no FAIR podem contribuir diretamente para que a integração semântica ocorra, como por exemplo, o uso de vocabulários FAIR, metadados ricos que favorecem a interoperabilidade, trazendo valor para os dados.

Para Dias, Anjos e Rodrigues (2019, p.181) "para que seja possível uma efetiva interoperabilidade entre conjuntos de dados, é importante que existam instrumentos para padronizar semanticamente os sistemas envolvidos no processo". Nesse contexto, os SOC carecem estar atrelados aos princípios FAIR, isso se deve ao fato de que cada SOC é utilizado em um determinado contexto, e a integração e compatibilização não é uma tarefa fácil. Porém ao se pensar numa aplicação dos princípios FAIR, a integração pode ocorrer de forma padronizada, sendo pensada no objetivo final, que é o reúso dos dados.

Quando voltasse para a primícia I2, que versa sobre "metadados com vocabulários que seguem os princípios FAIR" (GOFAIR, [2022], não paginado), isso quer dizer que, o sistema de informação deve garantir que o vocabulário seja padronizado visando o seu uso futuro, adequando-se às mudanças que possam ocorrer, bem como que a recuperação da informação seja possível tanto por máquina quanto por humano.

Para Xu *et al.* (2022), um vocabulário FAIR é capaz de fornecer um espaço para acrescentar anotações em determinados casos, onde os dados são gerados em domínios diferentes, assim podem permitir a interoperabilidade dos dados. Essa interoperabilidade só seria possível se nesse espaço para anotações coubesse um mapeamento dos conceitos dos vocabulários utilizados. Aqui, voltamos aos teóricos dentro da Ciência da Informação, que versam sobre compatibilização semântica, que pregam que não é necessário a construção de novos SOC, mas sim permitir a sua integração., mas sim elaborar estratégias que permitam a compatibilização semântica sem perda de informação.

Com a finalidade de apresentar algumas práticas para garantir a aplicação dos princípios FAIR em SOC, Garijo e Poveda-Villalón (2020) discorrem que para que os SOC se tornem FAIR, efetivamente, deve-se atentar aos metadados e seus registros, no sentido de que os metadados serão responsáveis por garantir a interpretação correta da informação, além de

apresentar uma visão de como ocorre a divisão dos SOC, como por exemplo em ontologias que podem ser divididas em classes, propriedades e objetos. Assim sendo, a informação se tornaria mais acessível para os usuários e futuros gestores.

Dessa forma, os princípios FAIR e sua aplicação em sistemas de informação podem garantir a recuperação da informação de forma eficiente e, posterior a isso, um aprimoramento na capacidade de reúso dos dados.

Isto posto, a presente pesquisa pode contribuir no aprimoramento de aspectos ligados ao FAIR que versam sobre vocabulários e interoperabilidade. Acredita-se também poder contribuir (mesmo que não seja o objetivo central) nas discussões que ocorrem na iniciativa GO FAIR Brasil, não só em Ciência da Informação, bem como na Biodiversidade, dado que as instituições fontes dos dados da presente pesquisa possuem pesquisadores envolvidos na iniciativa GO FAIR Brasil Biodiversidade.

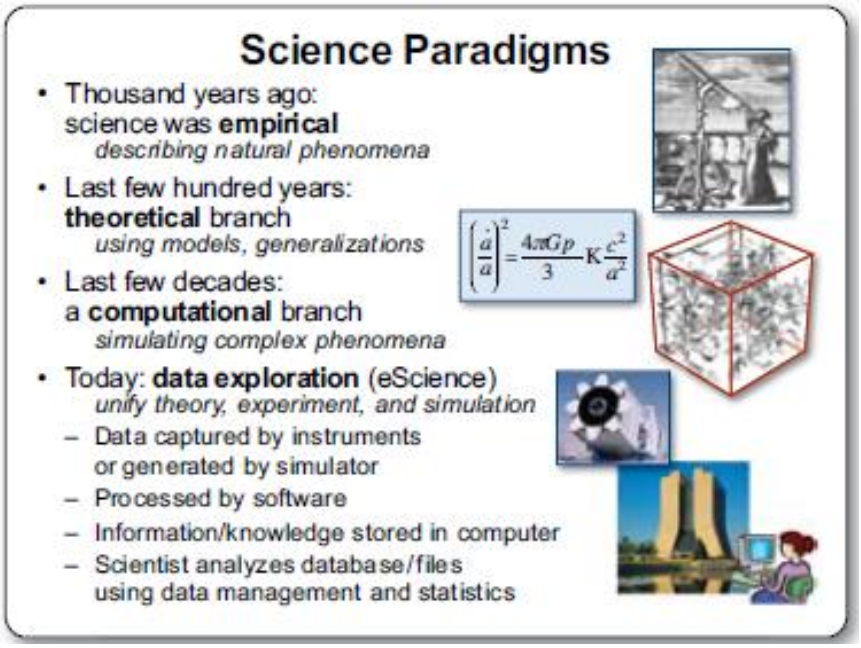
Serão apresentadas a seguir as Seções que constituem o embasamento teórico desta pesquisa, a começar com a apresentação dos conceitos de Gestão de dados, Metadados e Compatibilização Semântica e seguidas dos Procedimentos Metodológicos (natureza da pesquisa, tipo de pesquisa, instrumentos utilizados para a coleta de dados), Resultados e Discussão, Proposta final de compatibilização semântica e as Considerações Finais.

2 GESTÃO DE DADOS CIENTÍFICOS

No contexto da ciência atual, surge uma nova forma de fazer ciência, onde a produção de dados em larga escala ocorre e necessita estar organizada e disponível para a sua utilização. O fenômeno dentro da ciência, onde os dados são produzidos em larga escala, a ciência é intensiva em dados e o uso de máquinas é bastante utilizado, é definido como *e-Science*. A *e-Science* é um dos paradigmas da ciência. Em suma, os paradigmas da ciência são classificados em quatro fases, de acordo como ela ocorre.

Segundo Gray (2009), o primeiro paradigma é conhecido como a fase empírica da ciência, ou seja, os estudos eram desenvolvidos por meio dos experimentos e ocorreu há mais de mil anos. O segundo paradigma tratava da ciência teórica. O terceiro foi caracterizado pelo fenômeno computacional, onde começaram a ocorrer as simulações na ciência, e o quarto, que é o atual, é conhecido como *e-Science*, é onde ocorre o grande uso e geração dos dados. É uma ciência orientada por dados. A Figura 1, de autoria de Gray (2009), representa bem está divisão:

Figura 1 – Os paradigmas da ciência



Fonte: Gray (2009, p. 18).

Segundo Gray (2009), a *e-Science* é caracterizada pela grande exploração dos dados, onde esses dados são gerados por instrumentos ou produzidos por simulação, processados por *softwares*, em que a informação fica armazenada nos computadores e os cientistas passam a ser analistas de dados, fazendo essas análises muitas vezes por métodos estatísticos. Dessa forma, pode-se afirmar que a *e-Science* é um novo modo de se fazer ciência, onde o uso das máquinas

e a geração dos dados é algo intensivo. Segundo Büttner, Hobohm e Müller (2011, p. 13, tradução nossa),

> Em conexão com as tecnologias de informação, os dados tornam-se simultaneamente base e resultado do processo do conhecimento científico. Os dados vão relacionando a dados de outras fontes e formam um novo banco de dados para uso posterior. Isso geralmente acontece em um ambiente técnico, ambiente que apoia a pesquisa em rede e colaborativa. O nome disso agora é *e-Science* (ciência aprimorada) [11].

Assim, pode-se afirmar que os dados passam a ser os protagonistas no processo de desenvolvimento da ciência e a *e-Science* é uma ciência cujas tecnologias facilitam o seu aperfeiçoamento.

Mesmo com a recente mudança de paradigma no modo de se fazer ciência, muitos dos dados produzidos ainda estão somente no computador ou na instituição de quem os criou. O movimento da Ciência Aberta trabalha para que isso seja transformado.

As mudanças na ciência, alinhadas ao grande uso das tecnologias de informação, favorecem o surgimento do movimento Acesso Aberto da ciência, cujo foco é facilitar o acesso à ciência para os cientistas e para a sociedade. Para Albagli, Clinio e Raychtock (2014, p. 435), a Ciência aberta é:

> [...] um termo guarda-chuva, que engloba diferentes significados, tipos de práticas e iniciativas, bem como envolve distintas perspectivas, pressupostos e implicações. Aí estão incluídas desde a disponibilização gratuita dos resultados da pesquisa (acesso aberto), até a valorização e a participação direta de não cientistas e não especialistas no fazer ciência, tais como "leigos" e "amadores" (ciência cidadã).

Nesse contexto, as autoras apresentam ações que tanto englobam atividades realizadas pelo cientista, quanto pela sociedade. São elas o acesso aberto a publicações científicas, educação aberta, elaboração de ferramentas e materiais científicos abertos, ciência cidadã, cadernos de pesquisa abertos e dados abertos de pesquisa.

Sobre os dados abertos de pesquisa, vem ganhando destaque a abertura desses dados no âmbito científico e os cientistas já têm se mobilizado com relação a isso, o que fica claro ao se fazer buscas acerca da temática de gestão de dados e os teóricos discutirem sobre a possibilidade do reúso. Alonso-Arévalo (2019, p. 77, tradução nossa) define dados abertos como:

---

[11]Texto original: "*In Verbindung mit Informationstechnologien werden Daten somit gleichzeitig Grundlage und Ergebnis wissenschaftlicher Erkenntnisprozesse. Die Daten werden zu Daten aus anderen Quellen in Beziehung gesetzt und bilden eine neue Datenbasis für weitere Berechnungen. Dies geschieht häufig in einer technischen Umgebung, die vernetztes und kooperatives Forschen unterstützt. Die inzwischen gängige Bezeichnung dafür ist eScience (enhanced science)*".

> [...] são definidos como abertos quando podem ser usados livremente, modificados e compartilhados por qualquer pessoa para qualquer finalidade, criando um bem comum do qual todos podem participar. Os dados abertos são uteis para milhões de pessoas em todo mundo, pesquisadores, empresas e cidadãos[12].

Nesse contexto, a abertura dos dados faz com que a ciência avance sem necessariamente reproduzir os custos, pois a disponibilização desses dados garante o acesso sem refazer os mesmos passos do primeiro cientista coletor dos dados. Com relação à reprodução de pesquisas por meio dos dados abertos, Curty (2017, p. 2) afirma que as tecnologias impulsionam essa ação "ferramentas computacionais avançadas para compartilhamento e distribuição de dados estão pavimentando o caminho para uma melhor reprodutibilidade nas investigações científicas". Ou seja, as ações de abertura dos dados contribuem para a criação de ambientes colaborativos que favorecem a reutilização dos dados nas pesquisas.

Para Albagli, Clinio e Raychtock (2014, p. 440), além do reúso, a abertura dos dados permite uma maior qualidade no fazer científico:

> No campo científico, trata-se da publicização de dados primários de uma pesquisa, considerada uma ação fundamental para sua reprodutibilidade e reutilização em pesquisas derivadas ou não, além de permitir o amplo escrutínio -- o que pode contribuir para expor inconsistências, baixa qualidade, plágio ou fraude.

Um exemplo de que a não disseminação dos dados fez com que a pesquisa fosse questionada, foi o do artigo publicado no periódico britânico Lancelot (Bertoni, 2020), sobre o uso da cloroquina e sua derivada hidroxicloroquina, para o tratamento do Sars Covid-2, durante a pandemia de 2020. O artigo em questão afirmava que foram utilizados dados de 96 mil pacientes, internados em 671 hospitais de seis continentes, entre 20 de dezembro de 2019 e 14 de abril do ano de 2020. Porém, um jornal descobriu que a empresa que fornecera os dados contava apenas com seis funcionários. A comunidade científica começou a questionar a qualidade da pesquisa e, inclusive, o próprio periódico que tinha publicado o artigo. Após isso, no mês de junho, os três autores publicaram uma retratação no periódico afirmando que não tiveram acesso aos dados completos e por isso não poderiam garantir a autenticidade dos resultados da pesquisa. Se os avaliadores do periódico Lancelot tivessem tido acesso aos dados no momento da sua avaliação, esse transtorno poderia ter sido evitado.

---

[12]Texto original: "*Los datos se definen como abiertos cuando se pueden utilizar libremente, modificar y compartir por cualquiera para cualquier propósito, incluyendo la creación de un bien común en el que cualquiera puede participar. Los datos abiertos son de utilidad para millones de personas en todo el mundo, investigadores, empresas y ciudadanos*".

Segundo Borgman (2012), os dados são o estímulo que impulsiona a pesquisa em qualquer área do conhecimento. Seu formato dependerá da sua origem, finalidade e área do conhecimento. Como exemplo, podemos citar uma pedra, que pode ser dado tanto para a História quanto para a Geologia. Uma fotografia pode ser dado para a Antropologia, História, Medicina, Biodiversidade e outras disciplinas. A *Organisation for Economic Co-operation and Development* (OECD) (2007, p. 13, tradução nossa) define dados de pesquisa como

> [...] registros factuais (números, registros textuais, imagens e sons) usados como fontes primárias para a pesquisa científica, e que são comumente aceitos na comunidade científica como necessário para validar resultados da pesquisa[13].

Como se pode ver, dados de pesquisa são fundamentais para legitimar a pesquisa, pois são coletados por meio de um método específico. Borgman (2012) disserta que na Física ou na Biologia os dados podem ser experimentais, dados de observação ou modelos. Bertin, Visoli e Druker (2017, p. 38) discorrem que "Dados de pesquisa, por sua vez – e também de modo simplificado –, são todo o tipo de registro produzido, compilado ou utilizado no decorrer da pesquisa". Isto é, tudo aquilo que é gerado e empregado no âmbito da pesquisa é dado de pesquisa. Estevão, Arns e Strauhs (2019, p. 4) corroboram com esse pensamento:

> Esses dados são produtos de pesquisa que cobrem uma ampla gama de tipos de registros e podem ser estruturados e armazenados em vários formatos de arquivos. Embora a maioria desses dados seja originado em formato digital, todos os dados de pesquisa são incluídos na conceituação, independentemente do formato.

Nesse sentido, os dados de pesquisa não são gerados apenas em formatos digitais, cadernos de anotações, fotos analógicas, desenhos, cálculos feitos a mão, além de outros, são dados de pesquisa desde que tenham sido gerados em seu contexto. Podem ser produzidos das mais variadas formas. Segundo Córdula e Araújo (2019, p. 191),

> os dados são criados ou produzidos de várias maneiras: por meio de observação, de visualização, de monitoramento e de sensores, na criação de metadados, por análise de comportamento, por cálculos matemáticos e estatísticos, etc.

Assim sendo, sua forma final dependerá do modo e instrumento com o qual foi gerado. Por sua vez, Sales *et al.* (2019, p. 306) descrevem os dados evidenciando a sua importância para a

[13]Texto original: "*'research data" are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings*".

ciência “dados científicos são ativos informacionais imprescindíveis para o progresso da ciência e para a viabilização de novas descobertas que vão das ciências exatas às humanidades, arte e cultura”. Corroborando com esse pensamento, Curty (2017, p. 4) disserta que:

> Dados de pesquisa constituem matérias-primas importantes para a ecologia da ciência e são essenciais para novos ciclos de criação de conhecimento científico, pois fornecem insumos para um processo interativo no ciclo de vida da investigação, permitindo continuidade da descoberta científica e da inovação tecnológica.

Na visão da autora, os dados de pesquisa são elementos primordiais para todo o desenvolvimento da ciência e tecnologia, é por meio deles que a ciência estará sempre se renovando. Segundo Bertin, Visoli e Druker (2017, p. 41), os dados de pesquisa podem ser apresentados em uma variedade de suporte:

- Documentos (texto, Word), planilhas.
- Atas de laboratório, cadernos de campo, diários de pesquisa.
- Questionários, transcrições, tabelas de codificação.
- Fitas/CDs/DVDs de áudio e vídeo.
- Fotografias, filmes.
- Resultados de ensaios.
- Slides, artefatos, espécimes, amostras.
- Coleção de objetos digitais adquiridos e produzidos durante o processo de pesquisa.
- Arquivos de dados estatísticos ou de outra natureza.
- Conteúdo de bancos de dados (vídeo, áudio, texto, imagens).
- Modelos, algoritmos, scripts.
- Conteúdo de uma aplicação (inputs, outputs, arquivos de log para análise de software, softwares de simulação, esquemas).
- Metodologias e *workflows*.
- Procedimentos operacionais padrão e protocolos.
- Mapas e arquivos de dados espaciais, como *shapefiles* e imagens de satélite.

Como foi visto, dados de pesquisa são insumo para o funcionamento da ciência e podem ser descritos das mais variadas formas. Enquanto insumo para a ciência, e no contexto da ciência aberta, este dever ser aberto e posteriormente compartilhado. Segundo Borgman (2012, p. 1067-1071, tradução nossa), as razões e benefícios para o compartilhamento dos dados são:

> Reproduzir ou verificar pesquisas;
> Disponibilizar os resultados de pesquisas financiadas com dinheiro público ao público;
> Permitir que outras pessoas façam novas perguntas sobre dados existentes;

> Avançar no estado de pesquisa e inovação[14].

Um exemplo sobre o avanço das pesquisas com relação à abertura dos dados por parte do trabalho colaborativo é o da vacina contra a gripe suína (H1N1). Segundo González *et al.* (2013), a vacina foi gerada em três meses após o surto, porém as pesquisas e o compartilhamento dos dados já ocorriam há 10 anos antes do surto. O compartilhamento dos dados se faz importante para todas as áreas da saúde e sobretudo na Medicina. Segundo o Departamento de Saúde e Serviços Humanos dos EUA, em sua página na internet, ao tratar sobre o compartilhamento de dados afirma que "o compartilhamento de dados é essencial para a tradução rápida dos resultados da pesquisa em conhecimento, produtos e procedimentos para melhorar a saúde humana" (US Departamento of Health and Human Services, [2020], não paginado). O trabalho colaborativo que existe na abertura dos dados contribui de maneira significativa para a geração de benefícios de toda a sociedade.

Além disso, auxilia no progresso das pesquisas. Estevão, Arns e Strauhs (2019) afirmam que a abertura dos dados auxilia no entendimento do desenvolvimento das pesquisas:

> A importância da disponibilidade dos dados da pesquisa se dá principalmente pela possibilidade de compreender o processo de geração daquele conhecimento, sem obscuridades. Nessa perspectiva, a disponibilidade de dados da pesquisa ampliaria o reuso dos mesmos não apenas para validações, mas para novas inferências a partir de abordagens distintas, tornando o processo de construção científica transparente, aberto e democrático (Estevão; Arns; Strauhs, 2019, p. 10).

Nesse contexto, surgem estruturas de informação que são próprias para o compartilhamento dos dados, *data publishing* ou publicações de dados, em português. As publicações de dados são os *data paper* (artigo de dados), *data journal* (periódico de dados*)* e os repositórios de dados. Para Pampel e Kindling (2017), os debates acerca de *data publishing* ganharam força nos últimos anos por causa do movimento Acesso Aberto. Estes autores dissertam que:

> Para criar incentivos para os pesquisadores tornarem seus dados acessíveis, estratégias de publicação foram estabelecidas entre bibliotecários, editores e os próprios cientistas nos últimos anos, publicando estratégias que garantem o reconhecimento para quem disponibiliza os dados por outros pesquisadores[15] (Pampel; Kindling, 2017, p. 18).

---

[14]Texto original: "*To Reproduce or to Verify Research. To Make Results of Publicly Funded Research Available to the Public. To Enable Others to Ask New Questions of Extant Data. To Advance the State of Research and Innovation*".

[15]Texto original: "*Um Forschenden Anreize zur Zug¨anglichmachung ihrer Daten zu schaffen, haben sich im Zusammenspiel von Wissenschaft, Bibliotheken und Verlagen in den vergangen Jahren Publikationsstrategien etabliert, die den Forschenden, die Forschungsdaten Dritten bereitstellen, entsprechende*".

Portanto, as publicações de dados surgem como alternativas para quem quer compartilhar os dados e para quem quer ter acesso a eles. Todas as suas estruturas são pensadas nas formas como os dados são desenvolvidos. Repositórios, artigos e periódicos digitais já existiam, agora alguns são elaborados especialmente para os dados. Pampel e Kindling (2017) declaram que em algumas disciplinas científicas existem ofertas de estruturas de informação para dados desde a década de 50, do século passado.

Segundo a *Oregon State University Libraires* (2017, tradução nossa), um artigo de dados "descreve completamente os conjuntos de dados e geralmente não incluem qualquer interpretação ou discussão (uma exceção pode ser a discussão de diferentes métodos para coletar os dados, por exemplo)"[16]. São artigos que publicam somente os dados gerados na pesquisa, sem apresentar a estrutura de texto, como existem nos artigos científicos como introdução, metodologia e considerações finais.

Os periódicos de dados são outra forma de publicar os dados com informações específicas voltadas para a sua reutilização. Candela *et al.* (2015) informam que os periódicos de dados proporcionam a publicação de dados, tornando os conjuntos de dados acessíveis por meio de metadados, conforme padrões estabelecidos. Estes periódicos seguem o mesmo padrão de avaliação pelos pares como os tradicionais periódicos científicos.

Por sua vez, os repositórios de dados também são criados fornecendo informações e tecnologias para a publicação e preservação dos dados. O OpenAIRE (2018, não paginado, tradução nossa) define repositório de dados como "um arquivo digital que coleta e exibe conjuntos de dados e seus metadados" [17]. Os repositórios de dados podem estar ligados a instituições ou a grupos de pesquisa. Acerca dos repositórios de dados Sayão e Sales (2016, p. 96) declaram:

> São infraestruturas de base de dados desenvolvidas para apoiar todo o ciclo da gestão de dados de pesquisa, incluindo as ações mais dinâmicas e contundentes sobre os dados, que coletivamente são chamadas de curadoria de dados de pesquisa, que visam adicionar valor aos dados, avaliando, formatando, agregando e derivando novos dados.

Conforme dito acima, os repositórios são essenciais para a vida dos dados e seu compartilhamento. Porém só a criação de estruturas de publicação de dados não garante a

[16]Texto original: "*Data papers thoroughly describe datasets, and do not usually include any interpretation or discussion (an exception may be discussion of different methods to collect the data, e.g.)*".
[17]Texto original*:* "*a digital archive collecting and displaying datasets and their metadata"*.

sobrevivência dos dados, é necessário que ocorra a gestão dos dados de pesquisa. A gestão irá garantir uma melhor organização das informações e estruturas relacionadas aos dados, visando a preservação a longo prazo dos dados, garantindo o uso, compartilhamento e reúso dos mesmos. Segundo Majid, Zhang e Foo (2018, p. 2, tradução nossa), "dados de pesquisa gerenciados de maneira adequada fornecem credibilidade ao processo de pesquisa, bem como a integridade de suas descobertas"[18]. É a gestão que trará segurança e organização para o compartilhamento e preservação dos dados de pesquisa. Ainda segundo Majid, Zhang e Foo (2018), a gestão de dados é algo extenso, isso porque engloba várias atividades, elaboradas por diversos atores e influenciada pelos mais variados fatores. Conforme Tripathi, Shukla e Sonkar (2017, p. 418, tradução nossa),

> O gerenciamento de dados de pesquisa envolve todas as atividades e processos que são realizados ou feitos para garantir que os dados de pesquisa sejam devidamente documentados, organizados, armazenados, arquivados e com curadoria de modo que estejam disponíveis para acesso, uso e reutilização, sempre que houver necessidade após a pesquisa ter sido realizada e relatada[19].

Em linhas gerais, a gestão de dados será desenvolvida por meio de processos que visam garantir a reutilização dos dados. Para a Biblioteca Pública Estadual de Ciência e Tecnologia ([2019], não paginado, tradução nossa) a gestão de dados de pesquisa é

> parte integrante do ciclo de vida de um projeto científico, incluindo coleta, documentação, armazenamento, backup, compartilhamento, integridade, segurança, controle de versão, planejamento robusto e gerenciamento estratégico de dados. Escolher os formatos de dados corretos (estruturados e não estruturados), ontologias e ferramentas de software necessárias para conduzir experimentos ou criar um conjunto de dados, é uma etapa importante no ciclo de pesquisa. Os formatos e nomes de arquivos em conformidade com os padrões garantem que os dados possam ser identificados e acessíveis no futuro. Os dados geralmente requerem explicação, portanto, devem ser acompanhados por metadados (informações que descrevem os dados)[20].

---

[18]Texto original: "*Properly managed research data provides credibility to the research process as well as enhances the integrity of its findings*".

[19]Texto original: "*Research data management entails all activities and processes which are undertaken or done to ensure that research data is properly documented, organised, stored, archived and curated so that it is available for access, use and reuse whenever the need arises after the research has been done and reported*".

[20]Texto original do russo: "*Управление исследовательскими данными является неотъемлемой частью жизненного цикла научного проекта, включает в себя сбор, документирование, хранение, создание резервных копий, совместное использование, обеспечение целостности, безопасности, управление версиями, надежное планирование и стратегическое управление данными. Выбор корректных форматов данных (структурированных и неструктурированных), онтологий и программных средств, необходимых для проведения экспериментов или создания набора данных, является важным этапом исследовательского цикла. Соответствующие стандартам форматы и имена файлов гарантируют, что данные*

Portanto, a gestão de dados deve ser pensada durante toda a pesquisa e deve ser desenvolvida por meio de ações relacionadas à organização da informação dos dados e estrutura de compartilhamento, sempre visando que o dado possa ser utilizado mesmo após o final da pesquisa. Segundo Alonso-Arévalo (2019, p. 82, tradução nossa), a gestão de dados trabalha em cima de duas linhas: "preservação de longo prazo de conjuntos de dados nos sistemas de armazenamento e compartilhamento e reutilização de conjuntos de dados para a pesquisa[21]". À medida que ações, que tratam da preservação dos dados, são executadas, o compartilhamento e posteriormente seu reúso serão facilitados.

Explicitados o que vem a ser dados de pesquisa, bem como a função da gestão de dados, apresenta-se a seguir os tipos de dados desenvolvidos em Biodiversidade.

2.1 DADOS EM BIODIVERSIDADE

Conforme citado acima, 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 consequentemente a quantidade e a diversidade de dados necessários para a compreensão dessas mudanças, suas consequências 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, não paginado). Estes dados são coletados pelos biólogos, ecólogos e outros profissionais que tratam

---

могут быть идентифицированы и доступны в будущем. Данные нередко требуют пояснений, поэтому они должны сопровождаться метаданными (информацией, которая описывает данные). Использование соответствующих методов хранения и резервного копирования помогает защитить данные исследований от возможных потерь, а также обеспечивает доступ к ним в долгосрочной перспективе".

[21]Texto original: "*La preservación a largo plazo de los conjuntos de datos mediante sistemas de almacenamiento 2. Compartición y reutilización de los conjuntos de datos para la investigación [...]*".

da Biodiversidade e são armazenados nos museus e herbários. Por sua vez, ainda de acordo com Pereira e Peterson (2001, não paginado), "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 (Research Data Alliance, 2016) – 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.

Um outro tipo de dado gerado no contexto da Biodiversidade é o dado biogeográfico. Conforme König *et al.* (2019, p. 2, tradução nossa), "Os dados biogeográficos podem, portanto, estar ligados a uma ampla gama de organismos (por exemplo, taxonômicos, funcionais, filogenéticos) e de informações ambientais (por exemplo, clima, solo, topografia)"[22]. Esses dados são fundamentais para os pesquisadores compreenderem como o ocorre a disposição dos seres vivos na terra.

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.*

[22] Texto original: "*Biogeographical data can therefore be linked to a wide range of organismic (e.g., taxonomic, functional, phylogenetic) and environmental (e.g., climate, soil, topography) information*".

(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 (Quadro 1), podem ser classificados como:

Quadro 1 – Classificação dos dados em Biodiversidade

Classe	**Subclasse**	**Tipo**	**Exemplo**
Origem	Dados de pesquisa Dados para pesquisa	Observacionais Governamentais Computacionais	Dados de ocorrência de espécies; Dados sobre mudança climática; Dados biogeográficos.
Processamento	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.
Abordagem da pesquisa	Qualitativo Quantitativo	Multimidia Número	Imagens; Fotografia; Vídeo; Dados sobre conservação de espécies.
Materialidade	Digital Físico	Multimidia Amostra	Dados disponibilizados em repositórios digitais; Coleções de espécies.

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

O Quadro 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):

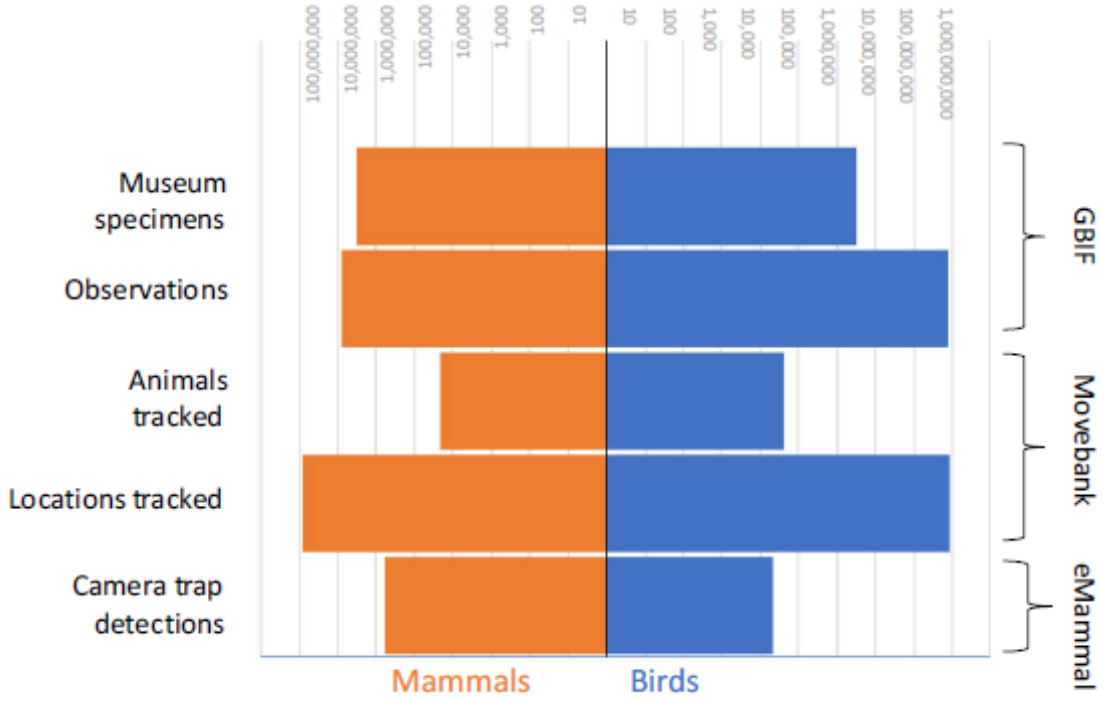
> a criação de bancos de dados por projetos que envolvem a coleta e o estudo de espécimes biológicos, as espécies a que pertencem e os locais em que ocorrem constitui-se num pré-requisito indispensável à conservação da biodiversidade e uso sustentável dos recursos naturais.

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, McShea e Wikelski (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. A Figura 2 mostra o quantitativo de dados de mamíferos e pássaros presentes em iniciativas que disponibilizam dados digitais sobre Biodiversidade.

Figura 2 – Tamanho total das bases de dados para pássaros e mamíferos

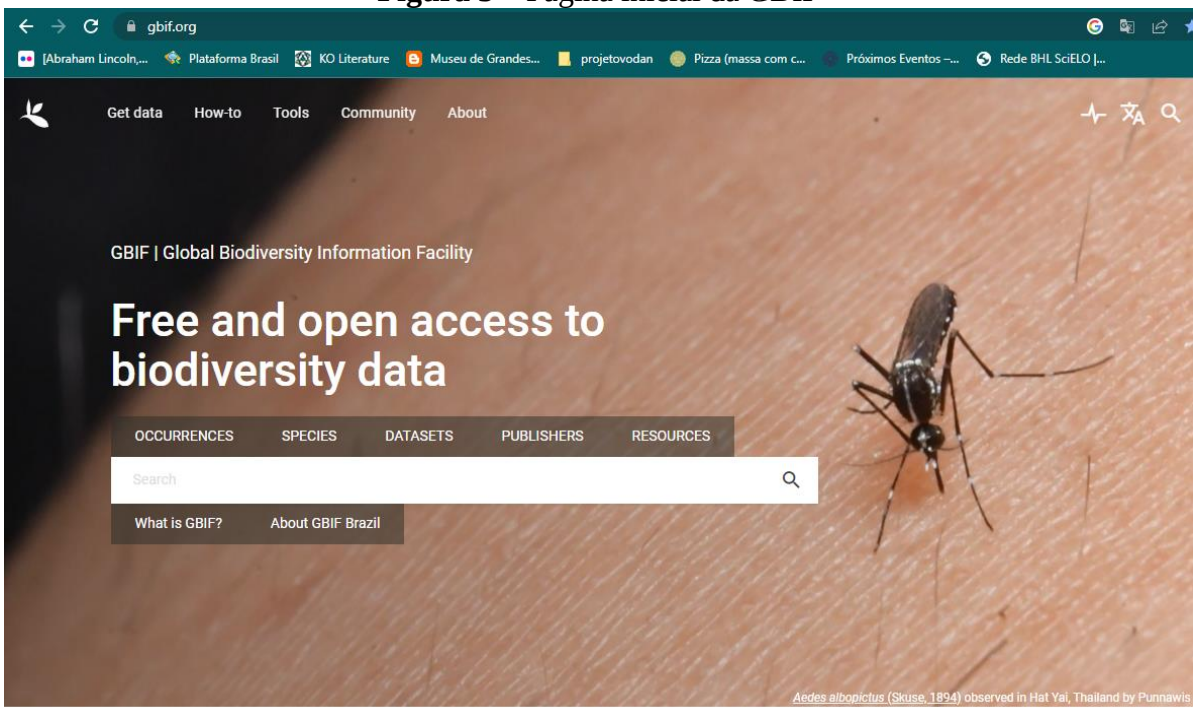


Fonte: Kays, McShea e Wikelski (2020, p. 645).

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. Cabe citar a iniciativa *Global Biodiversity Information Facility* (GBIF), que é uma rede internacional de infraestrutura de dados financiada pelos governos do mundo e destinada a fornecer a qualquer

pessoa, em qualquer lugar, acesso aberto a dados sobre todos os tipos de vida na Terra. Essa iniciativa conta com mais de 70 mil conjuntos de dados sobre a biodiversidade.

Figura 3 – Página inicial da GBIF



Fonte: *Global Biodiversity Information Facility* (2022).

O GBIF conta também com uma iniciativa brasileira, que atualmente disponibiliza mais de 4 mil conjuntos de dados. Um dos sistemas que se articula com o GBIF internacional é o Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr[23]). O SiBBr compartilha e integra dados de forma online e é formado por 161 instituições. O SiBBr é a ligação entre o Brasil e o GIBF. Ao se pesquisar sobre dados no SiBBr, pode-se perceber os mais variados tipos de dados presentes no sistema de informação, tais como: dados de ocorrência de espécies, dados taxonômicos, conservação de espécies, manejo de pragas, interação e comportamento de espécies, e outros. Os dados disponíveis no SiBBr estão apresentados nos mais variados formatos, podendo o usuário entrar em contato e citar o pesquisador primário que o coletou.

Sabe-se que esses dados, tantos 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. De acordo com Sayão e Sales (2022, p. 33):

> Isso significa dizer que estudar as propriedades do dado e como ele se manifesta em cada disciplina, é condição necessária para a construção de

[23] Disponível em: https://www.sibbr.gov.br/?lang=pt_BR. Acesso em: 20 maio 2022.

> critérios que tornarão o dado de pesquisa passível de ser selecionado, arquivado e preservado, de acordo com suas características.

Isto acontece porque o dado, assim como um documento, precisa de representação, para ser compreendido, recuperado, contextualizado e reusado. É essa representação que torna o dado informação, ou sob o olhar da teoria dalberghiniana (Dahlberg, 1978a), uma unidade de conhecimento formada pela tríade: referente, característica e representante.

Essas características, no âmbito da gestão de dados de pesquisa, se tornam metadados que descrevem o dado para além de sua estrutura, mas guardam também informações técnicas, administrativas e de preservação. Desta forma, o dado é visto como um conceito que, para ser gerenciado, precisa estar estruturado adequadamente dentro de um Sistema de Organização do Conhecimento (SOC).

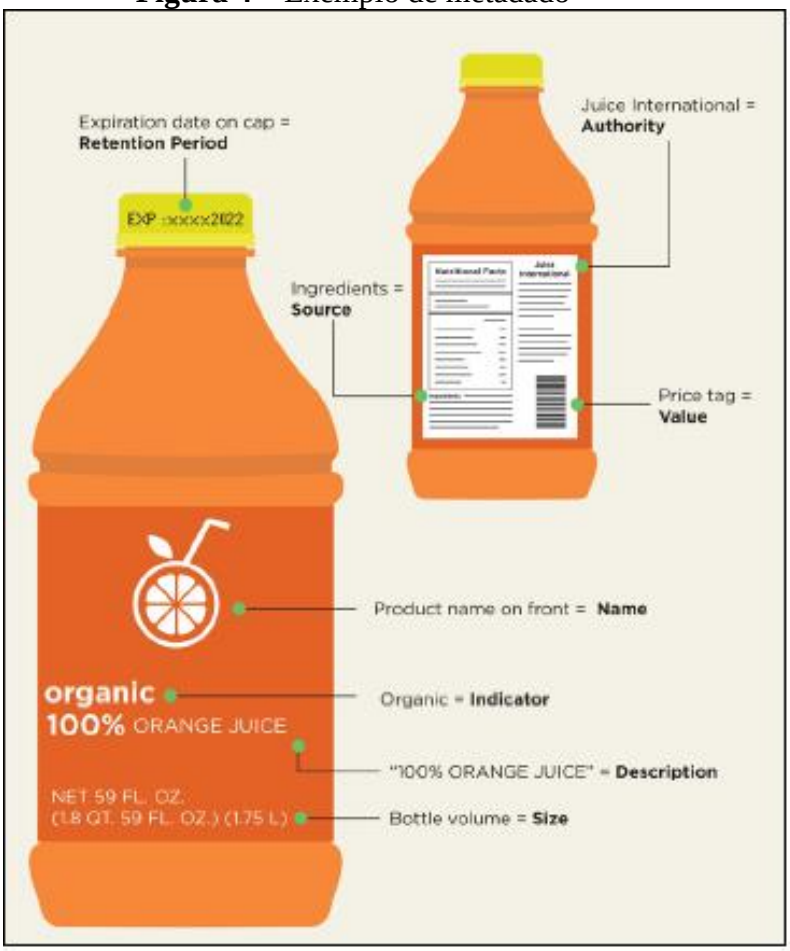
A seção a seguir trata desse elemento tão significativo para a gestão de dados que é o Metadado. No âmbito da gestão de dados, o Metadado é um elemento essencial, considerando que as suas funções foram se expandindo, ao passo que as tecnologias de informação também. O metadado está ligado aos mais variados aspectos da gestão, passando pela precisão das descrições dos dados, atribuição da proveniência dos dados, apresentação de informações que possam contribuir com a curadoria dos dados, além do seu papel na citação dos dados que forem reusados.

3 METADADOS

O ambiente digital proporcionou o surgimento dos mais variados tipos de sistemas de informação, com os mais variados tipos de acervos, tais como bibliotecas digitais, repositórios digitais e de dados, base de dados e outros. Deste modo, surgem demandas de gestão, descrição, preservação e recuperação da informação. Nesse contexto, os metadados que antes desempenhavam apenas função de descrição, passam a desempenhar funções específicas para atender outras necessidades do sistema de informação.

Metadados são mecanismos que transmitem informações acerca de informação. Segundo Mori e Carvalho (2004, p. 1), o termo originou-se do latim *metá* que significa “além”, “através de” ou “sobre”. Como um exemplo sobre o que são metadados, tem-se a Figura 4, a qual mostra como os dados sobre outros dados podem ser apresentados de maneira simples.

Figura 4 – Exemplo de metadado



Fonte: *National Forum on Education Statistics* (2021, p. 3).

Como se pode observar na Figura 4, tem-se uma garrafa de suco de laranja cujo rótulo da garrafa contém muitos metadados ou contexto que transmite informações importantes, tais como a quantidade de suco, o sabor do suco, o produtor e até mesmo os ingredientes que compõem o suco. Assim sendo, os dados sobre dados são fundamentais para o acesso à informação. Conforme o *National Forum on Education Statistics* (2021, p. 1), metadados são:

> [...] informações estruturadas que descreva, explique, localize ou torne mais fácil recuperar, usar ou gerenciar uma fonte de informação. Em outras palavras, os metadados fornecem o contexto no qual interpretar os dados.

Essa definição apresenta pontos e funções dos metadados atualmente, os quais são responsáveis não só por descrever informações, mas também gerenciar e facilitar o seu uso. Segundo Castro e Simionato (2020), o propósito do uso dos metadados tem seu início nos princípios da catalogação em bibliotecas, mas que com o advento das tecnologias de informação o seu uso foi ampliado. Isso porque, em ambiente digital, deve-se garantir não só a descrição, mas também a gestão e o uso.

> Assim, o desafio está em promover uma representação adequada dos recursos informacionais, garantindo não só sua recuperação, mas também seu acesso, sua preservação, seu uso e reuso, além de proporcionar a interoperabilidade dos dados entre os diferentes acervos e na web (Alves, 2017, p. 97).

Dessa forma, tem-se a ampliação do escopo das funções dos metadados e com isso, surgem metadados com funções específicas. No quadro a seguir, são apresentados os tipos de metadados e suas definições.

Quadro 2 – Definições de metadados

Tipo	Definição	Exemplos
Administrativo	Metadados usados no gerenciamento e administração de coleções e recursos de informação	Aquisição de informação; Rastreamento de direitos e reprodução.
Descritivo	Metadados usados para identificar e descrever coleções e afins recursos de informação	Registros de catalogação; Diferenciação entre versões.
De preservação	Metadados relacionados à gestão de preservação de coleções e recursos de informação	Documentação da condição física dos recursos; Documentação de ações tomadas para preservar versões físicas e digitais de recursos, por exemplo, atualização de dados e migração.
Técnicos	Metadados relacionados às funcionalidades dos sistemas	Documentação de hardware e software; Autenticação e dados de segurança, por exemplo, chaves de criptografia, senhas.

Tipo	Definição	Exemplos
De uso	Metadados relacionados ao nível e tipo de uso dos acervos e recursos de informação	Uso e rastreamento do usuário; Registros de pesquisa.
De proveniência	É um mecanismo para fornecer dados sobre entidades e seus relacionamentos com o recurso e com outras entidades.	Histórico de alterações do esquema;
De autenticação	Suportam a avaliação da integridade de um objeto de informação, legitimidade e qualidade geral genuína	Assinatura digital

Fonte: Elaborado pela autora baseado em Gilliland (2008), Sayão (2010) e Arakari (2019).

Como se pode observar, existem os mais variados tipos de metadados, cada um com uma função específica, que versam desde seu objetivo original de auxiliar a descrição em bibliotecas, como os metadados descritivos, até os sistemas que permitem assinatura eletrônica que funcionam com o suporte dos metadados de autenticação. Também é possível observar que, conforme os ambientes digitais vão se desenvolvendo, os metadados vão assumindo funções cada vez mais específicas. Segundo Sayão (2010, p. 3),

> Na medida em que a ideia de metadados se torne uma parte essencial do mundo digital, eles se mostram conceitualmente mais complexos e mais abrangentes, apoiando um espectro extremamente amplo de atividades. Essas novas dimensões de metadados são vitais para o acesso e para a interpretação dos recursos informacionais digitais.

Nesse contexto, além dos metadados, os esquemas de metadados são fundamentais em ambientes digitais. Um esquema de metadados estabelece como a descrição dos ambientes serão realizadas. Como exemplo de esquema de metadados tem-se o Dublin Core, considerado, para alguns autores, como um dos mais importantes esquemas existentes e sua origem, sendo um marco na história dos metadados, uma vez que ele pode dar origem a outros esquemas mais específicos.

O Dublin Core apresenta 15 elementos básicos que favorecem a descrição em ambientes digitais. Os elementos básicos do Dublin Core são: título, criador, assunto, descrição, editor, colaborador, data e hora, formato, identificador, fonte, idioma, relação, cobertura, direitos (Souza; Vendrusculo; Melo, 2000). Algumas das características do Dublin Core são a simplicidade e a interoperabilidade semântica, fazendo com que seu uso seja bastante amplo em ambientes digitais.

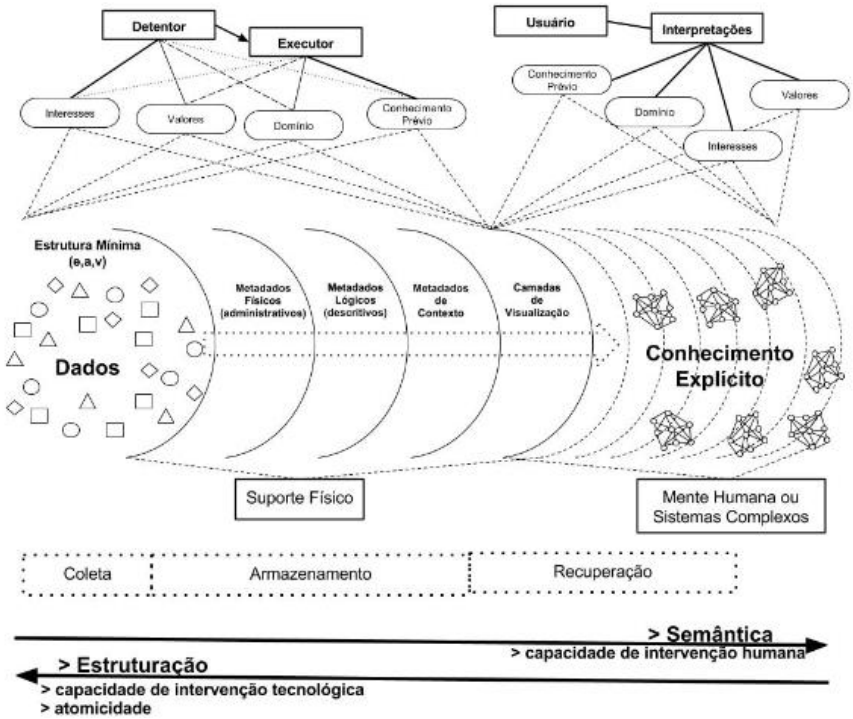
Essas características foram fundamentais para a criação de outros esquemas de metadados baseados no Dublin Core voltados para a gestão de dados de pesquisa. A expansão dos ambientes de publicação de dados de pesquisa fez com que os metadados se tornassem peças-chave nesses sistemas de informação. O uso dos metadados garantem o acesso de forma descomplicada por parte dos pesquisadores e da comunidade que desejam obter e utilizar os

dados de pesquisa. Se faz essencial o uso de metadados na gestão de dados de pesquisa, uma vez que os metadados possuem função de descrever, preservar e, assim, garantir sua reutilização (Farnel; Shiri, 2014). Veja-se este exemplo:

> Em um estudo de ciências da natureza, por exemplo, a comparação de dados coletados em um mesmo local, mas em diferentes datas, pode trazer à luz questões ligadas ao impacto ambiental que uma região sofre ao longo do tempo. E esse mesmo conjunto de dados pode ser valioso não apenas para pesquisadores de ciências da natureza, mas também para pesquisadores em ciências sociais, por exemplo, que procuram entender de que forma uma dada comunidade se organiza e como utiliza os recursos naturais disponíveis (Rocha; Sales; Sayão, 2017, p. 194).

Como se pode observar, a descrição do mesmo conjunto de dados com as variadas datas só é possível devido à existência de metadado. É o metadado que irá permitir que a descrição com datas diferentes possibilite a quem for utilizar, recuperar a informação com variados pontos de vista. Conforme Santos e Sant'Ana (2019, p. 58), "a interpretação dos dados, especialmente os não estruturados, solicita a utilização de metadados para ser representados". Essa afirmação retoma ao conceito básico de metadado, que é dado sobre dado. Uma vez que, sem a organização dos dados, não é possível sua utilização e reutilização, e a organização só é possível porque os metadados existem para fornecer informações acerca de outros dados. A Figura 5, demonstra alguns elementos de acesso aos dados, sendo um desses elementos, os metadados.

Figura 5 – Elementos, fatores e fases no cenário de acesso a dados



Fonte: Santos e Sant'Ana (2019, p. 62).

Como se pode observar na figura acima, os metadados são uma ponte entre quem está disseminando o dado e quem vai utilizá-lo. Dessa forma, estes contribuem para a coleta, armazenamento e a recuperação dos dados dentro dos sistemas de disseminação dos dados. Para Treloar e Wilkinson (2008), os metadados podem ser inseridos ao longo do ciclo de vida dos dados, trazendo inúmeros benefícios para as instituições e pesquisadores. Os tipos de metadados inseridos vão depender do tipo de dado e da necessidade da instituição, podendo se de descrição, preservação e outros.

Sobre as funções dos metadados no contexto da gestão de dados, Sayão e Sales ([2023?]) informam que os metadados devem cumprir um papel ativo que vai além da descrição dos conjuntos de dados, acesso, preservação, interoperabilidade e de outros aspectos.

Pensando a reutilização dos dados e, posteriormente, a citação dos dados reutilizados, surge o *DataCite*. O *DataCite* é um consórcio internacional cujo objetivo é otimizar o acesso aos dados de pesquisa na internet além de oferecer um auxílio no arquivamento dos dados. Por consequência disso, o *DataCite* oferece um esquema de metadados para os provedores de dados. O esquema oferecido pelo *DataCite* foi pensado para a identificação de quaisquer recursos de informação, mas em especial os conjuntos de dados, sobretudo dados de pesquisa (DataCite, 2015).

O esquema possuí propriedades que podem ser aplicadas em conjunto e sua utilização vai depender da necessidade do sistema de informação. As propriedades são divididas em três conjuntos: Obrigatórias, Recomendadas e Opcionais. Para o próprio *DataCite* (2015), a utilização dos três conjuntos e sobretudo o uso dos metadados opcionais trazem uma descrição rica para o conjunto de dados.

A utilização dos metadados são tão essenciais no contexto da gestão dos dados, que os princípios FAIR - acrônimo para *Findable* (encontrável), *Acessible* (acessível), *Interoperable* (interoperável) e *Reusable* (reutilizável) – trazem recomendações criadas para melhorar a gestão de dados por parte de pesquisadores, além de apresentarem orientações acerca dos metadados. Dentro de cada princípio, existem desdobramentos que são voltados para os metadados, os quais englobam os quatro princípios gerais. Para observar melhor como se estruturam essas orientações e suas respectivas subdivisões, segue o Quadro 3:

Quadro 3 – Os princípios orientadores FAIR

FAIR: Princípios orientadores
F - Ser encontrável (Findable)
F1. Os (meta)dados são atribuídos a um identificador persistente, único e global.
F2. Os dados são descritos com metadados ricos (definidos por R1 a seguir).
F3. Os metadados incluem, de forma clara e explícita, o identificador dos dados que descrevem.
F4. Os (meta)dados são registrados ou indexados em um recurso pesquisável.
A - Ser acessível (Accessible)
A1. Os (meta)dados são recuperáveis por seu identificador, usando-se um protocolo de comunicação padronizado.
A1.1 O protocolo é aberto, gratuito e universalmente implementável.
A1.2 O protocolo possibilita um procedimento de autenticação e autorização, quando necessário.
A2. Os metadados são acessíveis, mesmo quando os dados não estão mais disponíveis.
I - Ser interoperável (Interoperable)
I1. Os (meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para representar o conhecimento.
I2. Os (meta)dados usam vocabulários que seguem os Princípios FAIR.
I3. Os (meta)dados incluem referências qualificadas para outros (meta)dados.
R - Ser reutilizável (Reusable):
R1. Os meta(dados) são ricamente descritos com uma pluralidade de atributos precisos e relevantes.
R1.1. Os (meta)dados são disponibilizados com uma licença de uso de dados clara e acessível.
R1.2. Os (meta)dados estão associados a uma proveniência detalhada.
R1.3. Os (meta)dados estão de acordo como padrões comunitários relevantes para o domínio

Fonte: Dias, Anjos e Rodrigues (2019, p. 181).

Assim, pode-se verificar que os gestores precisam utilizar os metadados para garantirem a identificação do objeto de maneira única, assegurar a indexação de forma recuperável e certificar a disponibilidade dos metadados, ainda que os dados não estejam disponíveis para assegurar a troca de informação. Os metadados devem garantir a interoperabilidade entre as máquinas e garantir a descrição dos dados para que possam ser reutilizados. Os gestores que implementarem as diretrizes presentes no FAIR, podem viabilizar de maneira efetiva a reutilização dos dados.

Porém, cabe ressaltar que apenas as diretrizes apontadas no FAIR não serão suficientes para a recuperação da informação, para além disso é preciso que os gestores atentem para o seu possível público, não só com os que compartilham os dados, mas também com aqueles que os reutilizarão. Dessa forma, é necessário a utilização de metadados voltados para domínios específicos. Assim, no âmbito da Biodiversidade, os pesquisadores se preocuparão em desenvolver metadados e esquemas de metadados com características específicas, próprias para auxiliar a disseminação da informação nessa área.

3.1 METADADOS PARA DADOS EM BIODIVERSIDADE

Assim como existem SOC voltados para domínios específicos, existem padrões de metadados específicos. A seguir serão apresentados alguns padrões em Biodiversidade, onde são utilizados e suas características.

O padrão Darwin Core[24] é um padrão de dados e metadados desenvolvido pela organização *Biodiversity Information Standards*, uma comunidade voltada para o avanço de informações padronizadas em Biodiversidade. Foi publicado formalmente no dia 9 de outubro de 2009 (Ministério do Meio Ambiente, 2015). O mesmo foi elaborado com o intuito de contribuir com o compartilhamento de informações sobre a Biodiversidade, oferecendo identificadores e definições, sendo um compilado de informações além de padrão de metadados. Segundo a página do próprio Darwin Core (*Biodiversity Information Standards*, 2022ª, não paginado), ele "é baseado principalmente em táxons, sua ocorrência na natureza documentada por observações, espécimes, amostras e informações relacionadas". Como a própria nomenclatura diz, é baseado no padrão Darwin Core, só que todo desenvolvido para atender as demandas específicas de informações em Biodiversidade. Nesse contexto, o Darwin Core abrange:

> - Coleções de qualquer tipo de objetos ou dados biológicos.
> - Terminologia associada aos dados de coleta biológica.
> - Buscar a compatibilidade com outros padrões relacionados à biodiversidade.
> - Facilitando a adição de componentes e atributos de dados biológicos. (*Biodiversity Information Standards*, 2022a).

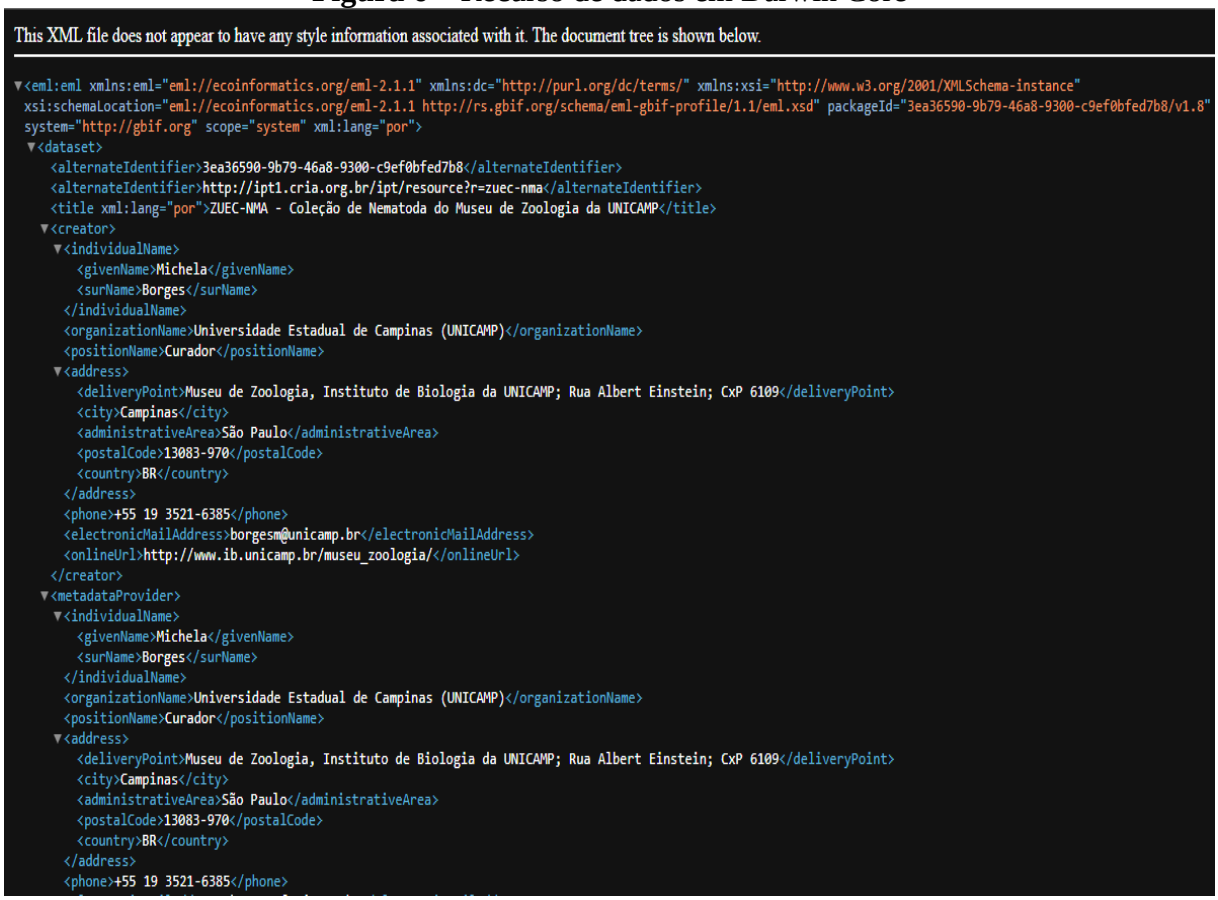
Logo, o Darwin Core é todo voltado para a descrição de informações em Biodiversidade. Segundo o Ministério do Meio Ambiente (2015), o padrão Darwin Core é diverso e amplo, além de oferecer um glossário que pode ser utilizado para descrever informações básicas de uma determinada espécie. A ZUEC-NMA – Coleção de Nematoda do Museu de Zoologia da UNICAMP, é um dos exemplos em Biodiversidade que utiliza o Darwin Core em seus dados (ZUEC-NMA, 2022). A Figura 6 mostra um recurso de dados, o Darwin Core Archive do museu da UNICAMP.

A Figura 6 demostra uma parte da descrição de informações em Darwin Core, nos quais é possível identificar os metadados para a instituição da guarda dos dados, bem como o

[24]*Biodiversity Information Standards* (2022b).

depositante dos dados. Além do museu da UNICANP, instituições como o Jardim Botânico do Rio de Janeiro e o SiBBr utilizam os Darwin Core em seus sistemas.

Figura 6 – Recurso de dados em Darwin Core



Fonte: ZUEC-NMA (2022).

Por sua vez, o GBIF *Metadata Profile* (GMP), foi criado com o intuito de uniformizar os conjuntos de dados disponíveis no portal do GBIF. O GMP permite a descrição dos dados em aspectos mais gerais de informações, tais como projetos, instituições e pessoas envolvidas, desmembradas em classes de metadados, tais como título, criador e contato, como exemplo. E os aspectos mais específicos para Biodiversidade, como informações taxonômicas, geográficas e temporais, que podem se subdividir em classificação taxonômica, descrição geográfica e date, como exemplo.

O GMP (*Global Biodiversity Information Facility*, 2011) foi desenvolvido baseado no *Ecological Metadata Language* (EML), que também foi pensado em facilitar a disseminação da informação em Biodiversidade. Segundo Jones *et al.* (2019, não paginado, tradução nossa), o EML “define um vocabulário abrangente e uma sintaxe de marcação XML legível para

documentar dados de pesquisa"[25]. Isso quer dizer que essa marcação vai permitir que a informação seja organizada de forma padronizada. Ainda conforme Jones *et al.* (2019, não paginado, tradução nossa),

> A EML inclui módulos para identificar e citar pacotes de dados, para descrever a extensão espacial, temporal, taxonômica e temática dos dados, para descrever métodos e protocolos de pesquisa, para descrever a estrutura e o conteúdo dos dados dentro de pacotes de dados às vezes complexos e para precisamente ir anotando dados com vocabulários semânticos. O EML inclui campos de metadados para detalhar totalmente os artigos de dados publicados em periódicos especializados em compartilhamento e preservação de dados científicos[26].

Ou seja, é um padrão de metadado abrangente, que permite não só a descrição dos dados, mas também contribui para a preservação dos mesmos. Conforme o Ministério do Meio Ambiente (2015, p. 57), o EML foi desenvolvido com bases nas seguintes premissas:

> •Aberto: para possibilitar a leitura humana e facilitar o arquivamento de dados em longo prazo;
> • Modular: para promover o reuso das estruturas/sessões dos metadados;
> • Extensível: para possibilitar a inclusão de metadados que não estão definidos originalmente no EML;
> • Estruturado: para possibilitar o processamento computacional realizado por aplicações de análise e outras aplicações de software;
> • Fácil de implementar.

Ainda apresentando esquema identificado pela comunidade internacional, tem-se o *Access to Biological Collections Data* (ABCD). O ABCD também é baseado em XML, voltado em garantir o acesso e a disponibilização de dados primários sobre espécimes e observações (Access to Biological Collection Data, 2007). O objetivo de seus desenvolvedores é permitir que o ABCD seja amplo e estruturado para que suporte os mais variados bancos de dados em Biodiversidade.

Segundo o JOINUP (2022, não paginado), um braço da União Europeia voltada para soluções em Tecnologia da Informação, o ABCD está sendo utilizado por nove redes de disseminação de informações sobre Biodiversidade, a saber:

- Global Biodiversity Information Facility (GBIF);
- Biological Collection Access Service (BioCASe);

[25]Texto original: "*The Ecological Metadada Language define a comprehensive vocabular and a readable XML markup syntax for documenting research data*".

[26]Texto original: "*EML includes modules for identifying and citing data packages, for describing the spatial, temporal, taxonomic, and thematic extent of data, for describing research methods and protocols, for describing the structure and content of data within sometimes complex packages of data, and for precisely annotating data with semantic vocabularies. EML includes metadata fields to fully detail data papers that are published in journals specializing in scientific data sharing and preservation*".

- OpenUp! natural history aggregator for Europeana;
- Global Genome Biodiversity Network;
- Geological Collection Access Service (GeoCASe);
- Australian Virtual Herbarium;
- Italian Biodiversity Data Network;
- Biodiversity Network of the Humboldt institutions (BiNHum);
- German Virtual Herbarium (VH/de).

Isso demonstra que o ABCD é aceito e amplamente utilizado em sistemas que são voltados para a Biodiversidade.

Identificou-se ainda a existência do *Audubon Core*, também voltado para a Biodiversidade, mais especificamente para o "gerenciamento do arquivo ou da coleção, descrições do seu conteúdo, sua taxonomia, geografia, cobertura temporal e a forma indicada para recuperação, reprodução e utilização dos mesmos" (Ministério do Meio Ambiente, 2015, não paginado). O *Audubon Core* foi desenvolvido para descrever recursos que estão no formato multimidia, tais como imagens, som, ilustrações, animações e outros.

Segundo o Ministério do Meio Ambiente (2015, p. 65) "o Audubon Core foi projetado para facilitar a publicação e descoberta de recursos multimídia relacionados à biodiversidade, por meio da descrição consistente de um recurso ou de um conjunto de recursos". Isso significa que este padrão se torna uma opção a mais para a publicação de dados em formato multimidia. Em sua página principal, o Audubon Core apresenta seus objetivos, guia de uso e suas características. É um padrão que também possui um vocabulário próprio. No contexto brasileiro, especificamente, existe o catálogo de metadados Geoespaciais desenvolvido pela Infraestrutura Nacional de Dados Espaciais (INDE), criado para atender as necessidades do INDE com relação aos recursos informacionais nos quesitos de localização e descrição. Atende os padrões ISO[27], visando assegurar uma maior padronização nos registros informacionais.

Segundo a própria página do INDE, o objetivo da elaboração do catálogo de metadados é "a catalogação, integração e harmonização de dados produzidos ou mantidos e geridos nas instituições de governo brasileiras – incluindo ministérios, universidades, agências reguladoras e outras" (INDE, 2021, não paginado). Instituições como a Agência Nacional de Mineração, o IBGE, o IBAMA, o Instituto de Cartografia da Aeronáutica, o Instituto Estadual do Ambiente (RJ) e a EMBRAPA, utilizam o catálogo de metadados Geoespaciais.

As instituições JBRJ, ICMBio e MMA criaram um padrão para complementar o Darwin Core, é o MMA-DwC. O intuito desse padrão também é facilitar a circulação de dados sobre Biodiversidade entre as três instituições. Assim, o MMA-DwC é constituído por algumas

[27]A A ISO - *International Organization for Standardization* – é uma organização cujo objetivo é elaborar e produzir normas que sejam aplicadas no mundo todo.

extensões e termos que visam acrescentar mais informações ao Darwin Core. Na página da wiki elaborada pelo Jardim Botânico do Rio de Janeiro (JBRJ), é possível observar algumas das extensões elaboradas pelas instituições.

Quadro 4 – Exemplo de extensão do MMA-DwC

Nome do Termo	grupo
Padrão	MMA-DwC
Descrição	Forma de organização de grupos taxonômico formais de interesse do MMA
Domínio	À definir
URL	http://dadoswiki.jbrj.gov.br/doku.php?id=mma-dwc#grupo
OBS	

Fonte: Jardim Botânico do Rio de Janeiro (2023).

O Quadro 5 demonstra um exemplo de campo que foi criado dentro da extensão do Darwin Core. Como se pode observar, mesmo o Darwin atendendo as necessidades das instituições de Biodiversidade, ainda assim se fez necessário desenvolver extensões já que as mesmas têm interesse em intercambiar os seus dados.

O Quadro 5, ilustra, de forma resumida, as diferenças e semelhanças encontradas nos padrões de dados citados acima.

Quadro 5 – Comparação entre os Padrões de metadados para a Biodiversidade

Padrão de metadado	Características	Tipos de dados cobertos
Darwin Core	Possui glossário	Taxômicos, ocorrência, dados amostrais
GBIF *Metadata Profile (GPM)*	Criado especificamente para atender um sistema de informação: o GIBF	Informações mais gerais ligadas aos dados; dados geográficos, de espécies.
Ecological Metadata Language (EML)	Possui vocabulário Baseado em XML	Dados ecológicos, dados sobre extensão espacial, temporal, taxonômica e temática dos dados
Access to Biological Collections Data (ABCD)	Baseado em XML Amplamente utilizado	dados primários sobre espécimes e observações em coleções
Audubon Core	Voltado para mídias Possuí vocabulário	Taxonômicos, geográficos, temporais em coleções
Geoespaciais INDE	Possui vocabulário possui sim, desenvolvido baseado na ISO	Dados geoespacias
MMA-DwC	Baseado no Padrão Darwin, apresenta-se como uma extensão personalizada	Dados sobre flora e fauna (que também podem ser de ocorrência e taxonômicos)

Fonte: Elaborado pela autora baseado nos dados da pesquisa (2023).

Como se pode observar, o quadro acima demonstra que alguns padrões foram criados inspirados nos outros, como uma espécie de aprimoramento ou complemento no registro da informação. Nem todos os padrões possuem vocabulários, o que pode dificultar numa possível representação temática da informação. Com exceção do Geoespaciais INDE, todos os outros padrões podem ser utilizados para mais de um tipo de dado no contexto da Biodiversidade. Mesmo que não deixem claro como o Audubon Core em suas páginas da internet, cujo objetivo é auxiliar na representação dos dados em multimidias, os outros padrões também podem ser aplicados nesse contexto, uma vez que, onde são utilizados, são disponibilizados dados nos mais variados tipos de mídia.

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 além das questões acerca da Compatibilização Semântica. O capítulo a seguir apresenta o as teorias acerca da Compatibilização Semântica que serviram de elaboração das análises apresentadas nos resultados.

4 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 contribuir para uma gestão de dados efetiva, em particular para a controvérsia esta pesquisa se debruça em busca de uma possível proposta de solução.

O conhecimento é parte fundamental na vida do homem e está totalmente ligado ao funcionamento da sociedade. Sendo assim, aspectos como ciência, religião e política apenas podem ser desenvolvidos mediante um conhecimento partilhado, e para tal precisa ser registrado. Contudo, o conhecimento registrado não garante sua circulação. Logo, deve ser organizado e acessível para todos.

Segundo Guimarães (2014, p. 14):

> [...] para que esse conhecimento socialmente produzido possa ter uso social, necessário se torna um processo mediador, de organização, em que se estabelecem "substitutos do conhecimento" (*surrogates of knowledge*), de modo a que os contextos de produção e de uso possam ser colocados em diálogo.

Nesse quadro, surge a Organização do Conhecimento, que como disciplina visa garantir que o conhecimento seja alcançável à recuperação da informação e fundamental no âmbito da Ciência da Informação e Biblioteconomia. Sendo assim, Guimarães (2014, p. 14) disserta que:

> Na atualidade, esse macroprocesso mediador, entre um conhecimento socialmente produzido e seu posterior uso social constitui área de estudos – a denominada organização do conhecimento – que transcende a ciência da informação, mas que hoje nela ocupa um dos mais significativos espaços de reflexão teórica, metodológica e, mais recentemente, vem sendo abordada a partir de seu contexto cultural.

Diante disso, a Organização do Conhecimento se apresenta como uma área central nos campos da Ciência da Informação e da Biblioteconomia, já que o desenvolvimento desses campos é pautado na construção de técnicas de recuperação de informação cujas bases se encontram nas teorias de organização do conhecimento. Sobre isto, Smiraglia (2012, p. 225, tradução nossa) alude que:

> A Organização do Conhecimento (também conhecida pela sigla KO, do inglês) é o domínio onde o ordenamento do conhecimento é o paradigma

> principal de investigação científica, cuja aplicação básica é o desenvolvimento de sistemas"[28].

Assim, a partir das teorias que embasam a disciplina são criados instrumentos que favorecem a representação e a recuperação da informação, tais como os sistemas de organização do conhecimento (KOS em inglês, ou SOC, em português).

Tradicionalmente, quanto ao funcionamento da Organização do Conhecimento, pode-se dividi-la em duas aplicações: representação descritiva e representação temática. A primeira está voltada para a descrição física dos documentos, também denominada Catalogação; e a segunda relativa à descrição de assuntos, também denominada, Indexação. Dessa maneira, o foco da presente pesquisa estará direcionado para a representação temática, especificamente no que tange à compatibilização das linguagens usadas como instrumentos de padronização terminológica dos termos usados como indexadores de assunto.

Conforme Rabelo e Pinto (2019, p. 67):

> A representação temática/indexação de assuntos tem como objetivo extrair ou associar os assuntos que melhor representam os conteúdos ou as temáticas registradas nos documentos, de modo a identificá-los de forma particular em meio a outros documentos independentemente, se textos verbais ou não verbais e de suportes de registros analógicos ou digitais. Ela se efetiva por meio de palavras-chave, conceitos, descritores, termos, resumos entre outros.

Então, a representação temática/indexação representa os documentos nos aspectos relativos aos assuntos/conteúdos. Para isso, termos são atribuídos a fim de representar de forma padronizada o conteúdo do documento que será indexado, possibilitando uma recuperação de informação mais precisa. Para que a indexação melhore a precisão da informação recuperada é essencial o uso de instrumentos que padronizem os termos que representam o assunto da informação. Tais instrumentos são conhecidos como linguagens de indexação, linguagens documentárias e/ou sistemas de organização do conhecimento. Para fins deste trabalho, empregar-se-á o termo Sistemas de Organização do Conhecimento (SOC).

Os SOC têm entre suas funções, auxiliar na representação temática quando um documento deve ser representado por palavras-chave. Os documentos são escritos em linguagem natural (linguagem do documento), enquanto os SOC oferecem termos que podem identificar o registro de forma padronizada. Os SOC podem ser de diversos tipos, entre eles se encontram, os tesauros, as listas de cabeçalho de assunto, os glossários, os vocabulários

[28] Texto original: "*Knowledge organization (also well-known by its acronym KO) is the domain in which the order of knowledge is both the primary paradigm for scientific investigation and the primary application in the development of systems*".

controlados, ontologias e taxonomias, entre outros. Suas configurações mudam de acordo com o objetivo, que pode ser melhorar a recuperação da informação, mas também pode ser melhorar a navegação, melhorar a comunicação entre especialistas, mapear domínios de conhecimento, auxiliar a máquina no processamento inteligente, etc. Desta forma, os SOC apresentam particularidades específicas que os diferenciam entre si.

O vocábulo "tesauro" vem do grego e significa tesouro. De acordo com N. Barbosa (2021, p. 35-36), "o tesauro é um tipo de SOC em que os conceitos que cobrem um domínio do conhecimento e são representados por termos organizados de forma sistemática em três níveis de relacionamento: hierárquico, associativo e de equivalência". Em sua estruturação é possível observar a ligação e a hierarquia de cada termo presente no tesauro tem e como ele se relaciona com os outros termos constituintes do tesauro. Esse tipo de SOC é ideal para uso em bases de dados especializadas, em que a busca pode ser feita de forma pós-coordenada, isto é, a combinação dos termos para formar um assunto se dá no momento da busca.

Sendo o oposto deste, as listas de cabeçalhos de assuntos são desenvolvidas para abarcar várias áreas do conhecimento, isto é, são multidisciplinares. Sua origem remete nos Estados Unidos com a produção de catálogos alfabéticos pela *Library of Congress* (Tôrres, c2021), e, até o presente momento, são estruturados de maneira alfabética. Esse tipo de SOC foi criado para atender uma tecnologia pouco usada atualmente que é o catálogo em fichas 13x7, cuja busca por assunto era uma busca do tipo pré-coordenada – neste caso a combinação dos termos era feita no momento da indexação e as fichas organizadas no catálogo em ordem alfabética. No entanto, muitas bibliotecas mantem o uso de suas listas de cabeçalhos-de assunto, dada a dificuldade de reindexar seus gigantes acervos já catalogados em bases de dados automatizadas.

Os glossários, por sua vez, podem ser definidos como listas de termos que apresentam as definições dos termos que os constituem (Tackabery, 2005). Geralmente, são temáticos e estruturados em ordem alfabética. Para Bräscher (1986, p. 137), "os termos podem ser organizados alfabeticamente ou classificados de acordo com a estruturação conceituai da área". Glossários são instrumentos construídos para auxiliar na comunicação entre especialistas, ajudando as áreas na padronização dos termos usados para representar cada conceito de uma área.

No tocante aos vocabulários controlados, Schiessl e Shintaku (2012, p. 55) registram que "vocabulário controlado pode ser sintetizado com um rol de termos apresentados explicitamente, no qual esses termos devem ser definidos de forma não ambígua e não redundante". Deste modo, devem constar em todos os termos presentes no vocabulário sua definição, evitando, assim, ambiguidade.

Por sua vez, as ontologias, que tem seu conceito com origem na Filosofia também utilizada na Ciência da Computação, em Ciência da Informação surge como instrumento de representação da informação e tem seu uso passou a ser mais amplamente difundido no contexto da web semântica, sendo responsável por padronizar significar os termos de determinado domínio. Para Moreiro Gonzáles (2011, p. 77):

> uma ontologia é uma descrição explícita e formal de conceitos em um domínio de discurso (classes, também chamadas conceitos), propriedades de cada conceito, descrevendo várias características e atributos do conceito (slots – funções ou propriedades), e restrições sobre slots (facetas – restrições de uma função).

À vista disso, uma ontologia pode ser definida como a junção de termos que formam a representação de determinado domínio, sendo constituída por axiomas (conceitos presentes na ontologia) e suas relações.

Com origem na Biologia, o termo taxonomia passou a transitar por empréstimo linguístico na área da Ciência da Computação e Ciência da Informação com o surgimento das tecnologias de informação. Tanto na Biologia quanto nas áreas da informação é usada por analogia ao sentido de "classificar". A Biblioteconomia também sempre teve em sua essência o estudo e a elaboração de sistemas de classificatórios. No entanto, segundo Mendes e Pinto (2019, p. 38): "Com o avanço do sistema web a taxonomia ganha maior importância para a representação, organização e recuperação da informação, particularmente no ambiente web semântica". Além de auxiliar na representação em sistemas da web e na navegação de páginas, está presente na construção de sistemas de classificação bibliográficas e SOC. Por analogia também a Biblioteconomia e a Ciência da Informação começam a se valer das Teorias da Classificação e sua experiência no desenvolvimento de sistemáticas de conceitos para a construção de Taxonomias.

O surgimento de novos sistemas e juntamente com eles novos SOC, nos coloca diante de um desafio até então inédito, particularmente quando o objetivo do sistema é interoperar com outros sistemas que se valem de linguagens diferentes. No contexto atual da *e-science*, pode-se citar como exemplos de sistemas que vem sendo desenvolvido com bastante frequência, os repositórios e os bancos de dados de pesquisa. A necessidade de interoperabilidade entre esses sistemas para além do nível técnico, mas também no nível semântico leva a necessidade de compatibilização entre as linguagens nesses sistemas, que na maioria dos casos, são distintas. Isso acontece no nível dos metadados e no nível do conteúdo. Neste contexto, a compatibilização semântica surge como uma alternativa que parece viável para a solução do problema.

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, N., 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 tesauros multilíngues. De todas as iniciativas criadas, o mais famoso foi o do UNIST, que "serviu como um mecanismo de integração em diferentes classificações e tesauros 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.

Por conseguinte, antes de adentrar nos aspectos referentes à compatibilização dos SOC, faz-se preciso explanar brevemente quanto aos estudos terminológicos, os quais estão totalmente ligados às questões da Organização do Conhecimento e da elaboração dos SOC e podem contribuir para as pesquisas tocantes à compatibilização dos SOC.

Concernente aos estudos sobre terminologia, Bräscher (1986, p. 135) inscreve que:

> A sistematização dos estudos terminológicos tornou-se indispensável para acompanhar esse crescimento de terminologias específicas, na medida em que eles fornecem critérios ao estabelecimento de termos precisos para representar os conceitos em determinado campo do conhecimento.

Esses estudos fornecem subsídios para a elaboração dos SOC, visto disporem de metodologias ligadas aos conceitos e aos termos que representam determinado domínio. Bräscher (1986, p.135) alude que esses estudos versam sobre a "representação sem ambiguidade no âmbito das linguagens especializadas", isto porque as linguagens são construídas por termos que podem caracterizar mais de um conceito, como, por exemplo, o termo manga, que pode denotar à camisa ou à fruta. Diante disso, estudar os SOC no âmbito dos estudos terminológicos garantirá uma maior precisão na representação da informação.

Para Gomes, Campos e Guimarães (2010) faz-se preciso o uso dos princípios da terminologia para garantirem a sistematização da informação essencial para o indexador. Essa sistematização auxilia o indexador a escolher o melhor termo para representar determinado assunto em um sistema de informação.

Posto isso, a compatibilização de linguagens visa melhorar a representação e a recuperação da informação em sistemas de informação. Dahlberg (1981, p. 87, tradução nossa[29]) afirma que a compatibilização de linguagem será capaz de

> (1) pesquisar qualquer termo de qualquer problema em uma variedade de arquivos;
> (2) recuperar as informações rotuladas com um específico conceito, indexado por qualquer um dos sistemas envolvidos;
> (3) informar ao usuário de uma determinada linguagem de indexação sobre a disponibilidade de informações sobre o tema de seu interesse, indexadas por qualquer outra linguagem de indexação;
> (4) obter os equivalentes de um sistema em um idioma natural (ex.: inglês) em outro idioma natural (ex.: alemão), portanto, o sistema de comutação também pode ser usado como um dicionário de sinônimos multilíngue.

Portanto, quando acontece a compatibilização entre as linguagens, a recuperação da informação por parte do usuário ocorre de maneira eficaz. Mesmo quando as linguagens são desenvolvidas em idiomas distintos, a compatibilização favorece a recuperação do termo em outra língua.

Para a UNISIT (1971 *apud* Zeng, 1992, p. 170, tradução nossa), a compatibilidade pode ser definida como "a qualidade de sistemas cujos produtos podem ser usados indistintamente,

[29]Texto original: "*(1) to search with any term on any problem in a variety of different files. (2) retrieve the information labelled with a specific concept from any store, 'indexed by any of the systems involved, (3) inform the user of a certain IL on the availability of information on the topic of his interest, indexed by any other IL, (4) get the equivalents of a system in one natural language (say English) in another natural language (say German), thus the switching system can also be used as a multilingual thesaurus*".

não obstante as diferenças na notação”[30], isto é, mesmo que as notações sejam apresentadas distintamente não impede da compatibilização de ocorrer se esses sistemas exibirem tal este atributo.

Para Glushkov *et al.* (1978 *apud* 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. Ainda, para Hammond (1965 *apud* Svenonius, 1983, p. 2, tradução nossa), a:

> “Compatibilidade” aplicada a sistemas de informação pode ser definida como “a capacidade de um sistema de informação aceitar a indexação e dados de um outro sistema em qualquer assunto, na qual a cobertura ocorra em ambos”[31].

Os sistemas ou as linguagens teriam a possiblidade 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. Sobre o conceito de conversão, foi definido primeiramente pelo UNISIST (1971 *apud* Dahlberg, 1981, p. 86, tradução nossa) como “o processo de transformar registros de informações orientado à codificação da transcrição, estrutura de dados etc., de modo a torná-los intercambiáveis entre dois ou mais serviços ou sistemas que usem diferentes convenções e mídia” [32].

Vale destacar que a elaboração de instrumentos, a conversão ou a própria compatibilização não deve ser algo realizado de uma hora para outra. Faz-se primordial observar vários aspectos, tais como: a cobertura do assunto do sistema, a própria linguagem de

---

[30]Texto original: “*[...] a quality of systems whose products can be used interchangeably, notwithstanding diferences in notation [...]*”.

[31]Texto original: “*Compatibility" applied to information systems was defined as "the ability of one information system to accept the original indexing and abstracting data of another information system for any given subject coverage that is common to both systems*.

[32]Texto original: “*As the process of transforming information records J with regard to transcription encoding, data structure, etc., so as to make them interchangeable between two or more services or systems using different conventions and media*”.

indexação e sua estrutura. Em vista disso, Zeng (1992) sugere que quanto mais célere for sucedida a compatibilização entre linguagens, melhor será o desempenho e o resultado.

Para Svenonius (1983, p. 2, tradução nossa), a compatibilidade entre sistemas geralmente ocorre de duas formas: "1) construção de uma linguagem de comutação ou 2) mapeamento ou tradução direta do vocabulário de uma linguagem de índice para a de outro" [33]. Uma linguagem de comutação pode ser percebida como uma linguagem intermediária, capaz de abarcar termos de determinada linguagem e transformá-los entendíveis em outra linguagem/ múltiplos idiomas. Ainda conforme a autora, ela apresenta outra possibilidade para alcançar a compatibilidade entre sistemas:

> O outro método de obter compatibilidade é simplesmente para traduzir ou mapear um idioma de índice para outro e para execute esta tradução para cada par de línguas onde for desejado. Este método requer a construção de uma tabela de conversão[34] (Svenonius, 1983, p. 2, tradução nossa).

Dessa maneira, a tabela de conversão apresentaria os termos correspondentes no idioma original e naquele que se quer traduzir. A proposta de Svenonius parece ser útil na aplicação de sistemas desenvolvidos por linguagens não consideradas de fácil compreensão, como russo, mandarim, japonês, árabe e outros. Nesse caso, um dos mais importantes estudos já realizados sobre compatibilização de linguagens é o de Neville (1970, 1972) que teve seu estudo direcionado para os tesauros. 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 tesauros, o autor sugere a compatibilização mediante reconciliação de tesauro, e afirma que

> O processo de reconciliação como um todo, resulta na criação de um conjunto de códigos numéricos. Cada palavra-chave em cada tesauro participante [da compatibilização] recebe um código, equiparando as palavras-chave" (Neville, 1970, p. 314).

Para melhor exemplificar, no Quadro 6 ilustra-se como funciona a proposta.

Quadro 6 – Demonstração simples do método de Neville

	Source thesaurus	***Thesaurus B***	***Thesaurus C***
Original entries	AIRFIELDS	AIRFIELD	FLUGPLÄTZE
Reconciled entries	AIRFIELDS (0101)	AIRFIELD (0101)	FLUGPLÄTZE (0101)
Keys to code numbers	0101=AIRFIELDS	0101=AIRFIELD	0101 = FLUGPLÄTZE

[33]Texto original: "*1) constructing a switching language or 2) mapping or directly translating from the vocabulary of one index language to that of another*".

[34]Texto original: "*The other method of achieving compatibility is simply to translate or map one index language to another and to perform this translation for every pair of languages where it is desired. This method requires the construction of a conversion table*".

Fonte: Neville (1970, p. 316).

No Quadro 6, percebe-se que mesmo com o emprego de termos expressos em idiomas diferentes, mas utilizados da mesma forma pelos tesauros, 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 considerada como correspondendo exatas, e o uso de diferentes linguagens não tem importância, desde que os conceitos subjacentes sejam os mesmos" [35]. 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 diferente em dois ou em mais tesauros, a reconciliação ocorre, dispensando a adição de novos termos:

> A reconciliação também envolve fazer certas adições a cada tesauro, principalmente de natureza de referências cruzadas, mas nenhuma palavra-chave existente é alterada, na forma ou escopo, ou de qualquer outra maneira, nenhuma palavra-chave precisa ser excluída ou novas palavras precisam ser adicionadas, embora em certos casos os participantes possam aproveitar a oportunidade para adicionar novas palavras-chave, se desejarem (Neville, 1970, p. 314-315, tradução nossa[36]).

À vista disso, não se deve modificar a estrutura nos tesauros envolvidos, apenas a elaboração de vínculos entre os termos. Segundo Sales (2022, p. 82):

> No que tange o reuso de dados de pesquisa, acredita-se que o método de Neville possa ser útil, em especial, para permitir que repositórios disciplinares interoperem, mantendo a origem semântica de seus dados, mas também podendo receber novas interpretações no campo em que será reusado.

Por não promover mudança estrutural nos sistemas (mas apenas uma forma de elaborar uma referência cruzada entre os termos), os SOC usados para representar o assunto em repositórios podem utilizar da metodologia de Neville para promover a compatibilização terminológica.

Sobre as pesquisas inerentes à possibilidade de compatibilização de linguagens, Dahlberg (1981) relata que após a segunda guerra mundial houve significativa proliferação de

---

[35] 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*".

[36] Texto original: "*The reconciliation also involves making certain additions to each thesaurus mostly of the nature of cross-references, but no existing keyword is altered, in form, in scope, or in any other way, no existing keyword need be deleted, and no new keywords need be added, although in certain cases participants can take the opportunity to add new keywords if they wish*".

linguagens de indexação, que gerou a falta de cooperação e a troca de informações entre os sistemas. A autora ainda sinaliza:

> nesta situação, um mecanismo deve ser encontrado para corrigir os elementos das diferentes linguagens de indexação uns com os outros, a fim de ser capaz de alternar entre eles durante a pesquisa em bancos de dados com suas indexações, ou na troca literatura indexada" (Dahlberg, 1981, p. 86, tradução nossa[37]).

Logo, no momento que o usuário realizar uma pesquisa, a compatibilização das linguagens permitirá a busca pelo termo presente em uma linguagem e será capaz de buscar o termo correlato na outra linguagem compatibilizada.

A autora acima referenciada 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.)

[37] Texto original: "*In this situation a mechanism must be found to corre· late the elements of the different ILs with each other in order to be able to switch between them when searching in databases with their indexations, or when exchanging indexed literature*".

> (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[38]).

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.

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.

Campos (2005, 2007, 2009) apresenta em seus estudos a possibilidade de integração de ontologias. A autora disserta que a recuperação da informação não ocorre de forma suficiente, se não houver ferramentas com controle terminológico que facilitem o acesso de maneira adequada, conforme abaixo:

> Para garantir esta precisão verifica-se a necessidade de ferramentas taxonômicas e terminológicas para o tratamento semântico de informações contidas em bases de dados, viabilizando entre outros processos a integração de informações como auxílio ao desenvolvimento de pesquisa em domínios de conhecimento (Campos, 2005, p. 2).

Isso acontece em razão das informações contidas nessas bases de dados serem heterogêneas, o que pode causar uma dificuldade na recuperação da informação. Assim, Campos (2009) defende que para a informação ser recuperada de forma homogênea é essencial que haja certo grau de interseção entre as áreas.

A autora apresenta as primeiras etapas acerca da possibilidade de integração de ontologias. A primeira seria a escolha do domínio em que se quer integralizar as ontologias. Posteriormente, deve-se elaborar uma "definição da estratégia de levantamento para obtenção de conceitos,

---

[38]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*".

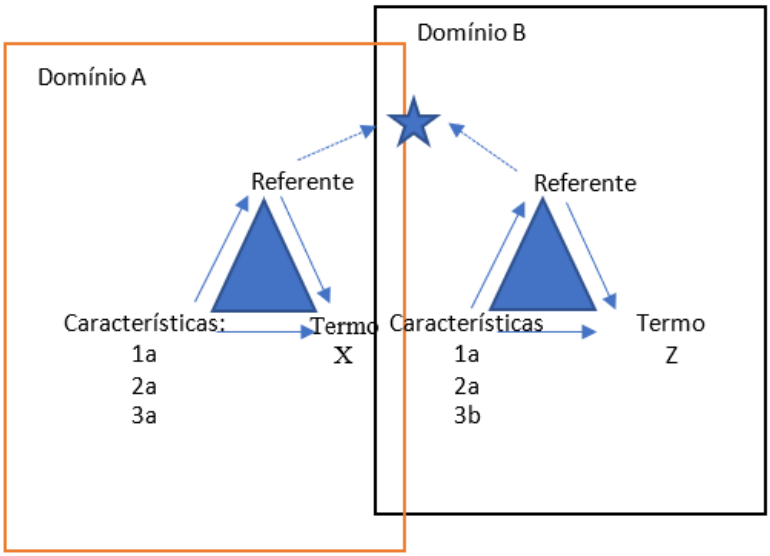
mapeamento das temáticas e desenvolvimento de ferramentas de software" (Campos, 2009, p. 11). Cabe sublinhar que o desdobramento de ferramentas só é possível caso ocorra a parceria com profissionais de outras áreas, como Tecnologia da Informação e Ciência da Computação.

No que diz respeito ao levantamento dos conceitos, sua posição na ontologia e as relações com os outros conceitos, a autora menciona que se apoiou nos estudos terminológicos de Dahlberg (1978b) e Wuester (1981), autores fundamentais no tocante ao estudo do conceito e da terminologia.

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 consequentemente 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 4) que representa graficamente uma proposta de compatibilização semântica para dados de pesquisa, baseado no triângulo conceitual de Dahlberg (1978b).

Figura 7 – Modelo Teórico de Compatibilização Semântica



Fonte: 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.

Assim sendo, com o problema de pesquisa e a hipótese apresentados, bem como a fundamentação teórica discorrida, na próxima seção apresenta o percurso metodológico utilizado para a realização da pesquisa.

5 PROCEDIMENTOS METODOLÓGICOS

Nesta seção são apresentadas as etapas de consecução da pesquisa, distribuídas mediante a seguinte dinâmica: caracterização da pesquisa, levantamento bibliográfico, análise de domínio, estudo empírico, seleção da amostra e análise dos dados.

5.1 CARACTERIZAÇÃO DA PESQUISA

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.

De acordo com a sua natureza, é classificada como teórica e empírica, visto que empregou e buscou em teóricos subsídios para o seu desenvolvimento, bem como recorreu à prática para complementar o seu desenvolvimento.

Em conformidade com Silveira e Córdova (2009, p. 35), a pesquisa exploratória objetiva "proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses". Deste modo, intenta-se aprofundar o entendimento sobre a temática investigada, não apenas sobre a representação da informação em Biodiversidade, mas buscando perceber os princípios de integração de SOC.

A pesquisa descritiva, por sua vez, possibilita apresentar a descrição da compreensão do que foi investigado na representação da informação em Biodiversidade e nos princípios de integração dos SOC, além de permitir a descrição de diretrizes para a compatibilização terminológica na área de Biodiversidade.

Assim, com a finalidade de auxiliar a pesquisa exploratória e a pesquisa descritiva, foram desenvolvidas pesquisa bibliográfica e documental. Segundo Fontana (2018, p. 66), a pesquisa bibliográfica "vincula-se à leitura, análise e interpretação de livros, periódicos, manuscritos, relatórios, teses, monografias, etc. (ou seja, na maioria das vezes, dos produtos que condensam a confecção do trabalho científico)". Assim, a presente pesquisa foi conduzida através da abordagem bibliográfica, pois se valeu de levantamento de princípios e diretrizes para a compatibilização semântica de vocabulários e de outros SOC.

Neste sentido, a pesquisa bibliográfica foi desmembrada em duas etapas: 1ª) busca nos *websites* Google Acadêmico, Portal de Periódicos Capes, Brapci e LISTA, objetivando recuperar produções sobre: dados de pesquisa, gestão de dados, organização do conhecimento, representação temática e compatibilização semântica, a fim de apoiar a elaboração das seções

apropriadas 2ª) análise de domínio para a compreensão do que vem a ser Biodiversidade. Os procedimentos desta fase serão mostrados no item 5.1.2

No tocante à pesquisa documental, Prodanov e Freitas (2013, p. 55) registram que ela "baseia-se em materiais que não receberam ainda um tratamento analítico ou que podem ser reelaborados de acordo com os objetivos da pesquisa". No presente estudo, esse tipo de pesquisa serviu como base para selecionar os termos que serviram de base para a proposta final das diretrizes.

Para fins de ilustração, as etapas metodológicas mencionadas estão associadas com os objetivos da pesquisa, conforme descrito no Quadro 7:

Quadro 7 – Articulação dos objetivos específicos com a caracterização da pesquisa

Objetivos	**Pesquisa**					
	teórica	**empírica**	**bibliográfica**	**documental**	**exploratória**	**descritiva**
Identificar, na literatura, o panorama acerca dos SOC voltados para o domínio da Biodiversidade;	■		■		■	■
Compreender os principais conceitos estudados na área da Biodiversidade		■		■	■	■
Estudar os conceitos de compatibilização semântica propostos pela Ciência da Informação	■		■	■	■	■
Investigar as possibilidades de aplicação das técnicas de compatibilização semânticas na integração de bases de dados de biodiversidade		■		■	■	■

Fonte: Elaborado pela autora (2021).

5.1.1 Levantamento Bibliográfico

Para atender o primeiro objetivo (**Identificar, na literatura, o panorama acerca dos SOC voltados para o domínio da Biodiversidade),** foi realizada a busca em bases

internacionais, a saber: 1) *Library, Information Science & Technology Abstracts* (LISTA) *with Full Text, 2) Web of Science* e *3) BioOne*. O critério de escolha das três bases foi de poder abarcar não só estudos desenvolvidos no âmbito da Ciência da Informação, mas também identificar a produção científica no âmbito da Biodiversidade e como os SOC podem favorecer a recuperação da informação em tal domínio.

Por sua vez, para alcançar o terceiro objetivo, que também versa sobre estudos bibliográficos (**Estudar os conceitos de compatibilização semântica propostos pela Ciência da Informação),** foi realizado uma busca em bases de informação específicas da área da CI como BRAPCI e LISTA. A escolha das bases forneceu a identificação de um panorama acerca da temática, além da possibilidade de observar o desenvolvimento histórico da temática e a identificação de autores seminais que auxiliam no desenvolvimento da pesquisa.

5.1.2 Análise de domínio

Para atender o segundo objetivo (**Compreender os principais estudos na área de biodiversidade)** foi realizado um estudo de Análise de Domínio, elaborada de acordo Hjørland (2002) – por ser uma abordagem muito usada no âmbito das pesquisas em Ciência em Informação, especificamente no âmbito da Organização do Conhecimento, subárea da CI, em que se encontra esse trabalho, conforme explicitam Freitas e Albuquerque (2017, p. 2),

> a abordagem teórico-metodológica da Análise de Domínio, especificamente nos estudos relacionados à Ciência da Informação, tem como um de seus objetivos, auxiliar no processo de pesquisa e construção de instrumentos para a organização do conhecimento.

A Análise de Domínio permite conhecer as teorias, instituições e necessidades informacionais de dada comunidade. Trata, segundo Smiraglia (2011), sobre o estudo das áreas do conhecimento, geralmente por meio de sua literatura e comunidade científica, objetivando compreender como sucedem os novos conhecimentos de dada comunidade. Para Hjørland (2002), essa análise pode ser desenvolvida através das seguintes abordagens:

> 1. Produção de guias de literatura e de entradas de assunto [índices]; 2. Produção de classificações especiais; 3. Pesquisa em indexação e recuperação [em áreas de] especialidades; 4. Estudos de usuários empíricos; 5. Estudos bibliométricos; 6. Estudo históricos; 7. Estudos de gênero e de documentos; 8. Estudos críticos e epistemológicos; 9. Estudos terminológicos de linguagem para propósitos especiais, estudos do discurso; 10. Estudos em estruturas e instituições em comunicação científica; 11. Análise de domínio em cognição profissional e inteligência artificial (Hjørland, 2002, p. 422).

Todas essas 11 abordagens possibilitam conhecer algum aspecto de um domínio quando aplicadas e podem ser usadas conjuntamente, a depender do objetivo da investigação. No entanto, nem sempre elas precisam ser usadas em sua totalidade, em alguns casos, a combinação de algumas delas pode ser suficiente para que o domínio seja analisado de forma eficiente. Tudo vai depender do objetivo da investigação e do nível de conhecimento que se necessita ter daquele domínio. Além disso, pode-se dizer também que mais vale poucas abordagens desenvolvidas em profundidade do que 11 abordagens desenvolvidas de forma superficial. Nesta pesquisa, foram utilizadas em profundidade três abordagens, a saber: abordagem de estudos **epistemológicos e críticos**; abordagem de **estudo bibliométrico**; e abordagem de **estudo terminológico**. Do ponto de vista de Hjørland (2002), a combinação de uma ou mais abordagens é o suficiente para garantir uma análise de domínio bem-feita.

Conforme Hjørland (2002), os estudos epistemológicos e críticos permitem conhecer as teorias, as metodologias e as aplicações de um domínio. Esta abordagem parece mais básica, mas necessária. Consoante o autor, quando esta abordagem não é empregada, a análise de domínio tende a ser mais rasa.

Como versa de uma abordagem mais básica, a partir de uma "trajetória de construção, seus paradigmas" (Guimarães, 2016, p. 18), em um primeiro instante foi pensado em direcionar a pesquisa para cursos de graduação, ao passo que é o contexto para o conhecimento desses aspectos, porém não existe curso de graduação em Biodiversidade no Brasil. Desse modo, para o desenvolvimento da abordagem de estudos epistemológicos e críticos, aplicou-se a análise dos escopos dos Programas de Pós-Graduação em Biodiversidade. Esse ato sucedeu por não haver cursos de graduação específicos para a Biodiversidade no país e a Pós-graduação concentrar grande número de pesquisadores de uma área. Na maioria dos casos, são responsáveis pela elaboração e pela manutenção de eventos científicos e periódicos científicos.

Posto isto, a busca foi efetivada na Plataforma Sucupira[39]. Foram selecionados os programas que oferecem cursos de mestrado e doutorado acadêmicos, uma vez que ao conter um curso de doutorado, entende-se que o programa já está mais consolidado. No total, foram encontrados 98 programas, distribuídos nas áreas de Botânica, Ecologia, Oceanografia e Zoologia. Posterior à pesquisa, os programas foram priorizados, e aqueles que não apresentaram ementas das áreas de concentração, desconsiderados. Sendo assim, foram analisados 62 programas.

[39] A Plataforma Sucupira é um sistema de informação responsável por coletar dados para a padronização e avaliação do Sistema Nacional de Pós-Graduação brasileira. Disponível no endereço eletrônico: https://sucupira.capes.gov.br/sucupira/.

A segunda abordagem de Análise de Domínio sugerida por Hjørland, utilizada nessa pesquisa, foi a do estudo bibliométrico. Essa abordagem busca demonstrar como acontece a produção científica de dado domínio e/ou assunto. Assim, neste momento, a pesquisa terá uma abordagem quantitativa, optando-se pelo banco de dados *Scopus*[40], que apresenta significativa relevância no meio acadêmico, sendo considerada uma das maiores e mais importantes bases de dados existentes na contemporaneidade. Conforme Sweileh (2020, p. 2, tradução nossa), a "Scopus é a maior base de dados científica disponível. Tem mais de 23.000 revistas indexadas em todas as disciplinas[41]". Dessa maneira, a investigação permitiu compreender a Biodiversidade numa perspectiva global, visto que são indexados periódicos de todo o mundo. A busca ocorreu com o termo *Biodiversity*, no campo Título, porque tem a função de apresentar ao leitor o tema central da pesquisa. Não se aplicou recorte temporal nesta busca.

A pesquisa apontou 27.304 materiais. Como o objetivo do estudo é compreender como a área se desenvolve, efetuou-se o recorte de artigos de periódicos, excetuando outros tipos de fontes de informação, tais como livros e *preprints*. Assim, desenvolveu-se a análise a partir de 18.547 artigos, sendo considerados as/os: áreas de conhecimento dos artigos, anos de publicação, periódicos mais destacados e palavras-chave usadas nos artigos. Cabe frisar que a própria *Scopus* gera os gráficos para o diagnóstico dos dados recuperados, sendo a exploração sucedida por meio dessas ilustrações.

Por fim, no que concerne à terceira abordagem de Análise de Domínio proposta por Hjørland, e escolhida para esta pesquisa, ou seja, a abordagem do estudo terminológico, optou-se pelo estudo do Sistema de Organização do Conhecimento muito utilizado na área de Biodiversidade: o ThesBio[42]. Este tesauro foi desenvolvido pela rede BHL Scielo, que objetiva facilitar a representação da informação e a recuperação da informação na Biodiversidade. De acordo com informação contida em sua página, "A Rede BHL SciELO tem como objetivo contribuir para a indexação, qualificação, publicação, acesso e interoperabilidade de informação científica em Biodiversidade. Integra as Redes gBHL e SciELO"[43](documento não datado). Além de disponibilizar informações sobre a Biodiversidade, a rede desenvolveu também o ThesBio, visando auxiliar a organização da informação sobre Biodiversidade. O tesauro apresenta em sua estrutura definição do termo, hierarquia e relação entre os termos, além de apresentar indicações de

[40]Disponível em: https://www.elsevier.com/pt-br/solutions/scopus. Acesso em: 1 set. 2021.

[41]Texto original: "*Scopus is the largest scientific database available. It has more than 23,000 indexed journals in all disciplines*".

[42]Disponível em: http://thesaurus.bhlscielo.org/vocab/index.php. Acesso em: 1 set. 2021.

[43]Disponível em: https://www.bhlscielo.org/. Acesso em: 1 set. 2021.

como utilizar o termo selecionado. Nesta parte da pesquisa, analisou-se as classes temáticas, as disciplinas e os termos que formam o tesauro. Procurou-se, ainda, identificar se os termos presentes no tesauro têm ligação com as temáticas identificadas, tanto nos programas de pós-graduação, quanto na análise, com base na *Scopus*.

Após esse estudo de análise de domínio, a fim de registrar o apreendido sobre o que vem a ser Biodiversidade, incluiu-se, no final da análise de domínio, um mapa conceitual, que é uma ferramenta que permite apresentar toda uma área, disciplina ou assunto e suas relações, mediante diagramas. Acerca deste artefato, para Rodrigues e Cervantes (2018, p. 726), "[...] os mapas conceituais são ferramentas que, por meio de diagramas auxiliam na ordenação que auxiliam na organização e representação do conhecimento".

Para a elaboração do mapa conceitual foi utilizada a proposta metodológica elaborada por Rodrigues e Cervantes (2018), que disponibilizam o passo a passo de como fazê-lo. Para auxiliar nessa etapa da pesquisa, foi utilizado o programa *Cmap Tools*[44], específico para a elaboração desse mapa.

5.1.3 Estudo empírico

Para atender o quarto objetivo específico (**Investigar as possibilidades de aplicação das técnicas de compatibilização semânticas na integração de bases de dados de biodiversidade)**, 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:

[44]O *Cmap Tools* é um software que permite a elaboração de mapas conceituais, desenvolvido pelo *Florida Institute for Human and Machine Cognition*, disponível para download em: https://cmap.ihmc.us/.

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) SiBBr, com SiBBr;
d) Ministério do Meio ambiente.

5.1.3.1 Seleção da amostra

As presentes instituições utilizam padrões de metadados, para descrever os dados tais como Darwin Core e GBIF Metadadata[45], 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 englobado 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.

[45]Esses esquemas serão apresentados no capítulo que trata sobre metadados para a Biodiversidade.

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", "*identCer*t", "*identified_by*", "nome", "nome completo", "s*cientificNameAuthorship*" não foram considerados válidos para a análise[46]. A vista disso, permaneceram 67 termos válidos para a pesquisa.

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.

[46]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.

5.1.3.2 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)[47].

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 no Quadro 8, sobre os quais Dahlberg (1981) sugere que seja elaborada:

Quadro 8 – 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: Dalhberg (1981, p. 87).

Percebe-se no Quadro 8, 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, o Quadro 9 foi construído para o registro dos conceitos da área de Biodiversidade, a partir de uma adaptação do Quadro 8, por esta autora.

[47]Essas teorias serão explicadas na seção de compatibilização semântica.

Quadro 9 – 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 Dalhberg (1981).

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.

6 RESULTADOS

A presente seção visa apresentar os resultados obtidos com essa pesquisa que teve como questão de partida: **Como melhorar a integração semântica de dados e informações em Biodiversidade?** e como objetivo geral "propor diretrizes para a compatibilização conceitual entre diversos vocabulários usados na indexação de bases de dados em Biodiversidade" promovendo a integração semântica neste domínio. O quadro a seguir é um resumo do caminho percorrido até aqui e um link para o que se pretende apresentar nesta seção.

Quadro 10 – Retrato da Pesquisa

Objetivo Específico	Procedimento Metodológico	Resultado
Identificar, na literatura, o panorama acerca dos SOC voltados para o domínio da Biodiversidade;	Levantamento Bibliográfico	**Revisão da literatura sobre SOC e uso no domínio da biodiversidade** apresentada na subseção 6.1
Compreender os principais conceitos estudados na área da Biodiversidade;	Análise de domínio segundo Hjørland (2002) utilizando 3 das 11 abordagens	Resultado apresentado na subseção 6.2 **O que é Biodiversidade**
Estudar os conceitos de compatibilização semântica propostos pela Ciência da Informação	Apresentação de teorias: Teoria do Conceito – Dalhberg (1978a); Reconciliação de tesauros - Neville (1970); registro do conceito – Dalhberg (1981); Compatibilização Semântica para dados – Sales (2022).	Estudo apresentado na seção 3, intitulado **Compatibilização semântica**
Investigar as possibilidades de aplicação das técnicas de compatibilização semânticas na integração de bases de dados de biodiversidade	Desenvolvimento de testes com base em de Dalhberg (1981), Neville (1970) e Sales (2022) com conceitos selecionados dentro do projeto Pró-Espécies	Apresentado na subseção 6.3 intitulado **Resultados da proposta de compatibilização**

Fonte: Elaborada pela autora (2023).

6.1 REVISÃO DA LITERATURA SOBRE SOC E USO NO DOMÍNIO DA BIODIVERSIDADE

A presente seção é fruto do levantamento realizado nas bases internacionais: 1) *Library, Information Science & Technology Abstracts* (LISTA) *with Full Text*, 2) *Web of Science* e *3) BioOne*. A escolha pela LISTA se deu por ser uma base que conta com a produção de mais de 730 periódicos na área de Ciência da Informação, sendo uma base com grande cobertura temática. Na área da CI, existe também outra base importante que é a *Library & Information Science Abstracts* (LISA), mas que não foi possível utilizar, pois seu acesso não é mais possível através do Portal de Periódicos da Capes.

Como o objetivo era identificar estudos que pudessem servir para o embasamento teórico dessa pesquisa, relacionados aos SOC e a Biodiversidade, os termos empreendidos foram os mesmos utilizados na problemática, porém, no idioma inglês, visto que todas a bases selecionadas são internacionais. Desta forma, a busca na LISTA foi realizada com o intuito de compreender os estudos que ligam os SOC e a Biodiversidade, sob a perspectiva da Ciência da Informação.

Por sua vez, a *Web of Science* e *a BioOne* foram selecionadas com o objetivo de compreender como a temática vem sendo desenvolvida sob a perspectiva da Biodiversidade. A escolha da *Web of Science* se deu pela sua relevância no meio acadêmico, sendo umas das maiores bases de dados científicas da atualidade. No que concerne à *BioOne*, é uma base específica acerca da Biodiversidade e Ciências Ambientais, englobando mais de 200 títulos de periódicos sobre a temática. A busca e a análise do material selecionado para compor o corpus foram realizadas durante o período de 15 a 30 de outubro de 2021.

A primeira base onde a pesquisa foi realizada foi a LISTA, utilizando-se os termos. *Thesaurus*, *Vocabulary* e *Ontology*, aplicados em conjunto com o termo *Biodiversity* e em consonância com o operador booleano *AND*. Os termos estão presentes no vocabulário controlado da própria base.

A primeira busca foi realizada a partir das palavras *Thesaurus* e *Biodiversity* nos campos título, resumo e palavras-chave, todavia não houve menção a qualquer artigo. A escolha pelos campos título, resumo e palavras-chave se deu porque entende-se que se o tema central de cada trabalho estará presente em pelo menos um desses campos.

Em seguida, foram usados os termos *Ontology* e *Biodiversity* nos campos título, resumo e palavras-chave. Essa busca retornou dois artigos, sendo que um dos trabalhos foi desenvolvido por pesquisadores brasileiros, na área da Ciência da Computação, em 2007, e trata acerca de um serviço web de ontologias para interoperabilidade em sistemas de biodiversidade chamado Aondê. Este serviço, ou proposta de ontologia, não foi citado em nenhum momento no curso de Gestão da Informação sobre Biodiversidade e não foi encontrado qualquer registro na web.

A justificativa dos autores para o desenvolvimento do Aondê é baseada na afirmação de que as ontologias podem resolver a falta de padronização e de interoperabilidade entre os dados gerados na área de Biodiversidade. Daltio e Medeiros (2008, p. 725, tradução nossa) afirmam que “apesar de fornecer aos cientistas funções de análise sofisticadas, os [sistemas de informações] exigem padronização de dados e de modelos, e a interoperabilidade nos sistemas

ainda é um problema em aberto[48]". Desta forma, o uso de ontologias favoreceria a padronização da informação presente nos sistemas de informação, além de promover a interoperabilidade semântica.

O outro trabalho apresentou uma avaliação semântica entre uma ontologia e três glossários que podem se tornar ontologias e concluiu que esta não é uma tarefa fácil (Cui, 2010). O objetivo deste autor era descobrir se as ontologias englobam os termos em conformidade com a literatura do domínio. Cui aponta que "os resultados obtidos pela verificação das ontologias em comparação com a literatura sugerem que mais trabalho é necessário para melhorar a cobertura das ontologias e proporcionar um acordo melhor entre elas[49]" (Cui, 2010, p. 1160, tradução nossa). Assim, pode-se dizer que no contexto investigado, as ontologias ainda precisam melhorar sua cobertura temática, mostrando que as ontologias ainda não cobrem tudo o que é produzido de acordo com a literatura científica.

Em sequência, efetuou-se na LISTA a busca pelos descritores *Vocabulary* e *Biodiversity*, também nos campos título, resumo e palavras-chave. A investigação aludiu a dois artigos: um abordou a respeito do trabalho citado anteriormente (ontologia intitulada Aondê), que foi desenvolvido por brasileiros da Ciência da Computação; no outro trabalho recuperado, intitulado "*Organizing Our Knowledge of Biodiversity"*, a questão da organização do conhecimento se refere a um estudo sobre metadados, indexação colaborativa e taxonomia. Esta taxonomia, porém, visa a representação descritiva dos dados, que é a descrição física da informação, como a apresentação de título, autor, data de coleta, e não a representação temática, que é a representação do assunto.

Destarte, na *Web of Science*, os termos empreendidos foram os mesmos usados na base LISTA (*Thesaurus*, *Vocabulary* e *Ontology*), visto que, diferente daquela base, não oferece um vocabulário com termos pré-definidos para a busca. Assim, a busca foi realizada a partir dos campos título, palavras-chave e resumo, sendo analisados o título, o resumo, as palavras-chave e a introdução de cada artigo recuperado, para uma melhor compreensão do tema. Para a análise, considerou-se os artigos que conectam a Biodiversidade e algum tópico relacionado à Organização do Conhecimento.

No tocante à investigação alcançada com os termos *Thesaurus* e *Biodiversity*, foram recuperados 15 artigos. Deste total, seis foram descartados, por tratar da biodiversidade em

[48]Texto original: "*In spite of providing scientists with sophisticated analysis functions, BIS require data and model standardization, and interoperability across BIS is still an open problem*".

[49]Texto original: "*The findings obtained by checking the ontologies against the literature suggest that more work is needed to improve the coverage of the ontologies and bring them into better agreement with each other*".

geral e dos aspectos sobre informática, todavia não relativos à OC. Sendo assim, considerou-se nove artigos válidos para o estudo.

Dentre os artigos válidos, o trabalho intitulado "*SWI: A Semantic Web Interactive Gazetteer to support Linked Open Data"* trata da arquitetura para tesauros geográficos, que usam ferramentas da web semântica, como ontologias e *Linked Open Data*, e também apresenta um dicionário geográfico (*Semantic Web Interactive Gazetteer* - SWI), de modo a implementar a mesma arquitetura do tesauro para mostrar que a ferramenta pode ser usada para adicionar coordenadas geográficas ausentes nos registros de biodiversidade, isto é, todos os tipos de informações devem estar representados nos SOC, visto que são necessários para os pesquisadores e fundamentais para a recuperação da informação. Conforme Schiessl e Shintaku (2012, p. 50), "concretamente, [SOC] fornecem uma base de conhecimento que apoia a busca e recuperação de informação pelo usuário final".

Cinco artigos focaram na criação de um tesauro para a representação de plantas, sendo dois sobre tesauro colaborativo – formado por vários atores que favorecem a colaboração, como o profissional da informação e o acesso aberto.

Por sua vez, os artigos "*ThesauForm—Traits: A web based collaborative tool to develop a thesaurus for plant functional diversity research"* e "*A thesaurus for phytoplankton trait-based approaches: Development and Applicability"* apresentaram propostas de ferramentas e criação de tesauros colaborativos. O primeiro artigo apresenta o *Thesauform – Traits*, ferramenta baseada na web dedicada à construção colaborativa de um tesauro sobre a diversidade funcional de plantas na França. A justificativa para a elaboração desse artefato, parte da necessidade de unificação e de integração dos termos que representam as plantas. Desse modo, os conceitos intitulados 'chave' necessitam ser unificados para facilitar a disseminação da informação. Já o segundo artigo versa a respeito da elaboração de um tesauro colaborativo para os estudos sobre fitoplâncton, mais especificamente dados de pesquisa. A armação de um tesauro evitaria a ambiguidade na descrição dos dados. Conforme Rosati *et al.* (2017, p. 131, tradução nossa), a "Ambiguidade terminológica retarda o progresso científico, leva a esforços de pesquisa redundantes, e, em última análise, impede avanços em direção a uma base unificada para ciências ecológicas" [50]. A estruturação de um SOC, como um tesauro, pode favorecer a padronização dos termos, evitando a ambiguidade na representação do conceito.

Ainda no contexto das plantas, um artigo esteve focado na criação de ontologia para planta, chamada *Flora Phenotype Ontology* (FLOPO). Neste trabalho, Hoehndorf *et al*. (2016)

[50]Texto original: "*Terminological ambiguity slows down scientific progress, leads to redundant research efforts, and ultimately impedes advances towards a unified foundation for ecological Science*".

afirmam haver uma limitação na descrição dos dados sobre flora e suas características. Eis por qual razão a necessidade de uma ontologia.

Por fim, o artigo de Lenters *et al.* (2021) intitulado "*Integration and harmonization of trait data from plant individuals across heterogeneous sources*" descreve um fluxo de trabalho que permite a integração e a harmonização de dados de características de plantas individuais. Segundo os autores, esse fluxo de trabalho foi desenvolvido baseado em padrões de metadados, vocabulários e ontologias voltados para a biodiversidade, como *Ecological Trait-data Standard* (ETS), *Darwin Core* (DwC), *Thesaurus of Plant characteristics* (TOP) e *Plant Trait Ontology* (TO).

Com base nesses SOC, o pesquisador, ao descrever os dados, poderia acrescentar mais informações quanto ao termo de maneira padronizada. Assim, percebe-se que no contexto apresentado, tanto aspectos da representação descritiva, quanto do tema, foram usados no tratamento da informação.

Com relação aos descritores *Ontology* e *Biodiversity*, ainda *na Web Of Science*, foram recuperados 123 artigos, tendo sido descartados: 21 artigos com acesso restrito; 1 capítulo de livro, cujo acesso livre era somente ao resumo e não ao texto completo; e 4 artigos cujos links de acesso apresentavam erros. Para a análise foram selecionados apenas 35 artigos.

Dos artigos recuperados, dois também estavam presentes na base LISTA – o artigo Aondê: *An ontology Web service for interoperability across biodiversity applications*, de Daltio e Medeiros (2008), e o desenvolvido por Cui (2010), intitulado "*Competency Evaluation of Plant Character Ontologies Against Domain Literature"*. A maioria dessas produções versa acerca da criação de ontologias para dados em Biodiversidade, seja para dados sobre peixes, insetos, fenótipos, plantas, seja para dados em Biodiversidade no geral.

A justificativa para o desenvolvimento de ontologias presentes nas pesquisas é a de que os dados em Biodiversidades são heterogêneos, o que poderia dificultar sua representação. Desta maneira, a criação de ontologias poderia promover a semântica dos dados em Biodiversidade. Conforme Walls *et al.* (2014, p. 2, tradução nossa), "padrões e ontologias serão centrais nesta abordagem e irão ajudar os cientistas a fazer uso de dados heterogêneos de maneira confiável e harmonizada" [51].

Outra justificativa encontrada nas pesquisas é a de que as taxonomias usadas para a representação de espécies passam por várias mudanças, cujos termos podem ficar desatualizados. Neste caso, uma ontologia poderia auxiliar na atualização da representação.

[51]Texto original: "*Resilient standards and ontologies will be central in addressing this need and will help scientists make use of heterogeneous data in a reliable, harmonized manner*".

Diante disto, no artigo intitulado "*Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to the Biological Collections Ontology and Related Ontologies"*, Walls *et al*. (2014) propõem a criação de uma ontologia fundada na primícia que os metadados operados no padrão *Darwin Core* são insuficientes para promover a semântica entre os dados, e algumas ontologias do domínio também atendem essa necessidade. Consoante os autores,

> uma revisão do panorama atual de padrões de metadados e ontologias em Ciência da Biodiversidade sugere que os padrões existentes, tais como a terminologia Darwin Core é inadequada para descrever dados de biodiversidade de uma forma semanticamente significativa e computacionalmente maneira útil (Walls *et al.*, 2014, p. 1, tradução nossa)[52].

Esta afirmação demostra que mesmo com o uso de metadados, os pesquisadores precisam descrevê-los, visto não ser o objetivo central do *Darwin Core*. A representação do assunto de um dado e suas interligações se faz necessária no contexto da Biodiversidade e deve ser feita conforme o conceito do termo usado para representar esse assunto. Isto somente será possível com o emprego de um vocabulário conceitual ou um tesauro-com-base-em-conceito.

Segundo Campos e Gomes (2006, p. 349), "tesauro conceitual seria, então, um tesauro com base em conceitos: seu nome indica que cada termo denota um conceito, ou seja, uma unidade de conhecimento". Assim, para que este seja desenvolvido deve-se estabelecer o termo central e a relação entre termos.

O artigo titulado "*The Bari Manifesto: An interoperability framework for essential biodiversity variables"* apresenta 10 princípios para a melhoria nas práticas informáticas ligadas à Biodiversidade. O trabalho de Hardisty *et al.* (2019) cita a necessidade do uso de metadados e de ontologias para representar dados em Biodiversidade.

Concernente às ontologias, os autores relatam que elas devem permitir que os termos dos vocabulários sejam acessíveis e legíveis por máquina, que apresentem relações conceituais e sejam apresentados de modo simples para promover um vasto uso. Assim, entende-se que não é suficiente disponibilizar os dados em sistemas de informação, posto que as informações devem estar presentes de maneira padronizada, facilitando a compreensão da informação, seja pelo uso de metadados ligados a representação descritiva, seja pelas ontologias tocantes à representação conceitual do dado.

---

[52]Texto original: "*A review of the current landscape of metadata standards and ontologies in biodiversity science suggests that existing standards such as the Darwin Core terminology are inadequate for describing biodiversity data in a semantically meaningful and computationally useful way*".

Outro artigo apresentou o desenvolvimento de um vocabulário e de uma ontologia para a modelagem de dados, qual seja, o trabalho desenvolvido por Stucky *et al.* (2019), que trata dos dados sobre a história natural dos insetos e defende que esses dados são heterogêneos. Para isso, para uma integração dos dados faz-se necessário, primeiramente, a elaboração de um vocabulário com a definição dos termos e, posteriormente, o desenvolvimento de uma ontologia que forneça uma semântica, de maneira que os computadores possam compreender o que está escrito. A diferença entre os dois SOC está apenas na sua estruturação, visto que o vocabulário foi a base com os termos para o desenvolvimento da ontologia e, a posteriori, acessada pelos pesquisadores.

Dois artigos recuperados tratam do desenvolvimento de ontologias voltadas a capturar o conhecimento gerado pela experiência dos pesquisadores. Os artigos "*Elicitation Process and Knowledge Structuring: a Conceptual Framework for Biodiversity*", desenvolvido por Albuquerque e Santos (2015), e o artigo "*OpenBiodiv-O: ontology of the OpenBiodiv knowledge management system*", elaborado por Senderov *et al.* (2018), abordam a necessidade de um mecanismo – nos casos específicos, ontologias – para abarcar não só o conhecimento gerado na literatura, mas, também, o que é originado da experiência dos profissionais. A justificativa para a estruturação dessas ontologias é a de que elas são capazes de transformar a percepção tácita em explícita, e é esse saber que deve ser transmitido para os pesquisadores da área.

Ainda foi possível recuperar trabalhos que dissertam sobre a construção de glossários, com vistas à padronização da informação. O artigo "T*axaGloss - A Glossary and Translation Tool for Biodiversity Studies*", elaborado por Collin *et al.* (2016), e "*Building the "Plant Glossary"— A controlled botanical vocabulary using terms extracted from the Floras of North America and China*", produzido por Endara *et al.* (2017), apresentam a criação de glossários tencionando auxiliar a representação da informação, sendo um voltado para área da conservação, e o outro para plantas.

Segundo Endara *et al.* (2017, p. 954, tradução nossa), "o glossário revela ambiguidade no uso da terminologia e pode ajudar a orientar novos autores para escolher os termos mais apropriados para usar em novas descrições" [53]. Isto porque ele foi elaborado para demostrar qual o melhor vocábulo a ser empreendido conforme seu significado, facilitando, assim, a representação da informação de forma semântica.

Somado aos artigos citados acima, o artigo "*Environments and EOL: identification of Environment Ontology terms in text and the annotation of the Encyclopedia of Life*", estruturado

[53]Texto original: "*the glossary reveals ambiguity in terminology usage, and can help guide new authors to choose the most appropriate terms to use in new descriptions*".

por Pafilis *et al.* (2015), apresenta o uso de *taggers* para a alimentação de uma ontologia. O *Environments* é usado como um marcador para identificar termos de maneira rápida em grandes volumes de informação.

No contexto da integração de ontologias, recuperou-se dois artigos. Na pesquisa, Haendel *et al.* (2014) apresentam o *Uberon* como um recurso de ontologia único que permite a interoperabilidade entre dados voltados para os vertebrados. Os autores apresentam o desenvolvimento de várias ontologias sobre a temática por diversos grupos de trabalho, o que contribuiu para a existência de conteúdos misturados e sobrepostos. No tocante aos resultados, as ontologias foram integradas ao *Uberon* e descontinuadas, e as informações contidas nas ontologias encontradas neste recurso tornou o *Uberon* uma única ontologia, cujos termos a respeito do tema foram padronizados e únicos.

Foi possível, também, recuperar um artigo cujos autores propõem uma estrutura para englobar ontologias para dados em agronomia. O trabalho desenvolvido por Jonquet *et al.* (2018) disserta sobre o AgroPortal, uma plataforma que abarca todas as ontologias tocante à agronomia. Segundo os pesquisadores, "oferecemos um portal que oferta hospedagem, pesquisa, controle de versão, visualização, comentário e recomendação de ontologias; permite a anotação semântica; armazena e explora alinhamentos de ontologias; e permite a interoperação com a web semântica" (Jonquet *et al.*, 2018, p. 126, tradução nossa[54]). Neste caso, o projeto poderia ajudar na busca por vocabulários e ontologias para representar dados em Agronomia.

O projeto aludido mostra a variedade de vocabulários existentes na área da Agronomia e ressalta o problema que a presente pesquisa visa investigar: a necessidade de uma metodologia para integração semântica de vocabulários. Diante disso, mesmo propondo a reunião de ontologias, o projeto *Uberon* não foi capaz de resolver o problema da compatibilização semântica, posto que as ontologias reunidas no projeto foram descontinuadas. Assim, o AgroPortal torna-se uma espécie de repositório de ontologia.

Já na pesquisa com os descritores *Vocabulary* e *Biodiversity* foram recuperados 48 artigos. A supressão de alguns destes foi a mesma apresentada anteriormente, isto é, abordavam a Biodiversidade no geral e ligados à informática. Cerca de 35 artigos foram eliminados com esta abordagem. Artigos de acesso restrito mesmo utilizando o Periódicos Capes, também foram eliminados, mais precisamente 6 trabalhos. Ademais, desconsiderou-se os artigos recuperados anteriormente e já analisados na busca. Assim, analisou-se apenas duas produções: *OTO:*

[54]Texto original: "*We offer a portal that features ontology hosting, search, versioning, visualization, comment, and recommendation; enables semantic annotation; stores and exploits ontology alignments; and enables interoperation with the semantic web*".

Ontology Term Organizer, desenvolvida por Huang *et al.* (2015), e o artigo *Hackathon-Workshop on Darwin Core and MIxS Standards Alignment*, estruturado por Tuama *et al.* (2012). No primeiro artigo, os autores tratam da criação de um software para a construção de ontologias. Visto ser desenvolvida por profissionais que não são da Biodiversidade, a ontologia pode não cobrir tudo o que os pesquisadores necessitam para representar a informação. Huang *et al.* (2015, p. 1, tradução nossa) declaram que "As opiniões diversas de diferentes especialistas no domínio na terminologia usadas na literatura como fontes, raramente são tratados pelo software existente" [55]. Desta maneira, elas devem ser capazes de englobar os termos presentes na literatura do domínio e tentar abarcar todas as informações possíveis quanto ao tema e ao conhecimento gerado por esses atores.

Além da presença de um artigo que trata da elaboração de extensões para o Darwin Core nos idiomas japonês e chines, o artigo "*Hackathon-Workshop on Darwin Core and MIxS Standards Alignment"* (Tuama *et al.*, 2012) registra os resultados de um Workshop voltado para discussões sobre o padrão Darwin Core. O trabalho desenvolvido pelos autores concerne a uma extensão do padrão Darwin Core para o uso nesses idiomas.

Na *BioOne*, por sua vez, usou-se termos mais específicos sobre os sistemas de organização do conhecimento, visto que a base é específica sobre a Biodiversidade. Os termos empregados foram *Vocabulary controlled*, *Thesaurus* e *Ontology* e constavam presentes no catálogo da base.

A busca remeteu a dois artigos relativos à Biodiversidade, aos aspectos da OC. Assim, no artigo desenvolvido por Lozano-Fuentes *et al.* (2013), os autores discutiram sobre a criação de uma ontologia para representar vetores artrópodes e patógenos. Para eles, "a falta de padronização na terminologia pode resultar no uso de vários nomes para o mesmo processo, conceito (artefato de informação) ou entidade física" (Lozano-Fuentes *et al.*, 2013, p. 1, tradução nossa[56]).

Outro trabalho que debateu acerca da criação de um tesauro para nomes de pássaros é o artigo intitulado "*A Thesaurus of Bird Names: Etymology of European Lexis Through Paradigms*", cujo pesquisador (Olson, 2001) apresenta um tesauro desenvolvido para abranger a nomenclatura de pássaros com nomes comuns, e não científicos. Ainda, oferece a quem o consulta informações a

[55]Texto original: "*The differences in the opinions of different domain experts and in the terminology usages in source literature are rarely addressed by existing software*".

[56]Texto original: "*Lack of standardization in terminology can result from multiple names being in use for the same process, concept (information artifact), or physical entity*".

respeito da localização das espécies apresentadas. Nesse contexto, o tesauro mostra-se uma rica fonte de informação para os pesquisadores que exploram a temática dos dados sobre pássaros.

Pode-se observar que os SOC são importantes para a Biodiversidade e os pesquisadores estão preocupados com questões de padronização, representação da informação e interoperabilidade semântica. Rosati *et al.* (2017, p. 1., tradução nossa) defendem que "o uso de vocabulários controlados e tesauros é uma boa prática reconhecida para estabelecer a base para interoperabilidade semântica, um requisito crítico para reutilização e compartilhamento de dados[57]". Ainda, aludem aos dados de fitoplâncton.

> Na verdade, tesauro, coletivamente construído, contorna as questões de ambiguidade na linguagem natural, facilitando a identificação e integração de informações disponíveis em várias fontes de dados e permitindo que cientistas e aplicativos interpretem mais efetivamente o significado dos dados (Rosati *et al.*, 2017, p. 129, tradução nossa[58]).

Desta forma, os SOC são importantes ferramentas no que toca à representação da informação, pois são capazes de excluir as ambivalências dos termos em determinados domínios. Segundo Albuquerque *et al.* (2010), as ontologias são apresentadas como uma base semântica para alcançar informações de relevância mostradas em um conjunto de documentos não ordenados, contribuindo, assim, para uma recuperação da informação eficiente.

Antes de apresentar a proposta final da pesquisa, verifica-se a necessidade de contextualizar o domínio da Biodiversidade.

6.2 O QUE É BIODIVERSIDADE

Visando atender o objetivo 2 desta pesquisa, esta Seção apresenta a área de Biodiversidade com base em três das 11 abordagens propostas por Hjørland (2002). Cada uma de suas subseções relatará o resultado da análise realizada com o foco em uma abordagem. Com esta análise, ao observar os Programas de Pós-Graduação no Brasil, a produção indexada na *Scopus* e o ThesBio, pode-se compreender, de forma específica, quais as particularidades presentes nas pesquisas sobre Biodiversidade e assim, compreender o que vem a ser

[57]Texto original: "*The use of controlled vocabularies and thesauri is an acknowledged good practice to establish the foundation for semantic interoperability, a critical requirement for reuse and sharing of data*".

[58]Texto original: "*In fact, thesauri, collectively constructed, bypass ambiguity issues in natural language, facilitating the identification and integration of the information available in multiple data sources and allowing both scientists and computer applications to interpret more effectively the meaning of data*".

Biodiversidade. Esta seção está dividida de acordo com as análises, e por fim, será mostrado o mapa conceitual constituído pela junção dos resultados das análises. A subseção a seguir apresenta os resultados da análise de domínio realizada a partir do estudo dos programas de pós-graduação em Biodiversidade, existentes no Brasil.

6.2.1 Primeira abordagem: a biodiversidade de acordo com os programas de pós-graduação brasileiros

A primeira abordagem da Análise de Domínio escolhida dentre as 11 apontadas por Hjørland (2002), é a análise dos estudos históricos, epistemológicos e críticos. Esta abordagem versa sobre compreender a base do domínio, os fundamentos e a epistemologia.

Como mencionado na seção sobre os procedimentos metodológicos, uma busca realizada na plataforma Sucupira, no mês de setembro de 2021, às 9:28h, retornou como resultado 98 Programas de Pós-Graduação em Biodiversidade, existentes no Brasil, divididos em grandes áreas do conhecimento: **Oceanografia, Botânica, Zoologia e Ecologia**. Nesta seção estão sendo apresentadas as áreas de concentração que formam cada área do domínio.

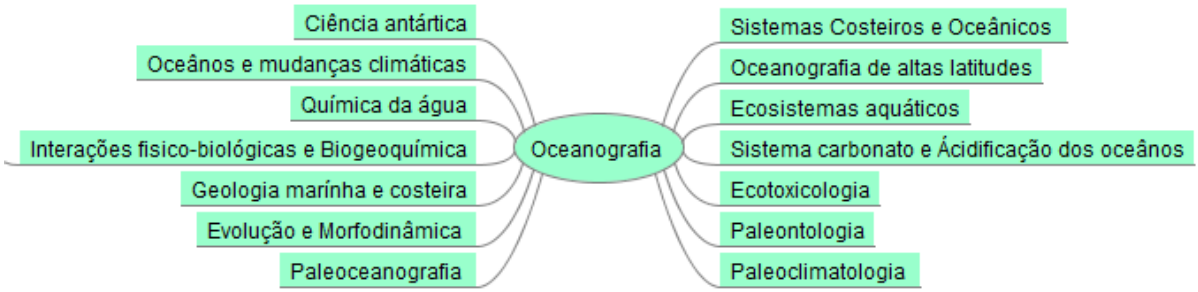
Acerca da área de concentração **Oceanografia**, os temas recuperados são os sistemas costeiros e oceânicos, numa perspectiva de recursos renováveis e não renováveis, a sua dinâmica buscando a quantificação e previsão das características físicas dos oceanos e das regiões estuarinas, além dos aspectos bióticos e abióticos. Ainda foram recuperados outros temas, como a Ciência Antártica e Oceanografia de Altas Latitudes, que busca verificar inter-relação entre os processos oceanográficos e climáticos, nas alterações dos oceanos polares; os Oceanos e as Mudanças Climáticas com o foco no papel dos oceanos no clima da Terra e a influência das alterações do clima na variabilidade natural dos processos dos oceanos, bem como os instrumentos utilizados pela disciplina para a coleta de dados de campo.

Constam, ainda, os temas sobre Ecossistemas Aquáticos; Química da Água; Sistema Carbonato e Acidificação dos Oceanos; Interações físico-biológicas e biogeoquímica; além da contaminação da água e ecotoxicologia, que buscam entender as interações com os processos físicos, químicos e biológicos e os efeitos que alguns agentes podem trazer para o meio biológico.

Tem-se ainda os temas voltados para a Geologia Marinha e Costeira. Estes estudos estão divididos em Geoquímica Ambiental e Marinha, que pesquisam as composições e processos químicos das rochas, solos, sedimentos dos corpos d'água continentais, costeiros e marinhos; Evolução e Morfodinâmica dos Ambientes Deposicionais Costeiros, pesquisando em zona costeira de transição e seus estuários, zonas de arrebentação e praias adjacentes, com o foco nas

modificações morfodinâmicas e nos recursos naturais dessas áreas; também os estudos sobre Paleontologia, Paleoclimatologia e Paleoceanografia, baseados em fósseis de animais vertebrados e invertebrados e nas mudanças paleoceanográficas e paleoclimáticas. Segue, abaixo, a Figura 8 representando os principais tópicos discutidos pela Oceanografia.

Figura 8 – Oceanografia



Fonte: Plataforma Sucupira, Programas de Pós-Graduação em Oceanografia (2021)[59].

No que concerne à área de concentração **Botânica**, foi possível identificar 4 grandes áreas de estudos: Ecologia vegetal, Sistemática vegetal, Morfologia vegetal e Biologia. Os temas de Ecologia vegetal tratam acerca da diversidade, estrutura e dinâmica de comunidades vegetais, interações biológicas, conservação, restauração e funcionamento de ambientes, estudos sobre restauração de comunidades naturais e o efeito de mudanças climáticas globais na dinâmica de populações. São ainda encontrados estudos florísticos, fitossociológicos, fenológicos, etnobotânicos e fisiológicos.

Os temas da área Sistemática vegetal englobam as pesquisas sobre evolução das espécies, bem como a evolução de características anatômicas, morfológicas, reprodutivas, ecológicas e fisiológicas. Também são encontrados estudos filogenéticos, biogeográficos, filogeográficos, etnobotânicos e fenológicos. Dentro desta área, percebe-se a presença de pesquisas em taxonomia, com o intuito de conhecer as espécies e a descrição de novas espécies, bem como entender o funcionamento de hierarquias e a organização da biodiversidade.

Por sua vez, os estudos sobre Morfologia vegetal, abarcam as pesquisas sobre a anatomia das plantas, anatomia ecológica: estrutura de adaptação, estrutura e desenvolvimento de órgãos vegetativos e reprodutivos, fisiologia das plantas, estudos organográficos e palinológicos visando a caracterizar espécies e grupos taxonômicos. Os estudos sobre fisiologia aparecem dentro de Morfologia, com o intuito de subsidiar pesquisas sobre desenvolvimento vegetativo e reprodutivo "*in vitro*".

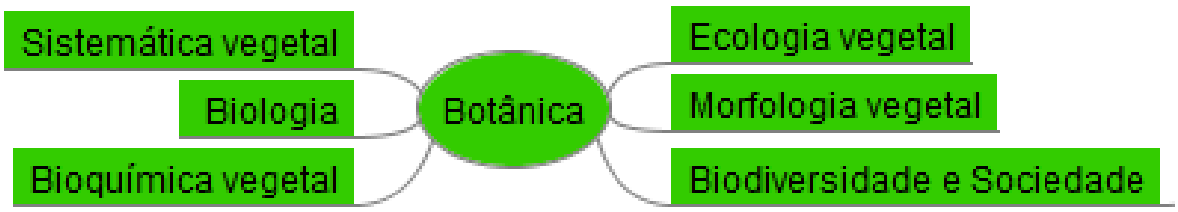
[59] Disponível em: https://sucupira.capes.gov.br/sucupira/. Acesso em: 30 nov. 2021.

No que tange às pesquisas em Biologia, abrangem a Biologia Molecular, que tem seu foco nos processos gênicos e genômicos das espécies, bem como fornece aos pesquisadores subsídios para o desenvolvimento de ferramentas em biologia computacional; a Biologia reprodutiva, com cerne em sistemas reprodutivos e fenologia; Coleções Biológicas, com o foco na curadoria e manejo de coleções biológicas para o desenvolvimento de atividades de restauro e diagnostico de coleções; além dos estudos sobre Bioquímica vegetal, Etnobiologia e o Ensino da Biologia.

Os temas dentro desta área de concentração (Botânica) incluem espécies tanto terrestres quanto aquáticas, bem como a interação das espécies com o solo e a água, além dos estudos sobre fungos e cianobactérias.

Além dessas quatro grandes áreas acima citadas, ainda foram identificados, dentro dos programas relacionado à Botânica, estudos sobre Biodiversidade e Sociedade. Esta área de concentração tem como foco a caracterização da biodiversidade, sua conservação e a relação da sociedade com seu patrimônio ambiental. Além disso, estudos da relação dos vegetais, como recursos econômicos e atividades socioeconômicas para o aproveitamento de plantas, e como recursos naturais e sua utilização de maneira sustentável. A Figura 9 demonstra as principais abordagens da Botânica.

Figura 9 – Botânica



Fonte: Plataforma Sucupira, Programas de Pós-Graduação em Oceanografia (2021).

Os temas da área de estudos **Zoologia** concentram-se em Zoologia Aplicada; Etomologia; Sistemática e Biogeografia; Morfologia e Fisiologia Animal; Ecologia Animal; e Conservação. Sobre a Zoologia aplicada, é correto afirmar que estas pesquisas são desenvolvidas com o foco no cultivo, manejo e controle populacional das espécies, e pesquisas sobre a utilização da Zoologia na interação e integração com as atividades humanas, em ambientes naturais e urbanos, serviços ecossistêmicos e saúde pública.

As linhas que incluem a Etonomologia têm seu foco nos insetos e nos artrópodos, incluindo pesquisas sobre agroecossistemas agrícola e florestal, conhecimento de espécies vetoras de doenças, parasitas, insetos associados às plantas de importância econômica,

especificamente da região amazônica, e estudo das espécies de artrópodos, com ênfase em insetos da região amazônica.

Sobre as pesquisas em Sistemática e Biogeografia, os estudos envolvem a descrição da diversidade biológica, junto com estudos sobre taxonomia, morfologia, estudos de filogenia e de evolução, padrões de evolução das espécies, além dos padrões e processos da distribuição geográfica das espécies.

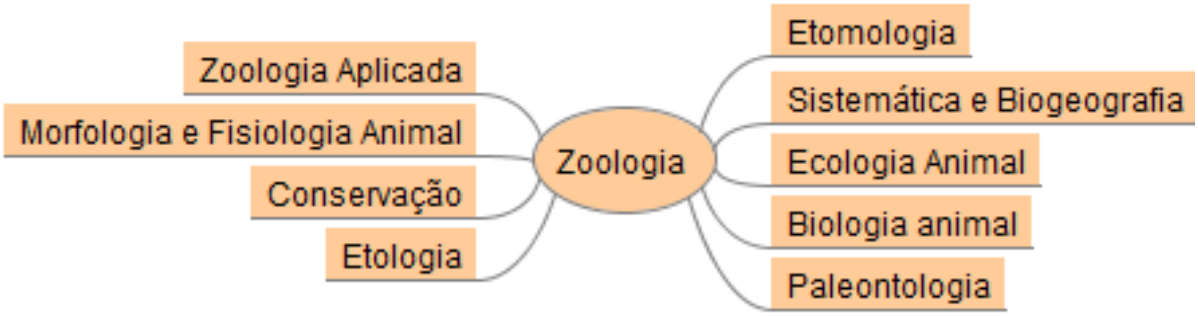
Acerca dos temas sobre Morfologia e Fisiologia Animal, estes se concentram em estudos morfológicos, morfométricos, histológicos, fisiológicos e/ou ecofisiológicos de grupos zoológicos, com enfoque comparado e/ou isolado, incluindo aqui também pesquisas sobre filogenia e sistemática.

Quanto às pesquisas em Ecologia Animal, tratam da relação dos animais com o meio ambiente, com foco no funcionamento dos ecossistemas. Estas pesquisas englobam estudos comportamentais, relações predador-presa, relações parasito-hospedeiro, reprodução das espécies e dinâmicas populacionais. Algumas linhas apresentam focos específicos em Ecologia Aquática e Marinha, e diversidade de insetos.

Os temas sobre Conservação incluem pesquisas sobre conservação das espécies, interações intra e interespecíficas e estratégias para a conservação, estudos referentes à história natural incluindo aqui pesquisas de cenários passados e presentes, desenvolvimento e comportamento das espécies, dieta e reprodução das espécies e fatores que impactam a sua manutenção no meio ambiente. A justificativa para esses estudos é a oportunidade de elaboração de ferramentas para a criação de Unidades de Conservação e o estabelecimento de políticas de conservação.

Ainda foi possível observar temas ligados à Biologia animal, em que são citados estudos sobre biologia do parasitismo, insetos, crustáceos e vertebrados. Pesquisas sobre Etologia, nas quais são encontrados estudos sobre o comportamento, evolução e adaptação das espécies. E ainda Paleontologia, com o foco na sistemática de fósseis e reconstituição paleoambiental. A seguir, a Figura 10 demostrando os principais tópicos tratados pela Zoologia, nos Programas de Pós-Graduação brasileiros, segundo a Plataforma Sucupira.

Figura 10 – Zoologia



Fonte: Plataforma Sucupira, Programas de Pós-Graduação em Oceanografia (2021).

A área de concentração **Ecologia** concentra suas pesquisas nas linhas que tratam sobre Ecologia Aplicada; Conservação; Ecologia de Ecossistemas e comunidades; Ecologia de Populações; Agroecologia; Limnologia; Ecologia e Biodiversidade Terrestre; Ecologia Aquática; Ecossistemas Marinhos; Ecologia Humana; Ecologia e Manejo de Recursos Pesqueiros; Biogeoquímica; Ecologia de Paisagem; Contaminação Ambiental; Sistemática e Evolução; Biogeografia; Ecologia e Genética de Populações; Ecotoxicologia e Qualidade Ambiental; Biologia e Ecologia de Organismos Costeiros; Processos naturais e antrópicos em Sistemas Costeiros; Oceânicos e Gestão e Conservação da Zona Costeira e Oceânica, Recursos Naturais; Ecologia Vegetal e Animal; e Gestão e Educação Ambiental.

Nas pesquisas acerca da Ecologia Aplicada, encontram-se os temas com foco na Gestão Ambiental, que visa, por meio das pesquisas, propor soluções para problemas ambientais, econômicos, sanitários e/ou social, tais como a Ecotoxicologia, além da gestão de recursos ambientais. Também são desenvolvidas, neste contexto, pesquisas sobre o manejo de espécies, conservação e ecofisiologia, ecologia forense, biomonitoramento, uso sustentável da biodiversidade e outras demandas sociais. Em resumo, as pesquisas desenvolvidas nesta linha observam as interações entre os sistemas ecológicos e socioeconômicos. Esses sistemas podem ser tanto terrestres quanto aquáticos.

As pesquisas sobre Conservação têm como objetivo a preservação das Espécies e dos Ecossistemas. Para isso, são realizados estudos sobre Biodiversidade e formas de redução das taxas de extinção, monitoramento de espécies, o efeito de perturbações naturais e antrópicas na estrutura e dinâmica de populações e comunidades animais e vegetais, efeitos de perda e fragmentação de habitats, delineamento de reservas, manejo de espécies, recuperação de áreas degradadas, sistemas impactados pelas ações do homem, além do desenvolvimento de propostas de políticas públicas para a conservação da Biodiversidade.

No que concerne à linha de Ecologia de Ecossistemas e Comunidade, as temáticas de pesquisa tratam sobre: o funcionamento dos ecossistemas e comunidades; processos ecológicos e os fatores que afetam a magnitude e a estabilidade destes processos; bem como as interações de populações biológicas com outras espécies e com o ambiente físico, químico, geológico e o ambiente abiótico e biótico; somam-se a esses tópicos: estruturas de paisagens; multiplicidade de interações entre populações de plantas, animais e microrganismos e entre essas populações e o ambiente físico e químico; modificação de comunidades biológicas e sua integração; e também as interações ecológicas no contexto de comunidades naturais e agrícolas; além de estudos sobre a dinâmica da distribuição e abundância de espécies ou grupos funcionais, fluxos de energia e ciclagem da matéria (decomposição) por meio de comunidades ecológicas, em sistemas aquáticos ou terrestres; fatores físico-químicos do ambiente e sua influência sobre a biodiversidade; também organização de comunidades: competição, predação, parasitismo, mutualismo, alopatia, facilitação; coexistência de espécies: modelos e predições.

A Ecologia de Populações abrange temáticas sobre a biologia, o comportamento, dinâmica, movimento individual, seleção do hábitat, "*life-History*", história natural, ecofisiologia e a descrição e estimativa de modulação de parâmetros demográficos de populações e metapopulações de espécies e sua variação (Ex. tamanho populacional, reprodução, movimento, ocupação) no espaço e no tempo. Essas populações podem ser de plantas e animais em ambientes terrestres, aquáticos e pesqueiros. Algumas linhas apresentam estudos regionais do país sobre estrutura de populações e comunidades vegetais e animais, terrestres e aquáticas (mata atlântica, manguezal, restinga, rios, lagoas e praias), em situações naturais e/ou artificiais.

No que tange à Agroecologia, tem-se os estudos do impacto de práticas agrícolas sobre as condições ambientais e de como as características ambientais podem ser manejadas eficientemente em agroecossistemas.

Sobre a Limnolgia, as pesquisas são voltadas para as comunidades aquáticas, parâmetros físicos, químicos e biológicos, incluindo pesquisas sobre padrões de variação de parâmetros ecológicos do fluxo de energia e ciclagem de nutrientes, qualidade de água e impacto, e efeitos de substâncias poluentes em ambientes e organismos aquáticos.

Nas linhas que tratam sobre a Ecologia e Biodiversidade Terrestre, são encontrados estudos acerca dos organismos, populações, comunidades e paisagens em sistemas terrestres, incluindo também acesso a informações sobre as características químicas, físicas e ecológicas destes ecossistemas. Pertencem a esta linha estudos sobre escala espaço-temporal, padrões ecológicos e processos determinantes e, consequentemente, mantenedores da biodiversidade.

incluindo: história natural, padrões de distribuição das populações, estrutura, conservação, as perdas e o favorecimento de biodiversidade e agrobiodiversidade, bem como a dinâmica e o funcionamento de populações e de comunidades naturais, além de processos associados à degradação ambiental. Algumas linhas têm seu foco em uma determinada região como, semiárido, nordeste do Brasil e ecossistemas tropicais. Além de especificamente apresentarem pesquisas sobre: diversidade e Ecologia de Insetos; Ecologia, Funcionamento e Restauração de Ecossistemas; Ecologia e Conservação de Vertebrados Terrestres; Efeitos de Estresse sobre a Anatomia e Fisiologia de Plantas; Diversidade, Evolução e Estudo Populacional de Plantas; Ecologia Humana; Manejo sustentável de populações naturais; ecologia de invasões biológicas.

As linhas que pesquisam Ecologia e Biodiversidade Aquática englobam pesquisas sobre os organismos, populações, comunidades em ecossistemas de água doce e de estuários, incluindo pesquisas sobre ambientes aquáticos, sua integração e relações com os processos físicos, químicos e geológicos. Os estudos abarcam estrutura e conservação das espécies, história natural, incluindo padrões espaciais e temporais, padrões de distribuição, dinâmica e funcionamento das populações e comunidades e o efeito da intervenção humana na biodiversidade, incluindo aqui a contaminação, poluição e degradação ambiental, e estudos sobre as mudanças climáticas. Além de especificamente apresentarem pesquisas sobre: Uso sustentável e conservação dos recursos aquáticos; Avaliação Ecológica de Organismos não Nativos; Biologia e Conservação de Invertebrados Aquáticos; Diversidade, Evolução e Biogeografia de Peixes; Ecofisiologia de Organismos Aquáticos, Ecologia e Conservação de Vertebrados Aquáticos. Algumas linhas de pesquisa são voltadas especificamente para a Biologia Animal em Ambientes Aquáticos Continentais.

Os Ecossistemas Marinhos englobam pesquisas sobre o ambiente marinho e suas nuances. Os estudos envolvem ecologia de populações e ecossistemas, macroecologia e biogeografia e etnoecologia (uso humano de recursos marinhos). Além das pesquisas que visam a conservação e manejo das espécies.

Os temas em Ecologia e Manejo de Recursos Pesqueiros têm seu foco tanto nos Ecossistemas Marinhos quanto nos aquáticos (incluindo estuários). São estudos sobre recursos pesqueiros e sustentabilidade, além de investigar os instrumentos sociais e políticos que auxiliam a conservação dos recursos pesqueiros por meio de processos formais ou informais de manejo.

A Ecologia Humana inclui estudos acerca do impacto humano no meio ambiente e os resultados destes impactos sobre a própria qualidade de vida do homem, além do uso atual e histórico de recursos naturais para subsistência e a exploração de alternativas econômicas ambientalmente sustentáveis.

Os estudos sobre Biogeoquímica envolvem as fontes e processos que regem a migração dos elementos em ambientes terrestres e aquáticos, interfaces, origem, mobilização, transporte, modificações e destino de materiais orgânicos e inorgânicos durante seu trânsito na paisagem. Estes estudos incluem pesquisas sobre Geoquímica de Metais Pesados e Poluição Ambiental; Biogeoquímica de Ecossistemas Aquáticos; Ciclagem de Nutrientes em Ecossistemas Costeiros e Terrestres; Ecologia de Ecossistemas e Biogeoquímica Aquática e Limnologia Geral e Ecologia de Lagoas Costeiras.

Quanto à Ecologia de Paisagem, reúne estudos sobre o funcionamento das paisagens (estrutura e dinâmica). Apresenta também estudos sobre Metapopulações e metacomunidade, Histórico da ecologia de paisagens, Sistemas agroflorestais e fauna e flora associadas, geoprocessamento no estudo de populações e comunidades, Modelagem computacional de paisagens e generalizações teóricas, mitigação, regeneração e recuperação de paisagens. Além de incluir pesquisas sobre a qualidade e alterações na qualidade dos habitats disponíveis para as espécies, após eventos de fragmentação.

Apesar de algumas linhas incluírem estudos sobre contaminação e interferência do homem, existem linhas que estudam especificamente esses aspectos. A linha Contaminação Ambiental pesquisa sobre a interferência humana na manutenção da biodiversidade, visando compreender como ocorre a contaminação e a degradação ambiental, e o impacto que é causado em ecossistemas e seus organismos, população e comunidades que os formam.

Assim como a Botânica, a Ecologia apresenta uma área de pesquisa sobre Sistemática, mas aqui ela está ligada à Evolução. As linhas de pesquisa que versam sobre Sistemática e Evolução envolvem pesquisas sobre os processos evolutivos da biodiversidade, incluindo estudos de sistemática (morfológica e molecular), distribuição (filogeografia), taxonomia e códigos de nomenclatura, morfologia e anatomia comparada, paleontologia, genética evolutiva e genômica, de plantas e animais.

Sobre as linhas que tratam sobre Biogeografia, estas incluem pesquisas sobre os padrões e fatores que influenciam a distribuição de espécies, história natural, aspectos geológicos e paleoambientais, sistemática das espécies e genética.

Acerca da Ecologia e Genética de Populações, suas pesquisas são voltadas para a diversidade genética e manejo de populações. Assim, os estudos abordam as pesquisas em genética populacional em diferentes proporções (indivíduos, locais, paisagens e demes), incluindo também a Biologia Molecular e Filogenia Molecular.

Os estudos sobre Ecotoxicologia e Qualidade Ambiental apresentam pesquisas voltadas para o vínculo entre os fatores, consequências e alterações causadas por xenobiontes sobre

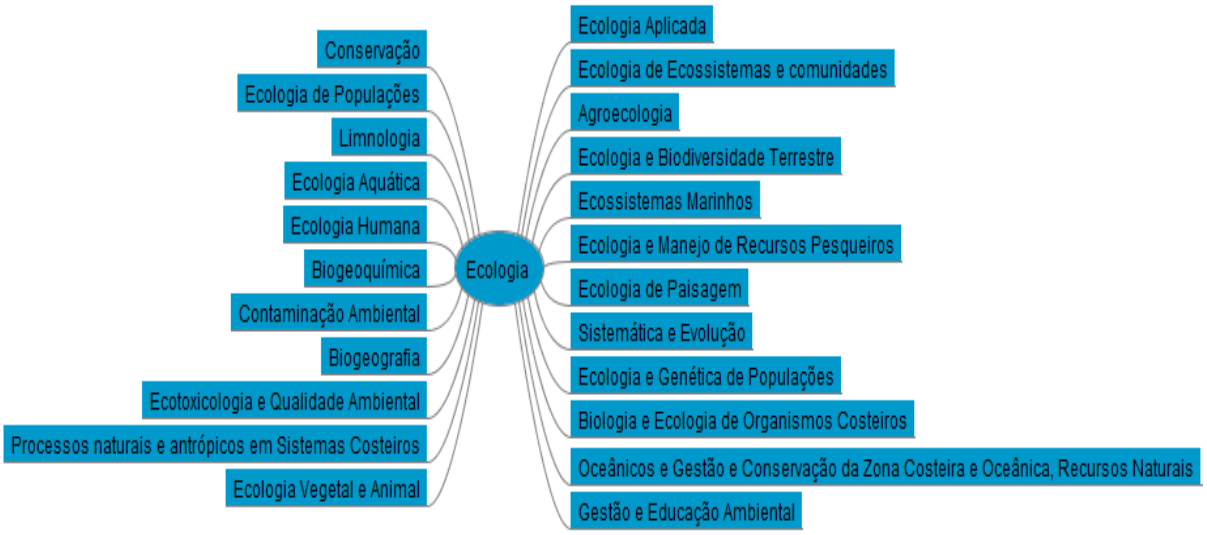
indivíduos, populações e comunidades. Por meio dessas pesquisas é possível observar o nível de toxicidade no ambiente. São encontradas linhas específicas para os Sistemas Costeiros e Oceânicos, tais como a Biologia e Ecologia de Organismos Costeiros, Processos naturais e antrópicos em Sistemas Costeiros e Oceânicos, e Gestão e Conservação da Zona Costeira e Oceânica. A Biologia e Ecologia de Organismos Costeiros incluem pesquisas sobre os mares e oceanos, bem como os organismos vivos presentes nesses ecossistemas, além das atividades de pesca e aquicultura. Pesquisas dos organismos em laboratório e no ambiente e os processos biotecnológicos. Em processos naturais e antrópicos, em sistemas costeiros e oceânicos, as pesquisas buscam relacionar a estrutura química, física e geológica dos sistemas costeiros e oceânicos a estudos de natureza ecológica. Visam, ainda, compreender as circunstâncias de fenômenos climáticos e mudanças ambientais promovidas pelo avanço das atividades humana em ambientes marinhos. Por sua vez, a Gestão e Conservação da Zona Costeira e Oceânica versa sobre aspectos socioambientais e de gestão em zonas costeiras e oceânicas, com o foco na gestão sustentável e uso de recursos naturais, bem como compreender como ocorre a resiliência socioecológica. Esta linha também engloba estudos sobre conservação.

Apesar de os estudos que envolvem os Recursos Naturais ocorrerem nas mais variadas linhas dentro da Ecologia, foi encontrada uma linha específica para esta temática. Essa linha foca sobre a utilização racional e sustentável dos recursos naturais, incluído a bioprospecção da biodiversidade.

Mesmo com a presença de estudos sobre Ecologia, no geral, foram encontradas linhas sobre Ecologia Vegetal e Ecologia Animal. Na Ecologia, as pesquisas abarcam aspectos ecofisiológicos, reprodutivos, biogeográficos, bem como aspectos da biologia das espécies, incluído os estudos sobre população e comunidades. Por sua vez, a linha de Ecologia Animal versa sobre os mesmos aspectos da Ecologia Animal, excluindo a Ecofisiologia, mas inclui estudos sobre Taxonomia e relação das espécies com os mais variados elementos ambientais.

Tem-se ainda a linha de Gestão e Educação Ambiental, com o enfoque na utilização de métodos e teorias da ecologia no manejo e recuperação ambiental e soluções para problemas ambientais, como as mudanças climáticas. A Figura 11 demonstra os principais tópicos inseridos na Ecologia.

Figura 11 – Ecologia



Fonte: Plataforma Sucupira, Programas de Pós-Graduação em Oceanografia (2021).

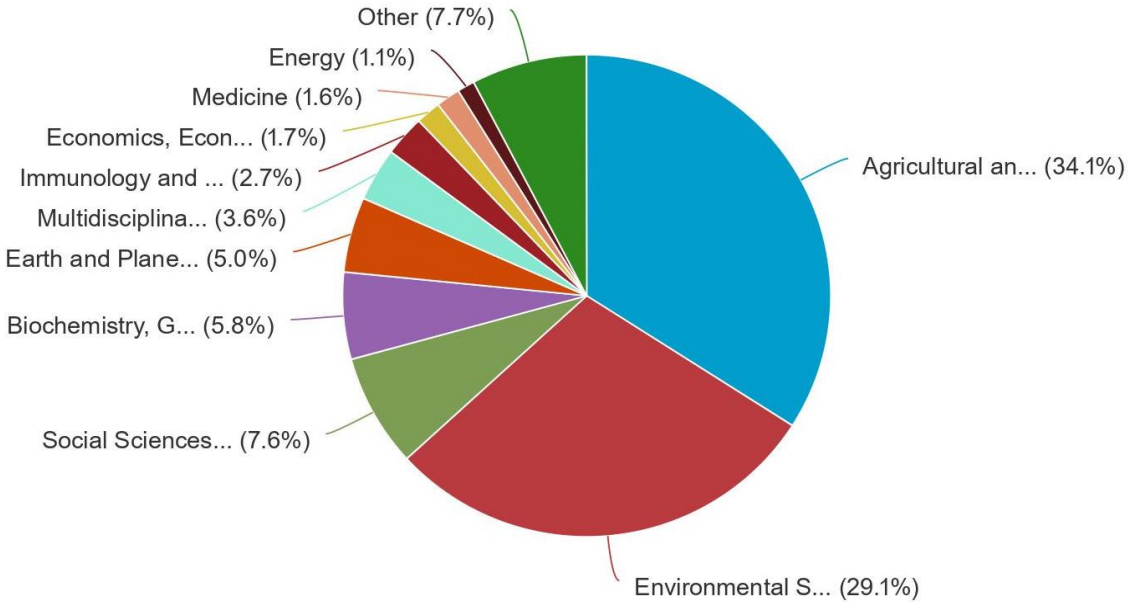
Após apresentar a Biodiversidade no contexto da Pós-Graduação brasileira, se faz necessário conhecer o que está refletido na literatura da área, numa perspectiva global, posto que a busca foi realizada na *Scopus*, base internacional considerada ser uma das maiores e mais importantes bases de dados existentes na contemporaneidade. Desta maneira, a próxima subseção revela o que é Biodiversidade no contexto da produção literária.

6.2.2 Segunda abordagem: a biodiversidade de acordo com a literatura

A segunda abordagem proposta por Hjørland (2002), escolhida nesta pesquisa para a análise do domínio de Biodiversidade, são os estudos bibliométricos. Esta abordagem permite ao pesquisador identificar como ocorre a produção cientifica do domínio investigado. Como a primeira abordagem selecionada é uma abordagem mais básica, o estudo bibliométrico serve para complementar a análise e preencher lacunas deixadas pela abordagem anterior.

O estudo bibliométrico realizado demonstra uma ligação com as temáticas dos programas de pós-graduação, porém com algumas informações novas, como a presença das Ciências Sociais e da Medicina, ligadas à Biodiversidade. A seguir as áreas do conhecimento presentes nos artigos recuperados que estudam a Biodiversidade (Gráfico 1).

Gráfico 1 – Documentos por área temática sobre Biodiversidade



Fonte: Banco de dados *Scopus* (2021).

Conforme o Gráfico 1, a área do conhecimento que mais produziu sobre Biodiversidade foi Ciências Agrícolas e Biológicas, com 34,1% dos artigos recuperados, seguida de Ciência Ambiental, com 29, 1% dos artigos, e na terceira posição Ciências Sociais, com 7,6%. A Biologia aparece aqui apareceu como a maior incidência. Presume-se que seja porque, assim como nos dados dos Programas de Pós-Graduação, também está presente na maioria das pesquisas, sendo que nesta abordagem apareceu em conjunto com as Ciências Agrícolas. Acredita-se que as Ciências Agrícolas se fazem presente na literatura acerca da Biodiversidade porque versa sobre plantas e alimentação e pode impactar os aspectos ligados à mudança climática. Por sua vez, o que o banco *Scopus* denomina Ciência Ambiental, tem total ligação com as temáticas encontradas nos Programas de Pós-Graduação.

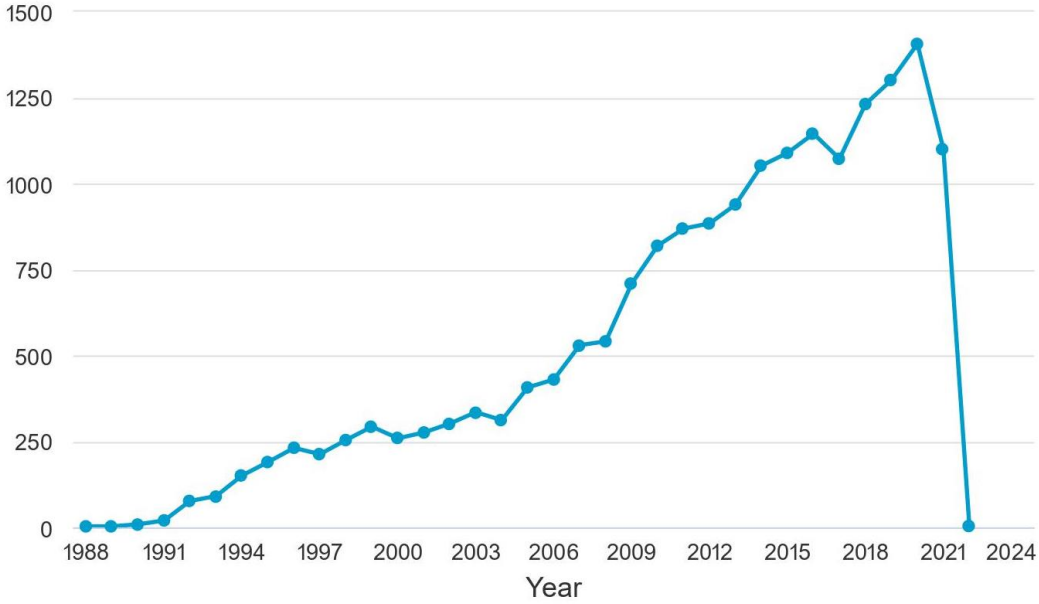
Também são encontradas áreas como Bioquímica, junto à Genética e Biologia Molecular (5,8 %), Ciências da Terra e Planetárias (5,0 %), áreas Multidisciplinares (3,6 %), Imunologia e Microbiologia (2,7 %), Economia, Econometria e Finança (1,7 %), Medicina (1,6), Energia (1,1 %). Com exceção de Medicina, Economia e Energia, todas as outras áreas estão presentes nos estudos dos Programas de Pós-Graduação. Não foi possível descobrir quais são as áreas multidisciplinares que estão presentes no gráfico.

Com os dados da pesquisa, ainda foi possível constatar a presença de outras áreas, apresentadas como Outras, que juntas formam cerca de 7,7% da produção, essas áreas são: Artes e Humanidades, Ciências da Decisão, Negócios, Gestão e Contabilidade, Ciência da Computação, Química, Matemática, Engenharia Química, Farmacologia, Toxicologia e

Farmacêutica, Física e Astronomia, Neurociência, Veterinária, Ciência de Matérias, Psicologia, Enfermagem e áreas ligadas à Saúde. Acredita-se que estas áreas, que não possuem ligação direta com as temáticas dos Programas de Pós-Graduação, estão presentes como áreas secundárias, uma vez que não têm o foco nos estudos ambientais, porém fazem com que os estudos ambientais possam funcionar.

No que concerne ao ano de produção sobre Biodiversidade, foi investigado com o intuito de compreender como ocorreu ao longo dos anos, por isso não foi aplicado um recorte temporal. O Gráfico 2 apresenta o desenvolvimento da área desde 1988 a 2021, em anos, sob o ponto de vista do *Scopus*. É necessário registrar que a busca foi realizada no mês de setembro de 2021 e o ano de 2021 ainda estava em curso, consequentemente a totalidade anual das produções relativa ao ano de 2021 não está adequadamente representada, devendo ser visualizada de forma parcial.

Gráfico 2 – Distribuição dos artigos por ano

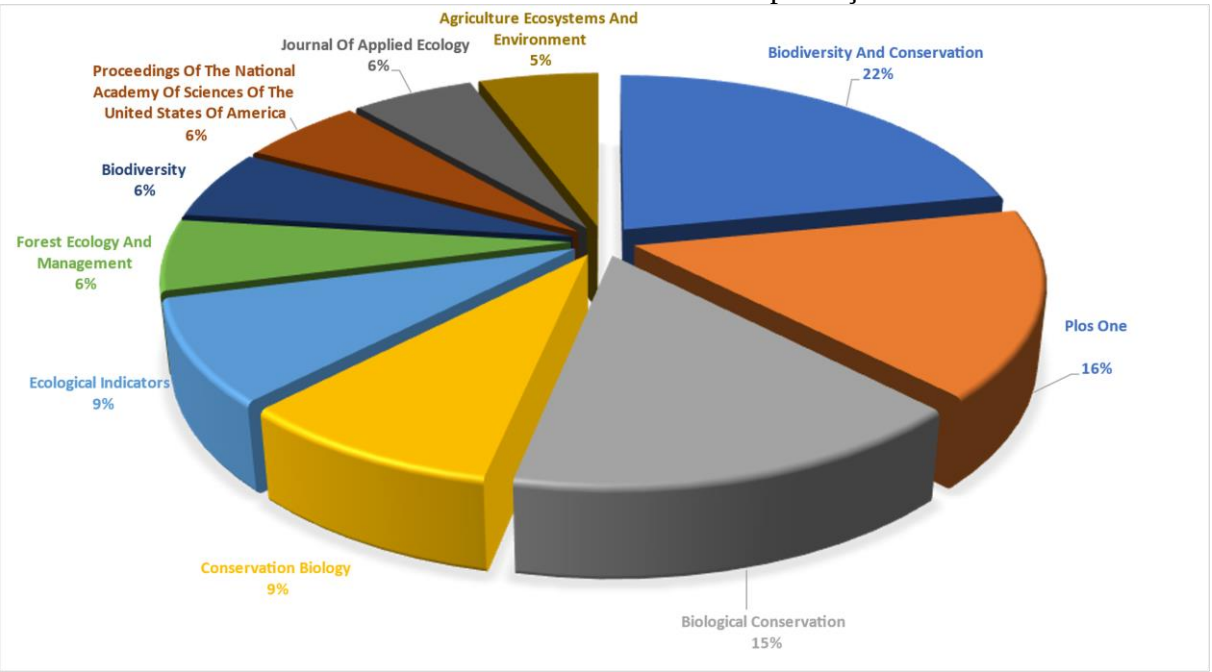


Fonte: Banco de dados *Scopus* (2021).

Como se pode constatar, a produção sobre Biodiversidade surge no final da década de 1980, no século passado, e vem ganhando força com o passar do tempo. Isso confirma o que disserta Wilson (1997), que afirma que Biodiversidade é um termo novo, que até o ano de 1986 não existia formalmente e que começou a ser utilizado no final da década de 1980 do século passado. Não se pode afirmar o porquê do avanço da produção sobre Biodiversidade, porém infere-se que as mudanças climáticas e suas consequências podem ter relação com o aumento da produção.

Os periódicos que apresentam maior incidência de produção são: *Biodiversity and Conservation*; *Plos One*; *Biological Conservation*; *Conservation Biology* e *Ecological Indicators*, todos com mais de 200 artigos recuperados. O único periódico multidisciplinar é o *Plos One*, os outros periódicos estão voltados para os estudos ambientais. Observou-se ainda os periódicos: *Journal of Applied Ecology*, *Forest Ecology and Management*; *Biodiversity; Proceedings of The National Academy of Sciences of The United States of America* e o *Agriculture Ecosystems and Environment*, com mais de 150 trabalhos recuperados. Percebeu-se a presença de um periódico intitulado Biodiversidade, que segundo o *Scopus*, versa sobre Ciências Ambientais e trata dos temas Conservação, Ecologia e Mudanças globais e planetárias. O gráfico 3 abaixo mostra a porcentagem de produção dos periódicos mais citados.

Gráfico 3 – Periódicos com maior produção



Fonte: Banco de Dados *Scopus* (2021).

Por fim, foi realizada a identificação dos termos mais utilizados como palavras-chave nos artigos. A análise dessas palavras-chave é importante para observar quais os termos mais empreendidos pelos pesquisadores na Biodiversidade. Para a elaboração da nuvem de palavras (Figura 12), termos gerais que não se encaixavam no objetivo da pesquisa foram excluídos. Os termos excluídos foram: artigo, Europa, França, Reino Unido, Euroásia, África, China, Ásia, Estados Unidos, Itália, Brasil, América do Norte e África do sul. Também foi excluído o termo Biodiversidade na elaboração da nuvem, uma vez que foi o termo utilizado para a busca, tendo sido recuperado em todos os trabalhos.

Figura 12 – Palavras-chave dos artigos

Fonte: Banco de dados *Scopus* (2021).

Como se pode observar, os termos com o maior número de citações, são termos ligados às ciências ambientais, tais como: riqueza de espécies, animais, conservação, diversidade de espécies, não humanos, proteção ambiental, ecossistema, gestão de conservação, alterações climáticas, conservação de recursos naturais, taxonomia, aves, conservação da biodiversidade, ecologia, humano, floresta, classificação, planejamento de conservação, filogenia, serviço de ecossistema, genética, entre outros. As palavras recuperadas nos artigos têm fusão com as quatro grandes áreas encontradas nos programas de Pós-Graduação e suas pesquisas. Também são encontrados termos ligados à Agroecologia e à Química, é possível que ocorra porque as pesquisas em Ecologia envolvem essas temáticas e os estudos ligados à Química ocorrem na Oceanografia, Ecologia e Botânica.

Após conhecer a produção de literatura da Biodiversidade na base *Scopus*, se fez importante entender como ocorre a representação da informação na área. Sabe-se que a Biodiversidade faz uso dos mais variados Sistemas de Organização do Conhecimento (SOC), mas quais termos e aspectos do domínio são contemplados nesses Sistemas? A análise da representação foi elaborada com base em um tesauro, que é um tipo de SOC.

6.2.3 O que é biodiversidade de acordo com a representação do domínio

Como citado na proposta metodológica, a terceira abordagem de análise de domínio, de Hjørland (2002), adotada, foi a de estudo terminológico, cuja função é identificar como o domínio apresenta e define seus conceitos. Essa análise da representação do domínio foi desenvolvida com base no ThesBio, um tesauro desenvolvido pela rede *Scielo*. A escolha do ThesBio se deu por ser desenvolvido especificamente para a área da Biodiversidade e porque está atrelado a uma fonte de informação importante no contexto do acesso aberto da informação, a Scielo. Como explicado na seção de compatibilização semântica, o tesauro em muitos casos pode ser específico para um determinado domínio, esse é o caso do ThesBio. Um tesauro é um SOC muitas vezes desenvolvido em forma de lista com os termos que descrevem um determinado domínio do conhecimento ou apresenta subdivisões para o uso dos termos.

O ThesBio apresenta em sua estrutura os termos que devem ser usados, além de oferecer ao leitor notas de escopo, com o objetivo de proporcionar melhor compreensão do uso dos termos e de suas aplicações. As notas podem ser encontradas em espanhol, inglês e português. Pela análise do ThesBio, pode-se perceber que a temática da Biodiversidade está sendo desenvolvida por meio de disciplinas ligadas ao Meio Ambiente, assim como às Ciências Humanas, Médica e Tecnológica.

Na capa do tesauro, encontram-se o que podemos identificar como classes temáticas que formam o tesauro, são elas: Biossegurança e Biotecnologia; Ciências Agrárias; Ciências Ambientais; Denominações Geográficas; Disciplinas; Fenômenos e Processos; Organismos e Política e Gestão. Essas classes temáticas não possuem notas de escopo para a sua definição, alguns dos termos, quando apresentados isolados, como por exemplo Biossegurança, possuem uma nota de escopo e a fonte bibliográfica de onde foi retirada a informação. Quando se analisou a classe temática Disciplina, foram recuperados os seguintes termos: Agronomia, Biologia, Ciências Médicas, Direito, Educação, Engenharia Química, Engenharia Genética, Oceanografia, Pesquisa, Tecnologia da Informação e Zootecnia. Foram encontradas disciplinas que ainda não tinham sido citadas nem pelos programas de Pós-Graduação e nem pelos artigos do *Scopus*, como Direito, Educação, Engenharia Química, Engenharia Genética, Pesquisa, Tecnologia da Informação e Zootecnia. A seguir, no Quadro 11, a apresentação das disciplinas com as notas de escopo presentes no ThesBio.

Quadro 11 – Disciplinas presentes no ThesBio

Disciplina	Definição
Agronomia	Área das Ciências Agrárias que estuda e desenvolve novas técnicas agrícolas de modo a otimizar a produção no âmbito econômico, técnico, social e ambiental.
Biologia	Área da ciência que estuda os seres vivos, em diversos níveis de complexidade: desde o molecular até os ecossistemas. Neste estudo, destaca-se a interação dos organismos vivos entre si e com o meio em que vivem atualmente e ao longo do tempo evolutivo.
Ciências Médicas	Área da Ciência associada ao estudo da saúde humana, desenvolvimento de tratamentos médicos e a prevenção de doenças.
Direito	Sistema de regras e normas implementadas por instituições sociais com o objetivo de governar e gerir o comportamento dos membros de uma sociedade.
Educação	É um processo de atuação de uma comunidade sobre o desenvolvimento do indivíduo a fim de que ele possa atuar em uma sociedade pronta para a busca da aceitação dos objetivos coletivos. Para tal educação, devemos considerar o homem no plano físico e intelectual consciente das possibilidades e limitações, capaz de compreender e refletir sobre a realidade do mundo que o cerca, devendo considerar seu papel de transformação social como uma sociedade que supere nos dias atuais a economia e a política, buscando solidariedade entre as pessoas, respeitando as diferenças individuais de cada um.
Engenharia Bioquímica	Ramo da engenharia que promove o estudo dos fenômenos químicos que se passam nos seres vivos e tem como objetivo a obtenção de produtos de grande interesse para a sociedade.
Engenharia Genética	Área da ciência que estuda os processos de manipulação de genes em um organismo.
Oceanografia	Estudo dos oceanos sob todos os seus aspectos: seus componentes bióticos e abióticos, como também no que diz respeito aos processos que atuam nestes ambientes.
Pesquisa	Processo sistemático de construção do conhecimento por meio de uma metodologia, no qual os resultados devem gerar novos conhecimentos e/ou corroborar ou refutar algum conhecimento pré-existente.
Tecnologia da Informação	Área de conhecimento que desenvolve técnicas e projetos para gestão de informação por meio de sistemas e equipamentos para acesso, operação e armazenamento dos dados que fomente tomadas de decisão.
Zootecnia	Ciência aplicada ao estudo de animais úteis ao homem, buscando produção e produtividade, com base em conceitos de sustentabilidade econômica, ambiental e social.

Fonte: ThesBio (Rede BHL Scielo, 2021).

Pode-se perceber que algumas áreas, isoladamente, como Pesquisa, Direito e Tecnologia da Informação, parecem não ter ligação com os estudos sobre “Meio Ambiente e áreas afins”, porém quando apresentados os termos gerais e termos específicos para a representação dos documentos, a conexão com o que se pensa sobre Biodiversidade é revelada. Quando olhados os termos ligados ao “Direito” presentes no tesauro, verifica-se a ligação com os estudos do meio ambiente, como “Direitos de pesca e Legislação Ambiental”. Na disciplina “Pesquisa”, observou-se a presença de termos como “Estudo de Caso e Experimentação”, que são métodos utilizados nos estudos sobre Biodiversidade. Em “Tecnologia da Informação”, são

categorizados conceitos que representam ferramentas para a circulação da informação em Biodiversidade, como os termos “Gestão da Informação” e “Redes de Informação”.

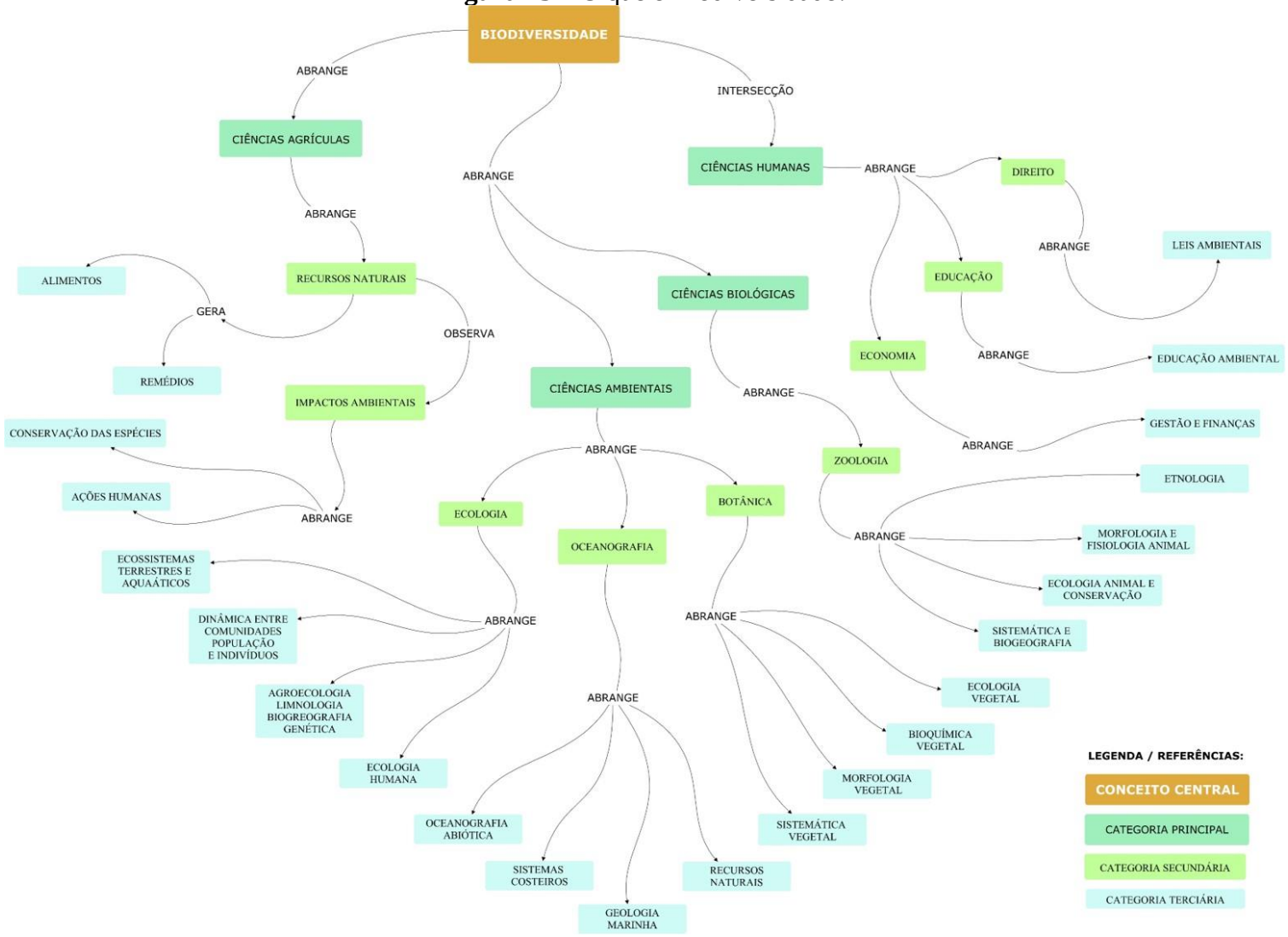
Quanto à observação e comparação dos termos presentes no ThesBio, nas áreas dos Programas de Pós-Graduação e nos artigos no *Scopus*, aparentemente percebeu-se um cruzamento entre as temáticas sobre Ecologia, Oceanografia, Botânica, Zoologia, Ciências Médicas e Saúde, Agronomia e Ciências Sociais, aqui representadas pelo Direito e os aspectos da Tecnologia. No entanto, seria preciso um estudo conceitual mais aprofundado para identificar esses cruzamentos com mais clareza.

Por ser tratar de um tesauro cujo objetivo é a representação da informação, os termos são apresentados de maneira mais específica do que nas ementas dos Programas. Como exemplo, o termo “Conservação”, que está presente como objeto de estudo nos Programas de Pós-Graduação e como objetivo de produção científica presente no *Scopus*, no ThesBio este termo pode ser encontrado como: Biologia da Conservação, Conservação da Diversidade Biológica, Conservação da Natureza, Conservação de Recursos, entre outros.

Percebe-se que a Biodiversidade se apresenta sempre centralizada nas questões ambientais. Juntando o estudo realizado a partir das 3 abordagens de Hjørland (2002), um mapa conceitual foi construído a fim de demonstrar um panorama do que foi apreendido sobre o domínio da Biodiversidade.

Um mapa conceitual é um instrumento para a representação da informação de forma sintetizada. Para Rodrigues e Cervantes (2014, p.158), “os mapas conceituais constituem-se de uma técnica para cumprir vários objetivos, porque representam relações entre os conceitos de uma área, disciplina ou assunto”. Para a elaboração do mapa, a seguir, foi utilizado o software chamado *CMP*tols, que foi desenvolvido com a finalidade específica de auxiliar na elaboração de mapas conceituais. Com a junção dos dados coletados a partir das 3 abordagens de análise de domínio de Hjørland (2002) e sua estruturação através de um mapa conceitual (Figura 13).

Figura 13 – O que é Biodiversidade?



Fonte: A autora (2021).

Como se pode observar no mapa, a Biodiversidade tem seu ponto focal nos estudos ambientais, englobando todos os aspectos da Ecologia, Oceanografia e Botânica. Além disso, também versa sobre os estudos agrícolas, uma vez que esses estudos lidam diretamente com os recursos naturais. Da mesma forma, abarca a Ciência Biológica em seu escopo com os estudos de Zoologia e com as pesquisas sobre os animais. Além do mais, faz intersecção com as Ciências Humanas, uma vez que as Ciências Ambientais e Agrícolas se valem de alguns aspectos das Ciências Humanas para funcionar.

Neste sentido, com base na Teoria da Terminologia de Wüster (1981), estudo voltado para o entendimento das definições, dentro da Organização do Conhecimento, cujo objetivo é permitir uma comunicação mais efetiva acerca das terminologias e o desenvolvimento de termos entre pares e especialistas das mais variadas áreas do conhecimento, arrisca-se aqui propor uma definição para o domínio da Biodiversidade como a ciência que se ocupa de toda a diversidade da terra, com foco nos estudos ambientais, agrícolas e biológicos e que se utiliza de alguns tópicos das Ciências Humanas, para ter êxito. Cabe frisar que esses resultados estão baseados no que foi desenvolvido na análise de domínio, os resultados poderiam ser diferentes ou não, a depender da utilização de outras abordagens.

6.3 RESULTADOS DA PROPOSTA DE COMPATIBILIZAÇÃO

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.

O Quadro 12 contém a lista com os 41 termos que participaram da análise.

Quadro 12 – 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

Termos	Termos
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 foram selecionados seis conjuntos de conceitos, que foram analisados em duplas, trios e até mesmo em quartetos, aqui tentando apresentar termos que tentam contemplar as mais variadas letras do alfabeto, para ficar uma representação dinâmica. A complementação da análise dos outros termos pode ser vista no Apêndice A, no final do trabalho. Com base nos resultados conseguidos, será proposto um modelo com regras para que a compatibilização semântica seja realizada.

➢ COMPATIBILIZAÇÃO DO PRIMEIRO CONJUNTO

O primeiro conjunto de conceitos analisados está relacionado ao metadado *Basis of Record*. O Quadro 13 mostra o primeiro registro do conceito do referido termo.

Quadro 13 – Registro do conceito "*Basis of Record"*

Aspecto analisado	Resultado
Nome do conceito	Basis of Record
Identificador	415
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	A natureza específica do registro de dados
Outros nomes para o conceito ou classe	Basis Of Record – original Basis of record badly formed BasisOfRec
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr-ALA

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

Como se pode observar, o referido metadado ainda possui mais 3 conceitos que podem ser passiveis de compatibilização, mas antes de estabelecer qualquer sugestão, faz-se necessário conhecer as definições dos outros conceitos. No Quadro 14, consta a análise do conceito *Basis of Record* – original.

Quadro 14 – Registro do conceito "*Basis of Record* – original"

Aspecto analisado	Resultado
Nome do conceito	Basis Of Record – original
Identificador	416
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	A natureza específica do registro de dados
Outros nomes para o conceito ou classe	Basis Of Record Basis of record badly formed BasisOfRec
Fonte do conceito	SiBBr-ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr-ALA

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

Tal como mostra o quadro acima, o conceito apresentado tem a mesma definição do conceito apresentado anteriormente no Quadro 13 e pertence ao mesmo sistema de informação, o SiBBr. Cabe frisar que o SiBBr, caso queira-se, pode fazer uma revisão em seus metadados e estabelecer qual será utilizado formalmente. Como os conceitos foram identificados pelo consultor, serão considerados para uma tentativa de compatibilização junto às outras ocorrências que talvez sejam correspondentes. O Quadro 15 a análise de outro metadado suscetível à compatibilização com os conceitos *Basis of Record*.

Quadro 15 – Registro do conceito "*Basis of record badly formed"*

Aspecto analisado	Resultado
Nome do conceito	Basis of record badly formed
Identificador	425
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Base de registro mal-formada
Outros nomes para o conceito ou classe	Basis Of Record Basis Of Record - original BasisOfRec
Fonte do conceito	SiBBr-ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr-ALA

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

No Quadro 15, pode-se perceber que o conceito examinado se trata de outro conceito, com outra função, podendo ser utilizado quando o pesquisador, que for acrescentar as informações sobre os dados, não possuir as informações precisas sobre o registro dos dados. Ainda dentro da planilha do consultor, foi identificado um quarto conceito que pode compor esse primeiro conjunto de conceitos. A seguir (Quadro 16), a sua análise por meio do registro do conceito.

Quadro 16 – Registro do conceito "*BasisOfRec"*

Aspecto analisado	Resultado
Nome do conceito	BasisOfRec
Identificador	121
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	A natureza específica do registro de dados.
Outros nomes para o conceito ou classe	Basis Of Record Basis Of Record – original Basis of record badly formed
Fonte do conceito	Padrão IUCN
Observações e comentários	Termo ou conceito pertence ao sistema de informação CNC-Flora

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

Apesar do termo *BasisOfRec* ter uma origem diferente e ser empregado em um outro sistema de informação, ele possui a mesma definição que os dois conceitos do conjunto analisado. Para uma melhor proposição de compatibilização, cabe a análise de acordo com Sales (2022), para a comparação do referente dos conceitos analisados.

Quadro 17 – Análise Semântica Comparativa do conjunto *Basis Of Record* conforme Sales

	Termo 01 – Basis Of Record	**Termo 02 – Basis Of Record – original**	**Termo 03 – Basis of record badly formed**	**Termo 04 - BasisOfRec**
Referente	Registro dos dados	Registro dos dados	Complemento para o registro dos dados	Registro dos dados
Característica 01	Voltado para os dados de maneira geral	Voltado para os dados de maneira geral	Voltado para os dados de maneira geral	Voltado para os dados de maneira geral
Característica 02	Presente no SiBBr	Presente no SiBBr	Presente no SiBBr	Pertence ao sistema de informação CNC-Flora

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

Logo, é possível que ocorra a compatibilização entre os conceitos "*Basis Of Record"* e "*Basis Of Record* – original", que constam no SiBBr e "*BasisOfRec"*, que consta no CNC-Flora, pois os conceitos, mesmo apresentando grafias diferentes, possuem o mesmo referente. Dessa forma, considera-se os conceitos citados equivalentes, pois contemplam o mesmo objetivo. Com relação ao conceito *Basis of record badly formed*, sugere-se que sua utilização seja voltada para a complementação de informações, uma vez que o SiBBr o considera importante, os outros sistemas de informação podem analisar o seu potencial de uso futuro. Essa recomendação versa sobre a teoria do conceito que fala que os enunciados formam o

conceito e no caso, o conceito *Basis of record badly formed* não pode ser considerado equivalente semanticamente.

Conforme apresentado na metodologia, foi realizada, também, por meio da proposta de Neville (1970), que é um método voltado para o signo, com foco no termo, a aplicação de códigos (Quadro 18). Assim, os conceitos correspondentes puderam receber um identificador único dentro do sistema, que remeta ao mesmo metadado, não precisando que o sistema de informação altere o termo presente dentro do sistema.

Quadro 18 – Atribuição do Código 0101 ao conjunto *Basis Of Record* conforme o Método de Neville

	Termo 01	***Termo 02***	***Termo 03***
Termo original	*Basis Of Record*	*Basis Of Record – original*	*BasisOfRec*
Identificador	*Basis Of Record* (0101)	*Basis Of Record – original* (0101)	*BasisOfRec* (0101)
Termos com os identificadores	0101= *Basis Of Record*	0101= *Basis Of Record – original*	0101 = *BasisOfRec*

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

A aplicação da proposta de Neville (1970), no Quadro 18, foi realizada com os conceitos que eram considerados aptos para receber o mesmo identificador. Sendo assim, os conceitos foram aptos para receber o mesmo código, mesmo com grafias diferentes, e conforme a metodologia do autor, tornam-se possíveis de compatibilização. Cabe frisar que esta análise e as outras que virão a seguir, foram desenvolvidas em um contexto acadêmico, cabendo sua utilização, ou não, por parte dos pesquisadores do projeto Pró-Espécies.

- **COMPATIBILIZAÇÃO DO SEGUNDO CONJUNTO**

Na sequência, será apresentada a análise do conjunto de conceitos formado pelo metadado *Class* (Quadro 19). A primeira parte de análise seguirá a ordem apresentada como no conjunto anterior, o registro do conceito.

Quadro 19 – Registro do conceito "*Class" termo 01*

Aspecto analisado	**Resultado**
Nome do conceito	Class
Identificador	4
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	O nome científico completo da classe em que o táxon é classificado.
Outros nomes para o conceito ou classe	class classname

Aspecto analisado	Resultado
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação Flora e Funga do Brasil e Catálogo da Fauna

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

O metadado acima é oriundo do padrão Darwin Core e está presente em dois sistemas de informação. Além disso, o metadado apresenta dois conceitos que formam o conjunto para análise, que podem ser sujeitos à compatibilização.

No Quadro 20, a seguir, tem-se o registro do conceito do termo *Class* presente no sistema de informação SiBBr.

Quadro 20 – Registro do conceito "*Class"termo 02*

Aspecto analisado	Resultado
Nome do conceito	Class
Identificador	392
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	O nome científico completo da classe em que o táxon é classificado
Outros nomes para o conceito ou classe	Classname Class
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

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

Como se pode observar, o conceito é oriundo do mesmo padrão de metadado, *Darwin Core*, e com isso trata do mesmo termo utilizado pelos sistemas de informação do JBRJ. Abaixo, pode-se conferir ainda o registro de outro termo presente nos sistemas de informação que possui semelhança com os conceitos anteriores.

Quadro 21 – Registro do conceito "*Classname"*

Aspecto analisado	Resultado
Nome do conceito	Classname
Identificador	46
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Nome da classe taxonômica ao qual a espécie pertence.
Outros nomes para o conceito ou classe	Class Class
Fonte do conceito	IUCN
Observações e comentários	Termo ou conceito pertence ao sistema de informação CNCFlora e SALVE

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

Neste Quadro 21, na leitura do registro do conceito, ao observar sua definição, percebe-se que os termos *class* e *classname* podem corresponder ao mesmo metadado, apenas apresentam definições escritas de formas diferentes, mas apresentam o mesmo significado.

No Quadro 22 os mesmos conceitos serão analisados com base na proposta de Sales (2022), para quem usuários diferentes ao identificar qual é o referente podem perceber que estão falando do mesmo objeto.

Quadro 22 – Análise do conjunto *Class* conforme Sales

	Termo 01 – Class	**Termo 02 – Class**	**Termo 03- Classname**
Referente	Nome da classe taxonômica	Nome da classe taxonômica	Nome da classe taxonômica
Característica 01	Voltado para a classe das espécies	Voltado para a classe das espécies	Voltado para a classe das espécies
Característica 02	Presente nos sistemas Flora e Funga do Brasil e Catálogo da Fauna	Presente no SiBBr	Presente nos sistemas CNCFlora e SALVE

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

O Quadro 22 mostra que 4 dos sistemas de informações presentes no projeto são geridos pela mesma instituição, que é o JBRJ e ainda assim, utilizam metadados diferentes. Mas, ao se fazer a análise conforme Sales (2022), é possível identificar que mesmo os metadados sendo diferentes, eles possuem o mesmo referente, pelo fato de possuírem a mesma característica (01), que no exemplo é o nome da classe taxonômica. Além disso, o outro sistema de informação (SiBBr) também utiliza o conceito em comum para a mesma função, logo, é possível que ocorra a compatibilização semântica entre os conceitos. A este conjunto *Class* foi atribuído o código 0202 (Quadro 23), com base em Neville (1970).

Quadro 23 – Atribuição do Código 0202 ao conjunto *Class* conforme o Método de Neville

	Termo 01	***Termo 02***	***Termo 03***
Termo original	Class	Class	Classname
Identificador	Class (0202)	Class (0202)	Classname (0202)
Termos com os identificadores	0202= Class	0202= *Class*	0101 = Classname

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

Como os conceitos possuem o mesmo referente, a aplicação dos códigos para identificar a mesma função do termo, dentro dos sistemas de informação, no projeto, é algo possível.

➢ COMPATIBILIZAÇÃO DO TERCEIRO CONJUNTO DE CONCEITOS

Na sequência, será apresentado um conjunto de conceitos composto por apenas dois termos: *Ocurrence Status* (Quadro 24). Diferente dos conjuntos anteriores, este conjunto pertence a sistemas de informação geridos pelo JBRJ e o SiBBr.

Quadro 24 – Registro do conceito "*Ocurrence Status*"

Aspecto analisado	Resultado
Nome do conceito	Ocorrence Status
Identificador	417
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Status da Ocorrência
Outros nomes para o conceito ou classe	Occurrence status assumed to be present Ocurrence remarks
Fonte do conceito	SiBBr
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBR

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

O conceito acima refere-se à ocorrência da espécie encontrada e seu status. É utilizado apenas em um sistema de informação, o SiBBr No Quadro 25, o registro do outro termo do *Occurrence status assumed to be present*.

Quadro 25 – Registro do conceito "*Occurrence status assumed to be present"*

Aspecto analisado	Resultado
Nome do conceito	Occurrence status assumed to be present
Identificador	424
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Status da ocorrência considerado presente
Outros nomes para o conceito ou classe	Ocurrence Status Ocurrence remarks
Fonte do conceito	SiBBR
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBR

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

Como se pode observar, os conceitos acima não parecem ter o mesmo objetivo, mas poderiam ser compatibilizados, caso o sistema informe ao pesquisador que todas as ocorrências de espécies, mesmo que não tenha certeza da informação podem ser descritas no metadado *Ocurrence Status*.

➢ COMPATIBILIZAÇÃO DO QUARTO CONJUNTO DE CONCEITOS

Na sequência, será apresentado um conjunto de conceitos composto por quatro termos: *order* e *ordername*. São termos presentes tanto nos sistemas do JBRJ quanto no SiBBR. O primeito termo analisado aparece no Flora e Fauna do Brasil.

Quadro 26 – Registro do conceito "*Order"termo 01*

Aspecto analisado	Resultado
Nome do conceito	Order
Identificador	5
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Order Ordername Order
Outros nomes para o conceito ou classe	O nome científico completo da ordem em que o táxon está classificado.
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação Flora e Fauna do Brasil

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

O conceito acima refere-se ao nome da ordem ao qual o táxon está classificado. Essa é uma importante informação para os especialistas que trabalham classificando as espécies. Em seguida temos o registro do homônimo *order* no quadro 27.

Quadro 27 – Registro do conceito "*Order"termo 02*

Aspecto analisado	Resultado
Nome do conceito	Occurrence status assumed to be present
Identificador	97
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	O nome científico completo da ordem em que o táxon está classificado.
Outros nomes para o conceito ou classe	Order Ordername Order
Fonte do conceito	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).

Como se pode observar, os conceitos acima parecem ter o mesmo objetivo, mesmo com a fonte de informação de origem distintas. Mas ainda falta a análise de dois termos para podermos avaliar uma compatibilização possível.

Quadro 28 – Registro do conceito "*Ordername"*

Aspecto analisado	Resultado
Nome do conceito	Ordername
Identificador	54
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Nome da ordem taxonômica ao qual a espécie pertence.
Outros nomes para o conceito ou classe	Order Order Order
Fonte do conceito	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).

Após o registro do conceito acima, pode-se observar que ele tem a origem, objetivo e uso igual ao conceito apresentado no quadro 27. Dessa forma, acredita-se que são passiveis de compatibilização ou a instituição, caso queira, pode escolher qual termo utilizar daqui por diante. Para esse conjunto ainda falta análise de mais um termo que será apresentado no quadro 29.

Quadro 29 – Registro do conceito "*Order"termo 03*

Aspecto analisado	Resultado
Nome do conceito	Order
Identificador	393
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Nome da ordem taxonômica ao qual a espécie pertence.
Outros nomes para o conceito ou classe	Order Ordername Order
Fonte do conceito	SiBBR
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBR

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

Ao examinar o Quadro 29, observa-se que o conceito em foco possui o mesmo significado que o termo apresentado nos quadros do conjunto analisado, onde apenas um conceito apresenta a grafia diferente. Sugere-se que o sistema ao qual o conceito pertence faça uma revisão e escolha qual o conceito mais apropriado para o sistema, além disso, sugere-se que os sistemas de informação que não utilizam o metadado façam uma análise e decidam se o conceito é pertinente para os seus usuários e informações disseminadas.

Em seguida, tem-se a análise do conjunto formado pelo metadado *Scientific Name*. Este conjunto de conceitos possui cinco ocorrências. A primeira análise foi empreendida por meio do registro do conceito de cada ocorrência.

➢ COMPATIBILIZAÇÃO DO QUINTO CONJUNTO DE CONCEITOS

Quadro 30 – Registro do conceito "*ScientificName"*

Aspecto analisado	Resultado
Nome do conceito	ScientificName
Identificador	10
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	O nome científico completo, com informação de data e autor de sabido. Se for uma identificação que não chega a nível de espécie, deve conter o nome do ranque taxonômico mais baixo determinado. Não deve conter qualificador de determinação, que por sua vez deve ser fornecido no campo IdentificationQualifier.
Outros nomes para o conceito ou classe	Scientific Name – original ScientificName ScientificNameID
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence aos sistemas de informação Flora e Funda do Brasil, CNC Flora e Catálogo da Fauna

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

A análise do conceito acima (Quadro 30) mostra que se origina do padrão Darwin Core e que está sendo utilizado por três sistemas de informação do JBRJ. No Quadro 31, a seguir, mostra o registro do conceito que outro conceito que pode ser suscetível à compatibilização.

Quadro 31 – Registro do conceito "*ScientificName* – original"

Aspecto analisado	Resultado
Nome do conceito	Scientific Name – original
Identificador	385
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Nome da espécie
Outros nomes para o conceito ou classe	Scientific Name Scientific Name Scientific NameID
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 não apresentando a definição tão completa como o conceito apresentado no Quadro 31, o conceito presente no Quadro 32 possui o mesmo significado. Além de apresentar uma grafia diferente e ter uma origem diferente, pode-se dizer que os dois conceitos possuem o mesmo objetivo. A seguir, outro registro de conceito presente no conjunto.

Quadro 32 – Registro do conceito "*Scientific Name"*

Aspecto analisado	Resultado
Nome do conceito	Scientific Name
Identificador	387
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	O nome científico completo, com informações de autoria e data, se conhecidas. Quando fizer parte de uma Identificação, este deve ser o nome no nível taxonômico mais baixo que pode ser determinado.
Outros nomes para o conceito ou classe	Scientific Name - Original Scientific Name Scientific NameID
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).

O conceito acima apresenta a descrição completa tal qual o conceito utilizado nos sistemas do JBRR, mas está sendo utilizado por outro sistema de informação, o SiBBr. Ao olhar as definições presentes nos quadros 30 e 32, percebe-se que ambas se apresentam de forma idêntica, dessa forma, infere-se que o conceito presente no quadro 31 tem sua origem baseada no conceito presente no Quadro 30, uma vez que o padrão Darwin Core é um padrão internacionalmente conhecido. No Quadro 33, a seguir, o registro do conceito de outro conceito presente no conjunto *Scientifc Name*.

Quadro 33 – Registro do conceito "*ScientificNameID"*

Aspecto analisado	Resultado
Nome do conceito	ScientificNameID
Identificador	326
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Um identificador para os detalhes nomenclaturais (não taxonômicos) de um nome científico.
Outros nomes para o conceito ou classe	ScientificNameAuthorship Scientific Name - original Scientific Name Scientific Name
Fonte do conceito	Padrão Darwim Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação CNCFlora

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

Após a análise do registro do conceito, percebe-se quanto ao conceito presente no Quadro 33 que se trata de um complemento de informações ligadas ao campo *Scientific Name*. Nesse caso, cabe ao gestor decidir se a utilização do campo é pertinente para o seu sistema de

informação como uma conexão com informações ao campo *Scientific Name*. Além disso, o conceito está presente em apenas um sistema de informação, o que cabe ainda aos outros gestores dos sistemas de informação optarem ou não por usar o conceito. Na sequência, será apresentado o último conceito (Quadro 34) deste conjunto analisado.

Quadro 34 – Registro do conceito "*scientificNameAuthorship"*

Aspecto analisado	Resultado
Nome do conceito	ScientificNameAuthorship
Identificador	12
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	As informações de autoria do campo scientificName formatadas conforme convenções do campo nomenclaturalCode indicado.
Outros nomes para o conceito ou classe	Scientific Name - original Scientific Name scientificNameID
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação Flora e Funga do Brasil

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

O conceito apresentado mostra que o seu foco é outro, o autor do nome científico, nesse caso, o metadado se apresenta como um metadado adicional para a informação relacionada ao nome científico. A seguir, a análise com base em Sales (2022). Pode-se provar que a compatibilização dos conceitos é possível, se os mesmos apresentarem as mesmas características.

Quadro 35 – Análise do conjunto *Scientific Name* conforme Sales

	Termo 01 – Scientific Name	**Termo 02 – Scientific Name – original**	**Termo 03 – Scientific Name**	**Termo 04 – ScientificNameID**	**Termo 05 – ScientificName Authorship**
Referente	Espécie	Espécie	Espécie	Informações extras	O autor
Característica 01	Voltado para a espécie	Voltado para a espécie	Voltado para a espécie	Voltado para informações não específicas	Voltado para o autor da espécie
Característica 02	Flora e Funga do Brasil, CNC Flora e Catálogo da Fauna	SiBBr	SiBBr	CNC Flora	Flora e Funga do Brasil

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

Percebe-se que ao utilizar a metodologia de Sales (2022), o tconceito *01- Scientific Name* é utilizado por todos os sistemas que compõem o projeto, e que o SiBBr o utiliza de duas formas, logo os metadados *Scientific Name e Scientific Name – original* são considerados equivalentes semanticamente. Constata-se também que existem dois termos, *ScientificNameID* e *ScientificNameAuthorship*, que fornecem informações complementares acerca desse assunto, assim, sugere-se uma revisão por parte dos gestores para decidir se os conceitos podem ser utilizados em seus respectivos sistemas. Como existe a presença de conceitos que possuem o mesmo significado, porém com a grafia diferente, será feita a análise conforme Neville (1970) (Quadro 34), com os conceitos que considera-se passíveis de compatibilização.

Quadro 36 – Atribuição do Código 0303 ao conjunto *Scientific Name* conforme o Método de Neville

	Termo 01	**Termo 02**	**Termo 03**
Termo original	*Scientific Name*	*Scientific Name - Original*	*Scientific Name*
Identificador	*Scientific Name* (0303)	*Scientific Name - Original*	*Scientific Name* (0303)
Termos com os identificadores	*0303=Scientific Name -Original*	*0303=Scientific Name -Original*	*0303=Scientific Name -Original*

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

Após a observação do Quadro 36, é possível deduzir que a atribuição de um código para os conceitos seria uma forma fácil de obter a compatibilização, basta apenas que os conceitos e códigos estejam registrados em algum documento para a descrição dos dados.

O próximo conjunto analisado refere-se ao termo *Species* (Quadros 37 a 39). Este termo possui três ocorrências entre os sistemas que participam do corpus.

➢ COMPATIBILIZAÇÃO DO SEXTO CONJUNTO DE CONCEITOS

Quadro 37 – Registro do conceito "*Species"*

Aspecto analisado	**Resultado**
Nome do conceito	*Species*
Identificador	56
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 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, no Quadro 38, a análise do conceito do mesmo conjunto para identificação do seu significado e origem.

Quadro 38 – 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 (Quadro 39), a análise do terceiro conceito presente no conjunto *Species*.

Quadro 39 – 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 (2030).

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 (Quadro 40).

Quadro 40 – 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, no Quadro 41, a análise conforme Neville (1970) e um exemplo de aplicação de códigos para o conjunto de conceitos.

Quadro 41 – Atribuição do Código 0404 ao conjunto *Species* conforme o Método de Neville

	Termo 01	***Termo 02***	***Termo 03***
Termo original	*Species*	*Species*	*SpeciesName*
Identificador	*Species* (0404)	*Species* (0404)	*SpeciesName* (0404)
Termos com os identificadores	*0404=Species*	*0404=Species*	0404=*SpeciesName*

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:

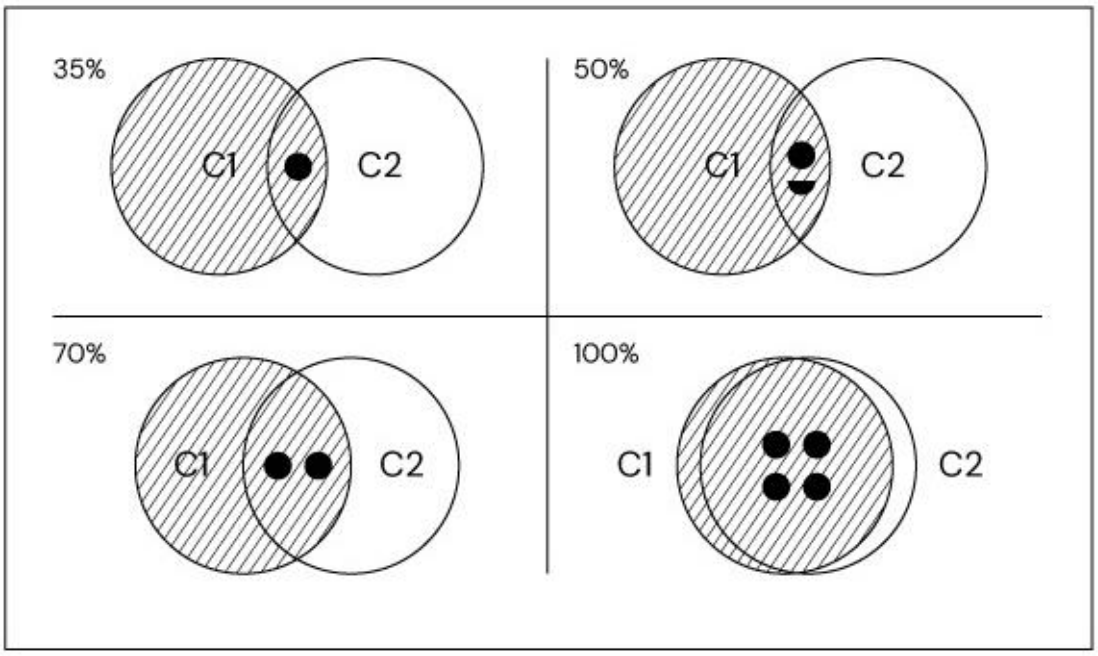
a) os sistemas possuem focos parecidos, mesmo quando voltados para objetos específicos, como flora e fauna. Os metadados utilizados por sistema mais abrangente, como o SiBBr, podem ser utilizados pelos mais específicos e vice-versa.

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.

Para complementar as informações já apresentadas, pretende-se aqui apresentar um modelo de análise da informação para compatibilização semântica com base em Neville (1970) e Sales (2022). Para Sayão (2001, p. 83) "[...] um modelo é uma criação cultural, 'mentefato', destinada a representar uma realidade, ou alguns dos seus aspectos, a fim de torná-los descritíveis qualitativamente e quantitativamente e, algumas vezes observáveis". Dessa forma, aqui pretende-se tentar descrever algo que possa auxiliar todos os profissionais que possam estar envolvidos em compatibilização semântica.

O foco do modelo proposto é estabelecer orientações para a análise em pares e/ou conjuntos de termos, uma vez que a compatibilização não pode ocorrer de forma unilateral. Para isso foram instituídos os padrões: A = Primeiro termo, B= Segundo termo, C= Terceiro termo, e assim por diante, a depender dos conjuntos analisado. Logo, o termo A é formado por características distintas aqui formadas por c1, c2 e c3. O termo B também pode ser constituído por c1, c2 e c3 e o termo C também c1, c2 e c3. Portanto, se possuírem as mesmas características ou pelo menos duas características em comum, pode-se realizar a compatibilização entre os termos (Figura 14).

Figura 14 – Modelo de análise para compatibilização semântica



Fonte: Elaborado pela autora (2023).

O modelo acima possui as seguintes regras para as análises dos termos:

a) Se os conceitos tiverem 1 característica em comum, estes são considerados 35% compatíveis;
b) Se possuírem 1 característica e meia são considerados 50% compatíveis;
c) Caso possuam 2 características em comum são 70% compatíveis e se possuírem 3 características ou mais são 100% compatíveis.

Recomenda-se que para considerar os conceitos compatíveis, precisa-se de 50% de compatibilização. Assim, retorna-se à Teoria do Conceito proposta por Dalhberg (1978a), onde a autora declara que os conceitos são constituídos de enunciados, ou seja, características que formam o conceito. Cabe frisar que este é um modelo que tem o objetivo de auxiliar no que consideramos a primeira parte para uma compatibilização semântica efetiva. A compatibilização dependerá dos SOC e só será realizada de acordo com o objetivo do usuário.

Acredita-se que, apenas a aplicação do modelo de compatibilização, não é o suficiente para estabelecer a compatibilização semântica entre sistemas de informações. 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.

6.4 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.

Como o objetivo final da pesquisa é propor melhorias na recuperação da informação para dados em Biodiversidade, por meio da compatibilização semântica, a seguir serão apresentadas diretrizes para 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. Segundo o Dicio - Dicionário Online de Português (Diretriz, 2023), significa "linha segundo a qual se traça um plano em qualquer estrada ou caminho". Assim sendo, servem como guia para o estabelecimento de um trabalho futuro. Por sua vez, Nurcan *et al*. (1999, p. 12, tradução nossa) afirmam que "uma diretriz executável corresponde a uma intenção operacionalizável que é diretamente aplicável através de um conjunto de atividades"[60].

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.

[60]Texto original: "*An Executable Guideline corresponds to an operationalizable intention which is directly applicable through a set of activities*".

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 TERMINÓLOGIA 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 como pode ser visto na subseção 6.3 (resultados da proposta de compatibilização), 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.

A identificação de conceitos com graus de equivalência pode partir do modelo apresentado na subseção 6.3, que estabelece um cálculo para identificar o grau de compatibilização semântica. Acredita-se que ao utilizar o modelo como base, não ocorrerá perda de informações e os conceitos que não puderem ser compatibilizados serão registrados do mesmo modo que os que foram deixando sempre claro a informação para o usuário e futuros profissionais envolvidos na compatibilização.

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.

7 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. Para isso, consta a elaboração dos SOC, dispositivos criados cujo objetivo é a padronização da informação. Estes podem ser tesauros, vocabulários controlados, taxonomias, ontologias, muitas vezes voltados para domínios específicos. Além da elaboração e aperfeiçoamento dos SOC, dentro da Organização do Conhecimento, tem-se os estudos relacionados à compatibilização semântica.

Os estudos que versam sobre a compatibilização semântica têm a origem na década de 60 do século passado. Estes estudos tentam orientar o indexador a promover uma integração entre os SOC, ocorrendo, assim, um intercâmbio de informação sem a necessidade de criar um outro SOC, apenas fazendo com que a informação seja recuperável em ambos os SOC que participam da compatibilização.

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.

Nessa perspectiva, a questão norteadora da presente pesquisa foi: **Como melhorar a integração semântica de dados e informações em Biodiversidade**? Esta questão serviu de ponto de partida para o estabelecimento da investigação, tendo como base a hipótese **de que as teorias que embasam a construção dos SOC podem melhorar o tratamento semântico de**

dados e informações em Biodiversidade, possibilitando a busca integrada e a interoperabilidade semântica entre os sistemas.

Para se chegar à conclusão e testar se a hipótese era válida, foi estabelecido como objetivo geral: Propor diretrizes para a compatibilização terminológica entre diversos vocabulários usados na indexação de bases de dados em Biodiversidade. Assim, fez-se o desdobramento da pesquisa em 4 objetivos específicos, a saber: Identificar, na literatura, o panorama acerca dos SOC voltados para o domínio da Biodiversidade; Compreender os principais conceitos estudados na área da Biodiversidade; 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.

Pode-se afirmar que, após a realização da pesquisa, os objetivos propostos foram alcançados, pois o estabelecimento dos procedimentos metodológicos desenvolvido foi fundamental para tal. A análise de domínio bem como a busca na literatura e o teste realizado no campo empírico, foram essenciais para se chegar a uma conclusão.

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.

A utilização dos metadados e as análises realizadas conforme as metodologias que versam sobre compatibilização serviram como base para propor um modelo de análise entre os termos para fins de compatibilização. Dessa forma, constatou-se que as teorias que versam sobre compatibilização semântica servem como subsídios para o processo inicial de desenvolvimento da mesma e podem ser aplicadas no contexto dos dados de pesquisa.

Porém, somente a aplicação dos mesmos não garantirá aos envolvidos uma compatibilização semântica efetiva, outros fatores e elementos devem ser empregados. Especialmente no contexto atual, onde os sistemas de informação, que promovem a disseminação de pesquisa, são ambientes digitais, aspectos específicos sobre esses ambientes devem ser levados em consideração no momento da proporção de uma compatibilização semântica. Ainda que o profissional da informação esteja presente na compatibilização semântica e seja especialista em tal tarefa, e as teorias que versam sobre tal possam auxiliar na compatibilização, isso não é o suficiente se a instituição e/ou sistema de informação realmente queira estabelecer a compatibilização semântica.

Assim, a presente pesquisa sugeriu a adoção de 5 diretrizes que podem auxiliar no desenvolvimento da compatibilização semântica, são elas: **Registro dos conceitos a serem**

avaliados em tabela terminológica individual guardando seu metadados; Registro dos conceitos em tabela terminologia comparativa; Análise comparativa das características do conceito; Identificação de conceitos com graus de equivalência; Atribuição dos mesmos códigos aos conceitos para efeitos de compatibilização no sistema.

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. Neste sentido, pode-se afirmar que a hipótese defendida por essa pesquisa é verdadeira. No entanto, seria interessante que essas diretrizes pudessem ser aplicadas também a outros domínios ou realidades para que possam ser melhor fundamentadas.

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.

Neste sentido, a presente tese será uma forte contribuição, dada a relevância da temática para as teorias elaboradas dentro da Organização do Conhecimento, bem como para os estudos que versam sobre Gestão de dados.

REFERÊNCIAS

ACCESS TO BIOLOGICAL COLLECTION DATA (ABCD). **Home**. 2007. Disponível em: https://abcd.tdwg.org/. Acesso em: 22 dez. 2022.

ALBAGLI, Sarita; CLINIO, Anne; RAYCHTOCK, Sabryna. Ciência Aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v.10, n.2, p. 434-450, nov. 2014. Disponível em: http://revista.ibict.br/liinc/article/view/3593. Acesso em: 29 set. 2020.

ALBUQUERQUE, Andréa Corrêa Flôres *et al*. A Negotiation Protocol for Data Integration Driven by Ontology. *In*: EUROPEAN CONFERENCE ON KNOWLEDGE MANAGEMENT, 2010, United Kingdom. **Anais** […]. United Kingdom: Academic Publishing Limited Reading, 2010. p.1-10.

ALBUQUERQUE, Andréa Corrêa Flôres; SANTOS, José Laurindo Campos dos. Elicitation Process and Knowledge Structuring: a Conceptual Framework for Biodiversity. *In*: MEXICAN INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE (MICAI), 14., 2015, Cuernavaca. **Anais** […]. Cuernavaca: IEEE, 2015, p. 67-72.

ALONSO-ARÉVALO, Julio. La gestión de datos de investigación en el horizonte de las bibliotecas universitarias y de investigación. **Cuadernos de Documentación Multimedia**, Madri, v. 30, p. 75-88, 2019. Disponível em: https://revistas.ucm.es/index.php/CDMU/article/view/62806. Acesso em: 25 jul. 2020.

ALVES, Rachel Cristina Vesu. Metadados para representação e recuperação da informação em ambiente web. *In*: MARINGELLI, Isabel Cristina Ayres da Silva. **Seminário Serviços de Informação em Museus**: informação digital como patrimônio cultural. São Paulo: Pinacoteca de São Paulo, 2017, p.95-106.

AMDOUNI, Emna; JONQUET, Clement. FAIR or FAIRer? An integrated quantitative FAIRness assessment grid for semantic resources and ontologies. *In*: INTERNATIONAL CONFERENCE ON METADATA AND SEMANTICS RESEARCH, 15., 2021, Madrid, ES. **Proceedings** […]. Madrid, ES: [*s.n.*], 2021. p. 67-80.

ARAKAKI, Felipe Augusto**. Metadados administrativos e a proveniência dos dados:** modelo baseado na família PROV. 2019. 139 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio Mesquita Filho, Marília, 2019.

BARBOSA, Everton Rodrigues. **Modelo colaborativo para a construção e a publicação de tesauros no contexto do Linked Open Data**: Aplicação no domínio da música. 2021. 373 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina, Florianópolis, 2021.

BARBOSA, Nilson Theobald. **Para uma economia da informação semântica:** a construção de ambientes semânticos para a recuperação inteligente da informação. 2021. 416 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-graduação em Ciência da Informação, Universidade Federal Fluminense, Niterói, 2021.

BATISTA, Gilda Helena Rocha. **Compatibilidade e Convertibilidade entre linguagens de indexação**: um estudo de caso. 1986. 176 f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Escola de Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia/Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1986.

BERTIN, Patrícia Rocha Bello; VISOLI, Marcos Cezar; DRUCKER, Debora Pignatari. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de p&d. **Ponto de Acesso**, Salvador, v. 11, n. 2, p. 34-48, ago. 2017. Disponível em: https://portalseer.ufba.br/index.php/revistaici/article/view/21449. Acesso em 20 jun. 2020.

BERTONI, Estevão. O impacto do erro da Lancet, segundo esta editora científica. **Nexo**, São Paulo, 5 jun. 2020. Disponível em: https://www.nexojornal.com.br/entrevista/2020/06/05/O-impacto-do-erro-da-Lancet-segundo-esta-editora-cient%C3%ADfica. Acesso em: 15 nov. 2020

BIBLIOTECA PÚBLICA ESTADUAL DE CIÊNCIA E TECNOLOGIA. **Guia de gestão de dados de pesquisa**. [2019]. Disponível em: http://www.spsl.nsc.ru/naukresursy-i-uslugi-gpntb-so-ran-dlya-nauki-i-biznesae-i-biznesu/rdm/vvedenie/. Acesso: 9 set. 2020.

BIODIVERSITY INFORMATION STANDARDS. **Darwin Core**. 2022a. Disponível em: https://www.tdwg.org/standards/dwc/. Acesso em: 1 nov. 2022.

BIODIVERSITY INFORMATION STANDARDS. **Standards**: from Darwin Core to WGSRPD standards aid the exchange of biodiversity information. 2022b. Disponível em: https://www.tdwg.org/standards/. Acesso em: 1 nov. 2022.

BORKO, Harold. Information Science. What is it? **American Documentation**, Santa Monica, v. 19, n. 1, p. 3-5, 1968.

BORGMAN, Christine L. The conundrum of sharing research data. **Journal of the Association for Information Science and Technology**, Syracuse, v. 63, n. 6, p. 1059-1078, June 2012. Disponível em: https://onlinelibrary.wiley.com/doi/epdf/10.1002/asi.22634. Acesso em: 2 fev. 2023.

BRASIL. Ministério do Meio Ambiente. **Projeto GEF Pró-espécies**. Brasília, DF: MMA, [2019]. Disponível em: https://antigo.mma.gov.br/biodiversidade/economia-dos-ecossistemas-e-da-biodiversidade/item/11642-projeto-gef-pr%C3%B3-esp%C3%A9cies.html. Acesso em: 20 jun. 2023.

BRÄSCHER, Marisa. Terminologia brasileira em Ciência da Informação: uma análise. **Ciência da Informação**, Brasília, v. 15, n. 2, p. 135-142, jul./dez. 1986.

BROWN, Christopher J. *et al*. Quantitative approaches in climate change ecology. **Global Change Biology**, Bethesda, v. 17, p. 3697-3713, Dec. 2011. DOI: https://doi.org/10.1111%2Fj.1365-2486.2011.02531.x. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3597248/. Acesso em: 12 ago. 2022.

BÜTNER, Stephan; HOBOHM, Hans-Christoph; MÜLLER, Lars. Research Data Management. *In*: BUTNER, Stephan; HOBOHM, Hans-Christoph; MÜLLER, Lars (org.). **Handbuch Forschungsdatenmanagement**. Bad Honnef: BOCK+HERCHEN Verlag, 2011. p.13-24.

CAMPOS, Maria Luiza de Almeida. A problemática da compatibilização terminológica e a integração de ontologias: o papel das definições conceituais. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 6., 2005, Florianópolis. **Anais** [...]. Florianópolis: UFSC, 2005. p. 1-9.

CAMPOS, Maria Luiza de Almeida. Integração de Ontologias: o domínio da Bioinformática. **RECIIS – R. Eletr. de Com. Inf. Inov. Saúde**. Rio de Janeiro, v. 1, n. 1, p. 117-121, jan./jun. 2007.

CAMPOS, Maria Luiza de Almeida. Aspectos semânticos da compatibilização terminológica entre ontologias no campo da Bioinformática. *In:* ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 10., 2009, João Pessoa. **Anais** [...]. João Pessoa: UFPB, 2009. p. 1-18.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Metodologia de elaboração de tesauro conceitual: a categorização como princípio norteador. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 3, p. 348-359, set./dez. 2006.

CANDELA, Leonardo *et al*. Data Journals: A Survey. **Journal of the association for information science and technology**, Syracuse, v. 66, n. 9, p. 1747–1762, Jan. 2015.

CASTRO, Fabiano Ferreira de; SIMIONATO, Ana Carolina. Revisitando ontologia e metadados à luz dos ambientes informacionais digitais**. Perspectivas em Ciência da Informação**, Belo Horizonte, v. 25, n. 4, p. 3-23, dez. 2020. Disponível em: https://www.scielo.br/j/pci/a/nr5YfqhgBbrjJHyHJhyjYLc/?lang=pt. Acesso em: 28 out. 2022.

CAVALCANTI, Márcia Teixeira; SALES, Luana Farias. Gestão de dados de pesquisa: um panorama da atuação da união europeia. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, Porto Alegre, v. 31, n. 1, p. 73-98, jan./jun. 2017. DOI: https://doi.org/10.14295/biblos.v31i1.5789. Disponível em: https://periodicos.furg.br/biblos/article/view/5789. Acesso em: 25 jul. 2020.

CAVALCANTI, Mauro José. Bancos de dados sobre biodiversidade na Amazônia: a experiência do Projeto Biotupé. *In*: SANTOS-SILVA, Edinaldo Nelson *et al.* (org.). **Diversidade Biológica e Sociocultural do Baixo Rio Negro, Amazônia Central**. Manaus: Editora INPA, 2005. p.199-2013.

CIAVATTA, Maria; RAMOS, Marise. A "era das diretrizes": a disputa pelo projeto de educação dos mais pobres. **Revista Brasileira de Educação**, Grajau, v. 17 n. 49, p. 11-37; 231-232, jan./abr. 2012. Disponível em: https://www.scielo.br/j/rbedu/a/nDS3v6XBFdjG3jQGLRk687m/?format=pdf&lang=pt. Acesso em: 1 fev. 2023.

COLLIN, Rachel *et al*. TaxaGloss - A Glossary and Translation Tool for Biodiversity Studies. **Biodiversity Data Journal**, Bruxelas, v. 4, p. 1-9, Dec. 2016. DOI:

https://doi.org/10.3897/BDJ.4.e10732. Disponível em: https://www.readcube.com/articles/10.3897%2Fbdj.4.e10732. Acesso em: 10 set. 2021.

CONVENTION ON BIOLOGICAL DIVERSITY (CDB). **Article 2.** Use of Terms. [*S.l.*]: UN Enviroment Programme, 2016. Disponível em: https://www.cbd.int/convention/articles/?a=cbd-02. Acesso em: 5 jun. 2022.

CÓRDULA, Flavio Ribeiro; ARAÚJO; Wagner Junqueira. O compartilhamento de dados científicos na era do E-science. *In*: DIAS, Guilherme Ataíde; OLIVEIRA, Bernadina Maria Juvenal Freire de (org.). **Dados científicos**: perspectivas e desafios. João Pessoa: Editora UFPB, 2019. p.189-206.

COSTA, Maíra Murrieta. **Diretrizes para uma política de gestão de dados científicos no Brasil**. 2017. 288 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília, Brasília, DF, 2017.

COX, Simon J. D. *et al*. Ten simple rules for making a vocabulary FAIR. **PLOS Computational Biology**, São Francisco, v. 17, n. 6, p. 1-15, June 2021. DOI: https://doi.org/10.1371/journal.pcbi.1009041. Disponível em: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009041. Acesso em: 1 abr. 2023.

CUI, Hong. Competency Evaluation of Plant Character Ontologies Against Domain Literature. **Journal of the American Society for Information Science and Technology**, Syracuse, v. 61, n. 6, p. 114-1165, 2010.

CURTY, Renata Gonçalves. O paradigma da publicação de dados e suas diferentes abordagens. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 18., 2017, Marília. **Anais** [...]. Marília: UNESP, 2017. p. 1-20. Disponível em: https://brapci.inf.br/index.php/res/v/105144. Acesso em: 20 ago. 2020.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da Informação**, Brasília, v. 7, n. 2, p. 101-107, 1978a.

DAHLBERG, Ingetraut. A referent-oriented analytical concept theory of interconcept. **International Classification**, [S.l.], v. 5, n. 3, p. 142-150, 1978b.

DAHLBERG, Ingetraut. Towards estabilishment of compatibility between indexing languages. **International Classifocation,** [S.l.], v. 8, n. 2, p. 88-91, 1981. Disponível em: https://www.nomos-elibrary.de/10.5771/0943-7444-1981-2-86.pdf?download_full_pdf=1. Acesso em: 10 set. 2021.

DALTIO, Jaudete; MEDEIROS, Claudia M. Bauzer. Um Servidor de Ontologias para Sistemas de Biodiversidade. *In*: SEMINÁRIO INTEGRADO DE SOFTWARE E HARDWARE, 31., 2007, Rio de Janeiro. **Anais** [...]. Rio de Janeiro, 2007. p. 2143- 2157.

DALTIO, Jaudete; MEDEIROS, Claudia M. Bauzer. Aondê: An ontology web service for interoperability across biodiversity applications. **Information Systems**, Oxford, v. 33, p. 724-753, 2008.

DATACITE. **DataCite metadata schema for the publication and citation of research data**, **2015**. Disponível em: https://schema.datacite.org/meta/kernel-4.3/doc/DataCite-MetadataKernel_v4.3.pdf. Acesso em: 30 nov. 2022.

DIAS, Guilherme Ataíde; ANJOS, Renata Lemos; RODRIGUES, Adriana Alves. Os princípios FAIR: viabilizando o reuso de dados científicos. *In*: DIAS, Guilherme Ataíde; OLIVEIRA, Bernadina Maria Juvenal (org.). **Dados científicos**: perspectivas e desafios. João Pessoa: Editora UFPB, 2019.p.177-188.

DIRETRIZ. *In*: DICIO, Dicionário Online de Português. Porto: 7Graus, 2023. Disponível em: https://www.dicio.com.br/diretriz/. Acesso em: 1 fev. 2023.

ENDARA, Lorena *et al*. Building the "Plant Glossary" - A controlled botanical vocabulary using terms extracted from the Floras of North America and China. **Taxon**, Wiley, v. 66, n. 4, p. 953-966, Aug. 2017. DOI: https://doi.org/10.12705/664.9. Disponível em: https://www.researchgate.net/publication/318923775_Building_the_Plant_Glossary-A_controlled_botanical_vocabulary_using_terms_extracted_from_the_Floras_of_North_America_and_China/link/5a78f3be45851541ce5c82f7/download. Acesso em: 1 out. 2021.

ESTEVÃO, Janete Saldanha Bach; ARNS, Elaine Mandelli; STRAUHS, Faimara do Rocio. Gestão de dados de pesquisa: uma prática para abrir a caixa preta da pesquisa científica. **RDBCI: Revista Digital Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 17, p. 1-26, jan. 2019. DOI: https://doi.org/10.20396/rdbci.v17i0.8656239. Disponível: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8656239/21458 . Acesso em: 9 set. 2020.

FARNEL, Sharon; SHIRI, Ali. Metadata for Research Data: Current Practices and Trends. *In*: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APLLICATIONS, 2014, Texas. **Proceedings** […]. Austin, TX: TDL.Org., 2014. p. 74-82.

FLORA e Funga do Brasil. Jardim Botânico do Rio de Janeiro. Disponível em: http://floradobrasil.jbrj.gov.br/ . Acesso em: 3 ago. 2022.

FONTANA, Felipe. Técnicas de pesquisa. *In*: MAZUCATO, Thiago (org.). **Metodologia da pesquisa e do trabalho científico**. Penápolis: FUNEPE, 2018. p. 59-80.

FREITAS, Lidiane Marques; ALBUQUERQUE, Ana Cristina de. As abordagens da análise de domínio como aporte metodológico para a classificação arquivística. **Tendencias da Pesquisa Brasileira em Ciência da Informação**, Belo Horizonte, v.10, n. 2, p. 1-19, ago./dez. 2017.

GARIJO, Daniel; POVEDA-VILLALÓN, María. **Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web.** Ithaca, NY: Cornell University**,** 2020. DOI: https://doi.org/10.48550/arXiv.2003.13084. Disponível em: https://arxiv.org/abs/2003.13084. Acesso em: 3 mar. 2022.

GLOBAL BIODIVERSITY INFORMATION FACILITY. **GBIF Metadata Profile** – How to Guide. Contributed by Ó Tuama, Eamonn, Kyke Braak, D. Remsen. Copenhagen: Global

Biodiversity Information Facility, 2011. Disponível em: http://links.gbif.org/gbif_metadata_profile_how-to_en_v1. Acesso em: 19 nov. 2022.

GILLILAND, Anne J. Setting the stage. *In*: BACA, M. (ed.). **Introduction to metadata**. 2nd ed. Los Angeles: Getty Research Institute, 2008. p. 1-19. Disponível em: https://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.pdf. Acesso em: 1 nov. 2022.

GLOBAL BIODIVERSITY INFORMATION FACILITY – GIBF. 2022. Disponível em: https://www.gbif.org/. Acesso em: 3 jun. 2022.

GOFAIR. **Fair Principles.** [2022]. Disponível em: https://www.go-fair.org/fair-principles/. Acesso em: 30 out. 2022.

GOMES, Hagar Espanha; CAMPOS, Maria Luiza de Almeida; GUIMARÃES, Ludmila dos Santos. Organização da Informação e Terminologia: a abordagem onomasiológica. **Data Grama Zero**, Rio de Janeiro, v.11, n. 5, p.1-12, 2010.

GRAY, Jim. Jim Gray on escience: a transformed scientific method. *In*: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (ed.). **The Fourth Data-Intensive Scientific Discovery Paradigm**. Washington: Microsoft Research, 2009. p.17-31.

GUIMARÃES, José Augusto Chaves. Análise de domínio como perspectiva metodológica em organização da informação. **Ci. Inf**., Brasília, DF, v. 41 n. 1, p.13-21, jan./abr. 2014.

HAENDEL, Melissa Anne *et al*. Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon. **Journal of Biomedical Semantics**, [s.l.], v. 5, p. 1-13, May 2014.

HARDISTY, Alex *et al*. The Bari Manifesto: An interoperability framework for essential biodiversity variables. **Ecological Informatics**, Toronto, v. 49, p. 22-31, Jan. 2019.

HJØRLAND, Binger. Domain analysis in information science: eleven approaches-traditional as well as innovative. **Journal of Documentation**, Leeds, v. 58, n. 4, p. 422-462, 2002.

HOEHNDORF, Robert *et al*. The flora phenotype ontology (FLOPO): tool for integrating morphological traits and phenotypes of vascular plants. **Journal of Biomedical Semantics**, [s.l.], v. 7, p. 1-11, 2016. DOI: https://doi.org/10.1186/s13326-016-0107-8. Disponível em: https://www.readcube.com/articles/10.1186%2Fs13326-016-0107-8. Acesso em: 31 mar. 2022.

HUANG, Fengqiong *et al*. OTO: Ontology Term Organizer. **BMC Bioinformatics**, [s.l.], v. 16, p. 1-18, Feb. 2015.

INDE. **IBGE atualiza perfil de metadados geoespaciais do Brasil**. 2021. Disponível em: https://inde.gov.br/Noticias/Detalhe/69. Acesso em: 2 fev. 2023.

ISO. International Organization for Standardization. **Information and documentation — Thesauri and interoperability with other vocabularies** — Part 1: Thesauri for information retrieval. ISO 25964-1. 2011

IUCN Species Information Service (SIS). 2023. Disponível em: https://www.iucnredlist.org/assessment/sis. Acesso em: 20 jan. 2023.

JARDIM BOTÂNICO DE BRASÍLIA. Saberes do Cerrado. 2023 Disponível em: https://www.jardimbotanico.df.gov.br/pesquisa/saberes-do-cerrado/#:~:text=O%20projeto%20Saberes%20do%20Cerrado,de%20S%C3%A3o%20Carlos%20(UFSCAR). Acesso em: 22 jun. 2022.

JARDIM BOTÂNICO DO RIO DE JANEIRO. dadoswiki.jbrj.gov.br. **MMA-Darwin Core (MMA-DwC)**. 2023. Disponível em: https://dadoswiki.jbrj.gov.br/doku.php?id=mma-dwc#mma-darwin_core_mma-dwc. Acesso em: 2 fev. 2023.

JOINUP. Access to Biological Collection Data (ABCD). **About Access to Biological Colletcion Data (ABCD)**. European Commission. 2007. Disponível: https://joinup.ec.europa.eu/collection/science-and-technology/solution/access-biological-collection-data-abcd/about. Acesso em: 22 dez. 2022.

JONES, Matthew B. *et al*. **Ecological Metadata Language version 2.2.0.** KNB Data Repository, 2019. DOI: https://doi.org/10.5063/F11834T2. Disponível em: https://eml.ecoinformatics.org/. Acesso em: 2 nov. 2022.

JONQUET, Clément *et al*. AgroPortal: A vocabulary and ontology repository for agronomy. **Computers and Electronics in Agriculture**, Washington, v. 144, p. 126–143, Jan. 2018.

KAYS, Roland; MCSHEA, William; WIKELSKI, Martin. Born-digital biodiversity data: Millions and billions. **Diversity and distributions**, [s.l.], v. 26, n. 5, p. 644-648, 2020. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1111/ddi.12993. Acesso em: 5 jun. 2021.

KOCH, Richard. **O Poder 80/20**: Os segredos para conseguir mais com menos nos negócios e na vida. São Paulo: Gutenberg, 2015.

KÖNING, Christian *et al*. Biodiversity data integration—the significance of data resolution and domain. **PLos Biology**, São Francisco, v. 17, n. 3, p. 1-16, Mar. 2019. DOI: https://doi.org/10.1371/journal.pbio.3000183. Disponível em: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000183. Acesso em: 10 maio 2021.

LENTERS, Tim P. *et al*. Integration and harmonization of trait data from plant individuals across heterogeneous sources. **Ecological Informatics**, Toronto, v. 62, p.1-10, May 2021.

LOZANO-FUENTES, Saul *et al*. Ontology for Vector Surveillance and Management. **Journal of Medical Entomology**, Oxford, v. 50, n. 1, p. 1-14, Jan. 2013. Disponível em: https://doi.org/10.1603/ME12169. Disponível em: https://academic.oup.com/jme/article/50/1/1/891648. Acesso em: 10 mar. 2021.

MAGAGNA *et al*. **InteroperAble Descriptions of Observable Property Terminologies (I-ADOPT) WG outputs and recommendations**. RDA, 22 Feb. 2022. Disponível em:

https://www.rd-alliance.org/group/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg/outcomes. Acesso em: 20 mar. 2023.

MAJID, Shaheen; ZHAG, Xue; FOO, Shubert. Research Data Management by Academics and Researchers: Perceptions, Knowledge and Practices. *In*: INTERNATIONAL CONFERENCE ON ASIA-PACIFIC DIGITAL LIBRARIES, ICADL, 20., 2018, Hamilton. **Anais** [...]. Hamilton, 2018. p. 166-178.

MENDES, Irlana; PINTO, Virgínia Bentes. Taxonomia nas áreas da Biblioteconomia e da Ciência da Informação: uma revisão sistemática. **Páginas a&b**, Porto, v.3, n. 12, p. 36-47, 2019.

MINISTÉRIO DO MEIO AMBIENTE. **Diretrizes para a Integração de Dados de Biodiversidade**. Brasília: 2015.100 p.

MIRANDA, Roberto Campos da Rocha. O uso da informação na formulação de ações estratégicas pelas empresas. **Ci. Inf**., Brasília, v. 28, n. 3, p. 286-292, set./dez. 1999. Disponível em: https://www.scielo.br/j/ci/a/r7L9msHr6FfrYpJ5PKk8fsS/?format=pdf&lang=pt. Acesso em 25 mar. 2023.

MOREIRO GONZÁLEZ, José Antonio. **Linguagens documentárias e vocabulários semânticos para a web**. Salvador: EDUFBA, 2011.

MORI, Alexandre; CARVALHO, Cedric Luiz de. **Metadados no Contexto da Web Semântica**. Relatório técnico RT-INF_002-04. Instituto de Informática, Universidade Federal de Goiás, 2004. Disponível em: https://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_002-04.pdf. Acesso em: 30 out. 2022.

NATIONAL FORUM ON EDUCATION STATISTICS. **Forum Guide to Metadata** (NFES 2021110). U.S. Department of Education. Washington, DC: National Center for Education Statistics, 2021.

NEVILLE, Henry Hargreaves. Feasibility study of a scheme for reconciling thesauri covering a commom subject. **Journal of Documentation**, Leeds, v. 26, n. 4, p. 313-336, Dec. 1970.

NEVILLE, Henry Hargreaves. Thesaurus reconciliation. **Aslib Proceedings**, Londres, v. 24, n.11, p. 620-626, Nov. 1972.

NURCAN, Selmin *et al*. Change process modeling using the EKD – Change Management Method. *In*: EUROPEAN CONFERENCE ON INFORMATION SYSTEMS, ECIS' 99, 7., 1999, Copenhagen, DK. **Proceedings** […]. Copenhagen, DK: HAL Open Science, 1999, p.1 - 13, 1999. Disponível em: https://hal.science/file/index/docid/707573/filename/ECIS_99.pdf. Acesso em: 1 fev. 2023.

OLSON, Storrs Lovejoy. A Thesaurus of Bird Names: Etymology of European Lexis Through Paradigms. **The Auk**, Nova York, v. 118, n. 3, p. 815-816, July 2001.

OPENAIRE – Open Access Infrastructure for Research in Europe. **FAQ (What are repositorie?)**. 2018. Disponível em: https://www.openaire.eu/where-can-i-read-more-about-fp7. Acesso em: 2 ago. 2020.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT [OECD]. OECD Principles and guidelines for access to research data from public data. Paris: OECD Publications, 2007. Disponível em: https://www.oecd.org/sti/inno/38500813.pdf. Acesso em: 20 jun. 2020.

OREGON STATE UNIVERSITY LIBRARIES. **Research Data Services**: data papers & journals. 20 ago. 2017. Disponível em: http://guides.library.oregonstate.edu/research-data-services/data-management-data-papers-journals. Acesso em: 16 out. 2020.

PAFILIS, Evangelos *et al*. Environments and EOL: identification of Environment Ontology terms in text and the annotation of the Encyclopedia of Life. **Bioinformatics**, Oxford, v. 31, n. 11, p. 1872–1874, June 2015.

PAMPEL, Heinz; KINDLING, Maxi. Informationsinfrastrukturangebote fur digitale Forschungsdaten. *In*: SCHIRMBACHER, Peter. **E(hren)-Journal**. Berlin: Angaben gemäß § 5 TMG, 2017. p. 15-33. Disponível em: https://edoc.hu-berlin.de/bitstream/handle/18452/2993/3.pdf?sequence=1&isAllowed=y. Acesso em: 15 jun. 2020.

PALETTA, Francisco Carlos; MALDONADO, Edison Puig. Informação e tecnologia: apropriação e produção de conhecimento na web 3.0. *In*: WORLD CONGRESS ON COMMUNICATION AND ARTS, 7., 2014, Vila Real, PT. **Anais** [...] Vila Real, PT: [s.n.], 2014. Disponível em: https://www.eca.usp.br/acervo/textos/Paletta_2014.pdf. Acesso em: 1 abr. 2023.p.343-346.

PEREIRA, Ricardo Scachetti; PETERSON, Andrew Towsend. **O uso de modelagem na definição de estratégias para a conservação da biodiversidade**. Campinas, SP: SBPC/LABJOR, 2001. Disponível em: https://www.comciencia.br/dossies-1-72/reportagens/biodiversidade/bio18.htm. Acesso em: 3 mar. 2021.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013. E-book

PRÓ-ESPÉCIES. **Pró-Espécies**: Todos contra a extinção. [*S.l.*], [8 jul. 2019]. Disponível em: https://proespecies.eco.br/projeto/. Acesso em: 10 jun. 2022.

PROJETO BIOTUPÉ. 2019. Disponível em: http://biotupe.org/site/node/2. Acesso em: 26 jun. 2022.

RABELO, Camila Regina de Oliveira; PINTO, Virgínia Bentes. Tendências nos estudos de Representação Temática da Informação: uma revisão integrativa dos artigos científicos indexados na Brapci. **Em Questão**, Porto Alegre, v. 25, n. 2, p. 66-88, maio/ago. 2019.

REDE BHL SCIELO. **ThesBio**: thesaurus em biodiversidade. São Paulo: MZUSP, 2021. Disponível em: http://thesaurus.bhlscielo.org/vocab/index.php. Acesso em: 10 ago. 2021.

RESEARCH DATA ALLIANCE. **About RDA**. 2016. Disponível em: https://www.rd-alliance.org/about-rda. Acesso em: 26 jun.2022.

ROCHA, Lucas de Lima; SALES, Luana Farias; SAYÃO, Luís Fernando. **Descrever para Preservar: Metadados como Ferramenta para Gestão de Dados de Pesquisa**. 2017. ISKO Brasil. p. 194-201. Disponível em: http://hdl.handle.net/20.500.11959/brapci/121924. Acesso em: 16 jan. 2023.

RODRIGUES, Maria Rosemary; CERVANTES, Brígida Maria Nogueira. Mapeamento conceitual na organização e representação do conhecimento. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais** [...]. Londrina: UEL, 2018. p. 722-740.
ROSATI, Ilaria *et al*. A thesaurus for phytoplankton trait-based approaches: Development and applicability. **Ecological Informatics**, Toronto, v. 42, p.129–138, 2017.

RUAS, Roberto; ANTONELLO, Cláudia Simone; BOFF, Luiz Henrique. Autodesenvolvimento e Competências: O caso do Trabalhador de Conhecimento como Especialista. In: RUAS, Roberto; ANTONELLO, Claudia Simone; BOFF, Luiz Henrique (org.). **Aprendizagem Organizacional e Competências**. Bookman: Porto Alegre, 2005. p. 70-86.

SALES, Luana Farias. Compatibilização semântica entre dados de pesquisa: promovendo fairificação em domínios interdisciplinares. *In*: SALDANHA, Gustavo; CASTRO, Paulo César; PIMENTA, Ricardo M. (org.). **Ciência da Informação**: sociedade, crítica e inovação. Rio de Janeiro: IBICT, 2022. p. 75-92. Disponível em: https://ridi.ibict.br/bitstream/123456789/1227/1/saldanha-castro-pimenta.pdf. Acesso em: 10 jan. 2023.

SALES, Luana Farias *et al*. Competências dos bibliotecários na gestão dos dados de pesquisa. **Ci. Inf**., Brasília, DF, v.48 n.3, p. 303-313, set./dez. 2019. Disponível em: https://www.arca.fiocruz.br/handle/icict/43074. Acesso em: 15 ago. 2020.

SANTOS, Plácida Leopoldina Ventura da Costa; SANT'ANA, Ricardo César Gonçalves. Camadas de representação de dados e suas especificidades no cenário científico. *In*: DIAS, Guilherme Ataíde; OLIVEIRA, Bernadina Maria Juvenal (org.). **Dados científicos**: perspectivas e desafios. João Pessoa: Editora UFPB, 2019. p.53-66.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SAYÃO, Luís Fernando. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação., Florianópolis, v. 15, n. 30, p.1-31, 2010. DOI: https://doi.org/10.5007/1518-2924.2010v15n30p1. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1. Acesso em: 30 out. 2022.

SAYÃO, Luis Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Inf. Inf**., Londrina, v. 21, n. 2, p. 90-115, maio/ago. 2016.

Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939. Acesso em: 15 set. 2020.

SAYÃO, Luis Fernando; SALES, Luana Farias. Afinal o que é dados de pesquisa**? Biblos**: Revista do Instituto de Ciências Humanas e da Informação, Rio Grande, v. 34, n. 2, p. 32-51, jul./dez. 2020. Disponível em: https://periodicos.furg.br/biblos/article/view/11875/8426 Acesso em: 20 set. 2020.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Another face of metadata:** metadata authoring model for describe information about context and provenance of disciplinary research objects. [2023?]. No prelo.

SAYÃO, Luiz Fernando. Modelos teóricos em ciência da informação-abstração e método científico. **Ciência da informação**, Brasília, v. 30, n. 1, p. 82–9, 2001.

SCHIESSL, Marcelo; SHINTAKU, Milton. Sistemas do Conhecimento. *In:* ALVARES, Lilian Maria Araújo de Resende (org.). **Organização da informação e do conhecimento**: conceitos, subsídios, interdisciplinares e aplicações. São Paulo: B4 Editores, 2012. p. 49-118.

SENDEROV, Viktor *et al*. OpenBiodiv-O: ontology of the OpenBiodiv knowledge management system. **Journal of Biomedical Semantics,** [s.l.], v. 9, n. 5, p.1-15, 2018.

SISTEMA DA INFORMAÇÃO SOBRE A BIODIVERSIDADE BRASILEIRA (SiBBr). 2023. Disponível em: https://www.sibbr.gov.br/?lang=pt_BR. Acesso em: 20 maio 2022.

SILVEIRA, Denise Tolfo; CÓRDOVA, Fernanda Peixoto. Pesquisa exploratória. *In*: GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (org.). **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009. p.35-36.

SMIRAGLIA, Richard P. Domain coherence within Knowledge Organization: People, Interacting Theoretically, Across Geopolitical and Cultural Boundaries. *In*: PROCEEDINGS OF ANNUAL CAIS/ACSI CONFERENCE, 39., 2011, Fredericton, Canada. [**ANAIS…]** Fredericton, Canada, 2011. p.1-6.

SMIRAGLIA, Richard P. Organización del conocimiento: algunas tendencias em un dominio emergente. **El profesional de la información**, Granada v. 21, n. 3. p. 225-227, maio/jun. 2012.

SMIRAGLIA, Richard P. **The elements of knowledge organization**. New York: Springer, 2014.

SOUZA, Marcia Izabel Fugisawa; VENDRUSCULO, Laurimar Gonçalves; MELO, Geane Cristina. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, Brasília, v. 29, n. 1, p. 93-102, jan./abr. 2000. Disponível em: https://www.scielo.br/j/ci/a/tcW3q4WvNBQNTqTyLK8qfFF/?format=pdf&lang=pt. Acesso em 15 nov. 2022.

SOUZA, Rosali Fernadez de. Organização do Conhecimento. *In*: TOUTAIN, Lídia Maria Batista Brandão (org.). **Para entender a ciência da informação**. Salvador: UFBA, 2007. p.103-123.

STUCKY, Brian J. *et al.* Developing a vocabulary and ontology for modeling insect natural history data: example data, use cases, and competency questions. **Biodiversity Data Journal**, Bruxelas, v. 7, p. 1-12, Mar. 2019.

SVENONIUS, Elaine. Compatibility of Retrieval Languages Introduction to a Forum. **Int. Classif.**, [s.l.], v. l0, n. 1, p. 2-4, 1983. Disponível em: https://www.nomos-elibrary.de/10.5771/0943-7444-1983-1-2.pdf. Acesso em: 10 set. 2021.

SWEILEH, Waleed M. Research trends and scientific analysis of publications on burnout and compassion fatigue among healthcare providers. **Journal of Occupational Medicine and Toxicology**, São Francisco, v. 15, n. 23, p.1-10, 2020. DOI: https://doi.org/10.1186/s12995-020-00274-z. Disponível em: https://occup-med.biomedcentral.com/track/pdf/10.1186/s12995-020-00274-z.pdf. Acesso em: 10 jun. 2021.

TACKABERY, Michelle Kidd. Defining Glossaries. **Technical Communication**, [s.l.], v. 52, n. 4, p. 427-433, 2005. Disponível em: https://www.researchgate.net/publication/233620484_Defining_Glossaries. Acesso em: 30 out. 2021.

TÔRRES, Lecy Maria Caldas. Sistematização da Sintaxe de Cabeçalho de Assunto. 2021. Disponível em: http://eooci.uff.br/sistematizacao-da-sintaxe-de-cabecalho-de-assunto/. Acesso em: 1 set. 2021.

TORRES, Ricardo da Silva. **Ambiente de Gerenciamento de Imagens e Dados Espaciais para Desenvolvimento de Aplicações em Biodiversidade**. 2004. 121f. Tese (Doutorado em Ciência da Computação) – Programa de Pós- Graduação em Ciência da Computação, UNICAMP, Campinas-SP. 2004.

TRELOAR, Andrew; WILKINSON, Ross. Rethinking Metadata Creation and Management in a Data-Driven Research World. *In*: IEEE INTERNATIONAL CONFERENCE ON eSCIENCE, 4., 2008, Indianápolis, IN. **Proceedings** […]. Indianápolis, IN: IEEE Computer Society, p. 782-789, 2008. DOI: https://doi.ieeecomputersociety.org/10.1109/eScience.2008.41. Disponível em: https://www.computer.org/csdl/proceedings-article/e-science/2008/04736899/12OmNyTfg4u. Acesso em: 9 dez. 2022.

TRIPATHI, Manorama; SHUKLA, Archana; SONKAR, Sharad Kumar. Research Data Management Practices in University Libraries: AStudy. **Journal of Library & Information Technology**, Delhi, v. 37, n. 6, p. 417-424, Nov. 2017. Disponível em: https://www.researchgate.net/publication/321212270_Research_Data_Management_Practices_in_University_libraries_A_study. Acesso em: 5 out. 2020.

TUAMA, Éamonn Ó *et al*. Hackathon-Workshop on Darwin Core and MIxS Standards Alignment. **Standards in Genomic Sciences**, [s.l.], v.7, n. 1, p. 166-170, Oct. 2012.

UC SANTA CRUZ. Biblioteca Universitária. **Gerenciamento de dados de pesquisa**. [201-?]. Disponível em: https://guides.library.ucsc.edu/datamanagement/. Acesso em: 9 set. 2020.

US DEPARTAMENT OF HEALT AND HUMAN SERVICES. National Institute of Health (NIH). **NIH Data Sharing Policy**. Maryland, [2020]. Disponível em: https://grants.nih.gov/grants/policy/data_sharing/. Acesso em: 20 out. 2020.

VIEIRA, Eliane Aparecida. **A gestão da informação na tomada das decisões gerenciais**: Estudo de caso na Organização Multinacional de Reflorestamento - V & M florestal. 2011. 79f. Dissertação (Mestrado em Administração). Programa de Pós-Graduação de Administração das Faculdades Integradas de Pedro Leopoldo, Pedro Leopoldo, 2011.

WALLS, Ramona L. *et al*. Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to the Biological Collections Ontology and Related Ontologies. **Plos One**, São Francisco, v. 9, n. 3, p. 1-13, 2014. DOI: https://doi.org/10.1371/journal.pone.0089606. Disponível em: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0089606. Acesso em: 10 jun. 2020.

WILSON, Edward O. Introduction. *In*: REAKA-KUDLA, Marjorie L.; WILSON, Don E.; WILSON, Edward O. (ed.). **Biodiversity II**: Understanding and Protecting our Biological Resources. Washington, DC: Joseph Henry Press, 1997. p.1-6.

WUESTER, E. L'Étude scientifique générale de la Terminologie, zone Frontalière entre la Linguistique, la Logique, l'Ontologie, l'Informatique et les Sciences des Choses. *In*: RONDEAU, G.; FELBER, F. (org.). **Textes Choisis de Terminologie**: I. Fondéments théoriques de la terminologie. Québec: GIRSTERM, 1981. p. 57-114.

XU, Fuqi *et al*. Features of a FAIR vocabular. *In*: INTERNATIONAL CONFERENCE ON SEMANTIC WEB APPLICATIONS AND TOOLS FOR HEALTH CARE AND LIFE SCIENCES, 13., 2022, Leiden. **Proceedings** [...]. Leiden, NL: CEUR, 2022. p. 118–148. Disponível em: https://ceur-ws.org/Vol-3127/paper-15.pdf. Acesso em: 9 maio 2023.

ZENG, Marcia Lei. Compatibility of Indexing Languages in an Online Access Environment: A Review of the Approaches *In*: ASIS SIG/CRCLASSIFICATION RESEARCH WORKSHOP**,** 3., 1992, Pittisburgh. **Proceedings** [...]. Pittsburgh, PA: [*s.n.*], 1992. p. 169-187. DOI: https://doi.org/10.7152/acro.v3i1.12604. Disponível em: https://journals.lib.washington.edu/index.php/acro/article/view/12604. Acesso em: 17 out. 2021

ZUEC-NMA. **Museu de Zoologia/Universidade Estadual de Campinas - ZUEC Nematoda**. 2022. Disponível em: http://ipt1.cria.org.br/ipt/resource?r=zuec-nma&v=1.8&request_locale=pt. Acesso em: 30 out. 2022.

APÊNDICE A – REGISTRO DO CONCEITO DOS TERMOS PARTICIPANTES DO EXPERIMENTO

Aqui serão apresentados os conjuntos de termos analisados que serviram para se chegar a uma conclusão, mas que não foram apresentados nos resultados. Cabe frisar que, as análises aqui apresentadas também serviram de inspiração para as diretrizes.

Quadro A.1 – Registro do conceito: catalogNumber

Aspecto analisado	Resultado
Nome do conceito	catalogNumber
Identificador	28
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Um identificador para o registro (de preferência único) no dataset ou coleção.
Outros nomes para o conceito ou classe	Catalogue Number
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação Flora e Funga do Brasil

Quadro A.2 – Registro do conceito: Catalogue Number

Aspecto analisado	Resultado
Nome do conceito	Catalogue Number
Identificador	383
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Um identificador (de preferência exclusivo) para o registro dentro do conjunto ou coleção de dados.
Outros nomes para o conceito ou classe	catalogNumber
Fonte do conceito	SiBBr-ALA – Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.3 – Registro do conceito collectID

Aspecto analisado	Resultado
Nome do conceito	collectID
Identificador	132
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Um identificador para a coleção ou conjunto de dados do qual o registro foi derivado.
Outros nomes para o conceito ou classe	Collection Collection Code Collection code not recognized Collection ID CollectionCode
Fonte do conceito	IUCN
Observações e comentários	Termo ou conceito pertence ao sistema de informação CNCFlora

Quadro A.4 – Registro do conceito Collection

Aspecto analisado	Resultado
Nome do conceito	Collection
Identificador	381
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Coleção ou conjunto de dados
Outros nomes para o conceito ou classe	CollectID Collection Code Collection code not recognized Collection ID CollectionCode
Fonte do conceito	SiBBr- ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.5 – Registro do conceito Collection Code

Aspecto analisado	Resultado
Nome do conceito	Collection Code
Identificador	398
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	O nome, acrônimo, código ou inicialismo que identifica a coleção ou conjunto de dados do qual o registro foi derivado.
Outros nomes para o conceito ou classe	Collection collectID Collection code not recognized Collection ID CollectionCode
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.6 – Registro do conceito Collection code not recognised

Aspecto analisado	Resultado
Nome do conceito	Collection code not recognised
Identificador	446
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	O nome, acrônimo, código ou inicialismo que identifica a coleção ou conjunto de dados do qual o registro foi derivado.
Outros nomes para o conceito ou classe	Collection collectID Collection ID collectionCode
Fonte do conceito	SiBBr- ALA – Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.7 – Registro do conceito Collection ID

Aspecto analisado	Resultado
Nome do conceito	Collection ID
Identificador	380
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Um identificador para a coleção ou conjunto de dados do qual o registro foi derivado
Outros nomes para o conceito ou classe	Collection collectID Collection code not recognised collectionCode Collection Code
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.8 – Registro do conceito collectionCode

Aspecto analisado	Resultado
Nome do conceito	collectionCode
Identificador	27
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	O nome, acrônico, código ou iniciais identificando a coleção ou data set de onde o registro é derivado.
Outros nomes para o conceito ou classe	Collection collectID Collection code not recognised Collection Code Collection ID
Fonte do conceito	SiBBr- ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.9 – Registro do conceito Country - parsed

Aspecto analisado	Resultado
Nome do conceito	Country - parsed
Identificador	406
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	País.
Outros nomes para o conceito ou classe	Country inferred from coordinates
Fonte do conceito	Padrão SiBBr-ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.10 – Registro do conceito ti inferred from coordinates

Aspecto analisado	Resultado
Nome do conceito	Country inferred from coordinates
Identificador	428
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	País inferido a partir de coordenadas.
Outros nomes para o conceito ou classe	Country - parsed
Fonte do conceito	Padrão SiBBr-ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.11 – Registro do conceito countryCode

Aspecto analisado	Resultado
Nome do conceito	CountryCode
Identificador	16
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	O código padrão para o país em que o campo dwc:Location está presente.
Outros nomes para o conceito ou classe	CountryCode
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação Flora e Funga do Brasil e CNCFlora

Quadro A.12 – Registro do conceito countryCode

Aspecto analisado	Resultado
Nome do conceito	CountryCode
Identificador	137
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Código padrão para país.
Outros nomes para o conceito ou classe	CountryCode
Fonte do conceito	Padrão IUCN
Observações e comentários	Termo ou conceito pertence ao sistema de informação CNCFlora

Quadro A.13 – Registro do conceito Longitude - original

Aspecto analisado	Resultado
Nome do conceito	Longitude - original
Identificador	401
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	coordenada original. termo darwin core decimalLongitude ou verbatimLongitude
Outros nomes para o conceito ou classe	Longitude
Fonte do conceito	Padrão SiBBr-ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.14 – Registro do conceito Longitude

Aspecto analisado	Resultado
Nome do conceito	Longitude
Identificador	404
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Coordenada original. termo darwin core decimalLongitude ou verbatimLongitude
Outros nomes para o conceito ou classe	Longitude - original
Fonte do conceito	Padrão SiBBr-ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.15 – Registro do conceito: *Vernacular name*

Aspecto analisado	Resultado
Nome do conceito	Vernacular name
Identificador	389
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Um nome comum ou vernacular.
Outros nomes para o conceito ou classe	Vernacular name Vernacular name - original
Fonte do conceito	Padrão SiBBr-ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

Quadro A.16 – Registro do conceito: *Vernacular name*

Aspecto analisado	Resultado
Nome do conceito	Vernacular name
Identificador	21
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Um nome comum ou vernacular associado.
Outros nomes para o conceito ou classe	Vernacular name Vernacular name - original
Fonte do conceito	Padrão Darwin Core
Observações e comentários	Termo ou conceito pertence ao sistema de informação Flora e Funga do Brasil e Catálogo da Fauna

Quadro A.17 – Registro do conceito: *Vernacular name – original*

Aspecto analisado	Resultado
Nome do conceito	Vernacular name - original
Identificador	386
Forma categorial do conceito: (O) Objeto, entidade	Metadado
Definição do conceito	Um nome comum ou vernacular associado.
Outros nomes para o conceito ou classe	Vernacular name Vernacular name
Fonte do conceito	Padrão SiBBr- ALA
Observações e comentários	Termo ou conceito pertence ao sistema de informação SiBBr

APÊNDICE B – COMPARTILHAMENTO DOS DADOS

A presente pesquisa se contextualiza e se compromete com os aspectos voltados para a Ciência Aberta. Durante o percurso de sua construção foi desenvolvido um plano de gestão de dados, com fins de empregar o compartilhamento e gerenciamento dos dados de forma precisa. Com o fim da pesquisa, verifica-se a necessidade de compartilhamento dos dados. Para isso, os dados da pesquisa estarão abertos e aptos para o reuso sem nenhuma restrição no repositório Dataverse do IBICT.