UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA – ENGENHARIA DE COMPUTAÇÃO

CLEVERTON FERREIRA BORBA

**Estratégias de computação para suportar interoperabilidade entre
modelos processados por ferramentas de modelagem de
distribuição de espécies**

São Paulo

2017

CLEVERTON FERREIRA BORBA

**Estratégias de computação para suportar interoperabilidade entre
modelos processados por ferramentas de modelagem de
distribuição de espécies**

Tese apresentada à Escola Politécnica da Universidade de São
Paulo para a obtenção do título de Doutor em Ciências.

Área de Concentração: Engenharia de Computação

Orientador: Prof. Dr. Pedro Luiz Pizzigatti Corrêa

São Paulo

2017

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 25 de Setembro de 2017

Assinatura do autor:

Assinatura do orientador:

Catálogo-na-publicação

Borba, Cleverton Ferreira

Estratégias de computação para suportar interoperabilidade entre modelos processados por ferramentas de modelagem de distribuição de espécies / C. F. Borba -- versão corr. -- São Paulo, 2017.

148 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Modelagem de distribuição de espécies 2.Informática para biodiversidade 3.Padrão de metadados 4.Interoperabilidade 5.Perfil de aplicação I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

BORBA, C. F. Estratégias de computação para suportar interoperabilidade entre modelos processados por ferramentas de modelagem de distribuição de espécies. Tese apresentada à Escola Politécnica da Universidade de São Paulo para a obtenção do título de Doutor em Ciências.

Aprovado em: 15 de Agosto de 2017

Banca Examinadora

Prof. Dr. Pedro Luiz Pizzigatti Corrêa

Instituição: Escola Politécnica da Universidade de São Paulo - USP

Julgamento: Aprovado

Prof. Dr. Edison Spina

Instituição: Escola Politécnica da Universidade de São Paulo - USP

Julgamento: Aprovado

Prof. Dr. Fernando Corrêa

Instituição: Universidade Metodista de Piracicaba - UNIMEP

Julgamento: Aprovado

Prof. Dr. Fuad Gattaz Sobrinho

Instituição: Centro Universitário Adventista de São Paulo - UNASP

Julgamento: Aprovado

Prof. Dr. Carlos Roberto Valêncio

Instituição: Universidade Estadual Paulista Julio de Mesquita Filho - UNESP

Julgamento: Aprovado

A minha esposa, Cintia, ao meu filho Pedro e aos meus pais, Paulo e Leda.

AGRADECIMENTOS

Agradeço a minha esposa Cintia pelo companheirismo, incentivo e dedicação para ajudar-me durante este período de estudos. Cada palavra de carinho e amor foi fundamental para a construção do meu desenvolvimento acadêmico.

Agradeço aos meus pais Paulo e Leda, que sempre se doaram por completo para que eu pudesse alcançar meus sonhos. Eles são exemplos de uma vida de dedicação e missão, e sem o auxílio deles nada disso seria possível.

Agradeço a minha irmã Kevelyn que me apoiou e incentivou na busca pelo conhecimento e estudo durante toda minha vida.

Agradeço ao meu pequeno filho Pedro, que mesmo em tão tenra idade é capaz de transmitir carinho e sabedoria para mim. Foi por ele que a cada dia eu me esforçava mais e procurava fazer o melhor.

Agradeço a toda minha família, tios, avós e primos que de alguma forma contribuíram na minha caminhada educacional e também pessoal. Agradeço em especial a minha tia Sara que se dedicou na correção deste trabalho.

Agradeço ao meu orientador Prof. Dr. Pedro Pizzigatti Corrêa, que desde a primeira vez que nos reunimos me incentivou a trabalhar nesta área e foi um mentor que esteve sempre presente para transmitir conhecimento para seus orientandos.

Aos professores, Mike Frame (UT – University of Tennessee), Jorge Becerra, Selma Meinkoff, Ana Realí Costa, Antonio Mauro Saraiva, Carlos Cugnasca, Edison Spina, pelas disciplinas ministradas com tanto carinho e dedicação, cada uma delas contribuiu grandemente para o avanço deste trabalho e com certeza definiram minha carreira na área de engenharia de computação.

Aos meus amigos de doutorado pelas conversas que direcionaram este trabalho, pelo incentivo e pelas palavras de apoio nos momentos de dificuldade.

Aos representantes da Universidade de São Paulo pela oportunidade de pesquisar em uma das maiores e mais respeitadas universidades da América Latina

e do mundo, assim como todo o corpo de servidores, professores e colegas têm meu respeito e admiração.

Agradeço ao Centro Universitário Adventista de São Paulo pelo apoio durante o período em que fiz parte do grupo de servidores desse estabelecimento. Como a missão não termina nunca, espero um dia fazer parte e contribuir uma vez mais para essa causa que tem auxiliado tantos alunos ao redor do Brasil.

Agradeço ao grupo de servidores do departamento de Engenharia de Computação da Escola Politécnica da Universidade de São Paulo, que em todos os momentos estavam dispostos a ajudar e a buscar o melhor para seus alunos e professores.

Aos professores da banca avaliadora, os meus agradecimentos pelas sugestões e críticas que contribuíram para o aprimoramento do trabalho.

Sobretudo, agradeço a Deus pela oportunidade da vida e por dedicar minha vida a trabalhar para Ele.

Epígrafe:

*“A educação é a arma mais poderosa que você pode usar para mudar o mundo.
Devemos promover a coragem onde há medo, promover o acordo onde existe
conflito, e inspirar esperança onde há desespero”.*

Nelson Mandela

RESUMO

BORBA, C. F. **Estratégias de computação para suportar interoperabilidade entre modelos processados por ferramentas de modelagem de distribuição de espécies**. 2017. 148 f. Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2017.

A área de pesquisa denominada Informática para Biodiversidade enfrenta o desafio de suprir a demanda por tecnologia de apoio à conservação da biodiversidade. Dentre as áreas da ecologia que se beneficiam de recursos tecnológicos, a modelagem de distribuição de espécies se destaca pelo número de ferramentas e de algoritmos desenvolvidos para os pesquisadores da área. No entanto, estudos demonstram que a modelagem tem se tornado mais complexa, bem como, a necessidade por melhorias da aplicação de novas técnicas e soluções computacionais. A interoperabilidade computacional para os dados de biodiversidade, que incluem os modelos gerados pelas ferramentas de modelagem, é citada pelos autores e pesquisadores da área como uma preocupação que demanda estudos, definição de melhores práticas e soluções de computação. Cada tópico que envolve a modelagem como, algoritmos, parâmetros, dados de ocorrência, camadas ambientais, ferramentas e técnicas para a elaboração de modelos, além dos padrões de metadados e perfis de aplicação, são objetos de estudo para que seja possível a interoperabilidade nesse campo de pesquisa. Este trabalho apresenta, portanto, uma estratégia computacional para apoiar a interoperabilidade entre os modelos gerados pelas ferramentas de modelagem de distribuição de espécies, através do uso de padrões de metadados e pelo desenvolvimento de um perfil de aplicação que apoie essa interoperabilidade. São apresentados dois estudos de caso com o objetivo de validar essas estratégias e uma arquitetura computacional para apoiar comunidades de desenvolvimento de ferramentas de modelagem a buscarem essa interoperabilidade entre os dados fornecidos e processados por elas. As contribuições deste trabalho envolvem a disponibilização de um perfil de aplicação voltado para a área de modelagem de distribuição de espécies, um protótipo de software para processar os modelos gerados pelas ferramentas de modelagem, o incentivo e o uso do digital object identification para disponibilização dos dados, além do processo de pesquisa e da validação de metodologias que foram utilizadas para se ancaçarem os resultados desta pesquisa.

Palavras-chave: Modelagem de distribuição de espécies. Informática para Biodiversidade. Padrão de Metadados. Interoperabilidade. Perfil de Aplicação.

ABSTRACT

BORBA, C. F. **Computational strategies to support interoperability between models processed by species distribution modeling tools**. 2017. 148 f. Thesis (Doctoral) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2017.

The research area called Informatics for Biodiversity faces the challenge of meeting the demand for technology to support biodiversity conservation. Among the areas of ecology that benefit from technological resources, the modeling of species distribution is notable for the number of tools and algorithms developed to researchers in this area. However, studies have shown that modeling has become more complex, as well as the need for improvements in the application of new techniques and computational solutions. The computational interoperability for biodiversity data, which includes models generated by modeling tools, is cited by authors and researchers in the field as a concern that demands studies, definition of best practices and computing solutions. Each topic that involves modeling such as algorithms, parameters, occurrence data, environmental layers, tools, techniques for the elaboration of models, standards of metadata and application profiles are objects of study of this dissertation. This thesis presents a computational strategy to support the interoperability between the models generated by the modeling tools of species distribution, through the use of metadata standards and the development of an application profile that supports this interoperability. Two case studies are presented with the objective of validating these strategies and we also present a computational architecture to support communities to develop modeling tools to seek this interoperability between the data supplied and processed by them. The contributions of this work involve the provision of an application profile for the area of species distribution modeling, a software prototype to process the models generated by the modeling tools, the incentive and use of the digital object identification to make data available, and in addition to the process of research and validation of methodologies that were used to summarize the results of this research.

Keywords: Species Distribution Modeling. Biodiversity Informatics. Metadata Standard. Interoperability. Application Profile.

LISTA DE ILUSTRAÇÕES

Figura 1 – Escopo atual da Informática para Biodiversidade.	20
Figura 2 - Objetivo do trabalho para oferecer uma interoperabilidade entre modelos gerados pelas ferramentas de modelagem de distribuição de espécies.	24
Figura 3 – Metodologia do trabalho.	28
Figura 4 – Ciclo de Vida dos Dados de Biodiversidade.....	32
Figura 5 - Modelagem de Distribuição de Espécies e seu escopo em modelos correlativos.....	36
Figura 6 - O diagrama apresenta uma combinação de fatores bióticos, abióticos e aonde a espécie tem acesso, para determinar a sua distribuição geográfica.	38
Figura 7 – Ciclo de Vida dos Dados de Biodiversidade.....	44
Figura 8 – Níveis de compatibilidade.....	54
Figura 9 - Estrutura de elaboração do arquivo DwC-A.....	63
Figura 10 - Visualização do <i>Schema</i> EML.....	65
Figura 11 – Principais padrões de metadados utilizados para área de Informática para Biodiversidade.	68
Figura 12 – Singapore Framework	76
Figura 13 - Dublin Core como padrão genérico e os demais como específicos.	82
Figura 14 - Modelo de Domínio do AP-SDM	92
Figura 15 - Estrutura do EML e suas aplicações para o modelo de domínio.	93
Figura 16 - Modelo de Domínio em associação com os padrões de metadados envolvidos.	94

Figura 17 - Atividades do design da arquitetura.....	97
Figura 18 - Caso de uso de cenário utilizando sistema para interoperar modelo gerado por ferramenta de MDE.	101
Figura 19 - Arquitetura proposta para interoperabilidade entre modelos gerados por ferramentas de MDE.....	103
Figura 20 - Principais diferenças entre os modelos gerados pelas ferramentas de modelagem openModeller e MaxEnt.	107
Figura 21 - Dados de ocorrência do estudo de caso.....	109
Figura 22 - Informações do Data Set do estudo de caso.	110
Figura 23 - Parâmetros utilizados para a modelagem do estudo de caso.	111
Figura 24 - Modelo gerado pela ferramenta openModeller	111
Figura 25 - Modelo gerado pela ferramenta openModeller para o estudo de caso.	112
Figura 26 - Parte do arquivo XML com o modelo gerado pela ferramenta openModeller.	112
Figura 27 - Arquivo EML no formato XML do estudo de caso utilizando openModeller para ser validado.	115
Figura 28 - Validação do arquivo EML gerado para o estudo de caso com a padronização do modelo gerado utilizando a ferramenta openModeller.	116
Figura 29 - Pacote com as informações geradas pelo ambiente de teste contendo todas os dados gerados pela ferramenta openModeller padronizados pelo perfil de aplicação AP-SDM.	117
Figura 30 - Parte dos Dados de ocorrência utilizados no estudo de caso.	118
Figura 31 - Conjunto de informações utilizadas na modelagem para este estudo de caso.	119

Figura 32 - Formato dos arquivos permitidos na ferramenta MaxEnt.	119
Figura 33 - Parâmetros de modelagem utilizados no estudo de caso.....	120
Figura 34 - Modelo de Distribuição de Espécie gerado pela ferramenta MaxEnt para este estudo de caso.....	121
Figura 35 - Arquivo no formato EML para validação da padronização do modelo gerado pela ferramenta MaxEnt.	123
Figura 36 - Validação do arquivo gerado pela ferramenta SIME com os dados do modelo gerado pela ferramenta MaxEnt.....	124
Figura 37 - Pacote com todas as informações do modelo padronizadas.....	124
Figura 38 - Tela do DataVerse com a publicação dos dados gerados pelos estudos de caso deste trabalho.....	126

LISTA DE TABELAS

Tabela 1 - Fatores que determinam a ausência e presença de espécies.....	41
Tabela 2 - Dados ecológicos com informações limitadas sobre os dados coletados.	58
Tabela 3 - Dados ecológicos com informações adicionais.....	58
Tabela 4 - Categoria dos sete termos do Darwin Core Simple e dos termos do Darwin Core Genérico.....	61
Tabela 5 - Relação entre pesquisas e a apresentação da metodologia utilizada para o desenvolvimento de um perfil de aplicação.	80
Tabela 6 - Análise das pesquisas relacionadas ao desenvolvimento e utilização de perfis de aplicação.	81
Tabela 7 - Aplicação do uso dos algoritmos e ferramentas de MDE.....	84
Tabela 8 - Abordagens de Interoperabilidade entre Dados de Biodiversidade.	86
Tabela 9 - Vocabulários utilizados para o desenvolvimento do perfil de aplicação e DataSet.	95
Tabela 10 - DSP - <i>Description Set Profile</i> do perfil de aplicação AP-SDM	96

LISTA DE ABREVIATURAS E SIGLAS

ACM - *Association for Computing Machinery*

ANN - *Artificial Neural Networks*

AP-SDM - *Application Profile for Species Distribution Modeling*

BioCASE - *Biological Collection Access Service*

CEN - Comitê Europeu de Normatização

CSM - *Climate Space Model*

DC - Dublin Core

DCAP - *Dublin Core Application Profile*

DCMI - *Dublin Core Metadata Initiative*

DOI - *Digital Object Identifier*

DSP - *Description Set Profile*

DwC-A - *Darwin Core Archives*

EML - *Ecological Metadata Language*

ENFA - *Ecological-Niche Factor Analysis*

GARP - *Genetic Algorithm for Rule Set Production*

GBIF - *Global Biodiversity Information Facility*

GBM - *Generalized Boosting Methods*

GEMINI - *Geo-spatial Metadata Interoperability Initiative*

GeoCASE - *GEosciences Collection Access Service*

HISPID - *Herbarium Information Standards and Protocols for Interchange of Data*

HTML - *Hypertext Markup Language*

IA - Inteligência Artificial

IABIN - *Inter-American Biodiversity Information Network*

IEEE - *Institute of Electrical and Electronics Engineers*

IMS - *Instructional Management System project*

ISO - *International Organization for Standardization*

KNB - *The Knowledge Network for Biocomplexity*

LTER - *Long-Term Ecological Research Network*

MARS - *Multivariate Adaptative Regression Splines*

MDE - Modelagem de Distribuição de Espécies

NASA - *National Aeronautics and Space Administration*

NBN - *National Biodiversity Network*

NCEAS - *National Center for Ecological Analysis and Synthesis*

NGDF - *National Geospatial Data Framework*

NSF - *National Science Foundation*

RFPA - Requisitos Funcionais do Perfil de Aplicação

SDM - *Species Distribution Modeling*

SIG - Sistemas de Informação Geográfica

SIME - Sistema de Interoperabilidade para Modelagem de Distribuição de Espécies

SVN - *Support Vector Machines*

TDWG - *Taxonomic Databases Working Group*

UML - *Unified Modeling Language*

USGS - *United States Geological Survey*

USP - Universidade de São Paulo

UT - *University of Tennessee*

W3C - *World Wide Web Consortium*

XML - *eXtensible Markup Language*

SUMÁRIO

1	INTRODUÇÃO	18
1.1	JUSTIFICATIVA E MOTIVAÇÃO	21
1.2	OBJETIVOS.....	23
1.3	MATERIAIS E MÉTODOS.....	25
1.4	ORGANIZAÇÃO DA TESE	29
2	ASPECTOS CONCEITUAIS.....	31
2.1	NICHO ECOLÓGICO.....	33
2.2	MODELAGEM DE DISTRIBUIÇÃO DE ESPÉCIE (MDE)	35
2.3	DADOS DE PRESENÇA-AUSÊNCIA DE ESPÉCIE.....	40
2.4	AS FERRAMENTAS PARA MODELAGEM.....	43
2.4.1	AS ABORDAGENS DE ALGORITMOS PARA MODELAGEM	47
3	INTEROPERABILIDADE APLICADA À INFORMÁTICA PARA BIODIVERSIDADE.....	52
3.1	INTEROPERABILIDADE DE DADOS.....	52
3.2	PADRÃO DE METADADOS	56
3.2.1	DUBLIN CORE	59
3.2.2	DARWIN CORE	60
3.2.3	DARWIN CORE A (DWC-A).....	62
3.2.4	EML (<i>ECOLOGICAL METADATA LANGUAGE</i>).....	63
3.2.5	OUTROS PADRÕES DE METADADOS	66
3.2.6	RELAÇÃO DA MODELAGEM DE DISTRIBUIÇÃO DE ESPÉCIES E OS PADRÕES DE METADADOS.....	68
3.3	PERFIL DE APLICAÇÃO	69
3.3.1	DESENVOLVIMENTO DE UM PERFIL DE APLICAÇÃO	74
3.3.2	METODOLOGIA DE DESENVOLVIMENTO DE UM PERFIL DE APLICAÇÃO.....	79
3.4	TRABALHOS RELACIONADOS	83
3.4.1	APRESENTAÇÃO DOS TRABALHOS.....	85
	FONTE: AUTOR.	86

4	DESENVOLVIMENTO DO PERFIL DE APLICAÇÃO E ARQUITETURA COMPUTACIONAL.....	88
4.1	PERFIL DE APLICAÇÃO PARA INTEROPERABILIDADE ENTRE FERRAMENTAS DE MODELAGEM DE DISTRIBUIÇÃO DE ESPÉCIES	88
4.1.1	ESPECIFICAÇÃO DOS REQUISITOS FUNCIONAIS	90
4.1.2	DESENVOLVIMENTO DO MODELO DE DOMÍNIO	92
4.1.3	DESENVOLVIMENTO DA DESCRIÇÃO DO CONJUNTO DE PERFIS (<i>DESCRIPTION SET PROFILE - DSP</i>)	95
4.2	ARQUITETURA COMPUTACIONAL PARA A INTEROPERABILIDADE ENTRE OS MODELOS GERADOS PELAS FERRAMENTAS DE MDE	96
4.2.1	REQUISITOS FUNCIONAIS DA ARQUITETURA.....	98
4.2.2	REQUISITOS NÃO FUNCIONAIS DA ARQUITETURA.....	100
4.2.3	CASO DE USO	100
4.2.4	ARQUITETURA PROPOSTA	102
5	ESTUDO DE CASOS	105
5.1	DESENVOLVIMENTO DO AMBIENTE PARA TESTES.....	105
5.2	ESTUDO DE CASO 1	108
5.2.1	RESULTADOS	113
5.3	ESTUDO DE CASO 2.....	117
5.3.1	RESULTADOS	122
5.4	DISPONIBILIZAÇÃO E PUBLICAÇÃO DE EXPERIMENTOS.....	125
6	CONSIDERAÇÕES FINAIS.....	127
6.1	CONTRIBUIÇÕES DA PESQUISA	127
6.2	TRABALHOS PUBLICADOS RELACIONADOS A ESTA PESQUISA.....	129
6.3	TRABALHOS FUTUROS	130
6.4	DISCUSSÕES FINAIS.....	132

1 INTRODUÇÃO

A computação está presente em todas as áreas da ciência e tem sido objeto de estudo integrado com diversas disciplinas do conhecimento. Na conservação e preservação da biodiversidade, a computação também tem papel importante, quer seja buscando, armazenando ou processando informação.

Na última década, tecnologias de apoio à conservação da biodiversidade têm sido desenvolvidas e aprimoradas, fazendo com que pesquisas na área sejam de vital importância para que essas tecnologias atendam de maneira satisfatória as necessidades dos pesquisadores da área. Peterson (1988) sugeriu que se necessitavam estabelecer mais programas interdisciplinares em colégios e universidades que estimulassem estudantes e professores na pesquisa nas áreas de ecologia e de computação (PETERSON, 1988; SOUZA MUÑOZ *et al.*, 2011).

Segundo George Bugliarello (BOTKIN *et al.*, 1989, p. 385), a partir do momento que “adquirimos mais informações e conhecimento sobre o funcionamento dos sistemas ecológicos, pode-se utilizar essas informações para melhores tomadas de decisão relacionadas à conservação da biodiversidade”.

No documento discutido e disponibilizado pela Comissão Mundial sobre Meio Ambiente e Desenvolvimento (WCED, 1987, p. 54), os membros discutem o fato de que esforços devem ser realizados para minimizar os impactos ambientais por meio de programas que incentivem o desenvolvimento de tecnologia inovadora ou a adaptação de tecnologias tradicional e existente.

Os sistemas computacionais possuem um grande potencial de apoio na monitoração do estado do ambiente, redução dos impactos das atividades humanas sobre os sistemas que mantêm a vida na Terra e de melhora da qualidade de vida (MANUAL GLOBAL DE ECOLOGIA, 2002).

Essa área integrada de estudos é conhecida como Informática para Biodiversidade, do inglês *Biodiversity Informatics*, esta tem o objetivo de atender a demanda por tecnologia para apoio à conservação e preservação da biodiversidade. Peterson ainda explica que esse é um campo de estudos novo, com a mais antiga

citação do termo realizada há 15 anos atrás, por isso a área ainda está em estágio inicial de desenvolvimento e está evoluindo à medida que novas soluções são trabalhadas e elaboradas (PETERSON *et al.*, 2010).

Ariño *et al.* (2011, p.1) afirma que Informática para Biodiversidade é uma disciplina relativamente nova que se estende da ciência e da engenharia de computação no contexto dos dados de biodiversidade e seu desenvolvimento, até então, não é uniforme ao redor do mundo¹.

Segundo Carl J. Walter, professor do Instituto de Ecologia de Recursos Naturais da Universidade da Columbia Britânica (*University of British Columbia*) - Canadá, a aplicação dos procedimentos de análise de sistemas na ecologia ficou conhecida como Ecologia de Sistemas, que está se convertendo em uma ciência principal, por duas razões (ODUM, 2012):

1. por disponibilizar novos instrumentos formais em termos de teoria matemática, cibernética, processamento eletrônico de dados etc.
2. pela simplificação formal de ecossistemas complexos que o homem não poderia mais continuar resolvendo por intermédio do ensaio e cometendo erros e a necessidade de existir uma solução própria para cada problema.

Outros autores como Recknagel, Yang, Goethal, também denominam essa área como Informática para Ecologia, do inglês *Ecological Informatics*, e que, em essência, tem o mesmo objetivo, de que recursos computacionais que auxiliem na preservação da biodiversidade (RECKNAGEL, 2011), (YANG, 2011) e (GOETHALS e CHON, 2013).

No entanto, a expressão Informática para Ecologia, tem um sentido mais amplo e define como um “*framework* interdisciplinar” que promove os princípios da

¹(Ariño *et al.*, 2011) diz que o estado da área de Informática para Biodiversidade é diferente para cada país do mundo e que sua composição pode ser definida através de três conceitos: (1) O potencial intrínseco da biodiversidade de um país (sua riqueza e fatores que favorecem a ele, biológica e ecologicamente); (2) A capacidade do país para gerar registros de biodiversidade e (3) A disponibilidade de infraestrutura técnica de um país para hospedagem, gerenciamento e compartilhamento de registros de dados de biodiversidade. Nesse artigo, os autores apresentam um índice no qual é possível qualificar os países de acordo com as características citadas acima, e o Brasil encontra-se em 2º lugar do ranking, atrás dos Estados Unidos e seguido pela França.

informática nos diversos níveis de complexidade dos ecossistemas, desde a genética até a sustentabilidade ecológica e o aquecimento global (RECKNAGEL, 2003). Já Michener e Jones (2012, p. 85) afirmam que a Informática para Ecologia disponibiliza ferramentas e abordagens para o gerenciamento de dados ecológicos a fim de transformá-los em informação e conhecimento.

Ambas as expressões são utilizadas na verdade com o mesmo intuito, de promover o uso da tecnologia, em diversos níveis de complexidade, para a conservação e preservação da biodiversidade. E sem o descrédito em nenhuma das expressões, este trabalho, utiliza ambas as definições com o mesmo objetivo, e tem como princípio dar suporte referencial para a pesquisa aqui apresentada.

Segundo Peterson *et al.* (2010), a Informática para Biodiversidade considera as seguintes etapas:

- a. Extração e captura dos dados;
- b. Compilação desses dados e serviços;
- c. Exibição e visualização dos dados coletados.

Conforme Recknagel (2003, p.9),

a integração de dados com todas as categorias de ecossistemas e níveis de complexidade, a inferência de padrões de dados para os processos ecológicos e bem como a simulação adaptativa e previsão de ecossistemas são características de atuação da Informática para Biodiversidade[...].

Figura 1 – Escopo atual da Informática para Biodiversidade.

A Figura 1 é representada por Recknagel (2003, p. 9) como

o escopo atual da informática para biodiversidade, os dados ecológicos são refinados em teorias de ecossistemas e em suporte a tomada de decisão através de ferramentas computacionais como arquivamento, recuperação e visualização de dados além da análise, síntese e previsão de ecossistemas existentes[...].

Vê-se, portanto, a importância em estudos dirigidos da engenharia e da ciência da computação em parceria com a área de ecologia, biodiversidade e afins.

1.1 Justificativa e Motivação

A área de pesquisa científica denominada Informática para Biodiversidade enfrenta o desafio de suprir a demanda por soluções computacionais e de tecnologia para o apoio à conservação da biodiversidade (PETERSON *et al.*, 2010; CORRÊA *et al.*, 2011).

De acordo com Peterson *et al.* (2010, p. 159), essa área de pesquisa “é uma área relativamente nova, portanto em seu estágio inicial de desenvolvimento, e está evoluindo à medida que novas ferramentas são elaboradas e desenvolvidas”. Dentre essas evoluções, podem-se citar as ferramentas de modelagem de distribuição de espécies ou MDE, do inglês (*species distribution modelling, SDM*), que têm como objetivos principais a predição da distribuição atual das espécies, entender fatores ambientais relacionados e realizar predição de abundância (ELITH *et al.*, 2011).

No entanto, estudos demonstram que a modelagem de distribuição de espécies tem se tornado mais complexa (SOBERÓN e NAKAMURA, 2009; HORTAL *et al.*, 2010). De igual modo, as ferramentas de modelagem necessitam de melhorias na aplicação de novas técnicas e estratégias de modelagem (PETERSON *et al.*, 2011), como também a necessidade de interoperabilidade de dados entre as ferramentas disponíveis.

Berendsohn *et al.* (2011, p.71) dizem que

um dos mais graves “estrangulamentos” nos projetos científicos das ciências da biodiversidade é a necessidade de integrar dados de diferentes

fontes, aplicações de software e serviços para análise, visualização e publicação e assim oferecer uma interoperabilidade de dados, informações, aplicações e ferramentas[...].

Para Wieczorek *et al.* (2012, p. 1), “um dos desafios dos dados de biodiversidade e seus componentes é o compartilhamento dessas informações, uma publicação coordenada de dados e também a integração de sistemas”.

Berendsohn *et al.* (2011, p. 72) ainda afirmam que

a captura e armazenamento de dados de biodiversidade não são suficientes; embora a maioria dos atributos seja compartilhada para o mesmo domínio, os conjuntos de dados não são facilmente ligados e integrados. A falta de vocabulários compartilhados e a diversidade das estruturas de dados utilizadas impedem o compartilhamento de dados[...].

Na conservação e preservação da biodiversidade, pesquisadores cada vez mais se utilizam das ferramentas de MDE para a definição e tomada de decisões relacionadas à conservação ambiental, no entanto, os pesquisadores processam informações nessas ferramentas e os dados processados por elas, ou seja, os modelos de distribuição de espécies são seu produto final, não disponibilizando estes para reúso, ou reprocessamento, nem mesmo tornando-os acessíveis para futuras pesquisas.

Existe, entretanto, uma lacuna a ser preenchida nos dias atuais, em que as informações compartilhadas apoiam a tomada de decisões com relação à conservação da biodiversidade. Esta pesquisa discute a importância da padronização para esses modelos gerados pelas ferramentas de MDE, e, portanto, proporcionando uma interoperabilidade entre elas, e entre os meios de disponibilização de dados já disponíveis na ciência.

Sobre os modelos gerados pelas ferramentas de MDE, Marshall *et al.* (2014, p. 1) afirmam que o

primeiro problema diz respeito aos metadados e a aparente falta de padrões de metadados para os resultados de MDE. Essas normas devem ser desenvolvidas para que qualquer pessoa que possua as saídas das MDE seja fornecida com informações suficientes sobre o modelo para permitir que utilize, de uma maneira apropriada, as saídas do modelo[...].

Logo, baseado nas discussões em grupos de pesquisa de biodiversidade e em citações publicadas pelos autores da área, é possível verificar que as ferramentas de MDE podem se beneficiar da interoperabilidade, visando à integração de sistemas, além de oferecer um meio de padronização para os modelos gerados por essas ferramentas. Pesquisadores e profissionais da área utilizam mais de uma ferramenta de modelagem para alcançar resultados satisfatórios para a tomada de decisão, que muitas vezes são insuficientes.

A necessidade de interoperabilidade é um dos desafios à criação de fluxos eficientes de trabalho que ajudam cientistas, pesquisadores e usuários de infraestruturas de Informática para Biodiversidade melhorar a qualidade e eficiência de seus processos de trabalho (BERENDSOHN *et al.*, 2011).

Este, portanto, é uma pesquisa que visa promover a discussão, o estudo e a pesquisa na área da engenharia da computação que apoia a Informática para Biodiversidade e define ferramentas computacionais que apoiem a interoperabilidade entre aplicações de MDE.

1.2 Objetivos

O objetivo principal desta tese é definir a interoperabilidade computacional entre modelos gerados por ferramentas de modelagem de distribuição de espécies.

Para alcançar o objetivo proposto nessa pesquisa, os seguintes objetivos específicos são estabelecidos a seguir:

1. Descrever os padrões de metadados que atendem à modelagem de distribuição de espécies;
2. Definir um perfil de aplicação baseado em padrões de metadados para padronização da comunicação entre ferramentas de modelagem e a disponibilização dos dados dos modelos gerados por essas ferramentas;
3. Especificar um protótipo de testes, baseado no perfil de aplicação desenvolvido e nos padrões de metadados, para verificar e validar a interoperabilidade entre ferramentas de MDE;

4. Realizar a divulgação da pesquisa por intermédio da publicação em periódicos da área e disseminar o estudo nas áreas de padrões de metadados, modelagem de distribuição de espécies e suas ferramentas.

Na Figura 2 é possível verificar de forma gráfica a ideia representada com base nos objetivos descritos anteriormente, em que determinada ferramenta de modelagem acessa os dados de espécies, que podem estar em bases locais, metadados ou base de dados compartilhados. A partir desses dados, as ferramentas de modelagem geram os modelos de distribuição de espécies. Com o modelo gerado e por intermédio do perfil de aplicação baseado nos padrões de metadados será gerado um arquivo único com todas as informações relacionadas ao modelo. Essas informações, em arquivo XML (*eXtensible Markup Language*), podem ser disponibilizadas para base de dados de biodiversidade ou reutilizadas pelas ferramentas de modelagem.

Figura 2 - Objetivo da pesquisa para oferecer uma interoperabilidade entre modelos gerados pelas ferramentas de modelagem de distribuição de espécies.

1.3 Materiais e Métodos

A metodologia utilizada para essa tese tem como passo inicial a avaliação das características de todos os artefatos que envolvem o modelo de distribuição de espécies. Primeiramente, os conceitos de biodiversidade, modelagem de distribuição de espécies, nicho ecológico, dados de ocorrência, de presença e ausência, são importantes, pois oferecem base para o entendimento dos fundamentos e do processamento de informações de dados de biodiversidade.

As ferramentas de MDE são fundamentais como objeto de estudo. Dessa maneira, foram estudados algoritmos empregados, mecanismos para o acesso aos dados de observação de espécies e o acesso aos dados de biodiversidade, bem como os padrões aceitos em cada ferramenta.

Trabalhos relacionados, expostos no Capítulo 3, foram estudados e analisados com o intuito de buscar dois aspectos que contribuíssem para a definição da proposta e dos objetivos dessa pesquisa: (1) Utilização das soluções de MDE para a tomada de decisão em diferentes frentes na conservação e preservação da biodiversidade com os algoritmos utilizados e (2) pesquisas e trabalhos relacionados à interoperabilidade dos dados de biodiversidade e suas ferramentas.

A justificativa e a motivação para esta tese estão baseadas nos fundamentos da modelagem de distribuição de espécies, bem como em suas principais características, tendo em vista ser um tema que forneça originalidade no meio científico e acadêmico. Por isso, esta pesquisa busca conceituar a importância da interoperabilidade para a modelagem de distribuição de espécies, mais especificamente para os modelos gerados pelas suas ferramentas. Para alcançar essa interoperabilidade é fundamental o entendimento dos padrões de metadados.

A investigação sobre padrões de metadados e ferramentas de modelagem, teve como objetivo responder às seguintes questões:

- Quais são os padrões de metadados existentes e de que maneira eles atendem a normatização de modelos gerados por ferramentas de MDE?

- Quais algoritmos e ferramentas são mais utilizados para a MDE? E por que essas ferramentas têm um grau de aceitação maior?

Os estudos computacionais sobre interoperabilidade na área de Informática para Biodiversidade têm como objetivo responder às seguintes perguntas:

- Quais são os dados que podem ser gerados por ferramentas de MDE?
- De que maneira é possível a disponibilização dos dados de modelos de distribuição de espécies?
- De que maneira é possível disponibilizar um conector para gerar dados de modelos gerados para as ferramentas existentes de MDE?

A definição de um padrão de metadados, apresentada na seção 2.6, como base para a interoperabilidade dos dados dos modelos demonstra o processo metodológico desta proposta. Com a avaliação dos trabalhos expostos no Capítulo 3, é possível verificar que apenas um padrão de metadados existente não satisfaz a padronização necessária para os modelos de MDE, por isso foi desenvolvido um perfil de aplicação.

O desenvolvimento desse perfil de aplicação foi baseado na metodologia apresentada no Capítulo 3 e essa pesquisa possibilitou a interoperabilidade pretendida entre os modelos de distribuição de espécies, tornando-se uma contribuição original e significativa para a comunidade da área de Informática para a Biodiversidade.

Além disso, baseado no modelo de interoperabilidade proposto, foi especificada e desenvolvida uma ferramenta de testes para a exportação dos modelos gerados pelas ferramentas de MDE.

Essa ferramenta de teste tem como objetivo validar o perfil de aplicação desenvolvido e disponibilizar, por intermédio dele, os dados dos modelos gerados pelas ferramentas e permitir o reúso e, portanto, uma interoperabilidade para que usuários finais possam interagir com os dados de biodiversidade desses modelos e, a partir disso, integrar informação, para realizar a tomada de decisões para a conservação e preservação da biodiversidade.

Esta tese define também um identificador único para os modelos gerados no *Digital Object Identifier* (D.O.I.)² de maneira a facilitar a utilização dos modelos gerados.

O procedimento metodológico descrito é também apresentado na Figura 3, e também vale o destaque para o processo de modelagem de distribuição de espécies utilizado por esta pesquisa, que foi proposto por (PINAYA, 2013) e envolve (1) definição da pesquisa ecológica; (2) coleta de dados; (3) análise dos dados em relação às variáveis preditoras candidatas e a posição georeferenciada; (4) seleção das variáveis preditoras; (5) seleção e execução do algoritmo de modelagem; (6) estimativa da distribuição de espécies para as novas condições e (7) análise do modelo gerado.

Como estudo de caso desta pesquisa, uma ferramenta de testes foi especificada e desenvolvida tendo como objetivo a utilização do perfil de aplicação para ser extraído todos os dados necessários de um modelo gerado por uma ferramenta de MDE, como openModeller (SOUZA MUÑOZ *et al.*, 2011) ou MaxEnt (ELITH *et al.*, 2006), disponibilizado num padrão em um formato XML, e ser reutilizado pela própria ferramenta ou por outra de modelagem.

²*Digital Object Identifier* é um identificador único de objetos de qualquer domínio que serve como um padrão para conteúdos digitais, maiores informações podem ser encontradas através do site: <https://www.doi.org/>

Figura 3 – Metodologia da tese.

Fonte: Elaborada pelo autor.

Em resumo, tendo obtido as informações dos modelos e realizados os testes de reúso dos modelos em diferentes ferramentas de MDE, é possível verificar a interoperabilidade entre diferentes ferramentas de MDE.

Foi realizado um estudo de caso utilizando duas ferramentas de MDE, openModeller e MaxEnt, buscando reproduzir todos os passos utilizados por pesquisadores da área de modelagem, que incluem a busca por dados de ocorrência, a definição dos algoritmos de modelagem utilizados e a produção de modelos de distribuição de espécies. Os dados das espécies (*furcata boliviana* e *bradypus variegatus*) foram disponibilizados pelas comunidades de desenvolvimento das duas ferramentas, bem como os dados climáticos e territoriais para a modelagem.

Todo o processo de modelagem, até a disponibilização dos dados por meio de um DOI, foi realizado com o conhecimento das parcerias entre a Universidade de São Paulo (USP), Ministério do Meio Ambiente da República Federativa do Brasil, Universidade do Tennessee (UT) e órgão de conservação e preservação da biodiversidade do governo americano (USGS – *United States Geological Survey*).

1.4 Organização da tese

Esta tese está dividida em capítulos, ordenados com o objetivo de esclarecer e servir de base para futuras pesquisas na área de engenharia de computação, aplicado à área de Informática para Biodiversidade. A divisão dos capítulos está disposta da seguinte maneira:

O Capítulo 2 apresenta os aspectos conceituais que dão base para a modelagem de distribuição de espécies. Nele são descritos os conceitos de nicho ecológico, de que maneira são adquiridos os dados de ocorrência das espécies, utilizados na MDE, as diferentes ferramentas de MDE, bem como os algoritmos utilizados e seus diferentes resultados e perspectivas.

No Capítulo 3 são apresentados também aspectos conceituais, no entanto, com o foco na interoperabilidade por intermédio de padrões de metadados e perfis de aplicação. Ainda nesse espaço são expostos os trabalhos relacionados a esta pesquisa, que buscam a definição das ferramentas, dos padrões de metadados e da metodologia para o desenvolvimento do perfil de aplicação.

Já no Capítulo 4, é detalhado o desenvolvimento do perfil de aplicação fundamentado nos padrões de metadados apresentados no Capítulo 3 e que dá base para os estudos de casos. Nesse capítulo, o desenvolvimento segue a metodologia apresentada também no Capítulo 3 e procura servir de base para a reprodução e outros perfis de aplicação para diferentes áreas, ou mesmo para a melhoria desta pesquisa.

Os testes de validação com o perfil de aplicação e o desenvolvimento da ferramenta de teste desenvolvida para esse fim são apresentados no Capítulo 5

deste trabalho. No capítulo 6, são apresentadas todas as conclusões e também as sugestões de trabalhos futuros baseados nos resultados obtidos. As referências e demais anexos são apresentados na sequência.

2 ASPECTOS CONCEITUAIS

A engenharia e os estudos de computação estão presentes em todas as áreas da ciência, portanto, a relação entre eles é multidisciplinar e abrange grande concentração de esforços acadêmicos e profissionais para a aplicação da ciência nas mais diversas frentes de pesquisa. Este capítulo apresenta aspectos conceituais de áreas distintas, entretanto, interligam-se por meio de soluções tecnológicas propostas por esta pesquisa.

Antes da apresentação dos conceitos específicos, no entanto, é necessário explicar o contexto do ciclo de vida dos dados de biodiversidade em que esta pesquisa está inserida. Ela foi baseada no conceito do ciclo de vida dos dados proposto pela organização DataONE³ em DataOne (2016, P. 1) que diz que este ciclo de vida fornece uma visão geral de alto nível dos estágios que envolvem o gerenciamento e a preservação dos dados para uso e reutilização. Segundo a própria DataONE, várias versões de um ciclo de vida dos dados existem com diferenças que variam de acordo com as práticas, os domínios e as comunidades.

Para esta pesquisa, foi adotado o ciclo de vida dos dados desenvolvido pela equipe de líderes da DataONE, e segundo eles, foi construído sobre o modelo de ciclo de vida apresentado pela NSF⁴ (*National Science Foundation*) e serve como uma estrutura para o desenvolvimento de ferramentas, serviços e materiais educacionais pela DataONE. Dentre os parceiros da DataONE, cita-se a Universidade do Tennessee, Universidade do Novo México, Universidade do Kansas, Laboratório Nacional de Pesquisas Oak Ridge, USGS (*United States Geological Survey*), NASA etc.

Na Figura 4, é possível ver o modelo do ciclo de vida dos dados proposto pela DataONE e adotado por esta pesquisa. O objetivo é contextualizar os aspectos

³DataONE é uma organização com sede física no Estado da Carolina do Norte nos Estados Unidos da América, a qual é composta por uma comunidade científica voltada para trabalhos relacionados com a conservação, preservação e manutenção dos recursos naturais da Terra. Por intermédio dessa comunidade são desenvolvidos diversas ferramentas, soluções e materiais com fim educacional. Para maiores detalhes com relação ao DataONE, acesse: <http://www.dataone.org>

⁴A NSF é uma agência governamental vinculada ao governo norte-americano, que busca promover pesquisas em diversos campos da ciência. Maiores detalhes sobre a NSF podem ser encontrados em: <https://www.nsf.gov/>

conceituais e o desenvolvimento desta pesquisa com as fases do modelo do ciclo de vida, isso faz com que seja possível uma melhor visualização do contexto que os dados de biodiversidade, mais especificamente, da modelagem de distribuição de espécies, se enquadra.

Figura 4 – Ciclo de Vida dos Dados de Biodiversidade.

Fonte: Adaptado de DATAONE, 2016.

O ciclo de vida dos dados de biodiversidade tem oito componentes, são eles (DATAONE, 2016):

1. **Planejar:** Fase que ocorre o planejamento e a descrição dos dados que serão compilados e também a forma que eles serão geridos e disponibilizados durante toda sua vida útil;
2. **Coletar:** Nesta fase as observações são realizadas manualmente ou com sensores ou outros instrumentos e posteriormente os dados são colocados em formato digital;
3. **Garantir:** A qualidade dos dados é assegurada por meio de verificações e inspeções;
4. **Descrever:** A descrição dos dados é feita de forma precisa e completa utilizando-se de padrões de metadados apropriados;

5. **Preservar:** Os dados são submetidos a um arquivo apropriado de longo prazo;
6. **Descobrir:** Os dados potencialmente úteis são localizados e obtidos, juntamente com as informações relevantes sobre este dado (metadado);
7. **Integrar:** Os dados de fontes diferentes são combinados para formar um conjunto homogêneo de dados que podem ser prontamente analisados;
8. **Analisar:** É realizada a análise dos dados.

Esta pesquisa aborda as fases de planejamento, preservação, descrição, integração e análise. Sendo um ciclo, este trabalho é dependente também das outras fases do ciclo de vida dos dados de biodiversidade. Durante a apresentação dos conceitos desse capítulo, será apresentada a sua ligação com as demais fases do ciclo de vida.

2.1 Nicho Ecológico

O nicho ecológico é o que cada espécie em uma teia alimentar realiza, ou em outras palavras, é o trabalho que cada espécie realiza para sobreviver ao seu meio, é como ela vive e interage com o ambiente físico e com as outras espécies em sua comunidade (ODUM e BARRET, 2011).

Hutchinson (1957) propôs que o nicho ecológico é um ambiente no qual um indivíduo ou espécie sobrevive indefinidamente, e que isso pode ser visualizado em um espaço multidimensional. Em outras palavras, o nicho proposto por Hutchinson pode ser avaliado matematicamente, pois as dimensões podem ser representadas através dos eixos x e y , e a cada camada, ou espécie que se queira sobrepor informação, é possível adicionar dimensões a esse plano.

Blonder *et al.* (2014, p. 1) afirmam que a abordagem de Hutchinson

oferece um conjunto de n variáveis que representam eixos biologicamente importantes que são identificados e a partir disso o hipervolume é definido por um conjunto de pontos em um espaço n -dimensional e reflete os valores adequados de variáveis como, por exemplo, temperatura ou tamanho de alimentos[...].

Em uma analogia, pode-se comparar o termo nicho com características da comunidade humana, em que o habitat seria o endereço de determinada pessoa e o nicho seriam as características que descrevem essa pessoa, como por exemplo, formação, profissão, alimentação, relacionamento etc.

O conceito de nicho é bastante discutido no meio acadêmico, pois existem diversas expressões que conceituam a palavra nicho e isso pode causar certa confusão por parte de pesquisadores, gestores e interessados na área. Em uma série de três artigos Greg Mcinerny e Rampal Etienne (MCINERNY e ETIENNE, 2012a), (MCINERNY e ETIENNE, 2012b), (MCINERNY e ETIENNE, 2012d) apresentam aspectos do uso correto e incorreto do conceito de nicho.⁵

Segundo Mcinerny e Etienneb (2012^a, p. 2096), “os conceitos originais de nicho não são diretamente aplicáveis aos aplicativos modernos (como a teoria dos conjuntos)”, “e a modelagem de distribuição de espécies, especialmente o uso da definição de nicho fundamental”, mas por outro lado, Soberón (2014, p. 414) argumenta que em certos tipos de MDE se pode “realmente dispensar alguns conceitos de nicho, mas que tal afirmação não é verdadeira para uma importante classe de atividades baseadas na MDE, incluindo a transferência de previsões no espaço e no tempo”. Soberón argumenta ainda que um único termo utilizado por Mcinerny e Etienne (2012a) para MDE não ajuda esclarecer questões relacionadas à utilização do conceito de nicho.

O nicho fundamental de uma espécie é representado por uma combinação de variáveis ambientais que habilitam a sobrevivência dela, e o nicho realizado é um subconjunto do nicho fundamental, o qual, afetado por fatores externos pode impedir o desenvolvimento eficiente de determinada espécie (SOBERÓN; PETERSON, 2005).

⁵Nestes artigos os autores analisam a utilidade do conceito de nicho na modelagem de distribuição de espécies e também na ecologia. A partir dessa análise, três pontos de vista foram considerados: o primeiro é que o termo nicho é tão antigo quanto a própria história da ecologia; o segundo afirma que a discussão é generalizada e que nenhuma área da ecologia é independente do conceito de nicho e o terceiro diz que o conceito é diverso em função da sua idade e influência no desenvolvimento da ecologia. Portanto, diferentes pesquisadores e disciplinas têm diferentes conceitos e convenções sobre o que o termo nicho é para cada um.

O nicho fundamental é uma construção teórica definida no espaço ambiental que descreve o conjunto de condições que permitem uma espécie sobreviver e ter uma taxa de crescimento ascendente. Assim, segundo Panzacchi *et al.* (2014, p. 1), “a identificação do nicho fundamental de uma espécie permitiria compreender as suas necessidades e prever a sua distribuição no espaço geográfico no passado e as condições ambientais para o presente e futuro”.

Um exemplo sobre a influência das variáveis no nicho ecológico é citado por Guisan e demais autores (2014), que mencionam o quanto a variável de clima pode determinar na distribuição de uma espécie e, de acordo com esses autores, a influência do clima na distribuição de espécie é uma clássica questão ecológica. Para algumas espécies, tem se tornado possível determinar o clima do nicho fundamental delas, baseando-se nas informações fisiológicas, mas para a maioria das espécies a busca pelo nicho fundamental é possível apenas por estudo empírico.

O problema básico da ecologia é determinar a causa da distribuição e abundância de espécies. Todo organismo vive em uma matriz de espaço e tempo que pode ser considerada uma unidade (KREBS, 1978). Considerando ainda o raciocínio de Krebs, os problemas de distribuição e abundância de determinada espécie podem ser analisados no nível de uma simples população de espécies ou em uma comunidade que contenha várias espécies. Com isso, a complexidade de análise aumenta mais e mais dependendo do número de espécies e parâmetros em foco (KREBS, 1978).

Fundamentado o nicho ecológico, na próxima seção serão apresentados os fundamentos da modelagem de distribuição de espécies e sua importância para a conservação e preservação da biodiversidade.

2.2 Modelagem de Distribuição de Espécie (MDE)

Baseado no conceito de nicho pode-se desenvolver a definição de modelagem de distribuição de espécies (MDE), sendo um termo comum nessa tese, será

utilizada uma sigla sempre que citado, MDE, na qual, através de modelos, é possível visualizar a evolução ou o declínio de determinada espécie.

Assim como a expressão “nicho” significa coisas diferentes para pessoas diferentes, a modelagem de distribuição de espécies também se refere a um conjunto heterogêneo de conceitos e objetivos onde o ambiente abiótico está correlacionado com as propriedades de distribuição de uma espécie. Deve-se focar na parte da espécie no diagrama apresentado na Figura 5 e a modelagem disso deve ser chamada modelagem de distribuição de espécies (SOBERÓN, 2014)⁶.

Figura 5 - Modelagem de Distribuição de Espécies e seu escopo em modelos correlativos.

Fonte: Adaptado de MCINERNY; ETIENNE, 2012d.

Modelos ecológicos podem ser comparados a mapas geográficos. Diferentes tipos de mapas servem para diferentes propósitos. Existem os mapas aéreos, fluviais, ferroviários, geológicos, arqueológicos etc. Todos eles são diferentes, pois focam em detalhes diferentes. Eles também estão disponíveis em diferentes escalas de acordo com a aplicação e o conhecimento que o mapa é utilizado (JORGENSEN, 1994). Da mesma forma, os modelos ecológicos são compostos por diversas camadas que fazem com que cada modelo seja único e com diferentes detalhes. Podem existir, por exemplo, diferentes modelos ecológicos para um mesmo

⁶Soberón (2014) afirma sua tese sobre nicho referindo-se à discussão apresentada por (MCINERNY; ETIENNE, 2012a; b; c) e diz que apesar da grande diversidade na utilização do termo nicho para a ecologia, na MDE esse conceito é claro e explica, utilizando a própria ilustração apresentada no artigo (MCINERNY; ETIENNE, 2012c).

ecossistema e as tomadas de decisão baseadas nesses modelos são relativas às variáveis envolvidas no processo de modelagem.

A área de modelagem ecológica tem se desenvolvido nos últimos 30 anos e isso se deve ao fato, principalmente de três fatores (FATH, 2011):

- O desenvolvimento tecnológico da computação, o que tem habilitado a utilização de complexos modelos matemáticos para o desenvolvimento de modelos ecológicos;
- Um entendimento geral dos problemas ambientais, incluindo que a completa eliminação da poluição não é algo viável. Em vez disso, um controle de poluição adequada com recursos econômicos limitados, o que requer uma séria reflexão sobre a influência dos impactos da poluição sobre os ecossistemas;
- O conhecimento dos sistemas ecológicos e ambientais tem aumentado significativamente, em especial, foi ganho mais conhecimento das relações quantitativas nos ecossistemas e entre as propriedades ecológicas e os fatores ambientais.

Dale (2003, p. 12), em seu livro intitulado “*Ecological Modeling for Resource Management*”, afirma que com a

expansão da computação, tem ocorrido uma explosão no desenvolvimento e uso de modelos ecológicos gerados por estes. Os computadores estão disponíveis para muitos gestores de recursos e de decisão, e muitos tipos de modelos matemáticos contribuem para a compreensão de questões relacionadas à gestão ambiental[...].

Esse autor ainda ressalta que os “modelos ecológicos estão disponíveis, tanto para descrever interações ecológicas, quanto para avaliar as implicações da utilização de recursos” (Dale, Idem). No entanto, ainda existem deficiências na utilização desses modelos matemáticos gerados pelas ferramentas por parte de gestores e usuários no que diz respeito à compreensão dos modelos e a sua relação com determinados projetos de preservação e conservação da biodiversidade. Em outras palavras, os modelos não são utilizados tão frequentemente quanto deveriam ser.

A Figura 6 apresenta o diagrama BAM, que exemplifica que uma espécie é passível de presença, caso satisfaça três principais condições: Na região (B), uma série de espécies deve estar presente e outras ausentes, como hospedeiros, plantas, polinizadores, doenças, predadores etc. O círculo representado pelo (M) significa que a espécie tem probabilidade de estar presente somente se for acessível. A região (A) intersectada com a região (B) é a área que simboliza a expressão geográfica do Nicho Realizado e a área de intersecção entre (B) e (M) é possível visualizar a Distribuição Geográfica da Espécie (SOBERÓN; PETERSON, 2005).

Figura 6 - O diagrama apresenta uma combinação de fatores bióticos, abióticos e aonde a espécie tem acesso, para determinar a sua distribuição geográfica.

Fonte: Adaptado de SOBERÓN; PETERSON, 2005.

Soberón e Peterson (2005) afirmam que alguns fatores são determinantes para determinar a área na qual uma espécie pode ser encontrada, são eles:

- **Fatores Abióticos:** incluem aspectos do clima, ambiente físico, condições que se encontra o solo etc., estes impõem limites fisiológicos sobre a capacidade das espécies de sobreviver em determinada área.
- **Fatores Bióticos:** é o conjunto de interações com outras espécies que modificam a capacidade dessas espécies de manter populações. Essas interações podem ser positivas ou negativas e ao se limitar ou melhorar

processos populacionais, as interações podem afetar a distribuição da espécie.

- **Habilidade de Dispersão da Espécie:** é a habilidade de definir regiões que são acessíveis para dispersão da espécie a partir de alguma área original. Utiliza-se esse fator para diferenciar a distribuição real da distribuição potencial de uma espécie, tudo isso com base na configuração da paisagem e das habilidades da dispersão da espécie.
- **Capacidade Evolutiva das Populações de Espécies:** que é a habilidade que uma espécie tem de se adaptar a novas condições. Esse fator é uma consideração adicional e importante para delinear as possibilidades de distribuição de espécies.

A combinação de todos esses fatores são variáveis importantes para a então definição da distribuição geográfica de espécies. Isso também é afirmado por (Soberón e Peterson, 2005) que dizem que esses fatores interagem de forma dinâmica e com diferentes pontos fortes e fracos para produzir a definição da distribuição de uma espécie.

Por sua vez, Austin (2002) examina três componentes necessários para a modelagem estatística de distribuição das espécies: (1) um modelo ecológico sobre a teoria ecológica que está sendo usada, (2) um modelo de dados relativo à coleta de dados e (3) um modelo estatístico sobre a teoria estatística.

A modelagem de distribuição das espécies faz com que seja possível verificar as alterações na distribuição das espécies, modificações nas populações e na sua diversidade durante um determinado período. Nas pesquisas de Broennimann *et al.* (2012) e Rodder *et al.* (2011), foi utilizada a modelagem na avaliação se a distribuição potencial de uma espécie pode aumentar ou diminuir com as alterações climáticas ou se pode até mesmo ser levada à extinção.

A modelagem de distribuição de espécies, seus métodos e seus algoritmos, pode ser usada para calcular (SOBERÓN, 2014):

- Mapas de áreas de distribuição ocupados;
- Potenciais áreas de distribuição;

- Estabelecer as condições ambientais correspondentes a essas áreas.

Em maiores detalhes, Elith diz que a modelagem de distribuição de espécies tem diversos objetivos, dos quais se podem destacar os seguintes (ELITH *et al.*, 2011):

- Predição da distribuição atual, como entrada para o planejamento da conservação, a avaliações de riscos ou novas pesquisas;
- Entender os fatores ambientais correlatos de ocorrência das espécies ou grupos de espécies;
- Predição das distribuições potenciais para espécies invasoras ou explorar a expansão da distribuição;
- Predição da abundância de espécies ou diversidade;
- Predição da distribuição atual para a compreensão morfológica/ diversidade genética, endemismo e dinâmica evolutiva de nichos;
- *Handcast*, também conhecido como *backtest*, de distribuições para entender os padrões de endemismo, que é o padrão de distribuição de organismos ou vicariante, que é a separação ou divisão de um grupo de organismo por uma barreira geográfica;
- Predição de distribuição de espécies para compreender as alterações em cenários específicos de mudanças climáticas; incluindo os estudos retrospectivos.

Na modelagem de distribuição de espécies, realizada manualmente ou pelo trabalho dos algoritmos, os dados coletados e utilizados para os cálculos podem ser tanto dados chamados de presença, quanto de ausência de algumas espécies. Esse conceito será explicado na próxima seção.

2.3 Dados de presença-ausência de espécie

Na modelagem de distribuição de espécies, os registros de ocorrências são em sua grande maioria relacionados à presença de espécies, no entanto, a ausência de também é um fator a ser considerado na modelagem.

Alguns fatores determinam a existência de registros de ausência ou presença de uma espécie, esses fatores são apresentados por Peterson *et al.* (2011) e Pinaya (2013) no Tabela 1:

Tabela 1 - Fatores que determinam a ausência e a presença de espécies.

Área	Presente/Ausente	Razão
Inadequada	Ausente	Área inadequada para sobrevivência.
Inadequada	Presente	Espécies estão presentes devido à dispersão a partir de áreas adequadas.
Adequada	Ausente	Espécies não têm sido capazes de alcançar a área por meio de sua capacidade de dispersão.
Adequada	Ausente	A área era adequada e foi ocupada pela espécie em um momento, mas a espécie foi extirpada da área desde então.
Adequada	Presente	A área pode ser ocupada pela espécie, mas não existe um pesquisador que já tenha visitado o local para visitaç�o e observa�o.
Adequada	Presente	A �rea pode ser ocupada, visitada e amostrada por pesquisadores, mas eles n�o detectam a esp�cie.
Adequada	Presente	A �rea pode ser ocupada, visitada, amostrada e esp�cies podem ser detectadas por pesquisadores, mas o registro n�o est� dispon�vel.
Adequada	Presente	A �rea pode ser ocupada, ter sido visitada e amostrada, as esp�cies foram detectadas, e um registro est� dispon�vel para o pesquisador.

Fonte: Adaptado de PETERSON *et al.*, 2011 e PINAYA, 2013.

Conforme Philips *et al.* (2006, p. 195), de uma forma geral, os modelos distribidos de espcies “so ajustes a uma funo entre os pontos de ocorrncia de uma espcie e um conjunto de informaes e camadas de dados ambientais”. De acordo com (MARCO JUNIOR; SIQUEIRA, 2009, p. 66) “como em geral s esto disponveis dados de presena, essas funes devem representar as caractersticas ambientais nos pontos de ocorrncia ou o nicho da espcie”.

Discussões sugerem que dados apenas de presença, em um sentido, libertam a problemática de registros de ausência não confiáveis, especialmente enfatizando que as ausências podem suportar impressões de interações bióticas, restrições de dispersão e distúrbios que podem impedir a modelagem de distribuição potencial de uma espécie (ELITH *et al.*, 2011).

No entanto, os registros de presença também influenciam de alguma maneira informações de ausência. Elith *et al.* (2011, p. 45) afirmam que se uma espécie está ausente de uma área ambiental apropriada, por exemplo, perturbações passadas causaram extinções locais, logo o sinal de ausência também será encontrado na distribuição dos registros de presença.

Outro detalhe com relação ao uso de dados de presença é que o registro de uma espécie pode variar de um local para o outro e isso pode gerar informações de falsa ausência de registro de uma espécie. Portanto, baseado em Elith *et al.* (2011), Pinaya (2013, p. 26) afirma que “a simples supressão de registros de ausência não soluciona as limitações nos registros de presença, causadas pelo fato de que as espécies não são perfeitamente detectáveis, e não podem ocupar toda a área do habitat adequado”.

Como solução para a utilização de dados de presença/ausência de espécies, Vanderwal *et al.* (2009, p. 589) afirmam que uma apropriada seleção de pseudoausência ou informações de *background* são essenciais para a modelagem de distribuição de espécies. Isso também é afirmado por Philips, que diz que a maioria dos métodos de MDE, realizados a partir de dados de ocorrência, necessita de dados adicionais que definam ou apresentem um conjunto de informações ambientais das espécies em estudo (PHILIPS *et al.*, 2009).

Portanto, como elemento vital para a modelagem de distribuição de espécies, a correta definição e utilização dos dados de ocorrência, tanto de presença/ausência, como da utilização de dados de *background*, devem fazer parte das informações de biodiversidade dos modelos gerados pelas ferramentas de MDE, e que a interoperabilidade dos dados de modelos gerados devem conter essas informações que influenciam nos resultados finais.

Visto os conceitos de modelagem de distribuição de espécies, na próxima seção serão apresentadas as abordagens sobre as ferramentas de modelagem de distribuição de espécies.

2.4 As ferramentas para modelagem

Uma vez que os recursos humanos e financeiros para o meio ambiente geralmente são escassos, frente à demanda atual, ferramentas de modelagem podem contribuir para a conservação ambiental. Tais instrumentos podem ser aplicados, por exemplo, na identificação de áreas com alto risco de perda da biodiversidade, possibilitando que os recursos sejam direcionados para essas regiões e suas espécies ameaçadas (RODRIGUES, E. S. C., 2012).

O uso de ferramentas de MDE tem facilitado a definição de distribuição geográfica de espécies por muitos pesquisadores e cientistas da área e também por usuários que não possuem tantos conhecimentos técnico/científicos da área biológica o que, segundo Peterson e Soberón (2012), tem infelizmente diminuído o genuíno potencial dessas ferramentas e causam desconfiança e incompreensão por parte de toda a comunidade científica de biodiversidade.

Conforme visto no início deste capítulo, no ciclo de vida dos dados de biodiversidade, as ferramentas de modelagem de distribuição de espécies se encaixam na fase de coleta dos dados, pois apesar de a modelagem já trabalhar com dados coletados, no processo de modelagem existe a coleta dos dados para geração de informação, que garante, descreve, preserva, descobre, integra e permite a análise dos dados gerados pelos modelos. Dessa maneira, o ciclo de vida dos dados de biodiversidade pode ser revisto na Figura 7.

Figura 7 – Ciclo de Vida dos dados de biodiversidade.

Fonte: Adaptado de DATAONE, 2016.

Dentre as ferramentas existentes para a modelagem de distribuição de espécies, pode-se destacar as seguintes: MaxEnt, DesktopGarp, openModeller, BIOMOD, DIVA-GIS, Bioensembles, entre outras.

MaxEnt⁷ é uma ferramenta de modelagem de distribuição geográfica das espécies que utiliza o método de entropia máxima. De acordo com Rodrigues, E. S. C. (2012, p. 48) “os métodos baseados em entropia máxima produzem modelos que não são funções de probabilidade de ocorrência das espécies, mas uma distribuição de probabilidade, isto é, o problema é estimar a distribuição com máxima entropia”.

Outra ferramenta é o DesktopGARP⁸, que é um pacote de software que permite ao usuário prever e analisar a distribuição das espécies. É uma versão *desktop* do algoritmo GARP (*Genetic Algorithm for Rule Set Production*) que cria modelos de nicho ecológico das espécies, descrevendo as condições ambientais em que as espécies devem ser capazes de manter populações (SCACHETTI-PEREIRA, 2002).

⁷ A ferramenta MaxEnt está disponível em <http://homepages.inf.ed.ac.uk/lzhang10/maxent.html>

⁸ A ferramenta DesktopGARP está disponível em <http://www.nhm.ku.edu/desktopgarp/>

Essas ferramentas citadas, MaxEnt e DesktopGARP são usadas em diversos centros de estudo e pesquisa na área de modelagem, no entanto, essas e outras ferramentas são desenvolvidas com o objetivo de modelar distribuição de espécies baseadas em apenas um algoritmo, entropia máxima e GARP respectivamente. No entanto, não existe um único algoritmo que possa atender todas as necessidades do pesquisador em biodiversidade, portanto, cresce a demanda por ferramentas de modelagem que integrem diferentes algoritmos, das quais podemos citar o openModeller⁹ (SOUZA MUÑOZ *et al.*, 2011) e (ELITH *et al.*, 2006).

A ferramenta openModeller¹⁰ fornece à comunidade científica um conjunto, robusto e flexível, de estratégias relacionadas à modelagem de distribuição de espécies e dentre duas vantagens destaca-se a união de vários algoritmos de modelagem em uma única arquitetura, como *Artificial Neural Networks* (ANN), *AquaMaps*, *Bioclim*, *Climate Space Model* (CSM), *Ecological-Niche Factor Analysis* (ENFA), *Envelope Score*, *Environmental Distance*, GARP, *GARP Best Subsets*, *Random Forests* e *Support Vector Machines* (SVN).

A ferramenta OpenModeller faz o acesso a portais de base de dados de biodiversidade utilizando padrões de metadados como Darwin Core (Wieczorek *et al.*, 2012). Os portais são repositórios de dados que integram várias coleções, observações e ocorrências de espécies disponíveis pelo mundo, eles permitem diferentes análises de dados sobre biodiversidade. Exemplos desses portais: GBIF (*Global Biodiversity Information Facility*) (GBIF, 2015) de abrangência mundial e IABIN (*Inter-American Biodiversity Information Network*) (IABIN, 2015).

O pacote de software BIOMOD¹¹ é também uma plataforma para realizar a previsão de distribuições de espécies, permitindo também o tratamento de uma série de incertezas metodológicas em modelos e do exame das relações espécie x ambiente (BIOMOD, 2015).

⁹ A ferramenta openModeller está disponível em <http://openmodeller.sourceforge.net/>

¹⁰ Por ser uma ferramenta com muitos recursos, o *openModeller* tem chamado a atenção de instituições como a NASA (*National Aeronautics and Space Administration*) e está sendo também utilizada por pesquisadores de muitos países, como Estados Unidos (*The Kansas University, California University at Berkeley*), Inglaterra (*Oxford University*), Itália, Taiwan, dentre outros (Rodrigues, E. S. C., 2012).

¹¹ O pacote de software BIOMOD está disponível em <http://www.will.chez-alice.fr/Software.html>

O pacote BIOMOD é implementado na linguagem R e inclui a capacidade de modelar distribuição de espécies utilizando várias técnicas e modelos de testes com uma vasta gama de abordagens e com diferentes condições ambientais (cenários climáticos, mudanças no uso da terra etc.) (BIOMOD, 2015).

A ferramenta DIVA-GIS¹² foi definida por Hijmans *et al.* (2012, p. 3) “como um programa de computador de acesso livre que serve para o mapeamento e análise de dados espaciais”. É particularmente útil para a análise da distribuição de espécies para elucidar os padrões geográficos e ecológicos.

De acordo com Hijmans *et al.* (2012, p. 2)

DIVA-GIS suporta vetores, imagens, e tipos de dados em *grid*. Ele pode melhorar a qualidade dos dados através da procura por coordenadas de localidades usando dicionários e checando coordenadas existentes usando consultas espaciais de coleções nos bancos de dados administrativos. Para calcular a modelagem de nicho ecológico, a ferramenta disponibiliza os algoritmos DOMAIN e Bioclim[...].

No trabalho de Araújo (2015) é apresentada a ferramenta Bioensembles como uma plataforma para modelagem de distribuição de espécies que inclui 13 diferentes técnicas de modelagem. A primeira citação da ferramenta Bioensembles foi apresentada por Diniz; Filho *et al.* (2009).

No artigo apresentado por Elith *et al.* (2006), intitulado “*Novel methods improve prediction of species distribution from occurrence data*”, através da pesquisa apresentada é possível concluir que um único algoritmo não atende todas as necessidades dos pesquisadores da área, logo, uma única ferramenta também não dá as soluções necessárias, e, portanto, garantir uma interoperabilidade entre os modelos gerados pelas ferramentas de modelagem de distribuição de espécies, possibilita aos pesquisadores da área de *Biodiversity Informatics*, gerar e disponibilizar informações de maior precisão e confiabilidade.

Nesta seção foram apresentadas algumas das ferramentas utilizadas para a modelagem de distribuição de espécies e cada uma delas apresenta características e funcionalidades diferentes que têm o mesmo objetivo de auxiliar a tomada de decisão com relação à conservação da biodiversidade.

¹²A ferramenta DIVA-GIS está disponível para download no site: <http://www.diva-gis.org/download>

A seguir serão apresentados alguns dos principais algoritmos de modelagem de distribuição de espécies. Essa parte da pesquisa visa mostrar as principais características de cada algoritmo para que na seção de desenvolvimento do perfil de aplicação (capítulo 4), possam ser discutidos que objetos farão parte do estudo de caso para interoperabilidade entre ferramentas de modelagem.

2.4.1 As abordagens de algoritmos para modelagem

Aspectos conceituais sobre algoritmos de modelagem são abordados nesta pesquisa pela importância que eles têm mediante o processo de modelagem de distribuição de espécies, e principalmente pela importância em conhecer a maneira que determinado algoritmo processa dados de biodiversidade, pois isso reflete diretamente no modelo gerado pelas ferramentas de modelagem e, portanto, deve ser estudado para que ocorra uma correta interpretação dos dados gerados e posteriormente reutilizados e disponibilizados como proposta desta tese.

De acordo com Rodrigues, F. A. (2012, p. 31),

os algoritmos de modelagem são responsáveis pela tarefa de inferência do processo de modelagem, cujo objetivo é encontrar uma função de probabilidade que representa a relação entre os dados previamente conhecidos (dados de entrada) e um valor de saída esperado para cada dado (classe). Os dados de entrada são parâmetros de um conjunto de vetores com característica ambientais de um determinado ponto de ocorrência[...].

Para o processamento dos algoritmos de modelagem de distribuição de espécies, apenas duas opções de dados são utilizadas. Quando existe um ponto de presença de uma espécie é utilizado o identificador “1”, ao contrário, como ponto de ausência, o identificador “0” é representado, o que varia de algoritmo para algoritmo é a inferência lógica adotada sobre o ponto de presença ou ausência.

Alguns algoritmos utilizados atualmente na modelagem são baseados em técnicas computacionais da área de aprendizagem de máquina, nas quais se destaca a capacidade de aprendizado através de exemplos, ou seja, adaptar conhecimento a partir de dados e observações já realizadas (RODRIGUES, F. A., 2012).

De acordo com Guo *et al.* (2005, p. 77),

com o avanço da eficiência computacional combinado com métodos estatísticos sofisticados, os métodos de aprendizado de máquina têm sido cada vez mais utilizados e se mostrado como ferramentas poderosas em uma ampla variedade de disciplinas científicas, incluindo as áreas de ciência da computação e ciências ambientais[...].

A definição de um algoritmo para realização da MDE depende do problema a ser resolvido e dos dados de biodiversidade selecionados. Soberón e Peterson (2005) confirmam que a principal atividade dos algoritmos de modelagem é a de encontrar regiões onde os valores das variáveis ambientais (camadas) utilizadas na entrada para o processo de modelagem sejam semelhantes aos valores das variáveis dos pontos de ocorrência de espécies fornecidos para o processamento.

Soberón e Peterson (2005, p. 4) ainda afirmam que “a qualidade do modelo gerado através do algoritmo utilizado depende da eficiência de representação do nicho abiótico e da capacidade de extrapolação do algoritmo utilizado”.

Os principais algoritmos de modelagem encontrados em utilização, por parte dos pesquisadores e pelas ferramentas de modelagem, serão citados e comentados nos tópicos a seguir. A definição dos algoritmos e a forma como cada um é utilizado é importante para esta pesquisa, pois para interoperabilidade dos dados dos modelos gerados pelas ferramentas de MDE, as informações dos algoritmos utilizados, bem como os parâmetros, são importantes para a composição de metadados desses modelos.

2.4.1.1 Entropia Máxima

No trabalho de Philips *et al.* (2004), foi proposta a aplicação da Entropia Máxima (MaxEnt) para a MDE e afirmaram que em MaxEnt é dado um conjunto de amostras a partir de uma distribuição sobre algum espaço, bem como um conjunto de características desse espaço. A ideia da MaxEnt é estimar a distribuição alvo através da distribuição da Entropia Máxima, sujeita à restrição de que o valor esperado de cada recurso corresponda à sua média empírica.

Philips *et al.*, (2006, p. 234) declaram que MaxEnt é um “método de propósito geral para elaborar predições ou inferências a partir de informações incompletas”. Nesse artigo os autores apresentam algumas vantagens do uso da Entropia Máxima, das quais se destacam as seguintes características:

1. Para determinar a área de estudo, MaxEnt requer apenas dados de presença, juntamente com a informação ambiental;
2. Pode-se utilizar tanto dados contínuos, quanto categóricos e podem incorporar as interações entre diferentes variáveis;
3. Algoritmos determinísticos eficientes foram desenvolvidos que garantem a distribuição da probabilidade ótima;
4. A distribuição de probabilidade MaxEnt tem uma definição matemática concisa, e portanto, permite análise;
5. MaxEnt também pode ser aplicado aos dados de espécie de presença/ausência usando um modelo condicional;
6. MaxEnt tem uma abordagem gerativa, ao invés de discriminativa, que pode ser uma vantagem inerente quando a quantidade de dados de treinamento são limitados;
7. A modelagem de Entropia Máxima é uma área ativa de pesquisa em estatística e aprendizagem de máquina;
8. Como um método geral estatístico flexível, ele pode ser utilizado para todas as aplicações e em todas as escalas.

Portanto, MaxEnt é um modelo estatístico e para aplicá-lo para a modelagem de distribuição de espécies com sucesso, deve-se considerar como ele se relaciona com outros componentes de modelagem (modelo de dados e modelo ecológico) (PHILIPS *et al.*, 2006).

2.4.1.2 GARP

O algoritmo GARP (*Genetic Algorithm for Rule-set Production*) foi descrito pela primeira vez por Stockwell e Noble (1992, p. 385) com a seguinte afirmação: “sistemas de modelagem devem ter pelo menos três características: suporte à predição, exploração, e explicação”, e por isso o GARP foi desenvolvido com o

objetivo de descrever regras para essas tarefas. Além disso, o conjunto de regras proposto pelos autores tem duas vantagens que definem um conjunto de regras robusto (importante para estabilizar perturbações de dados) e um conjunto de regras informativo, permitindo modelar informação complexa a partir de parâmetros simples.

2.4.1.3 Redes Neurais

As Redes Neurais Artificiais ou simplesmente Redes Neurais são uma área de estudo da Inteligência Artificial (IA) que, segundo Russell e Norving (2004, p. 713), a partir do estudo de um “neurônio que é uma célula no cérebro cuja principal função é coletar, processar e disseminar sinais elétricos”, a IA busca desde 1943 modelos detalhados e realistas do processamento do cérebro para levar ao campo moderno da neurociência computacional.

2.4.1.4 Máquinas de Vetores de Suporte

A técnica de Máquina de Vetores de Suporte do inglês (*Support Vector Machines* - SVMs) constitui uma técnica primeiramente elaborada e embasada na Teoria de Aprendizado Estatístico (VAPNIK, 1995).

De acordo com Guo *et al.* (2005, p. 77), da Universidade da Califórnia em Berkeley, entre muitos métodos de aprendizagem de máquina, “os algoritmos de máquinas de vetores de suporte são considerados como uma nova geração de algoritmos de aprendizagem”.

As SVMs têm várias características atraentes para a modelagem de distribuição, que incluem: (1) modelos com base estatística, ao invés de soltas com sistemas de aprendizagem natural e (2) teoricamente garantem uma ótima performance (GUO *et al.*, 2005).

2.4.1.5 Outros Algoritmos de Modelagem

Com o avanço da área de *Biodiversity Informatics*, os algoritmos existentes foram sendo modificados para que os seus desempenhos também sofressem uma melhoria. Muitas dessas variações e também a comparação com os principais algoritmos geraram outras vertentes e características, proporcionando aos cientistas e pesquisadores uma vasta gama de algoritmos para modelagem de distribuição de espécies.

O algoritmo Bioclim, foi citado por Henry Nix no artigo intitulado “*A biogeographic analysis of Australian elapid snakes*” em 1986 e é desenvolvido utilizando o conceito de envelope bioclimático, no qual, considerando os pontos de ocorrência para a modelagem, o algoritmo calcula a média e o desvio padrão para cada variável ambiental de entrada (BOOTH *et al.*, 2014).

Em comparação ao algoritmo Bioclim, Carpenter *et al.* (1993) apresentaram à comunidade científica o algoritmo DOMAIN, que é um procedimento que usa uma métrica de proximidade ponto-a-ponto para atribuir um valor de classificação para um habitat favorável baseado na proximidade do ambiente com outro mais similar, registrado anteriormente.

DOMAIN não define nenhum limite discreto para os envelopes climáticos, para os autores, para todos os pontos candidatos são atribuídos valores de similaridade e intervalos limiares definidos pelo usuário para determinar os intervalos reais mapeados (CARPENTER *et al.*, 1993).

Outros algoritmos também são mencionados por Wisz *et al.* (2008) como GBM (*Generalized Boosting Methods*), e o MARS (*Multivariate Adaptive Regression Splines*), algoritmos que trabalham com dados de presença, mas com alguma forma de tratamento para dados de ausência. No experimento apresentado por Wisz *et al.* (Idem), GBM foi o melhor algoritmo, comparado com MaxEnt, MARS, GARP etc., no que diz respeito a amostras de dados de tamanho entre 30 e 100 registros.

3 INTEROPERABILIDADE APLICADA À INFORMÁTICA PARA BIODIVERSIDADE

Conforme apresentado no Capítulo 2, e seguindo a justificativa e objetivos deste trabalho, a modelagem de distribuição de espécies através dos modelos gerados por suas ferramentas demandam uma interoperabilidade entre estes modelos para que a comunidade científica da área agregue ao seu poder de tomada de decisões para a conservação da biodiversidade, características de reúso e padronização da disponibilidade de informações geradas pelas pesquisas na área de MDE.

3.1 Interoperabilidade de Dados

O conceito de interoperabilidade é amplo e permite sua aplicação em muitas áreas da ciência. No dicionário on-line Priberam (DPLP, 2013), a palavra interoperabilidade vem da definição e da qualidade do que “interoperável”, que significa o que é capaz de operar, funcionar ou atuar com outro”, no entanto, o seu conceito mais abrangente vem da área computacional, no qual será apresentado nos próximos parágrafos.

O dicionário *Online* para Biblioteconomia e Ciência da Informação, do inglês *OnLineDictionary for Library and Information Science*, Reitz (2004) define interoperabilidade como sendo

a capacidade de um sistema de informática de hardware ou software de se comunicar e trabalhar eficazmente com outro sistema na troca de dados, normalmente um sistema de um tipo diferente, concebido e produzido por um fornecedor diferente[...].

Vernadat (1996, p. 25) define interoperabilidade como sendo a “habilidade de se comunicar com sistemas pares e acessar a funcionalidade desses sistemas. Estabelecer interoperabilidade significa relacionar dois sistemas juntos e remover qualquer incompatibilidade entre eles”.

Sayão e Marcondes (2008 p. 136) dão esta conceituação ao termo: “para a área de tecnologia da informação, há um consenso geral de que interoperabilidade é

algo como a capacidade de computadores e programas de fabricantes diferentes trocarem informações”. Ainda acrescentam (Idem, p. 137-138) que a interoperabilidade pode ter muitas “faces”, sendo que a mais conhecida é a técnica, normalmente relacionada à informática, portanto, além dessa, esse artigo apresenta outras, que seguem:

- a) Interoperabilidade técnica: esforços técnicos para assegurar interoperabilidade de comunicação, transporte, armazenamento e visualização de dados. Existem organizações que têm como objetivo assegurar essa interoperabilidade, como por exemplo: W3C¹³ (*World Wide Web Consortium*), ISO¹⁴ (*Internacional Organization for Standartization*), entre outras;
- b) Interoperabilidade semântica: são esforços para assegurar interoperabilidade “das informações originadas de diferentes recursos e é solucionada pela adoção de ferramentas comuns ou/e mapeáveis de representação da informação, como esquemas de metadados, classificações, tesouros e mais recentemente, ontologias”;
- c) Interoperabilidade política/humana: além das organizações e comunidades de desenvolvimento para interoperabilidade de dados é imprescindível o apoio político e humano para que exista uma parceria entre os esforços realizados por diferentes frentes para que uma interoperabilidade seja possível;
- d) Interoperabilidade intercomunitária: com o crescimento da interdisciplinaridade entre as comunidades científicas, é necessário um esforço para assegurar interoperabilidade também em um aspecto mais abrangente entre diferentes comunidades com áreas de conhecimento distintas;
- e) Interoperabilidade legal: “considera as exigências e as implicações legais de tornar livremente disponíveis itens de informação”;
- f) Interoperabilidade internacional: esse aspecto busca trabalhar com esforços internacionais em busca da interoperabilidade, que podem sofrer

¹³Maiores informações sobre as iniciativas de interoperabilidade da W3C, podem ser acessadas pelo endereço <https://www.w3.org/>.

¹⁴A ISO hoje é uma das maiores e mais respeitadas organizações controladoras de padrões ao redor do mundo. Maiores informações podem ser acessadas no endereço <http://www.iso.org/iso/home.html>

barreiras linguísticas, culturais e, portanto, necessitam de um apoio globalizado das comunidades interessadas.

Conforme Blair *et al.* (2011, p. 3), a interoperabilidade refere-se à “capacidade de dois ou mais sistemas, potencialmente desenvolvidos por diferentes frentes de trabalho, trabalhos em conjunto, incluindo a capacidade de trocar e interpretar pedidos de ação e requisitar conjuntos de dados associados”.

A interoperabilidade, em outras palavras pode ser definida como o nível de compatibilidade ou o nível de comunicação que um sistema tem com outro, em apoio a isso, a IEC TC 65/290/DC em (IEC, 2002) identifica níveis de compatibilidade, que podem variar de acordo com o nível de comunicação com os recursos da aplicação ou do dispositivo, conforme visto na Figura 8.

Figura 8 – Níveis de compatibilidade.

Fonte: Adaptado de IEC, 2002.

A IEEE (*Institute of Electrical and Electronics Engineers*), em seu dicionário de termos tecnológicos (IEEE, 2000), divide o conceito de interoperabilidade em quatro principais categorias, são elas:

1. A habilidade de dois ou mais sistemas ou elementos de sistemas de trocarem informação e a habilidade do uso dessa informação que foi trocada;
2. A capacidade de unidades de equipamentos trabalharem juntas com o objetivo de alcançarem funções úteis;
3. A capacidade promovida pela conformidade conjunta de um determinado conjunto de normas e padrões que permite que tecnologias heterogêneas trabalhem em conjunto por intermédio de um ambiente de rede;
4. A habilidade de dois ou mais sistemas ou componentes de sistemas de trocar informações em uma rede heterogênea e ser hábil na utilização dessa informação.

Interoperabilidade, portanto, é a capacidade que um sistema possui de se comunicar de forma transparente com outro sistema, através de recursos ontológicos e/ou padrões de dados.

De acordo com Mason e Galatis (2007, p. 51) “práticas têm mostrado que em um nível mínimo de interoperabilidade é assegurado através do uso de elementos comuns como título, autor, identificação, descrição e ou palavras-chave”.

Já o Comitê Europeu de Normatização (CEN) afirma que interoperabilidade, por ela mesma, tem sido definida de inúmeras maneiras, mas no contexto de metadados, ela usualmente é citada como a habilidade de um sistema processar instâncias de metadados produzidas por um sistema terceiro (CEN, 2006).

Carrier (2008, p. 6) menciona que “a importância da interoperabilidade, e particularmente como metadados podem melhorar interoperabilidade tem sido reconhecido em inúmeras iniciativas”, como verifica-se no próprio projeto relatado pela autora na sua dissertação.

A interoperabilidade tem sido uma preocupação atual para a área de Informática para Biodiversidade, pois Dubois *et al.* (2013, p. 2) explicam que os dados de biodiversidade “são enormes em seu âmbito espaço-temporal e dimensão, enquanto, ao mesmo tempo, eles são muitas vezes documentados e geridos de uma forma muito fragmentada e inconsistente”.

Ainda sobre a interoperabilidade e a importância dos metadados para tal, no âmbito das bibliotecas digitais, Castro e Santos (2007, p.13) afirmam que as

bibliotecas digitais, como ambientes propícios para a recuperação de informações, têm na utilização de metadados a padronização das formas de representação e a possibilidade de garantia de interoperabilidade entre sistemas, favorecendo a integridade e a acessibilidade dos recursos informacionais de forma eficiente pelo usuário final[...].

Na próxima seção, serão detalhados os conceitos sobre os padrões de metadados, bem como os mais utilizados na área de Informática para Biodiversidade. Também serão associados os conceitos de interoperabilidade às soluções para a problemática envolvendo os modelos gerados pelas ferramentas de MDE e a necessidade em produzir informações interoperáveis.

3.2 Padrão de Metadados

Metadados podem ser definidos como dados sobre dados, segundo Hillmann (2005), e um registro de metadados é composto por um “conjunto de atributos, ou elementos necessários para descrever o recurso em questão”. A autora ainda explicita que apesar do uso do conceito de metadados ser anterior à internet, o interesse mundial em normas e práticas de metadados “explodiu” com o aumento da edição eletrônica de bibliotecas digitais e ao que ela chama de “sobrecarga de informação”.

Conforme Weibel e Lagoze (1997, p. 176) “a associação de metadados descritivos e padronizados com objetos em rede, tem o potencial de melhorar substancialmente a capacidade de descoberta de recursos, permitindo pesquisas e indexação de objetos não textuais”.

Os autores Mason e Galatis (2007, p. 51) afirmam que “metadados serão associados com tudo que se move, suportando múltiplas plataformas, vários tipos de objetos de informação, coleções, serviços, pessoas, organizações, locais, termos, formatos, direitos, etc”.

Dentre as mais respeitadas organizações para a criação, divulgação e manutenção de metadados está a DCMI¹⁵ (*Dublin Core Metadata Initiative*). Ela mesma define-se como sendo uma “organização aberta que apoia a inovação no *design* de metadados e as melhores práticas em toda a ecologia de metadados”. Na página a internet da organização, eles descrevem que as atividades da DCMI “incluem trabalhos sobre arquitetura e modelagem, discussões e trabalho colaborativo em comunidades DCMI e grupos de trabalho DCMI, conferências globais¹⁶, reuniões e workshops e esforços educacionais para promover a aceitação generalizada de padrões de metadados e melhores práticas”.

Nesse contexto, os dados de biodiversidade se encaixam perfeitamente nesse conceito, sendo de grande importância a utilização de padrões de metadados para publicação, indexação, visualização.

Wieczorek *et al.* (2012, p. 1) afirmam que

dados sobre biodiversidade derivam de fontes inumeráveis, armazenados em diversos formatos e disponíveis em diversas plataformas de hardware e software distintamente. Um passo essencial para a compreensão dos padrões globais de biodiversidade é fornecer uma visão padronizada destas fontes de dados heterogêneos para melhorar a interoperabilidade que é fundamental, para esse avanço, definir termos comuns[...].

Metadados, portanto são as informações que descrevem “quem, o quê, onde, quando, por que, e como” um conjunto de dados ecológicos foi recolhido. Metadados são simplesmente dados sobre dados (Fegraus *et al.*, 2005).

¹⁵Todas as informações relacionadas ao Dublin Core Metadata Initiative podem ser encontradas no site oficial do órgão: <http://dublincore.org>

¹⁶ As conferências administradas pela DCMI são realmente bem abrangentes, no âmbito de comunidades científicas, com a oportunidade de participar em duas conferências, é possível ver o dinamismo que padrões de metadados são utilizados em diversas áreas, e também como estes têm contribuído com a ciência. Os trabalhos publicados relacionados a esta dissertação podem ser encontrados em (Borba e Corrêa, 2014), (Silva *et al.*, 2015) e (Borba e Corrêa, 2015).

Michener *et al.* (1997, p.330) argumentam que

o valor a longo prazo, a utilidade de dados ecológicos para o avanço e compreensão ecológica, e a solução de importantes problemas ambientais dependem da disponibilidade de metadados apropriados e adequados, ou informações descritivas que apresentam o conteúdo de dados, contexto, qualidade, estrutura e acessibilidade de uma informação ecológica[...].

Os autores em Feagraus *et al.* (2005, p. 159) ainda sustentam que a maioria dos ecologistas teve dificuldade para lembrar

detalhes importantes sobre os seus próprios dados, mesmo depois de apenas alguns meses, e a menos que os dados sejam devidamente documentados, essa dificuldade só aumenta com o tempo e que mesmo a análise mais simples requer algum nível de metadados[...].

Um exemplo simples da utilidade dos padrões de metadados pode ser visto nas Tabelas 2 e 3. Na Tabela 2 são apresentados dados aleatórios que não possuem padrões de metadados definidos.

Tabela 2 - Dados ecológicos com informações limitadas sobre os dados coletados.

AEDGY	15032015	-23.550520	-46.633309
AEDGY	16032015	-23.550530	-46.633319
Aedes aegypti	15032015	-23.550540	-46.633329
Coleoptera	01012014	-23.550550	-46.633339

Fonte: Elaborado pelo autor.

A Tabela 3 ilustra os mesmos dados da Tabela 2, no entanto com informações adicionais.

Tabela 3 - Dados ecológicos com informações adicionais.

Sp	Data	La	Lo
AEDGY	15032015	-23.550520	-46.633309
AEDGY	16032015	-23.550530	-46.633319
Aedes aegypti	15032015	-23.550540	-46.633329
Coleoptera	01012014	-23.550550	-46.633339

Fonte: Elaborado pelo autor.

Analisando as Tabelas 2 e 3, é possível verificar que ocorreu uma melhora no que diz respeito ao entendimento das informações coletadas, entretanto, ainda são limitadas. Informações relevantes como: coletor, dados da pesquisa, disponibilização dos dados etc., ainda são dados que estão ocultos nessas tabelas. Nesse contexto, a definição de um padrão de metadados para a pesquisa de dados ecológicos é de grande importância, pois define vocabulários e informações necessárias para a disponibilização, replicação e reutilização desses dados.

Nesse contexto e pela importância dos padrões de metadados para a compreensão, disponibilização e processamento de informações ecológicas por ferramentas tecnológicas e computacionais, alguns padrões de metadados aplicados para a área de biodiversidade e ecologia são apresentados a seguir.

3.2.1 Dublin Core

O padrão de metadados Dublin Core (DC) é definido por Hillmann (2005) como um “elemento simples, mas eficaz para descrever uma ampla gama de recursos”. Ele pode ser dividido em dois níveis, simples e qualificado. O nível simples compreende um conjunto de 15 elementos e o nível qualificado inclui três elementos adicionais, bem como um grupo de refinamentos para garantir uma semântica completa e útil para diversos elementos. A semântica do Dublin Core foi estabelecida por profissionais de biblioteconomia, ciência da computação, processamento de linguagem e outras áreas afins.

O Dublin Core tem quatro principais objetivos (HILLMANN, 2005):

- a) **Simplicidade de criação e manutenção:** o conjunto de elementos Dublin Core foi mantido pequeno e simples para permitir, quanto possível, que um não especialista crie registros simples com recursos de informação de fácil acesso e de baixo custo;
- b) **Semântica comumente entendida:** Dublin Core auxilia um pesquisador não especialista a encontrar “seu caminho” através do apoio comum de elementos por intermédio de uma semântica universalmente compreendida e apoiada;

- c) **Escopo internacional:** o escopo foi inicialmente definido na língua inglesa, no entanto versões estão sendo criadas em diversas línguas como: português, alemão, japonês, finlandês, norueguês, grego, indonésio e espanhol;
- d) **Extensibilidade:** Com o objetivo de descrever recursos digitais e com a necessidade de recuperação de informação, o Dublin Core tem reconhecido a importância de fornecer um mecanismo para estender os elementos do DC.

O padrão de metadados Dublin Core tem um modelo que permite que diferentes comunidades usem os seus elementos, permitindo extensões específicas de domínio que fazem sentido em uma área mais limitada de atuação (HILLMANN, 2005).

3.2.2 Darwin Core

Baseado no padrão Dublin Core, por volta do ano de 1999, o grupo de pesquisa TDWG (*Taxonomic Databases Working Group*) desenvolveu e mantém até hoje o padrão Darwin Core, especificamente para uma padronização dos dados de biodiversidade, oferecendo assim uma norma para pesquisadores da área. A filosofia principal do grupo é manter o padrão mais simples de acessibilidade, no entanto, permitindo uma elasticidade de termos de acordo com a necessidade de cada pesquisa.

Segundo Wieczorek *et al.* (2015), no website do TDWG, Darwin Core (DwC) é

um manual de normas que inclui um glossário de termos, conceitos, atributos ou propriedades que se destina a facilitar o compartilhamento de informações sobre a biodiversidade, fornecendo, portanto, definições que servem de referência. O Darwin Core se baseia na taxa de ocorrência de espécies na natureza através de documentos adquiridos através de observação e coleta de espécimes, amostras e informações relacionadas[...].

Em sua concepção inicial, o Darwin Core tinha o objetivo apenas de facilitar a descoberta, recuperação e integração de dados sobre espécimes biológicos modernos, sua ocorrência no espaço temporal e seus elementos de prova alojada

em coleções, tanto física quanto digital (WIECZOREK *et al.*, 2015). Entretanto, de acordo com a organização TDWG, no mesmo documento, os autores afirmam que hoje, o padrão se destina a fornecer definições semânticas estáveis, com o objetivo de ser maximamente reutilizáveis em uma variedade de contextos.

Wieczorek *et al.* (2012, p. 2) afirmam que o padrão Darwin Core “tem sua semântica bem definida que pode ser compreendida por pessoas ou interpretada por máquinas, o que torna possível determinar os usos adequados dos dados neles codificados”.

A Tabela 4 apresenta os principais termos que estruturam o padrão Darwin Core. Destes, sete instituem o chamado Darwin Core Simple, que é o mínimo para uma informação padrão, e dois termos que permitem aos pesquisadores a inclusão de novas e adicionais informações não previstas no Darwin Core Simple, permitindo que o Darwin Core tenha um uso mais amplo de suas características.

Tabela 4 - Categoria dos sete termos do Darwin Core Simple e dos termos do Darwin Core Genérico.

Termos de Nível de Registro	Termos do Dublin Core, instituições, coleções, registro de dados da natureza.	Darwin Core Simple
Ocorrência	Evidência de espécies na natureza, observação, mídia associada, referências e comportamento.	
Evento	Protocolos de amostragem e métodos, data, hora, notas de campo.	
Localização	Geográfica, descrição local, dados espaciais.	
Identificação	Conexão entre Táxons e Ocorrências.	
Taxonomia	Nomes científicos, nomes utilizados, conceitos taxonômicos e a relação entre eles.	
Contexto Geológico	Tempo geológico, cronoestratigrafia, bioestratigrafia, litoestratigrafia.	
Pesquisas Relacionadas	Relacionamentos explícitos entre pesquisas identificadas.	Darwin Core Genérico
Fato ou Medição	Medição, fatos, características, asserções e referências.	

Fonte: Adaptado de WIECZOREK *et al.*, 2012.

3.2.3 Darwin Core A (DwC-A)

O padrão de metadados Darwin Core Archives ¹⁷(DwC-A) foi apresentado por (GBIF, 2010) como uma norma para *Biodiversity Informatics*, internacionalmente reconhecida, que simplifica a publicação de dados sobre biodiversidade e foi desenvolvido com base no padrão Darwin Core, apresentado na seção anterior.

Berendsohn *et al.* (2011, p.74) explicam que

a diferença entre os dois padrões, DwC e DwC-A, é que o segundo trata-se de uma versão estendida que não se limita aos dados de ocorrência em si, mas também abrange nomes de organismos, informações taxonômicas de espécies, dados factuais, distribuições de espécies, mídia e literatura[...].

E segundo Robertson (GBIF, 2010, p. 2),

compartilhando conjuntos de dados inteiros no padrão de metadados DwC-A ao invés de usar serviços disponibilizados na web como TAPIR¹⁸ ou DiGIR¹⁹ permite de uma maneira muito mais simples e mais eficiente transferência de dados. Por exemplo, a recuperação de 260 mil registros via TAPIR podem levar cerca de 9 horas de processamento e envolve a emissão de 1300 solicitações HTTP para transferir 500MB de dados no formato XML, o mesmo conjunto de dados, quando codificado como DwC-A se torna em um arquivo zipado de 3 MB[...].

Outras características do padrão de metadados DwC-A é que para produzir um arquivo nesse formato não é necessário instalar qualquer software de edição de dados, o que o torna uma opção de fácil uso.

A ideia geral do padrão é que os seus arquivos de dados são logicamente organizados de uma maneira semelhante ao que os autores chamam de “estrela”, com um arquivo de dados com um núcleo rodeado por um número qualquer de arquivos de extensão. O arquivo principal, ou núcleo, e o arquivo de extensão contêm registros de dados em cada linha e cada registro de extensão aponta para um registro no arquivo principal, isso garante que muitos registros de extensão possam existir para cada registro no arquivo principal (Ver Figura 9) (GBIF, 2010).

¹⁷ É o formato recomendado para publicação no portal de biodiversidade GBIF (GBIF, 2010).

¹⁸ TAPIR - http://www.tdwg.org/dav/subgroups/tapir/1.0/docs/TAPIRNetworkBuildersGuide_2010-05-05.html#toc11 – Acesso em Março de 2015.

¹⁹ DiGIR - <http://digir.sourceforge.net/> Acesso em Fevereiro de 2015.

Figura 9 - Estrutura de elaboração do arquivo DwC-A.

Fonte: Adaptado de GBIF, 2010.

Portanto, o núcleo do padrão de metadados DwC-A pode consistir em uma pasta com arquivos de dados únicos ou múltiplos, dependendo apenas da extensão dos dados publicados.

3.2.4 EML (*Ecological Metadata Language*)

A EML tem se destacado como uma linguagem de metadados para informações ecológicas e de biodiversidade. EML é um método para formalizar e padronizar o conjunto de conceitos que são essenciais para a descrição de dados ecológicos (FEGRAUS *et al.*, 2005).

A EML surgiu de um estudo e esforço comunitário envolvendo pesquisadores ecológicos, gestores de informação e desenvolvedores de software, liderados pelo NCEAS (*National Center for Ecological Analysis and Synthesis*) e o LTER (*Long Term Ecological Research Network*).

KNB (*The Knowledge Network for Biocomplexity*) sub-divisão do grupo de pesquisa em biodiversidade DataONE, apresenta o padrão EML (KNB, 2015) como sendo

uma implementação de um conjunto de tipos de documentos XML que podem ser utilizados de uma forma modular e extensível para documentar dados ecológicos. Cada módulo EML é projetado para descrever uma parte lógica dos metadados totais que devem ser incluídos em qualquer conjunto de dados ecológicos[...].

A arquitetura EML foi projetada com base em trabalhos anteriores e em outras linguagens de metadados relacionados, como Darwin Core. Algumas das características da EML são descritas a seguir (KNB, 2015):

- Modularidade: a EML foi desenvolvida como um conjunto de módulos ao invés de um grande padrão, o que facilita o crescimento do padrão em amplitude e profundidade. Ao implementar EML com uma arquitetura extensível, é possível definir quais módulos são pertinentes para descrever determinado recurso de dados, literatura ou software.
- Estrutura detalhada: a EML se esforça para equilibrar a troca de informações com muitos detalhes em dados suficientes que permitem o processamento de dados por meio de uma análise dos metadados.
- Compatibilidade: a EML adota uma síntese de outros padrões de metadados que evoluíram a partir da experiência dos grupos de outras disciplinas. Sempre que possível a EML adota um conjunto inteiro de informações a fim de facilitar a conversão de elementos EML para outras linguagens de metadados.
- Tipagem forte: EML é implementada em Extensible Markup Language (XML), que é uma linguagem que define as regras que governam a sintaxe EML.
- Distinção entre o modelo de conteúdo (preocupação com os conceitos por trás da estrutura do documento, por exemplo, cardinalidade) e a

implementação sintática desse modelo (preocupação com a tecnologia usada para expressar os conceitos definidos no modelo de conteúdo).

Através da EML são descritos um conjunto de aspectos essenciais de dados ecológicos, tais como: nomes e definições de variáveis, unidades de medidas, data, hora, local da coleta de dados, identidade da pessoa que recolheu os dados, desenho amostral etc.

A EML tenta reduzir a ambiguidade e a incerteza ao formalizar esses conceitos de metadados em um conjunto abrangente e padronizado de termos e definições destinados especificamente para dados ecológicos (FEGRAUS *et al.*, 2005).

A Figura 10 apresenta um exemplo do conjunto de dados EML, na qual é possível verificar a presença de elementos padrões para metadados e também a possibilidade de uma extensão de um conjunto de metadados adicionais não previstos nas normas da EML.

Figura 10 - Visualização do *Schema* EML.

Algumas ferramentas podem ser utilizadas para criar arquivos de metadados EML:

Morpho²⁰: É um software de gerenciamento de metadados que funciona em sistemas operacionais modernos e permite que ecologistas e pesquisadores criem, editem e gerem os seus próprios dados de pesquisa. Morpho também fornece recursos especiais para pesquisa e consulta de arquivos de dados ecológicos baseados em EML, além de fornecer acesso a todo o conteúdo EML que inclui atualmente mais de 2000 mil conceitos de metadados ou termos para descrever os dados ecológicos (Fegraus *et al.*, 2005).

Metacat: É um repositório de dados e metadados de biodiversidade que auxiliam cientistas e pesquisadores a encontrar, entender e usar de maneira eficaz os conjuntos de dados gerados por eles ou que tenham sido criados por outros. Segundo o site da própria mantenedora do software Metacat ²¹, “milhares de conjuntos de dados são atualmente documentados de forma padronizada e armazenados em sistemas como esse, proporcionando à comunidade científica uma gama de dados consistentes, descritos de forma simples para serem reutilizados, comparados, mesclados etc”.

3.2.5 Outros padrões de Metadados

Outros padrões de metadados também foram objetos dessa pesquisa e estão descritos a seguir.

O padrão de metadados GEMINI (*Geo-spatial Metadata Interoperability Initiative*) da NBN (*National Biodiversity Network*) foi definido pela primeira vez em 2002 e segundo a NBN (2011) “ele fornece uma orientação editorial adaptada especificamente para a definição de informações contextuais para descrever um recurso de dados de vida selvagem”.

²⁰ Para mais informações sobre o software de controle de metadados e também para download do Morpho acesse: <https://knb.ecoinformatics.org/#tools/morpho>.

²¹ Metacat: <https://knb.ecoinformatics.org/knb/docs/>

GEMINI adotou o modelo do padrão de metadados definido anteriormente Glgateway (anteriormente conhecido como *National Geospatial Data Framework* [NGDF]) e o fez assim baseado em duas razões (NBN, 2011):

- O padrão Glgateway foi desenvolvido especificamente para documentar conjunto de dados geo-referenciados para o uso dentro de Sistemas de Informação Geográfica (SIG), ao fazer isso o padrão cobre informações importantes para a maioria dos conjuntos de dados biológicos e está alinhado com as normas internacionais de SIG.
- Na elaboração do padrão Glgateway os responsáveis trabalhavam em estreita colaboração com as normas nacionais da Inglaterra e internacionais de desenvolvimento e isso garantiu um nível de confiabilidade em relação à compatibilidade ou por ser facilmente mapeado para as outras normas e padrões internacionais existentes.

Portanto, o padrão de metadados GEMINI tem como objetivo apresentar um novo núcleo de elementos de metadados para apoiar a criação de metadados geoespaciais em todo o Reino Unido.

Já o “o padrão de metadados ABCD foi estabelecido para capturar todos os elementos possíveis em espécimes e dados de observação coletados que podem ser fornecidos por sistemas de informações de coleções” (BERENDSOHN *et al.*, 2011) p.73.

O ABCD é composto por cerca de 1200 elementos e atributos e, segundo Berendsohn *et al.* (2011, p. 73),

nenhuma coleção utiliza mais do que uma fração dos elementos definidos, e o conjunto de elementos utilizados pode variar consideravelmente. O padrão ABCD também é bastante utilizado pelo portal GBIF e pelo BioCASE (*Biological Collection Access Service*) e hoje ele possui uma versão estendida para apoiar a rede de bancos de DNA, o portal GeoCASE (*Geosciences Collection Access Service*) (<http://www.geocase.eu/>) é a última versão do HISPID (*Herbarium Information Standards and Protocols for Interchange of Data*)[...].

3.2.6 Relação da modelagem de distribuição de espécies e os padrões de metadados.

Conforme abordado, os padrões de metadados possuem uma abrangente utilização por diversos domínios e por diversas comunidades científicas, e de mercado, de diferentes áreas. A área de Informática para Biodiversidade e suas ramificações, incluindo a modelagem de distribuição de espécies e suas ferramentas, também se utilizam de padrões de metadados com o intuito de facilitar a disseminação da informação em todos os ciclos de vida dos dados de biodiversidade.

Baseado nos estudos e definições dos padrões de metadados apresentados nas seções anteriores é possível constatar que eles possuem relações de compatibilidade, pois a padronização das linguagens de cada um pode estar relacionada ao estudo de um padrão de metadados previamente definido, ou seja, existe um refinamento e uma reutilização do esforço de outras equipes de desenvolvimento de padrões para a criação ou definição de um padrão de metadados para um domínio específico.

Sobre esse assunto, Castro e Santos (2007, p. 16) afirmam que

formas diferenciadas de representação para um tratamento mais acurado sobre tais informações com a finalidade de obter uma recuperação eficiente dos recursos informacionais em ambientes digitais e que atendam aos requisitos da interoperabilidade entre os sistemas informacionais, apresentam-se como uma necessidade para a otimização de ambientes informacionais digitais[...].

Em outras palavras, a utilização de padrões de metadados para a recuperação e a disponibilização de dados em ambientes virtuais tem sido uma necessidade em diversas áreas da ciência e também não é diferente na modelagem de distribuição de espécies.

Sobre essa conexão, entre os principais padrões de metadados utilizados para a área de Informática para Biodiversidade e para a modelagem de distribuição de espécies, a Figura 11 apresenta de forma gráfica a relação entre os padrões de metadados estudados nessa pesquisa.

Figura 11 – Principais padrões de metadados utilizados para área de Informática para Biodiversidade.

Padrões de Metadados para a Área de Biodiversity Informatics

Fonte: Elaborada pelo autor.

Conforme citado na introdução desta tese, fazendo referência a Marshall, Glegg e Howeel (2014), e constatado nesta pesquisa, especificamente para a modelagem de distribuição de espécies, um dos principais problemas é a aparente lacuna na utilização de padrões de metadados para as saídas ou modelos gerados pelas ferramentas de modelagem. Existe sim uma padronização para os dados de entrada, principalmente para os de ocorrência, provenientes de base de dados internacionais como o GBIF.

3.3 Perfil de Aplicação

Os padrões de metadados trouxeram suporte e interoperabilidade para comunidades científicas, em que inexistia uma padronização para disponibilização e para troca de dados entre plataformas e comunidades de desenvolvimento. No entanto, com o crescimento e a especificidade de cada segmento da ciência, diversos padrões de metadados foram e estão sendo desenvolvidos e disponibilizados. Com isso, áreas específicas não possuem padrões de metadados que abranjam todo o seu conteúdo, e, portanto, torna-se necessária muitas vezes a

utilização de mais de um padrão de metadados para que seja possível uma interoperabilidade dos dados.

Como os padrões de metadados são genéricos, eles não atendem como um todo e/ou especificamente as áreas da ciência que necessitam a utilização deles, sendo assim, cada área acaba criando um padrão para atendê-los e, por conseguinte, existem muitos padrões que acabam sendo usados apenas de forma exclusiva, e o conceito de interoperabilidade é também exclusivo para cada área. Quem concorda com isso é Coyle e Baker (2009) ao afirmar que um “metadado de uma comunidade particular é muito particular e o resultado disso é uma proliferação de formatos de metadados e, até em aplicações da mesma comunidade, esses metadados não são capazes de trocar informação”, em outras palavras não existe interoperabilidade entre eles.

Nesse contexto, perfis de aplicação do inglês *application profiles*²² permitem que a comunidade científica de cada área desenvolva um perfil para aplicação de padrões de metadados especificamente para cada situação e necessidade de utilização de um ou de um conjunto de padrões de metadados.

De acordo com Nilsson *et al.* (2008, p. 1), o termo perfil (*profile*) é amplamente utilizado para se referir a “documentos que descrevem como padrões ou especificações são implantados para suportar os requisitos de uma determinada aplicação, função, comunidade ou contexto”. No mesmo artigo, os autores continuam dizendo que a expressão perfil de aplicação na comunidade de metadados, “tem sido aplicada para descrever a adaptação de padrões para aplicações específicas”.

A modelagem de distribuição de espécies com seus algoritmos e ferramentas se encaixam no que Nilsson e os demais autores apresentaram no parágrafo anterior como uma aplicação específica e, por conseguinte, também necessitam da adaptação de padrões existentes para a interoperabilidade de dados (NILSSON *et al.*, 2008).

²² Para este estudo, foi definida a utilização da expressão perfil de aplicação traduzido do inglês, pois disponibiliza os estudos realizados nesta pesquisa num contexto de maior visualização e também utiliza a mesma expressão já aceita pela comunidade científica de *Biodiversity Informatics*, tanto no Brasil quanto no mundo.

De acordo com Mason e Galatis (2007), é incerto especificar quando a expressão perfil de aplicação foi utilizada pela primeira vez, no entanto, esta recebeu uma adoção respeitada pela comunidade científica, a partir da iniciativa da DCMI (*Dublin Core Metadata Initiative*) em publicar um artigo de Heery; Patel (2000, p. 1), definindo perfil de aplicação como: “Esquemas que consistem em elementos de dados desenhados a partir de um ou mais espaços de nomes, combinados por implementadores e otimizados para uma aplicação local específica“. Em outras palavras, a combinação de elementos de padrões e nomenclaturas já utilizadas pela comunidade científica, seguindo uma metodologia de desenvolvimento pode-se chamar perfil de aplicação.

De forma mais prática, o que Heery e Patel (*Idem*, p. 3) afirmam é que os perfis de aplicação são úteis porque permitem que “comunidades de desenvolvedores (nisto, pode-se incluir, analistas, cientistas, desenvolvedores, usuários, etc.) declarem de que forma eles estão utilizando os esquemas de padrões de metadados já existentes”.

Em 2002, um artigo publicado pelos autores Duval *et al.* (2002), apresentaram uma definição que, segundo Mason e Galatis (2007, p. 45), passa a ser mais bem aceita e abrange uma explicação mais completa do que é um perfil de aplicação. De acordo com os autores, um perfil de aplicação é

um conjunto de elementos de metadados selecionados de um mais esquemas de metadados e combinados em um esquema composto com o objetivo de adaptar ou combinar esquemas existentes em um pacote adaptado aos requisitos funcionais de uma aplicação específica, mantendo a interoperabilidade com os esquemas de base originais[...].

A definição apresentada acima vem ao encontro do objetivo desta pesquisa que é apresentar uma interoperabilidade entre as ferramentas de modelagem de distribuição de espécies, utilizando-se de padrões de metadados já existentes e consolidados pela comunidade científica no âmbito da Informática para Biodiversidade. Os autores Mason e Galatis (2007, p. 45) confirmam isso, dizendo que existem muitas razões para se estabelecer perfis de aplicação, mas a principal delas “é a preocupação em atender requisitos específicos em uma comunidade enquanto esta mantém interoperabilidade”.

Quem afirma isso também são os autores do documento apresentado pela CEN²³ (Comitê Europeu de Normatização), em (CEN, 2006), dizendo que perfis de aplicação permitem uma “mistura e combinação” de elementos de padrões de metadados para enfrentar requisitos específicos para um contexto particular, como por exemplo, algumas comunidades podem necessitar que certos elementos sejam obrigatórios ou restringir valores para um elemento em particular.

De acordo com Mason e Galatis (2007, p. 45), o parecer oficial da DCMI sobre perfis de aplicação é que “este não tem por objetivo criar e declarar novos termos e definições em padrões de metadados, muito pelo contrário, a ideia principal é reutilizar termos a partir de conjuntos de elementos já existentes em padrões consolidados”.

Quanto a isso, Heery e Patel (2000, p. 4) fazem as seguintes considerações:

todos os elementos de um perfil de aplicação são desenhados de algum lugar, de um padrão distinto. Se um implementador de um perfil de aplicação deseja criar novos elementos que não existem em lugar nenhum (em nenhum outro padrão de metadados), este deve se tornar responsável por esse novo esquema, assumindo a responsabilidade de manter este na comunidade científica[...].

Esses autores (Idem) também afirmam que

definindo um perfil de aplicação, os implementadores deste devem compartilhar esta informação sobre o novo esquema em ordem de interagir com grupos de trabalho interessados no mesmo assunto. Através dessa iniciativa, os membros de uma mesma comunidade podem iniciar o processo de utilização do mesmo perfil de aplicação, fazendo com que as informações sejam interoperáveis entre grupos relacionados ao mesmo tema[...].

Conforme apresentado no capítulo 3.2, sobre padrões de metadados, para existir uma interoperabilidade entre ferramentas de modelagem de distribuição de espécies, mais de um padrão de metadados podem ser usados, nesse caso, Heery e Patel (2000, p. 5) afirmam que um perfil de aplicação é a melhor solução para isso, porque “perfis de aplicação provem uma base de diferentes padrões de metadados para que estes possam trabalhar juntos com o objetivo nos requisitos de implementação de uma solução”, possibilitando que seja genuína a utilização de

²³Maiores informações sobre o Comitê Europeu de Normatização podem ser encontradas no site: <http://www.cen.eu/pages/default.aspx>

perfis de aplicação para extrair de padrões existentes um conjunto de nomenclaturas para se alcançar uma interoperabilidade.

Sobre a importância de um perfil de aplicação para interoperabilidade, assim como Heery; Patel (2000), o CEN (2006) afirma que a principal iniciativa para comunidades desenvolverem perfis de aplicação ao invés de criar “acordos de implementação” ou “um conjunto de especificações”²⁴ é para que no futuro seja possível:

- Interoperar informação com parceiros de fora da comunidade de implementação, e
- Construir um *framework* com ferramentas que tenham sido desenvolvidas e que possam servir para um mercado mais global.

Segundo Carrier (2008, p. 6), os perfis de aplicação são usados para garantir a interoperabilidade de um esquema de metadados, como uma extensão dele. Ela também cita Heery e Patel (2000, p. 4), dizendo que na declaração dos perfis de aplicação os desenvolvedores podem começar a compartilhar informações sobre seus esquemas para interagir com agrupamentos mais amplos.

Portanto, como indicado acima, por meio de outros autores, o CEN também afirma em (CEN, 2006, p. 7) que a principal motivação para definir perfis de aplicação “é o desejo de enfrentar requisitos específicos para uma comunidade, enquanto se mantém a interoperabilidade”.

De acordo com Curado Malta e Baptista (2012, p. 1), “um perfil de aplicação é uma técnica usada para adaptar metadados para atender necessidades de

²⁴A discussão sobre a utilização ou não de perfis de aplicação e da diferença entre possíveis similaridades entre termos científicos também utilizados para fins semelhantes é melhor discutido em (CEN, 2006). Neste os autores apresentam a diferença entre perfil de aplicação e essas outras definições. Seguem as diferenças:

- **Application Profile**: um perfil que se aplica a um protocolo de aplicação em oposição a um protocolo de baixo nível.
- **International Standardized Profile (ISP)**: um document formal, aprovado internacionalmente que engloba um ou mais perfis.
- **Internationally Registered Profile (IRP)**: um perfil que tem sido desenvolvido para um grupo reconhecido de usuários com grande potencial de aplicabilidade, e que tem sido revisado por conformidades técnicas dos grupos responsáveis pelos padrões.
- **Perfis Harmônicos**: um perfil que atingiu um nível de consenso global.
- **Acordo de Implementação**: um acordo informal entre os implementadores de um conjunto com o objetivo da interoperabilidade.

comunidade específicas”. O que vem ao encontro do que afirma Heery e Patel (2000, p. 1) dizendo que “um perfil de aplicação utiliza elementos de dados de diferentes padrões de metadados e “coloca tudo junto” para que desenvolvedores possam ajustar estes a uma comunidade particular”.

Além da interoperabilidade já citada, como sendo um dos benefícios da utilização dos perfis de aplicação para um “refinamento” dos padrões de metadados, a IMS (*Instructional Management System project*)²⁵ (IMS *Global Learning Consortium*) em (IMS, 2005) apresenta também os seguintes benefícios pelo uso dos perfis de aplicação:

- Um perfil de aplicação concordando com um conjunto consistente de regras para a construção de um perfil irá limitar as mudanças que podem ser realizadas, garantindo assim uma maior interoperabilidade entre os arquivos;
- O fornecimento de documentação consistente dos perfis de aplicação permite que os fornecedores construam mais facilmente produtos e serviços que abranjam várias comunidades com configurações simples;
- Com o crescente número de perfis de aplicações publicamente documentos permitirá que as comunidades adotantes selecionem e reutilizem projetos existentes;
- Em última análise, o fornecimento de definições legíveis por máquina do perfil de aplicação permitirá a negociação de contexto de tempo de execução entre domínios para facilitar a troca de dados e a interoperabilidade entre as comunidades.

3.3.1 Desenvolvimento de um Perfil de Aplicação

O DCMI, como sendo uma das instituições mais respeitadas pelo tema de metadados e seus padrões, desenvolveu um *framework* para a elaboração de perfis

²⁵ IMS Global Learning Consortium: <https://www.imsglobal.org/>

de aplicação, DCAP (*Dublin Core Application Profile*), ou perfil de aplicação Dublin Core, o DCMI chama esse *framework de Singapore Framework*²⁶.

Já a abordagem de Coyle e Baker (2009, p.1) é que

um DCAP define registros de metadados que atendem a necessidades de aplicativos específicos, ao mesmo tempo em que fornece interoperabilidade semântica com outros aplicativos com base em vocabulários e modelos definidos globalmente[...].

Ainda seguindo o raciocínio de Coyle e Baker (2009, p. 2), é importante destacar que um

DCAP é uma construção genérica para projetar registros de metadados que não exigem o uso de termos de metadados definidos pelo DCMI. Um DCAP pode usar quaisquer termos que são definidos com base em RDF²⁷ (*Resource Description Framework*), combinando termos de vários *namespaces* conforme necessário[...].

Segundo Carrier (2008, p. 5), as “diretrizes do DCAP enfatizam a importância de descrições legíveis por humanos e, sendo assim, enfatizam esquemas como documentos de texto”. No entanto, as diretrizes dão recomendações para expressar perfis de aplicativos em RDF e XML.

O *framework* definido pela DCMI não exige que ele utilize os termos adotados pela instituição, desse modo, permite o uso de padrões de metadados desenvolvidos por outras organizações. Esse conceito será aplicado e desenvolvido no Capítulo 4 desta tese, no qual são utilizados mais de um padrão para a concretização dos objetivos propostos.

Um perfil de aplicação Dublin Core inclui uma orientação para criadores de metadados e especificações claras para desenvolvedores de metadados. Ao articular o que se pretende e pode ser esperado dos dados, os perfis de aplicação promovem a partilha e a ligação de dados dentro e entre as comunidades científicas existentes (COYLE; BAKER, 2009, p. 2).

²⁶*Singapore Framework*: NILSSON *et al.*, 2008.

²⁷RDF (*Resource Description Framework*) é mantido pela W3C e as informações relacionadas a este esquema podem ser encontradas em <https://www.w3.org/TR/rdf-schema/>

No guia para o desenvolvimento para perfis de aplicação Dublin Core, Coyle e Baker (Idem) definem DCAP como sendo “um documento (ou conjunto de documentos) que especifica e descreve os metadados usados em um aplicativo específico”. Para fazer isso, um perfil (o diagrama do desenvolvimento de um perfil de aplicação Dublin Core) pode ser visto na Figura 12 com as seguintes funções:

- Descreve o que uma comunidade quer realizar com sua aplicação (Requisitos Funcionais);
- Caracteriza os tipos de elementos descritos pelos metadados e suas relações (Modelo de Domínio);
- Enumera os termos de metadados a serem utilizados e as regras para seu uso (Descrição Definir Perfil e Diretrizes de Uso);
- Define a sintaxe da máquina que será usada para codificar os dados.

Figura 12 – *Singapore Framework*.

O primeiro passo para o desenvolvimento de um perfil de aplicação é bem conhecido dos profissionais da engenharia de software: o levantamento e a definição dos requisitos funcionais. A definição dos requisitos funcionais pode seguir diversas metodologias, no entanto a adotada pela DCMI é a da UML²⁸ (*Unified Modeling Language*). A seguir seguem as quatro etapas descritas pela DCMI para o desenvolvimento de um perfil de aplicação.

A) **Definindo Requisitos Funcionais:** O primeiro passo, considerado como essencial, é definir o propósito do perfil de aplicação. Deve ser feito de maneira clara, dando suporte para o uso e a aplicação. Os requisitos funcionais orientam o desenvolvimento desse perfil, fornecendo metas e limites, que são componentes imprescindíveis para um processo de desenvolvimento de perfil de aplicação bem-sucedido.

O objetivo básico de um perfil de aplicação é suportar requisitos específicos, de um contexto exclusivo, por intermédio de um perfil de padrão genérico. A fim de validar esse processo, é importante ter uma compreensão explícita desses requisitos específicos. No caso de um perfil de aplicação de metadados, isso significa ter um escopo e propósito claros (CEN, 2006).

B) **Modelo de Domínio:** Depois de definir os requisitos funcionais, a próxima etapa é selecionar ou desenvolver o modelo de domínio. Segundo Coyle e Baker (2009), o modelo de domínio é uma descrição dos objetos que seus metadados descreverão e das relações entre esses objetos. O modelo de domínio é o modelo básico para a construção do perfil de aplicação.

Uma vez que os requisitos são definidos, a primeira decisão importante no desenvolvimento real dos perfis de aplicação de metadados é a seleção de elementos de dados e, muitas vezes, os desenvolvedores do perfil de aplicação iniciarão a partir de um esquema de metadados que tem um escopo e finalidade semelhante à do perfil do aplicativo. Em outras palavras, isso

²⁸ Referências sobre a UML podem ser encontradas em: (Booch *et al.*, 2006) e (Guedes, 2008).

facilita o trabalho e evita redundância de informações na comunidade (CEN, 2006).

- C) **Perfil de Diretrizes de Uso:** Definido o modelo de domínio, é necessário definir as propriedades para descrever os objetos nesse modelo. O próximo passo, então, é verificar os vocabulários RDF disponíveis para ver se as propriedades necessárias já foram declaradas e estão disponíveis para uso. Usar propriedades existentes, quando apropriado, requer menos esforço e aumenta a interoperabilidade de seus metadados. De acordo com Coyle e Baker (2009, p.4), “a consideração mais óbvia na avaliação de termos a partir de vocabulários existentes é a sua definição. A propriedade título Dublin Core, por exemplo, é definida como um nome dado ao recurso”. Se a definição atender as necessidades, essa propriedade será uma candidata para uso no perfil. No entanto, a adequação de uma propriedade para uso em uma determinada aplicação também depende do tipo de valores que a propriedade pode ter. Os tipos de valores destinados às propriedades devem corresponder aos permitidos das propriedades existentes que se deseja usar.
- D) **Sintaxe:** O próximo e último passo é descrever o registro de metadados em detalhes. Na abordagem DCMI, um registro de metadados é baseado no modelo de conjunto de descrição, do inglês, (*Description Set Model*), e os detalhes de gravação são descritos no conjunto de descrição do perfil, do inglês DSP²⁹, (*Description Set Profile*). Para cada Descrição e Declaração em um registro, o DSP define um modelo e cada modelo contém restrições relevantes especificando detalhes técnicos, como a repetibilidade de elementos ou restrições sobre valores permitidos (COYLE; BAKER, 2009, p. 8).

Como mencionado anteriormente, o uso e a definição de um guia para os futuros desenvolvimentos de perfis de aplicação para a mesma área, facilitam os desenvolvedores na aplicação de “como fazer”, e isso permite que eles não redupliquem informação e reutilizem dados já existentes em um perfil de aplicação

²⁹Detalhes sobre o desenvolvimento de um conjunto de descrições de um perfil são encontradas pelo guia desenvolvido pela DCMI em (NILSSON, 2008).

em funcionamento, facilitando a interoperabilidade entre eles. De acordo com Coyle e Baker (2009, p.8), esse guia “oferece instruções para aqueles que criarão os registros de metadados”, idealmente eles explicam cada propriedade e antecipam as decisões que devem ser tomadas no decorrer da criação de um registro de metadados.

3.3.2 Metodologia de desenvolvimento de um perfil de aplicação

Nos estudos apresentados por Curado Malta e Baptista (2012), em que foram realizadas análises em 21 itens voltados para o desenvolvimento de perfis de aplicação, apenas nove dos artigos, trabalhos ou metodologias apresentados apresentam de alguma forma uma metodologia de desenvolvimento de um perfil de aplicação.

Baseando-se nessa pesquisa inicial, para esta pesquisa foram realizadas mais pesquisas com o intuito de buscar uma metodologia que se adequasse as necessidades da problemática sobre a interoperabilidade para ferramentas de modelagem de distribuição de espécies e que fosse detalhada o suficiente contribuir para a comunidade científica da área de Informática para Biodiversidade.

Para uma melhor compreensão, os dados apresentados por Curado Malta e Baptista (2012) foram tabulados e, a partir deste estudo, foi elaborado um refinamento de trabalhos realizados e acrescentados após a apresentação desta pesquisa, facilitando a visualização e também a definição de uma metodologia de desenvolvimento de perfil de aplicação a ser empregada nesta pesquisa. De acordo com Curado Malta e Baptista (2012), foram analisadas duas informações principais, a primeira é se a pesquisa apresenta as etapas seguidas para a definição de um perfil de aplicação (etapas), e a segunda, é se a pesquisa contém o método seguido por cada etapa (método). Na Tabela 5, é apresentado também um pequeno resumo sobre a aplicação da pesquisa.

Tabela 5 - Relação entre pesquisas e a apresentação da metodologia utilizada para o desenvolvimento de um perfil de aplicação.

Pesquisa	Etapas	Método	Detalhe	Padrão
(Chen e Chen, 2005)	Sim	Não	<i>Modelo para trabalhar com metadados em bibliotecas digitais.</i>	DCMI
(BSI, 2005)	Sim	Não	<i>Faz recomendações para organizações que desejam criar perfis de aplicação para domínios de ensino-aprendizagem.</i>	DCMI e IEEE LOM
(IMS, 2005)	Sim	Sim	<i>Consórcio com o objetivo de apresentar etapas para a construção de um perfil de aplicação.</i>	Genérico
(CEN, 2006)	Sim	Sim	<i>Comitê europeu para padronização apresenta um guia para desenvolvimento de perfis de aplicação para a área de e-learning.</i>	DCMI
(NILSSON <i>et al.</i> , 2008)	Sim	Não	<i>Singapore Framework define o que é um perfil de aplicação Dublin Core e as regras para desenvolvê-lo.</i>	DCMI
(DUVAL <i>et al.</i> , 2002)	Não	Não	<i>Práticas de metadados usadas para perfis de aplicação.</i>	DCMI e IEEE LOM
(FRIESEN <i>et al.</i> , 2002)	Sim	Não	<i>Apresenta tópicos para a implementação de um perfil de aplicação no domínio de objetos de aprendizagem.</i>	DCMI
(CURRIE <i>et al.</i> , 2002)	Não	Sim	<i>Provê um método para fazer interoperabilidade visível, agregando elementos de diferentes metadados.</i>	DCMI
(EESV, 2012)	Não	Sim	<i>Projeto europeu ISA que desenvolveu o perfil de aplicação ADMS.</i>	Genérico
(DCMI, 2017)	Não	Sim	<i>Proposta de uso de cenários para o desenvolvimento dos perfis de aplicação Dublin Core.</i>	DCMI e IEEE LOM
(ONYANCHA <i>et al.</i> , 2001)	Não	Não	<i>Reporta o desenvolvimento de um perfil de aplicação para a área de agricultura.</i>	DCMI
(AGOSTINHO <i>et al.</i> , 2004)	Sim	Sim	<i>Apresenta as etapas do desenvolvimento do perfil de aplicação LOMAP para o domínio de objetos de aprendizagem.</i>	IEEE LOM
(DE LA PASSADIERE e JARRAUD, 2004)	Sim	Não	<i>Apresenta o desenvolvimento do perfil de aplicação ManUel no domínio de objetos de aprendizado.</i>	IEEE LOM
(GÁRCIA-QUISMONDO <i>et al.</i> , 2006)	Sim	Não	<i>Apresenta o desenvolvimento do perfil de aplicação MIMETA no domínio de objetos de aprendizado.</i>	DCMI e IEEE LOM
(WILSON <i>et al.</i> , 2007)	Sim	Não	<i>Perfil de aplicação para área de biblioteconomia.</i>	DCMI
(BUONAZIA e MASCI, 2007)	Sim	Sim	<i>Desenvolvimento do perfil de aplicação PICO.</i>	DCMI
(EADIE, 2008)	Sim	Sim	<i>Apresenta que um grupo de trabalho de diferentes backgrounds compôs um perfil de aplicação.</i>	DCMI
(SALOKHE <i>et al.</i> , 2008)	Sim	Não	<i>Apresenta um perfil de aplicação para a área da agricultura.</i>	Genérico

Pesquisa	Etapas	Método	Detalhe	Padrão
(BOUNTOURI <i>et al.</i> , 2009)	Sim	Não	<i>Apresenta um perfil de aplicação para domínios de serviço público.</i>	DCMI
(PALAVITSINIS <i>et al.</i> , 2009)	Sim	Não	<i>Apresenta um perfil de aplicação para a área da agricultura.</i>	IEEE LOM
(ZSCHOCKE <i>et al.</i> , 2009)	Sim	Sim	<i>Perfil de aplicação CIGAR para o domínio de objetos de aprendizado.</i>	IEEE LOM

Fonte: Adaptado de CURADO MALTA; BAPTISTA, 2012.

A partir da pesquisa citada acima, foram acrescentados os estudos de oito artigos com o objetivo de verificar quais os esforços na área de metodologias para o desenvolvimento de perfis de aplicação e os resultados são apresentados na Tabela 6. As mesmas características de análise foram levadas em consideração para essas mais recentes pesquisas. O levantamento foi realizado utilizando a base de dados da IEEE, ACM e Google Scholar, baseando-se na relevância e sendo superior ao ano de 2013.

Tabela 6 - Análise das pesquisas relacionadas ao desenvolvimento e utilização de perfis de aplicação.

Pesquisa	Etapas	Método	Detalhes	Padrão
(ALEIXO <i>et al.</i> , 2015)	Não	Não	<i>Perfil de Aplicação em conformidade com as normas gerais de descrição arquivística.</i>	DCMI
(KRAUSE <i>et al.</i> , 2015)	Não	Não	<i>Perfil de aplicação para publicações escolares.</i>	DCMI
(SILVA, 2016)	Sim	Sim	<i>Utilização do perfil de aplicação como ontologia.</i>	DCMI
(CIASULLO <i>et al.</i> , 2013)	Não	Não	<i>Perfil de aplicação para serviços públicos.</i>	Genérico
(CASTRO <i>et al.</i> , 2013)	Sim	Não	<i>Desenvolvimento de um perfil de aplicação para diferentes domínios.</i>	DCMI
(BAIR E STEUER, 2013)	Sim	Não	<i>Perfil de aplicação para manuscritos modernos.</i>	DCMI
(BOMAN <i>et al.</i> , 2016)	Sim	Não	<i>Melhores práticas usando um perfil de aplicação criado para a Associação Americana de Bibliotecas.</i>	DCMI
(AMADOR <i>et al.</i> , 2016)	Sim	Não	<i>Perfil de aplicação para a área educacional.</i>	IEEE LOM

Fonte: Elaborado pelo autor.

Com a realização dessa análise de 29 artigos dos últimos 16 anos, é possível verificar que 72% das pesquisas apresentam as etapas de desenvolvimento do perfil de aplicação, no entanto, 65% não apresentam os métodos ou metodologias

adotadas nesta tese. Sobre a utilização dos padrões para desenvolvimento do perfil de aplicação, é possível verificar que 55% das pesquisas utilizaram os padrões da DCMI, 17% empregaram o padrão da IEEE LOM e 27% aplicaram padrões genéricos ou o padrão DCMI e IEEE LOM juntos. Em acréscimo à análise anterior, também foi realizada uma sondagem das principais referências, manuais ou guias utilizados pelas pesquisas, com o objetivo de definir um padrão e método para o desenvolvimento do perfil de aplicação desta pesquisa.

Segundo os autores Curado Malta e Baptista (2012), e também baseado nas pesquisas realizadas para esta pesquisa, o *framework* apresentado por Nilsson *et al.* (2008) com a definição de um perfil de aplicação para o padrão de metadados Dublin Core é o que retrata de forma mais abrangente uma metodologia de desenvolvimento para um perfil de aplicação, mas, mesmo assim não apresenta em detalhes cada passo do desenvolvimento.

Conclui-se, portanto que existem alguns modelos para o desenvolvimento de perfis de aplicação, contudo o mais utilizado, segundo Chen *et al.* (2003) e de acordo com os levantamentos apresentados nas Tabelas 5 e 6, o padrão Dublin Core é amplamente aceito como um padrão mais genérico, e os demais são utilizados e derivados para domínios mais específicos (Figura 13).

Figura 13 - Dublin Core como padrão genérico e os demais como específicos.

Fonte: Figura adaptada de CHEN *et al.*, 2003.

3.4 Trabalhos Relacionados

O objetivo desta seção é apresentar pesquisas e trabalhos relacionados à interoperabilidade entre ferramentas de modelagem de distribuição de espécies, principalmente os projetos relacionados aos modelos gerados pelas ferramentas de MDE.

Dentre os temas, a padronização de dados de biodiversidade, o uso de padrões de metadados e ontologias para definir domínios que garantam uma interoperabilidade de informações, bem como a disponibilização desses dados gerados, têm sido discutidos entre pesquisadores e estudantes da área e serão apresentadas agora, algumas das pesquisas de destaque.

No entanto, antes mesmo da apresentação de trabalhos relacionados à interoperabilidade de dados de biodiversidade, é apresentado um estudo sobre o uso dos algoritmos e ferramentas de MDE. O objetivo desse breve relato é destacar a finalidade do uso das ferramentas e apresentar em contrapartida estudos da área que comprovem a utilização desses recursos computacionais.

A problemática apresentada na Tabela 7 foi atualizada e adaptada, baseadas nos estudos de Marco Junior; Siqueira (2009), Peterson *et al.* (2011), Soberón; Peterson (2005) e Philips *et al.* (2006), nessa exposição são apontados alguns exemplos das aplicações da MDE e quais algoritmos e ferramentas são aplicadas a cada um dos problemas.

Essa Tabela retrata a abrangência de ferramentas e suas funcionalidades na modelagem de distribuição de espécies, bem como a utilização dos algoritmos de modelagem por cada ferramenta. Esse estudo é importante para que seja realizada uma apresentação das atuais ferramentas de modelagem e estimular o estudo da interoperabilidade entre elas.

Tabela 7 - Aplicação do uso dos algoritmos e ferramentas de MDE.

Problemática	Método		Referência
	Algoritmo	Ferramenta	
Predição de distribuição de espécies raras ou ameaçadas de extinção.	Bioclim, Domain, Entropia Máxima, Random Forest, GBM, MARS, ANN, GARP MaxEnt, Domain, GARP SVM	DIVA-GIS MaxEnt BIOMOD DesktopGARP Não Informado Não Informado	(TORRES <i>et al.</i> , 2012) (PADONOU <i>et al.</i> , 2015) (GUO <i>et al.</i> , 2005)
Detectar espécies novas ou raras e novos padrões de distribuição.	Entropia Máxima SDM, Bioclim, GARP	MaxEnt openModeller	(PEREIRA, 2014) (DELATORRE <i>et al.</i> , 2014)
Definição de espécie para recuperação de áreas degradadas.	Entropia Máxima	MaxEnt	(ADHIKARI <i>et al.</i> , 2012)
Definição de áreas prioritárias para conservação e influência das áreas de calibração sobre a modelagem.	Entropia Máxima GARP Bioclim, Domain, SVM, GARP, Entropia Máxima	MaxEnt DesktopGARP DIVA-GIS openModeller MaxEnt	(WILSON <i>et al.</i> , 2011) (ANACLETO; OLIVEIRA, 2014) (GIOVANELLI <i>et al.</i> , 2010)
Determinar áreas com maior risco de invasão por espécies exóticas.	Entropia Máxima GARP Entropia Máxima	MaxEnt openModeller MaxEnt	(SOBEK-SWANT <i>et al.</i> , 2012) (YANG <i>et al.</i> , 2013)
Análise do efeito das mudanças climáticas globais sobre a biodiversidade.	Entropia Máxima Entropia Máxima Bioclim, GARP, Entropia Máxima, Distância Euclidiana, GLM, Random Forest Entropia Máxima Entropia Máxima Distância Euclidiana	MaxEnt MaxEnt Bioensembles MaxEnt MaxEnt Bioensembles	(KHANUM <i>et al.</i> , 2013) (CONVERTINO <i>et al.</i> , 2014) (FORDHAM <i>et al.</i> , 2012) (IHLOW <i>et al.</i> , 2012) (OLIVEIRA; CASSEMIRO, 2013)
Estudo de espécies no passado ou extintas.	Bioclim, Distância Euclidiana, GARP, MaxEnt	Bioensembles	(LIMA-RIBEIRO; DINIZ-FILHO, 2012)

Fonte: Elaborado pelo autor.

Por intermédio da Tabela 7, é possível verificar dentre as pesquisas apresentadas pelo menos cinco diferentes ferramentas de modelagem de distribuição de espécies, nas quais, algumas delas oferecem uma gama de

algoritmos e técnicas de modelagem diferentes. Apesar das diferentes frentes e objetivos de cada pesquisa, as ferramentas atendem de forma satisfatória os resultados de cada estudo proposto.

Realizada essa apresentação de estudos na área de modelagem, e sobre a utilização das ferramentas e algoritmos para cada um deles, a seguir são apontados trabalhos relacionados à interoperabilidade dos dados de biodiversidade.

3.4.1 Apresentação dos Trabalhos

Vários trabalhos abordam temas relacionados ao tratamento de dados de biodiversidade, como padrão de metadados, ferramentas de modelagem, portais de disponibilização dos dados, melhores práticas entre os algoritmos de modelagem etc. Como o objetivo desta proposta é apresentar uma solução que garanta a interoperabilidade computacional entre os modelos gerados pelas ferramentas de MDE, os trabalhos selecionados por meio de uma pesquisa bibliográfica estão relacionados a este tema de interoperabilidade de dados de biodiversidade, como por exemplo, a padronização de dados coletados, a padronização de metadados e ontologias, a padronização para disponibilização de informações, a identificação de pesquisas e coletas de dados para reutilização e a interoperabilidade entre ferramentas de MDE.

A Tabela 8 apresenta trabalhos relacionados ao tema desta proposta. Todas as pesquisas estão preocupadas com soluções interoperáveis para dados de biodiversidade. Essas soluções envolvem, tanto modelos como padrões e dados de biodiversidade.

Portanto, a análise e o estudo dos trabalhos embasaram o entendimento a respeito de alguns aspectos importantes para a interoperabilidade de dados de biodiversidade e o mapeamento das seguintes informações:

- **Domínio Específico da Pesquisa:** Se a pesquisa está voltada para a interoperabilidade de modelos gerados a partir do processamento de MDE ou

se está focada na interoperabilidade de algum dos parâmetros anteriores necessários a MDE.

- **Definição de Padrão de Metadados ou Ontologia:** Se o trabalho apresenta um padrão de metadados ou ontologia para garantir a interoperabilidade de dados de biodiversidade.
- **Utilização de ferramenta para garantir interoperabilidade:** Quais são as ferramentas e ou plataformas utilizadas para definição dos padrões de metadados, para execução dos algoritmos de modelagem, bem como a apresentação dos modelos gerados pelas ferramentas.
- **Disponibilização dos dados de biodiversidade:** Se a pesquisa apresenta formas de disponibilização e reutilização dos dados de biodiversidade definidos como interoperáveis.

Tabela 8 - Abordagens de Interoperabilidade entre Dados de Biodiversidade.

Fonte	Abordagem	Domínio Específico	Desc. do Padrão de Metadados	Ferramenta ou Plataforma	Disponibilização dos Dados de Biodiversidade
(BERENDSON <i>et al.</i> , 2011)	Apresenta um estudo para garantir interoperabilidade entre dados de biodiversidade.	Não	Sim	ViBRANT	Sim
(DUBOIS <i>et al.</i> , 2013)	Apresenta uma ferramenta interoperável para dados multidisciplinares como modelagem ecológica.	Sim	Não	eHabitat	Sim
(NATIVI <i>et al.</i> , 2013)	Acesso a modelos ambientais e interoperabilidade para comparação entre modelos.	Não	Sim	GEO Model	Sim
(DUBOIS <i>et al.</i> , 2015)	Apresentação de um sistema com dados interoperáveis para áreas de conservação ao redor do mundo.	Não	Sim	DOPA	Sim
(ROBERTSON <i>et al.</i> , 2014)	Ferramenta para integração e disponibilização de dados interoperáveis de biodiversidade.	Não	Sim	IPT - GBIF	Sim

Fonte: Elaborado pelo autor.

Em geral, todas as pesquisas apresentaram conceitos e aplicações de interoperabilidade para dados de biodiversidade. A seguir serão apresentadas de forma detalhada as soluções interoperáveis para cada pesquisa e também a relação com a proposta desta pesquisa. Merecem destaque os três primeiros estudos, mas não menos importantes, os dois últimos também são mostrados de forma simplificada.

4 DESENVOLVIMENTO DO PERFIL DE APLICAÇÃO E ARQUITETURA COMPUTACIONAL

O perfil de aplicação desenvolvido para esta tese segue a metodologia apresentada no capítulo 3, tendo como base o processo de desenvolvimento da DCMI, o projeto *Singapore* (NILSSON et al., 2008) e também seguindo a metodologia proposta por Curado Malta e Baptista (2013) que como mencionado, tem sido utilizado como modelo para perfis de aplicação em diversas áreas da ciência. De acordo com essas pesquisas, um perfil de aplicação consiste em um conjunto de cinco componentes, três obrigatórios e dois opcionais:

1. **Requisitos funcionais** (obrigatório): apresentam os requisitos funcionais do perfil de aplicação, dando suporte para o desenvolvimento do modelo de domínio e futuras aplicações ao utilizar-se desse perfil.
2. **Modelo de Domínio** (obrigatório): apresenta as entidades básicas do perfil de aplicação, bem como o relacionamento entre elas.
3. **Descrição do Perfil** (obrigatório): apresenta os elementos de metadados que fazem parte do perfil de aplicação.
4. **Diretrizes de uso e Diretrizes de sintaxe** (opcionais): documentam e descrevem as regras de utilização do perfil de aplicação.

A seguir são apresentados e descritos os passos obrigatórios citados acima para o desenvolvimento do perfil de aplicação.

4.1 Perfil de Aplicação para Interoperabilidade entre Ferramentas de Modelagem de Distribuição de Espécies

O perfil de aplicação desenvolvido para a interoperabilidade entre as ferramentas de MDE recebe o nome de AP-SDM, do inglês, *Application Profile – Species Distribution Modeling*, e tem como objetivo principal disponibilizar para a comunidade científica um meio pelo qual os modelos gerados pelas ferramentas de

MDE possam ser interoperáveis, facilitando sua reutilização, disseminação e visualização dos dados de maneira padronizada.

O perfil de aplicação AP-SDM utiliza características e elementos dos três principais padrões relacionados à área de Informática para Biodiversidade, o Dublin Core, Darwin Core e o EML. Nesta seção serão apresentados os passos para o desenvolvimento desse perfil, iniciando pelos requisitos funcionais. Conforme mencionado no Capítulo 3, a metodologia de desenvolvimento adotado pelo DCMI no projeto *Singapore* (NILSSON *et al.*, 2008) é o da UML, no qual será exposto a seguir, utilizando-se de algumas etapas do processo, esse mesmo processo também é praticado por Curado Malta; Baptista (2013).

No processo de elaboração de um modelo de distribuição de espécies foram identificados elementos que compõem o entendimento do processo de uma modelagem e que devem estar disponíveis na interoperabilidade computacional proposta por esta pesquisa e também visualmente apresentadas no modelo de domínio na seção 4.2 deste capítulo, são eles:

- **Modelo de distribuição de espécies (*Species Distribution Model*):** Modelo gerado pela ferramenta de MDE, contendo as informações distribuídas das espécies e dependendo da ferramenta utilizada, mapas, gráficos etc.
- **Projeto (*Project*):** Esse tem como objetivo descrever o projeto, envolvendo a modelagem de distribuição de espécies e incorporando informações relevantes da pesquisa.
- **Pessoa (*Person*):** Tem como objetivo apresentar os pesquisadores envolvidos no projeto.
- **Ferramenta (*SDM Tool*):** Apresenta a ferramenta empregada na modelagem de distribuição de espécies, com suas peculiaridades.
- **Algoritmo (*Algorithm*):** tem como objetivo apresentar o algoritmo aplicado para a geração do modelo de distribuição de espécies, bem como os parâmetros utilizados para a execução dele.
- **Dados de ocorrência (*Occurrence Data*):** Disponibiliza todos os dados de ocorrência utilizados na MDE.

- **Data Set:** apresenta os dados que compõem o modelo de distribuição de espécies, podem ser climáticos, de solo, geográficos etc.

Definidos os elementos básicos do modelo de distribuição de espécies, o próximo passo é a definição dos requisitos funcionais do perfil de aplicação AP-SDM.

4.1.1 Especificação dos Requisitos Funcionais

Os requisitos funcionais para o perfil de aplicação AP-SDM foram elencados a partir de levantamentos realizados juntamente com o grupo de pesquisa coordenado pelo prof. Dr. Pedro Luiz Pizzigatti Corrêa, bem como por investigação realizada com dois pesquisadores da área de modelagem de distribuição de espécies, e também por meio dos artigos estudados e apresentados nesta pesquisa. Todo esse empenho auxiliou grandemente para o avanço e definição dos requisitos funcionais do perfil de aplicação (RFPA) que serão demonstrados a seguir.

Requisitos funcionais:

RFPA01 – Prover um perfil de aplicação com metadados ricos para modelos gerados pelas ferramentas de modelagem de distribuição de espécies.

- Cenário de uso: Esse perfil de aplicação oferece à comunidade de desenvolvimento de ferramentas de modelagem um meio de padronizar a saída dos dados das ferramentas de modelagem, provendo uma interoperabilidade entre os resultados e permitindo um reúso dela mesma ou de outras ferramentas que suportem o mesmo padrão.

RFPA02 – Facilitar a criação e disponibilização de metadados consistentes para os modelos gerados pelas ferramentas de modelagem de distribuição de espécies.

- Cenário de uso: Os elementos apresentados pelo perfil de aplicação são essenciais para o reúso de um modelo gerado por uma ferramenta de MDE. A padronização desses elementos em um metadado padrão

propicia a criação de arquivos com dados de modelos gerados e também facilita a disponibilização e a divulgação padronizada dos resultados.

RFPA03 – Ser compatível com padrões de metadados já existentes na comunidade científica.

RFPA04 – Suportar extensibilidade para perfis de aplicação de outros domínios.

- Cenário de uso: O perfil de aplicação deve ser extensível para outras funcionalidades e outras comunidades científicas.

RFPA05 – O perfil de aplicação deve ser capaz de adicionar serviços a futuras novas funcionalidades que possam ser implementadas nas ferramentas de MDE.

- Cenário de Uso: O perfil de aplicação deve buscar uma flexibilização com relação a futuros ajustes e acréscimos em informações de saídas dos modelos gerados pelas ferramentas de MDE.

RFPA06 – O perfil de aplicação pode ser versionado, facilitando o controle de versões com correções, adaptações e manutenções no perfil corrente.

- Cenário de uso: Um repositório com versões do perfil de aplicação será disponibilizado para controle de versões do perfil de aplicação.

RFPA07 – O repositório de versões deverá permitir uma navegação entre as versões.

- Cenário de uso: O repositório deve permitir que um usuário navegue entre as versões, disponibilizando um tutorial sobre cada uma das versões.

RFPA08 – O repositório de versões deverá oferecer informação sobre a última ou melhor versão para modelos gerados por ferramentas de MDE.

RFPA09 – Esse perfil de aplicação deverá ser disponibilizado de maneira aberta, oferecendo à comunidade científica da área total liberdade para utilizá-lo.

Esses foram os requisitos funcionais elencados durante o período de pesquisa desta pesquisa, a próxima etapa envolve a construção do modelo de domínio.

4.1.2 Desenvolvimento do Modelo de Domínio

Realizado o levantamento dos requisitos para o perfil de aplicação AP-SDM, para a interoperabilidade entre os modelos gerados pelas ferramentas de MDE, foi desenhado o modelo de domínio. O modelo de domínio do perfil de aplicação AP-SDM é apresentado na Figura 14. Toda a modelagem dos diagramas UML realizados nesta tese foram desenvolvidas com a utilização da ferramenta aberta Argo³⁰ UML versão 0.34.

Figura 14 - Modelo de Domínio do AP-SDM.

Fonte: Elaborada pelo autor.

³⁰O Argo UML está disponível para download através do link: <http://argouml.tigris.org/>

É possível verificar na Figura 14 os principais elementos necessários em um modelo de distribuição de espécies, portanto, as características necessárias para o modelo de domínio do perfil de aplicação AP-SDM.

O fluxo do modelo representado acima é detalhado da seguinte maneira: uma pessoa está associada a um projeto de pesquisa de distribuição de espécies. Um modelo de distribuição de espécies pertence a um projeto de pesquisa. A ferramenta de MDE gera um modelo de distribuição de espécies, que se utiliza de um ou mais algoritmos, dados de ocorrência de espécies, e um *dataset* de informações, que pode variar de acordo com a finalidade do projeto. Os principais elementos de cada classe foram elencados no modelo, no entanto a lista completa pode ser encontrada nos arquivos relacionados a este trabalho no doi:10.7910/DVN/MC76QC.

Realizada a definição do modelo de domínio, é apresentado na Figura 16 um modelo de domínio em associação com os padrões de metadados usados para a construção do perfil de aplicação, EML, Darwin Core e Dublin Core. Para isso, também foi levantado o que da estrutura do EML seria utilizado no modelo e domínio (Figura 15).

Figura 15 - Estrutura do EML e suas aplicações para o modelo de domínio.

Fonte: Elaborado pelo autor.

As informações gerais do projeto e padronizadas no Dublin Core estão relacionadas com a entidade “Project” no modelo de domínio dela, enquanto que as informações do projeto do EML estão associadas com a mesma entidade. Do padrão Darwin Core, foram selecionadas as entidades que armazenam informações taxonômicas e dos organismos envolvidos na pesquisa.

O modelo de domínio do perfil de aplicação AP-SDM oferece um conjunto de dados capaz de englobar três diferentes padrões de metadados relacionados à modelagem de distribuição de espécies, possibilitando que todas as informações geradas e inseridas nos modelos de distribuição de espécies de uma ferramenta de modelagem, sejam disponibilizadas em um único modelo de dados. A Figura 16 mostra o modelo de domínio associado com os padrões de metadados envolvidos nesta pesquisa.

Figura 16 - Modelo de Domínio em associação com os padrões de metadados envolvidos.

Concluído o modelo de domínio, o próximo passo para o desenvolvimento do perfil de aplicação AP-SDM é a definição do conjunto de perfis ou, do inglês, (*Description Set Profile*).

4.1.3 Desenvolvimento da Descrição do Conjunto de Perfis (*Description Set Profile - DSP*)

Realizado o modelo de domínio e também a apresentação da sua relação com os padrões de metadados aplicados nesta pesquisa, esta seção tem como objetivo apresentar a descrição do conjunto de perfis usados para o desenvolvimento do perfil de aplicação AP-SDM. Os vocabulários identificados e utilizados na construção desse perfil podem ser visualizados na Tabela 9.

Tabela 9 - Vocabulários utilizados para o desenvolvimento do perfil de aplicação e data set.

Nome	Termos	Prefixo
Ecological Metadata Language - EML	https://knb.ecoinformatics.org/#external//emlparser/docs/eml-2.1.1/index.html	eml
Darwin Core	http://rs.tdwg.org/dwc/terms/	dwc
Dublin Core	http://dublincore.org/documents/dcmi-terms/	dc

Fonte: Elaborado pelo autor.

Como exemplo, um conjunto de elementos é descrito na Tabela 10 e o conjunto completo com todos os elementos encontra-se no doi:10.7910/DVN/MC76QC.

Tabela 10 - DSP - *Description Set Profile* do perfil de aplicação AP-SDM.

Standard	Domain	Propriedade	Required
darwin core	organism	organismID	não
darwin core	organism	organismName	não
darwin core	taxon	taxonID	não
darwin core	taxon	scientificNameID	não
darwin core	taxon	scientificName	não
dublin core	general	dcterms: title	sim
dublin core	general	dcterms: created	sim
dublin core	general	dcterms: language	sim
eml	entity	physical	não
eml	literature	article	não

Fonte: Elaborado pelo autor.

As diretrizes e documentação de utilização do perfil de aplicação é uma sugestão para trabalhos futuros, e será realizada após a conclusão desta tese. O próximo passo apresentado é a arquitetura computacional para dar suporte a futuras aplicações que venham operar com o perfil de aplicação AP-SDM.

4.2 Arquitetura Computacional para a Interoperabilidade entre os modelos gerados pelas ferramentas de MDE

Com o objetivo de oferecer suporte para aplicações que visem à utilização do perfil de aplicação AP-SDM, para garantir interoperabilidade entre modelos gerados pelas ferramentas de modelagem de distribuição de espécies, a seguir é apresentada uma arquitetura de software.

A arquitetura de um software define todos os componentes, elementos, características, relacionamentos necessários para um sistema de computação, a arquitetura serve como um padrão para futuros desenvolvimentos de software (Shaw e Garlan, 1996).

A metodologia empregada para a construção da arquitetura computacional candidata é descrita por Hofmeister *et al.* (2007), na qual sugere um modelo geral

para construção de arquiteturas envolvendo as características de outras metodologias.

De acordo com Hofmeister *et al.* (2007, p. 110), o processo arquitetural no RUP começa com os seguintes artefatos: um documento de visão, um modelo de caso de uso (requisitos funcionais), e especificações suplementares (requisitos não funcionais, de qualidade etc.). Os três principais grupos de atividades são:

- **Definir a arquitetura candidata:** Normalmente se inicia com uma análise de casos de uso, focando os que influenciarão a arquitetura.
- **Apresentar uma síntese da arquitetura:** construindo uma prova de conceito, verificando sua viabilidade relacionada a suas funcionalidades.
- **Refinar a arquitetura:** identificar elementos de design e integrar com o protótipo da arquitetura.

No modelo para o desenvolvimento de arquitetura proposto por Hofmeister *et al.* (2007), primeiramente são classificadas todas as atividades durante a fase de design. A análise da arquitetura articula todos os requisitos significantes para a configuração dela, baseada nos interesses e no contexto. As atividades de design da arquitetura proposta são apresentadas na Figura 17.

Figura 17 - Atividades do design da arquitetura.

Fonte: Adaptado de HOFMEISTER *et al.*, 2007, p. 113.

Onde os **interesses arquiteturais** geralmente são descritos nos requisitos funcionais do sistema, o **contexto** determina as circunstâncias de desenvolvimento, operações, políticas etc. Os **requisitos arquiteturais relevantes** são um refinamento das principais características que devem ser descritas na arquitetura e a **análise dela** serve para definir os problemas a serem resolvidos. As **soluções de arquitetura candidatas** devem apresentar as alternativas ou soluções parciais para a arquitetura. A **síntese arquitetural** é a parte principal no desenvolvimento da arquitetura. A **arquitetura validada** consiste naquela que condiz com os requisitos apresentados e a **avaliação dela** assegura que ela é a correta.

O primeiro passo para a definição da arquitetura proposta é a definição dos requisitos funcionais, que serão apresentados na próxima seção.

4.2.1 Requisitos funcionais da arquitetura

Os requisitos funcionais e não funcionais da arquitetura foram elencados a partir de entrevistas e reuniões do grupo de pesquisa de biodiversidade da Escola Politécnica da Universidade de São Paulo. Os requisitos funcionais elencados são os seguintes:

RF01 – Arquivar identificação do usuário responsável pelo modelo de distribuição de espécie.

- A arquitetura deve arquivar a informação do usuário responsável pelo modelo de distribuição de espécie, permitindo também uma extensão para informações de equipes científicas.

RF02 – Arquivar os metadados utilizados e gerados pela ferramenta de MDE.

- Deve arquivar os metadados originais operados pela ferramenta de modelagem de distribuição de espécies.

RF03 – Armazenar os dados utilizados para definição do modelo de distribuição de espécies.

- Armazenar os dados relacionados ao algoritmo usado, os parâmetros e a ferramenta aplicada para gerar o modelo.

RF04 – Armazenar metadados sobre a autoria dos dados de ocorrência e dataset.

- Armazenar os metadados sobre a autoria dos dados de ocorrência e também dos *datasets* utilizados para o desenvolvimento do modelo de distribuição de espécies.

RF05 – Permitir o reuso e interoperabilidade do modelo de distribuição de espécies.

- Permitir o reuso do modelo de distribuição de espécies pela mesma ferramenta de modelagem ou por outra diferente, utilizando-se de um padrão de metadados e de um perfil de aplicação para definição do vocabulário padrão.

RF06 – Disponibilizar uma base de dados para consulta dos modelos gerados.

- Disponibilizar uma base de dados para consulta dos modelos gerados pelas ferramentas.

RF07 – Consultar metadados de dados de ocorrência e datasets disponibilizados na web.

- Possibilitar serviços para consulta a bases de dados de biodiversidade para obtenção de metadados de dados de ocorrência de espécies e datasets climáticos, geográficos etc.

RF08 – Importar os metadados gerados por ferramentas de modelagem.

- Deve ser capaz de importar os modelos e as informações geradas pelas ferramentas de modelagem, bem como, permitir a inclusão de informações não obtidas automaticamente, principalmente por ferramentas proprietárias.

RF09 – Exportar os metadados do modelo padronizado.

- A arquitetura deve exportar os metadados do modelo padronizado por meio do perfil de aplicação ou padrão de metadados para que possam ser lidos por outras ferramentas de modelagem.

RF10 – Consultar os metadados gerados.

- Permitir a consulta aos dados gerados pela ferramenta.

4.2.2 Requisitos não funcionais da arquitetura

RNF01 – Definir um modelo de dados padrão para o modelo de distribuição de espécies.

- Verificar um modelo de dados padrão para a definição dos dados gerados e obtidos através do modelo gerado pela ferramenta de MDE.

RNF02 – Permitir o uso do modelo gerado em ambientes variados.

- Permitir a utilização do modelo gerado pela ferramenta em ambientes variados, ferramentas proprietárias e abertas, permitindo que o metadado gerado seja editável para adaptação de acordo com a necessidade.

RNF03 – Estimar protocolos e padrões para disponibilização na web.

- Proporcionar metadados baseados em protocolos internacionais que facilitem a publicação e a visualização na internet.

4.2.3 Caso de Uso

Para o melhor entendimento e desenvolvimento da arquitetura, é apresentado na Figura 18 um diagrama de caso de uso com as principais atividades envolvendo a interoperabilidade entre ferramentas de modelagem de distribuição de espécies, baseados nos requisitos elencados nas seções anteriores.

Esse cenário apresentado no caso de uso representa a utilização após a modelagem ter sido realizada na ferramenta desejada. O usuário dá seguimento ao processo usando um sistema para recuperar o modelo gerado, padronizando-o com o perfil de aplicação desenvolvido com o intuito de deixar o modelo interoperável para o fim desejado: reutilização, disponibilização etc.

Figura 18 - Caso de uso de cenário utilizando sistema para interoperar modelo gerado por ferramenta de MDE.

O contexto parte da arquitetura proposta e é variável de acordo com a aplicação e a comunidade de desenvolvimento na qual utilizará a arquitetura proposta. No contexto de aplicação desta tese, no próximo capítulo, por intermédio do desenvolvimento de uma ferramenta de teste para validação, foi definido um ambiente de desenvolvimento local com acesso à base de dados também locais e utilizando-se de ferramenta de desenvolvimento proprietária.

Na próxima seção será apresentada uma solução de arquitetura candidata, seguindo o modelo apresentado por HOFMEISTER *et al.* (2007).

4.2.4 Arquitetura proposta

Baseado nos requisitos, nos casos de uso e também no modelo de domínio do perfil de aplicação AP-SDM desenvolvido para esta tese, uma arquitetura proposta para futuros desenvolvimentos, com o objetivo de se alcançar interoperabilidade entre os modelos de distribuição de espécies gerados por ferramentas de modelagem, é demonstrada na Figura 19.

Figura 19 - Arquitetura proposta para interoperabilidade entre modelos gerados por ferramentas de MDE.

Fonte: Elaborada pelo autor.

O usuário tem acesso aos metadados por meio de portais de biodiversidade ou através da pesquisa e da própria coleta de dados. Com os metadados, o usuário utiliza uma ferramenta de modelagem para alcançar como resultado um modelo distribuído de espécies, independente de ferramenta, algoritmo aplicado, parâmetros etc. Todas essas informações devem estar presentes no modelo gerado e, caso não estejam, devem ser adicionadas ao arquivo de metadados de saída.

Esse arquivo de metadados de saída, baseado na ferramenta utilizada, não possui padrão de metadados, portanto, só pode ser entendido e compreendido pela própria ferramenta operada. Por intermédio dos serviços de metadados, esse arquivo gerado passa por um processo de padronização e por meio do perfil de aplicação ou padrão de metadados de biodiversidade, é gerado um arquivo XML que passa ser interoperável, pois possui um padrão reconhecido. Esse arquivo pode ser

depositado em um repositório de arquivos locais e também disponibilizado novamente em portais de biodiversidade por uma identificação internacional como um D.O.I..

Conforme sugerido por Hofmeister *et al.* (2007), essa é uma arquitetura proposta, e pode, ou deve passar por um processo de revisão e refinamento até se alcançar uma arquitetura ideal para chegar-se ao objetivo proposto.

Este capítulo, portanto, apresentou os passos de desenvolvimento de um perfil de aplicação para interoperabilidade entre modelos gerados por ferramentas de MDE, e também uma proposta de arquitetura computacional para futuras comunidades de desenvolvimento interessadas. No próximo capítulo, serão vistos estudos de caso para validação do desenvolvimento apresentado neste capítulo.

5 ESTUDO DE CASOS

O objetivo deste estudo de caso é apresentar o desenvolvimento de um sistema de testes baseados no perfil de aplicação e na arquitetura apresentados no Capítulo 4, fazendo uso de duas ferramentas de modelagem de distribuição de espécies, o sistema de teste criado, para realizar uma validação da interoperabilidade dos modelos gerados pelas ferramentas de modelagem de MDE.

A primeira parte deste capítulo apresenta uma breve descrição do desenvolvimento de um ambiente de testes para modelagem, obtendo um modelo gerado com a utilização de dados de ocorrência: dataset, algoritmo de modelagem e ferramenta de MDE, usados no ambiente científico da Informática para Biodiversidade.

A segunda parte demonstra dois estudos de casos, valendo-se de dados de ocorrência, dataset, obtidos em portais de biodiversidade ou fornecidos pelas ferramentas de modelagem para testes, e também na validação dos arquivos gerados pelo ambiente de testes.

5.1 Desenvolvimento do Ambiente para Testes

Para o estudo e o desenvolvimento de um perfil de aplicação para a interoperabilidade entre as ferramentas de modelagem de distribuição de espécies foi desenvolvido também um protótipo de testes, a fim de obter os dados de saída, ou seja, o modelo de espécies distribuído de uma ferramenta.

A partir do modelo de distribuição de determinada ferramenta de MDE, como openModeller, MaxEnt etc., esse ambiente permite que se possa verificar se o perfil de aplicação desenvolvido para a interoperabilidade entre as ferramentas seja funcional e também possibilita constatar se é possível a interoperabilidade entre ferramentas de MDE.

Esse protótipo de teste, foi nomeado SIME (Sistema de Interoperabilidade para Modelagem de Distribuição de Espécies) e foi desenvolvido pelo grupo de

pesquisa³¹, organizado para estudar a interoperabilidade entre as ferramentas de MDE.

A escolha dos pares de ferramentas para a utilização desta pesquisa para validação do perfil de aplicação por intermédio do SIME, foi definida a partir do desenvolvimento e testes baseados nos modelos gerados pelas ferramentas de modelagem de distribuição de espécies, openModeller e MaxEnt. Elas foram selecionadas pelo critério de que a ferramenta openModeller é fruto de pesquisas do departamento de Engenharia de Computação da Escola Politécnica da USP e o MaxEnt é uma das ferramentas mais utilizadas para a modelagem de distribuição de espécies.

O código e a utilização do perfil de aplicação para o SIME têm como objetivo serem disponibilizados para a comunidade científica da área para futuras modificações e aplicações de saídas de modelos de outras ferramentas de modelagem, contribuindo para o estudo e o desenvolvimento de novas características relacionadas à interoperabilidade entre ferramentas de modelagem.

Antes do desenvolvimento do sistema, foi realizado um estudo para verificar as principais diferenças entre os arquivos de entrada e de saída das ferramentas de MDE openModeller e MaxEnt. O processo de utilização não difere muito de uma da outra, pode-se ver na Figura 1 que existem algumas diferenças básicas entre o processamento de cada ferramenta. No caso dos arquivos de dados de ocorrência, a openModeller reconhece o ponto-e-vírgula como tabulação entre os metadados de ocorrência, e o MaxEnt reconhece os metadados tabulados com uma vírgula. Outra diferença é que enquanto o openModeller exige que cada metadado listado no arquivo dos dados de ocorrência possuam uma identificação, no MaxEnt isso não ocorre. O reconhecimento do arquivo ocorre, caso ele esteja no formato “.txt” para o openModeller e “.csv” para o MaxEnt.

Ainda apresentando as principais diferenças entre as ferramentas openModeller e MaxEnt, no caso dos arquivos utilizados para a modelagem de

³¹O grupo de pesquisa citado faz parte de uma parceria entre a USP e o UNASP (Centro Universitário Adventista de São Paulo) e tem como principal desenvolvedor Agnei Silva. Foram desenvolvidos os estudos sobre os perfis de aplicação e também realizado o desenvolvimento do protótipo de teste para validação do perfil de aplicação.

distribuição de espécies apresentado na Figura 20 e no caso dos datasets, no openModeller deve estar no formato “.adf” e no MaxEnt no formato “.asc”.

Em relação aos arquivos de saída, ou seja, os modelos de distribuição de espécies processados, nosso principal material de estudo neste trabalho, existem muitas semelhanças entre o conteúdo disponibilizado, como mapas, HTML (*HyperText Markup Language*) com os dados do modelo, o algoritmo utilizado, o tempo de execução, a versão da ferramenta e os próprios dados de ocorrência e dataset. A única diferença encontrada, relacionada à informação disponibilizada e a quantidade de memória computacional utilizada, esse dado o MaxEnt apresenta de maneira detalhada.

Figura 20 - Principais diferenças entre os modelos gerados pelas ferramentas de modelagem openModeller e MaxEnt.

Fonte: Elaborado pelo autor.

Com essas informações foi possível iniciar o processo de desenvolvimento de um ambiente de teste que se proporciona a utilização dos arquivos gerados por ambas as ferramentas e também dos padrões de metadados EML, Darwin Core, e Dublin Core, por intermédio do perfil de aplicação AP-SDM, desenvolvido neste trabalho. Esse ambiente de teste não tem como objetivo ser a principal solução para interoperabilidade desta pesquisa, mas proporcionar um sistema que fosse possível

manipular arquivos de modelos de distribuição de espécies e validar esses arquivos nos ambientes de validação disponibilizados pelas comunidades de padrões de metadados.

Portanto, o principal objetivo do sistema SIME é realizar testes com saídas de dados de modelos gerados pelas ferramentas de modelagem de distribuição de espécies openModeller e MaxEnt e, a partir dessas saídas, gerar novas entradas para ambas as ferramentas, em outras palavras, o modelo gerado pela ferramenta openModeller pode ser utilizado para realizar outra modelagem na ferramenta MaxEnt, ou vice-versa, ou ainda disponibilizar que saídas de uma ferramenta sejam reutilizadas na mesma ferramenta com a adição de novos parâmetros.

Os dados mencionados no parágrafo anterior se referem a todo o conjunto de informações que compõe uma modelagem de distribuição de espécies, como por exemplo: os dados de ocorrência, as camadas climáticas utilizadas, o algoritmo utilizado para a modelagem, os parâmetros utilizados no algoritmo, o mapa gerado e outras informações adicionais disponibilizadas pelos autores do processo de modelagem.

Esse sistema percorre o seguinte fluxo: o SIME processa os dados fornecidos e disponibiliza uma saída em um padrão de metadados que se baseia no perfil de aplicação apresentado no Capítulo 4, disponibilizando um arquivo de dados capaz de ser reutilizado por outra ou pela mesma ferramenta. Se houver uma codificação para que seja possível uma releitura utilizando-se do perfil de aplicação desenvolvido nesta tese ou apenas avaliando os dados e disponibilizando-os para que sejam reutilizáveis em qualquer ferramenta de MDE, haverá uma interoperabilidade dos dados gerados de forma primária por uma ferramenta de modelagem.

5.2 Estudo de Caso 1

Para a realização do primeiro estudo de caso para esta pesquisa, todo o processo de modelagem foi realizado utilizando-se da ferramenta de MDE openModeller. O principal objetivo deste primeiro estudo é desenvolver todo o

processo de modelagem de distribuição de espécies empregando o openModeller e depois, através do ambiente de teste, acessar o modelo gerado no desenvolvimento e torná-lo interoperável, utilizando-se dos padrões de metadados, EML, Darwin Core e Dublin Core por intermédio do perfil de aplicação AP-SDM.

Os dados de ocorrência aplicados neste estudo de caso estão disponibilizados no doi:10.7910/DVN/MC76QC, bem como as informações do dataset de dados climáticos, geográficos e de temperatura. Os dados de ocorrência são da espécie *Furcata Boliviana*, disponíveis com o material da comunidade de desenvolvimento da ferramenta openModeller.

É importante salientar que este estudo de caso não tem seu foco na qualidade dos dados de ocorrência, nem na análise final do modelo gerado, pois o objetivo principal é a verificação da interoperabilidade possível de um modelo gerado por meio de um perfil de aplicação e de padrões de metadados para a área de biodiversidade, portanto os passos descritos a seguir visam esse objetivo.

A primeira etapa para o desenvolvimento do modelo de distribuição de espécies é a obtenção dos dados de ocorrência e a verificação se eles estão qualificados para o uso na ferramenta selecionada, no caso a openModeller. Na Figura 21 é possível ter um exemplo dos dados de ocorrência utilizados para este estudo de caso. É importante notar que para o openModeller é necessária uma identificação para cada dado de ocorrência, uma descrição, as posições longitudinal e latitudinal e a abundância da espécie. Não estando nesse formato, a ferramenta openModeller não realiza o processamento das informações.

Figura 21 - Dados de ocorrência do estudo de caso.

#id	label	long	lat	abundance
1	furcata boliviana	-68.85	-11.15	1
2	furcata boliviana	-67.38	-14.32	1
3	furcata boliviana	-67.55	-14.33	1
4	furcata boliviana	-67.58	-14.50	1
5	furcata boliviana	-68.17	-15.25	1
6	furcata boliviana	-67.75	-15.27	1
7	furcata boliviana	-68.46	-15.27	1
8	furcata boliviana	-68.27	-15.37	1
9	furcata boliviana	-67.80	-15.45	1

O segundo passo é a definição dos dados climáticos, geográficos e de temperatura que serão utilizados para a modelagem de distribuição de espécies. Um exemplo das características levadas em conta para este estudo de caso pode ser visualizado na Figura 22. Nele foram adquiridas informações de temperatura e de pluviosidade. Todos os arquivos do DataSet devem estar no formato “.adf”, padrão aceito pela ferramenta openModeller.

Figura 22 - Informações do DataSet do estudo de caso.

Fonte: Elaborada pelo autor.

Com os dados de ocorrência, as informações de DataSet disponíveis e qualificadas para o uso na ferramenta, o próximo passo é a execução da modelagem em si. A Figura 23 apresenta o uso da ferramenta desktop openModeller. Conforme mencionado no Capítulo 2, ela tem a capacidade de trabalhar com diversos algoritmos de modelagem e, para este estudo de caso, foi utilizado o algoritmo de Entropia Máxima, descrito também no Capítulo 2 deste trabalho.

Figura 23 - Parâmetros utilizados para a modelagem do estudo de caso.

Fonte: Elaborada pelo autor.

A ferramenta openModeller processa as informações e as apresenta de forma detalhada conforme pode-se ver nas Figura 24 e 25.

Figura 24 - Modelo gerado pela ferramenta openModeller

Fonte: Elaborada pelo autor.

Figura 25 - Modelo gerado pela ferramenta openModeller para o estudo de CASO.

Fonte: Elaborada pelo autor.

Todas as informações relacionadas ao modelo são geradas e localizadas na pasta selecionada, em um arquivo XML, conforme se pode visualizar na Figura 26. A descrição completa do modelo gerado encontra-se no doi:10.7910/DVN/MC76QC com todos os dados relacionados a esse estudo de caso.

Figura 26 - Parte do arquivo XML com o modelo gerado pela ferramenta openModeller.

```
<?xml version="1.0"?>
- <Experiment openModellerDesktopVersion="110" Description="Case Study created on Monday, 24 April, 2017" Name="Estudo de Caso 1 - Furcata"
  Guid="c1497865-1650-4661-9b68-ca8e5a7f7387">
  - <Models>
    - <Model Guid="39a87b74-f96c-40bc-9cae-da67d00cfd5c" Error="false" Completed="true">
      - <CoordinateSystem>
        <Name>Lat/Long WGS84</Name>
        <WKT>GEOGCS["WGS 84",DATUM["WGS_1984",SPHEROID["WGS 84",6378137,298.257223563,AUTHORITY
          ["EPSG","7030"]],TOWGS84[0,0,0,0,0,0],AUTHORITY["EPSG","6326"]],PRIMEM["Greenwich",0,AUTHORITY
            ["EPSG","8901"]],UNIT["degree",0.0174532925199433,AUTHORITY["EPSG","9108"]],AXIS["Lat",NORTH],AXIS
              ["Long",EAST],AUTHORITY["EPSG","4326"]]</WKT>
        </CoordinateSystem>
      - <Options>
        <OccurrencesFilter></OccurrencesFilter>
        </Options>
        <SpeciesFile>C:/Program Files (x86)/openModeller Desktop
          1.1.0/SampleData/EnvironmentLayers/SouthAmerica/furcata_boliviana.txt</SpeciesFile>
        <TaxonName>furcata boliviana</TaxonName>
      - <CreationLayerSet>
        + <Environment Description="Estudo de Caso 1" Name="LayerSet - Estudo de Caso 1" Guid="9c1ddf6d-8082-4e42-b617-bd519901262e"
          NumLayers="8">
        </CreationLayerSet>
      - <Algorithm Id="MAXENT" Version="0.2">
        <Name>Maximum Entropy</Name>
```

Fonte: Elaborada pelo autor.

É interessante notar que os dados disponibilizados pela ferramenta openModeller englobam as informações necessárias para uma reutilização do modelo e favorecem a obtenção das informações para a disponibilização em um formato padronizado.

A partir deste momento, com os dados do modelo adquiridos, o próximo passo é o processo de transformação da informação para um padrão de metadados utilizando-se do perfil de aplicação descrito no Capítulo 4. Para o processamento das informações foi desenvolvido um ambiente de testes, que neste estudo de caso tem como principal objetivo capturar as informações geradas pela ferramenta openModeller e disponibilizá-las padronizadas por meio do perfil de aplicação.

Como estudo de caso inicial e para validação dos dados, este primeiro estudo foi padronizado na estrutura do EML, pois a comunidade de desenvolvimento do EML disponibiliza uma ferramenta para validação de arquivos EML, o que valida nosso estudo especialmente para o padrão EML. Os demais dados propostos no perfil de aplicação AP-SDM serão adicionados posteriormente e disponibilizados para trabalhos futuros.

Os resultados com a utilização do ambiente de testes estão descritos e apresentados na próxima seção.

5.2.1 Resultados

Com os resultados do modelo selecionado, inicia a primeira fase para a utilização do ambiente de testes e da ferramenta SIME. Nela acontece a definição do responsável pela pesquisa, da forma que sugere o perfil de aplicação AP-SDM e de acordo com os padrões de metadados, Dublin Core, Darwin Core e EML. Abaixo segue a descrição do responsável por intermédio de um arquivo XML. Essa definição do arquivo responsável é utilizada também para o estudo de caso 2, descrito na seção 5.3.1.

O comando para criação do arquivo de responsável no SIME é: `sime /r`.

```

<?xml version="1.0" encoding="utf-16"?>
<responsible xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <title>Title - Required.</title>
  <keywords keyword="Perfil de Aplicação" />
  <keywords keyword="Padrão de Metadados" />
  <keywords keyword="Interoperabilidade" />
  <keywords keyword="Modelagem de Distribuição de Espécies" />
  <person salutation="Mrs." givenname="Cleverton" surname="Borba" />
  <address deliverypoint="Universidade de São Paulo - USP" city="São
Paulo" administrativeArea="SP" postalCodecountry="05508-900"
country="Brazil" phoneNumber="+55 11 30914700"
mailaddress="cleverton.borba@usp.br" />
</responsible>

```

A partir desse ponto, o ambiente de testes é acessado e a busca pelos modelos gerados é iniciada, bem como a padronização dos modelos selecionados para o padrão de metadados proposto por esta pesquisa.

Com o ambiente de testes em execução para este estudo de caso, o primeiro passo é a definição do padrão de metadados que será aplicado para interoperar o modelo gerado pela ferramenta openModeller. Nesse primeiro momento foi definido o padrão de metadados EML como teste para o arquivo, pois ele possui um ambiente de validação disponível pela comunidade de desenvolvimento do padrão EML e disponível no link: <https://knb.ecoinformatics.org/emlparser/>.

Nesta primeira etapa o SIME gera um arquivo no padrão EML para ser válido, operando com a ferramenta citada no paragrafo anterior. A partir desta etapa, é possível, por meio da definição do perfil de aplicação, inserir as informações relacionadas aos outros padrões de metadados definidos para esta pesquisa que são o Darwin Core e o Dublin Core.

A Figura 27 apresenta parte do arquivo gerado pela ferramenta SIME (o arquivo completo encontra-se no doi:10.7910/DVN/MC76QC). Este arquivo no padrão da linguagem de metadados EML fornece a este estudo de caso um primeiro estágio para ser validado.

Figura 27 - Arquivo EML no formato XML do estudo de caso utilizando openModeller para ser validado.

```
<?xml version="1.0" encoding="utf-8"?>
<eml:eml xmlns:schemaLocation="eml://ecoinformatics.org/eml-2.1.1 eml.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:stmml="http://www.xml-cml.org/schema/stmml-1.1" packageId="eml.1.1" system="SIME - Sistema de Interoperabilidade entre Sistemas de Modelagens de Espécies." xmlns:eml="eml://ecoinformatics.org/eml-2.1.1">
  <dataset>
    <title>Title - Required.</title>
    <creator id="cleverton.borba">
      <individualName>
        <salutation>Mrs.</salutation>
        <givenName>Cleverton</givenName>
        <surName>Borba</surName>
      </individualName>
      <address>
        <deliveryPoint>Escola Politecnica da Universidade de Sao Paulo</deliveryPoint>
        <city>Sao Paulo</city>
        <administrativeArea>SP</administrativeArea>
        <postalCode>05508-900</postalCode>
        <country>Brazil</country>
      </address>
      <phone>+1 269 3624295</phone>
      <electronicMailAddress>cleverton.borba@usp.br</electronicMailAddress>
    </creator>
    <keywordSet>
      <keyword>Interoperability</keyword>
      <keyword>Species Distribution Modeling</keyword>
      <keyword>Metadata Standard</keyword>
      <keyword>Application Profile</keyword>
    </keywordSet>
    <contact>
      <references>cleverton.borba</references>
    </contact>
    <dataTable id="om0.0ccurrenceData" system="openModeller">
      <entityName>Entity do arquivo - furcata_boliviana.csv</entityName>
      <entityDescription>Entity do arquivo - furcata_boliviana.csv</entityDescription>
      <physical>
        <objectName>furcata_boliviana.csv</objectName>
        <size unit="bytes">4348</size>
        <characterEncoding>ASCII</characterEncoding>
        <dataFormat>
          <textFormat>
            <attributeOrientation>column</attributeOrientation>
            <simpleDelimited>
              <fieldDelimiter>;</fieldDelimiter>
            </simpleDelimited>
          </textFormat>
        </dataFormat>
        <distribution>
          <offline>
```

Fonte: Elaborada pelo autor.

Realizado esse processo, a sequência é a validação do arquivo utilizando a ferramenta de validação disponibilizada pela comunidade de desenvolvimento da linguagem de metadados EML. A Figura 28 apresenta o resultado e a localização do ambiente de validação do arquivo gerado.

Figura 28 - Validação do arquivo EML gerado para o estudo de caso com a padronização do modelo gerado, utilizando a ferramenta openModeller.

Fonte: Elaborada pelo autor.

Por intermédio desse processo, foi possível validar o arquivo e verificar que o modelo gerado pela ferramenta openModeller foi padronizado para a linguagem de metadados EML. O próximo passo executado foi a inclusão das demais informações relacionadas no perfil de aplicação AP-SDM desenvolvido para este trabalho.

A Figura 29 apresenta o pacote completo das informações relacionadas ao modelo gerado pela ferramenta SIME, juntamente com os dados climáticos aplicados na modelagem, os de ocorrência e as informações relacionadas à ferramenta, como também ao algoritmo operado e os parâmetros relacionados a ele.

Figura 29 - Pacote com as informações geradas pelo ambiente de teste contendo todas os dados gerados pela ferramenta openModeller, padronizados pelo perfil de aplicação AP-SDM.

Fonte: Elaborada pelo autor.

Como resultado final tem-se um arquivo com todas as informações relacionadas ao modelo gerado pela ferramenta openModeller, padronizados por intermédio do perfil de aplicação AP-SDM, alcançando-se o objetivo proposto por esta pesquisa no contexto de disponibilizar um arquivo padronizado dos modelos gerados por ferramentas de MDE, proporcionando uma interoperabilidade entre as ferramentas de modelagem.

5.3 Estudo de Caso 2

Ao contrário do primeiro estudo de caso, neste segundo estudo, a ferramenta selecionada para o desenvolvimento do modelo de distribuição de espécies, é a desktop MaxEnt. Quanto ao principal objetivo deste estudo é também desenvolver

todo o processo de modelagem de distribuição de espécies, utilizando o MaxEnt e depois, por meio do ambiente de teste, acessar o modelo gerado no desenvolvimento e torná-lo interoperável, servindo-se dos padrões de metadados, EML, Darwin Core, Dublin Core, por intermédio do perfil de aplicação AP-SDM.

Da mesma forma que o primeiro estudo de caso, os dados de ocorrência e as informações de *DataSet* estão listadas no doi:10.7910/DVN/MC76QC. Os dados de ocorrência foram viabilizados pela comunidade de desenvolvimento da ferramenta MaxEnt e estão disponíveis para uso científico.

O principal objetivo deste segundo estudo é realizar uma modelagem de distribuição de espécies usando a ferramenta MaxEnt e, por meio do ambiente de testes, ter acesso ao modelo gerado e padronizá-lo para um padrão de metadados possível de validação, por intermédio do perfil de aplicação desenvolvido.

O primeiro passo para a modelagem de distribuição de espécies empregada neste estudo de caso é a preparação dos dados de ocorrência para a aplicação da ferramenta MaxEnt. De maneira diferente, os dados de ocorrência devem estar formatados de forma diversa ao do primeiro estudo, conforme visto na primeira seção deste capítulo. Um exemplo dos dados de ocorrência utilizados para este estudo de caso pode ser visualizado na Figura 30.

Figura 30 - Parte dos Dados de ocorrência utilizados no estudo de caso.

1	species,dd long,dd lat
2	bradypus_variegatus,-65.4,-10.3833
3	bradypus_variegatus,-65.3833,-10.3833
4	bradypus_variegatus,-65.1333,-16.8
5	bradypus_variegatus,-63.6667,-17.45
6	bradypus_variegatus,-63.85,-17.4
7	bradypus_variegatus,-64.4167,-16
8	bradypus_variegatus,-63.1667,-17.8
9	bradypus_variegatus,-56.7333,-2.6
10	bradypus_variegatus,-59.1333,-3.7
11	bradypus_variegatus,-60.0833,-3.1333
12	bradypus_variegatus,-64.8333,-0.5167

Fonte: Elaborada pelo autor.

A segunda etapa a ser realizada é a definição das camadas climáticas e de temperatura que serão utilizadas no processo de modelagem. Esse conjunto de

dados também foi oferecido pela comunidade de desenvolvimento da ferramenta MaxEnt e um exemplo dos DataSets utilizados podem ser visualizados nas Figuras 31 e 32.

Figura 31 - Conjunto de informações utilizadas na modelagem para este estudo de caso.

Fonte: Elaborada pelo autor.

Figura 32 - Formato dos arquivos permitidos na ferramenta MaxEnt.

Fonte: Elaborada pelo autor.

Na sequência é possível iniciar a execução da modelagem na ferramenta MaxEnt, por intermédio da versão desktop disponibilizada em: http://biodiversityinformatics.amnh.org/open_source/maxent/. A Figura 11 apresenta a tela de parâmetros de modelagem empregada neste segundo estudo de caso.

Para este estudo de caso a espécie selecionada foi a *bradypus variegatus* e as camadas ambientais são as seguintes: cld6190_ann dtr6190_ann.ecoreg

frs6190_ann h_dem pre6190_ann pre6190_l1 pre6190_l10 pre6190_l4 pre6190_l7 tmn6190_ann tmp6190_ann tmx6190_ann vap6190_ann. O ambiente de modelagem para ambas as ferramentas são bem amigáveis e permitem uma compreensão rápida e simples para os usuários.

Definidos os dados de ocorrência e os parâmetros, a modelagem pode ser iniciada. No caso da ferramenta MaxEnt, o único algoritmo disponível para realizar a modelagem é o de Entropia Máxima, também descrito no Capítulo 2 deste trabalho. A versão da ferramenta utilizada foi a 3.4.

Para este estudo de caso, foram utilizados 114 pontos de ocorrência da espécie selecionada, e 10112 pontos usados para determinar a distribuição de entropia máxima.

Figura 33 - Parâmetros de modelagem utilizados no estudo de caso.

Fonte: Elaborada pelo autor.

Após a modelagem ser realizada, os arquivos relacionados ao modelo de distribuição de espécies são disponibilizados na pasta definida pelo autor da pesquisa e na Figura 34 é possível verificar parte do modelo gerado.

A Figura 34 é a representação do modelo MaxEnt para a espécie *bradpus_variegatus*, sendo que as cores mais quentes representam as áreas com melhores condições de predição. Os pontos brancos apresentam a localização dos pontos de ocorrência usados neste estudo de caso.

Figura 34 - Modelo de Distribuição de Espécie gerado pela ferramenta MaxEnt para este estudo de caso.

Fonte: Elaborada pelo autor.

Realizado o processo de modelagem utilizando a ferramenta MaxEnt, o próximo passo é acessar o modelo gerado por meio do ambiente de testes da ferramenta SIME e produzir um modelo padronizado de distribuição de espécies através do padrão de metadados EML e do perfil de aplicação proposto por este trabalho.

Os resultados do estudo de caso 2, realizado com a ferramenta MaxEnt são apresentados na próxima seção.

5.3.1 Resultados

Assim como no primeiro estudo de caso apresentado nesta pesquisa na seção 5.2.1, os dados relacionados ao responsável pelo projeto e autor da modelagem realizada devem ser inseridos no arquivo correspondente para que o ambiente de testes gere o arquivo com as informações corretas.

O sistema desenvolvido para esta tese então acessa as informações relacionadas ao modelo gerado pela ferramenta MaxEnt e como primeiro passo gera um arquivo no formato EML que pode ser validado. A Figura 35 apresenta parte do arquivo XML gerado com as informações do modelo MaxEnt (o arquivo completo encontra-se disponível no doi:10.7910/DVN/MC76QC).

Figura 35 - Arquivo no formato EML para validação da padronização do modelo gerado pela ferramenta MaxEnt.

```

1  <?xml version="1.0" encoding="utf-8"?>
2  <eml:eml xmlns:schemaLocation="eml://ecoinformatics.org/eml-2.1.1 eml.xsd" xmlns:
3  <dataset>
4    <title>SDM - Bradipys - MaxEnt</title>
5    <creator id="cleverton.borba">
6      <individualName>
7        <salutation>Mrs.</salutation>
8        <givenName>Cleverton</givenName>
9        <surName>Borba</surName>
10     </individualName>
11     <address>
12       <deliveryPoint>Escola Politécnica da Universidade de Sao Paulo</delivery
13       <city>Sao Paulo</city>
14       <administrativeArea>SP</administrativeArea>
15       <postalCode>05508-990</postalCode>
16       <country>Brazil</country>
17     </address>
18     <phone>+1 269 3624295</phone>
19     <electronicMailAddress>cleverton.borba@usp.br</electronicMailAddress>
20   </creator>
21   <keywordSet>
22     <keyword>Interoperability</keyword>
23     <keyword>Species Distribution Modeling</keyword>
24     <keyword>Metadata Standard</keyword>
25     <keyword>Application Profile</keyword>
26   </keywordSet>
27   <contact>
28     <references>cleverton.borba</references>
29   </contact>
30   <dataTable id="me0.OccurrenceData" system="MaxEnt">
31     <entityName>Entity do arquivo - bradypus.csv</entityName>
32     <entityDescription>Entity do arquivo - bradypus.csv</entityDescription>
33     <physical>
34       <objectName>bradypus.csv</objectName>
35       <size unit="bytes">4130</size>
36       <characterEncoding>ASCII</characterEncoding>
37     </physical>
38     <dataFormat>
39       <textFormat>
40         <attributeOrientation>column</attributeOrientation>
41         <simpleDelimited>
42           <fieldDelimiter>;</fieldDelimiter>
43         </simpleDelimited>
44       </textFormat>
45     </dataFormat>
46     <distribution>
47       <offline>

```

Fonte: Elaborada pelo autor.

Por intermédio do validador de arquivos disponibilizado pela comunidade de desenvolvimento da linguagem de metadados EML, foi possível verificar a funcionalidade do arquivo gerado pela ferramenta SIME com as informações relacionadas ao modelo gerado pela ferramenta MaxEnt. Segue na Figura 36 a validação concluída.

Figura 36 - Validação do arquivo gerado pela ferramenta SIME com os dados do modelo gerado pela ferramenta MaxEnt.

Fonte: Elaborada pelo autor.

Realizada a validação do arquivo gerado, a situação seguinte é adicionar as informações relacionados ao perfil de aplicação AP-SDM e disponibilizar um pacote padronizado com todas as informações do modelo gerado pela ferramenta MaxEnt. A Figura 37 apresenta todas as informações geradas.

Figura 37 - Pacote com todas as informações do modelo padronizadas.

Fonte: Elaborada pelo autor.

Como resultado final tem-se um pacote de informações com os dados de ocorrência, os dados climáticos utilizados, os parâmetros do algoritmo de Entropia Máxima da ferramenta MaxEnt e o arquivo padronizado com todas as informações relacionadas.

5.4 Disponibilização e Publicação de Experimentos

Como parte do objetivo desta tese, é importante o incentivo com relação à disponibilização e publicação dos experimentos realizados na modelagem de distribuição de espécies e, neste caso, a publicação e a viabilização dos estudos utilizando-se do perfil de aplicação AP-SDM, desenvolvido nesta pesquisa e também nos estudos de caso apresentados nas seções 5.2 e 5.3.

A utilização de um ambiente que disponha um número DOI (*Digital Object Identifier*) foi fundamental para que a liberação seja realizada de forma padronizada e que os dados estejam disponíveis para toda a comunidade científica interessada na interoperabilidade entre ferramentas de modelagem de distribuição de espécies.

Durante o período de desenvolvimento desta pesquisa, foi realizado um estudo juntamente com o laboratório de pesquisas da USGS (*United States Geological Survey*) e a Universidade do Tennessee em Knoxville, sobre um ambiente que possibilite um DOI para os pesquisadores vinculados ao governo. Esse ambiente é o EZID, fornecido pela Universidade da Califórnia. No entanto, este não possui mais licenças disponíveis para o uso, sendo assim para este trabalho foi empregado outro ambiente para a publicação dos dados.

Existem alguns ambientes disponíveis para publicação dos dados e a disponibilização de um DOI, dentre eles pode-se citar: DataCite (datacite.org), Dryad (www.datacryad.org), Figshare (figshare.com), DataVerse (dataverse.harvard.edu) e o CrossRef (crossref.org). Dentre estes, o DataCite, o Dryad e o CrossRef são necessários cadastro e pagamento pelo uso do ambiente. O DataVerse faz parte de um ambiente acadêmico (Universidade Harvard) e, por isso, está disponível

gratuitamente para a publicação e disponibilização dos dados. Portanto, foi utilizado o DataVerse para gerar um DOI para os estudos de caso desta tese.

O sistema para publicação dos dados do DataVerse tem boa usabilidade e proporciona um ambiente de fácil interação e necessita apenas de um cadastro simples de usuário para permitir a publicação de dados. Para esta tese foi criado um metadado com o título “*Study Case of Interoperability between Species Distribution Modeling Tools*”.

Por intermédio do DataVerse é possível inserir todos os arquivos na pesquisa e no estudo de caso, permitindo acesso da comunidade a todas as informações relacionadas a este trabalho e também incentivando o uso desse tipo de ferramenta para que pesquisadores utilizem a padronização dos modelos gerados pelas ferramentas de MDE e que disponibilizem os dados em plataformas digitais para que outros reutilizem os arquivos interoperáveis.

A Figura 38 apresenta parte da plataforma DataVerse com os dados publicados dos estudos de caso desta pesquisa. O DOI com as informações deste estudo podem ser acessadas através de: doi:10.7910/DVN/MC76QC.

Figura 38 - Tela do DataVerse com a publicação dos dados gerados pelos estudos de caso desta pesquisa.

The screenshot displays the Harvard Dataverse web interface. At the top, the 'Dataverse' logo is on the left, and navigation links (Search, About, Guides, Support) and a user profile (Cleverton Borba) are on the right. Below the header, the 'Harvard Dataverse' logo and name are shown. The main content area features the dataset title 'Study Case of Interoperability between Species Distribution Modeling Tools' with a 'Draft' status and 'Unpublished' label. Below the title, the dataset description is provided: 'Borba, Cleverton, 2017, "Study Case of Interoperability between Species Distribution Modeling Tools", doi:10.7910/DVN/MC76QC, Harvard Dataverse, DRAFT VERSION, UNF:6:wpkBvFiptIC3StRUDnFP6w=='. A 'Cite Dataset' button is visible. The 'Description' section contains the text: 'This is part of a study for interoperability between Species Distribution Modeling tools. It has files of the species distribution modeling realized using openModeller tool and standardized using the AP-SDM application profile defined as part of my degree in Ph.D. in Computer Engineering.' The 'Subject' is listed as 'Computer and Information Science', and the 'Keyword' is 'Interoperability, Species Distribution Modeling, Application Profile'.

Fonte: Elaborada pelo autor.

6 CONSIDERAÇÕES FINAIS

As considerações finais sobre esta pesquisa estão expostas neste capítulo e seguem a seguinte ordem: as contribuições da pesquisa são apresentadas na seção 6.1, todos os trabalhos publicados e relacionados ao desenvolvimento desta pesquisa são apresentados na seção 6.2, as sugestões para trabalhos futuros encontram-se na seção 6.3 e por fim as discussões gerais sobre o desenvolvimento da pesquisa estão descritos na seção 6.4.

6.1 Contribuições da pesquisa

O desenvolvimento desta tese produziu contribuições para a comunidade científica da área de Informática para Biodiversidade, modelagem de distribuição de espécies e engenharia de computação. Sendo que as principais contribuições seguem nos parágrafos seguintes.

Como primeira contribuição, é possível citar o próprio conteúdo teórico apresentado nesta pesquisa, pois disponibiliza aos pares relacionados à área um material atualizado dos principais trabalhos e pesquisas relacionados a área de Informática para Biodiversidade e também para a modelagem de distribuição de espécies.

Uma análise dos padrões de metadados e suas contribuições para a área de MDE também são um aporte dessa pesquisa, pois acessibiliza um estudo detalhado dos benefícios do uso de padrões de metadados para tornar disponíveis os dados de biodiversidade.

Esta tese viabilizou não apenas a interoperabilidade entre ferramentas de MDE, que é uma das propostas da pesquisa, mais especificamente para a comunidade científica vinculada à área de modelagem de distribuição de espécies, mas também promoveu o acesso ao desenvolvimento de um perfil de aplicação, que une padrões de metadados existentes e formalizados, para atender uma lacuna existente na padronização dos modelos de distribuição de espécies gerados pelas ferramentas de modelagem. Em outras palavras, por meio da utilização do perfil de

aplicação AP-SDM é possível padronizar os modelos de distribuição de espécies e contribuir para o acesso dos dados processados em bancos de dados de biodiversidade para a reutilização, promovendo a interoperabilidade de modelos e a motivação para o reuso de pesquisas já existentes.

Uma arquitetura computacional para ferramentas que venham operar com o perfil de aplicação desenvolvido nesta pesquisa também é uma colaboração desta pesquisa, pois disponibiliza para os engenheiros de computação um modelo a seguir para o desenvolvimento de ferramentas associadas a essa área de pesquisa.

Em relação aos estudos de caso, é possível citar que com o desenvolvimento de dois estudos de caso, um envolvendo a ferramenta openModeller e outro a ferramenta MaxEnt, foi possível verificar o funcionamento e os processos que envolvem a modelagem de distribuição de espécies, bem como constatar as necessidades dos cientistas da área para obtenção, processamento e disponibilização de dados.

Dentro dos estudos de caso e também como subsídio desta pesquisa, com o protótipo de testes SIME, desenvolvido para esta pesquisa, foi possível verificar que é possível uma padronização dos dados dos modelos gerados pelas ferramentas de MDE e que com estes padronizados, é possível uma interoperabilidade entre essas ferramentas. A padronização por meio do perfil de aplicação AP-SDM e da arquitetura computacional, é possível que a comunidade de desenvolvimento de cada ferramenta de modelagem sinta-se interessada em ceder versões de seus softwares, capazes de assimilar arquivos gerados nesse padrão.

Outra cooperação desta pesquisa, pode-se citar também o incentivo à publicação e disponibilização dos dados científicos e de pesquisa pelas plataformas que geram um DOI, fornecendo para toda a comunidade acadêmica acesso às informações processadas e geradas durante a pesquisa e desenvolvimento da modelagem de distribuição de espécies.

Na próxima seção são apresentados alguns trabalhos publicados e submetidos à apreciação durante o decorrer do processo de desenvolvimento desta tese.

6.2 Trabalhos publicados relacionados a esta pesquisa

Os trabalhos publicados e também as participações em eventos durante o período de pesquisa encontram-se a seguir:

BORBA, C. F., CORRÊA, P. L. P. *Dublin Core to Ensure Interoperability between Models Generated by Tools of Species Distribution Modeling*. Proc. International Conference on Dublin Core and Metadata Applications. 2014.

BORBA, C. F., CORRÊA, P. L. P. *Use of Ontology for Connection between Models of Species Distribution Modeling Tools*. International Conference on Ecological Informatics - ICEI. 2014.

CORRÊA, A. SH., BORBA, C. F., SILVA, D. L., CORRÊA, P. L. P. *A Fuzzy Ontology-Driven Approach To Semantic Interoperability in e-Government*. International Journal of Social Science and Humanity. v.5. p. 178-181. 2015

BORBA, CLEVERTON, CORREA, PEDRO LUIZ PIZZIGATTI. *Application of Metadata Standards for Interoperability Between Species Distribution Models* In: Communications in Computer and Information Science.1 ed. : Springer International Publishing, 2015, p. 113-118.

BORBA, CLEVERTON, AGNEI SILVA, CORREA, P. L. P. *Process interoperability between biodiversity modeling systems: Engineering of complex systems* In: SDPS 2015 - Society for Design and Process Science, 2015, Dallas / Fort Worth. SDPS Conference Proceedings. , 2015.

AGNEI SILVA, BORBA, CLEVERTON, CORREA, P. L. P. *Metadata Standard for Models Generated by openModeller* In: International Conference on Dublin Core and Metadata Applications, 2015, São Paulo. DC-2015-São Paulo Proceedings. , 2015.

BORBA, CLEVERTON, CORREA, P. L. P. *Use of Metadata for Interoperability between Models of Species Distribution Modeling*. In: International Conference on Dublin Core and Metadata Applications, 2015, São Paulo.

Proc. International Conference on Dublin Core and Metadata Applications. , 2015.

BORBA, CLEVERTON, CORREA, P. L. P. *Computing Strategies to support interoperability for information sciences from species distribution modeling.* Information Sciences. 2017 – Submissão realizada em Junho de 2017.

6.3 Trabalhos futuros

Durante todo o período de desenvolvimento desta pesquisa, novas hipóteses foram surgindo de acordo com avanço deste estudo, no entanto, por questões de delimitações de escopo e objetivo, não foi possível se deter a todas as ideias e explorá-las a tal ponto que fizessem parte desta pesquisa. Portanto, seguem algumas sugestões de trabalhos futuros.

- Parceria com comunidades de desenvolvimento das ferramentas de modelagem de distribuição de espécies para a adaptação delas para que sejam capazes de utilizar e reconhecer arquivos gerados a partir do padrão de metadados proposto por esta tese ou outro padrão compatível com o perfil de aplicação AP-SDM.
- Desenvolvimento de um ambiente para manipulação de arquivos de modelos gerados por ferramentas de MDE, com o mesmo princípio do SIME, desenvolvido para esta pesquisa. Também usando como base a arquitetura computacional proposta, com o objetivo de disponibilizar para a comunidade científica um ambiente que seja possível e simplificado para manipular arquivos gerados pelos modelos.
- Desenvolvimento de um ambiente brasileiro para a publicação e disponibilização de dados científicos capazes de gerar um DOI para essas pesquisas, proporcionando uma plataforma em português para que pesquisadores da área sintam-se motivados para a publicação de seus resultados.
- Sugere-se que entre as metodologias existentes para o processo de modelagem de distribuição de espécies desperte-se o interesse em incluir nos

passos do desenvolvimento de um modelo, também a sua publicação e padronização por meio do perfil de aplicação apresentado nesta pesquisa ou pelo menos de algum dos padrões de metadados utilizados na área.

- Desenvolvimento ou adaptação das ferramentas de modelagem para que ao gerarem os modelos estes já sejam configurados e padronizados pelo perfil de aplicação AP-SDM, pois disponibilizaria um modelo já interoperável com outras ferramentas que venham a também utilizar o perfil de aplicação como padrão de entrada e saída de dados.
- Com o avanço dos dados em grande escala, um estudo de caso focando em plataformas que suportem outro tipo de dados, como o Big Data, é interessante para verificar-se de que maneira se comportaria o ambiente de testes SIME e também o perfil de aplicação para a modelagem de distribuição de espécies nesse ambiente.
- Realizar processo de desenvolvimento com as comunidades de padrões de metadados voltados para a área de informática para biodiversidade para que também suportem e verifiquem arquivos relacionados a modelos gerados por ferramentas de MDE.
- Incluir outras ferramentas de modelagem como estudo de caso para verificar o desempenho do perfil de aplicação e do ambiente de testes, e também pesquisar que tipo de dados e informações são relevantes para cada uma delas, bem como realizar uma comparação entre as ferramentas para verificar os dados semelhantes e incomuns de cada uma delas.
- No âmbito das ferramentas de modelagem de distribuição de espécies, sugere-se verificar e comparar o desempenho e a capacidade de processamento de informações relacionadas a um grande volume de informações. Realizar um estudo de caso capaz de gerar valores de comparação entre essas ferramentas de modelagem, verificando o desempenho dos algoritmos, de que forma cada um trabalha e como isso se compartilharia na padronização final do modelo gerado por essas ferramentas.

6.4 Discussões finais

Durante o processo de pesquisa desta tese, foi possível verificar a importância e o crescimento da área de informática para biodiversidade, tanto no Brasil quanto ao redor do mundo, ao focar o âmbito acadêmico, de mercado, e de pesquisa. Isso se deve ao fato da importância que a conservação e a preservação do meio ambiente também dependem das novas tecnologias e engenharias computacionais disponíveis atualmente.

Dentro da área de informática para biodiversidade, a modelagem de distribuição de espécies desempenha um importante papel em proporcionar resultados que auxiliam na tomada de decisão em níveis governamentais, profissionais e de pesquisa, oferecendo suporte para a conservação da biodiversidade.

Como em qualquer outra área da ciência, a modelagem de distribuição de espécies também se beneficia de novas tecnologias e engenharias, que a medida que avançam em termos de novas teorias e soluções computacionais, têm sido disponibilizadas para a sociedade. Nesse contexto, por meio desta pesquisa, procurou-se explorar as lacunas computacionais existentes para a comunidade da área e oferecer soluções que venham a acrescentar estímulo à pesquisa e também saciar esse espaço deixado pelos pares nos artigos apresentados nesta tese.

Para atingir-se os objetivos propostos, foi preciso entender todo o processo de modelagem e as características de cada artefato que faz parte do modelo gerado pelas ferramentas de MDE, dentre eles, o funcionamento das ferramentas mais utilizadas, o entendimento dos algoritmos de modelagem, de que forma os dados de ocorrência são coletados e disponibilizados, de que maneira os dados climáticos influenciam a modelagem e, também, de que forma cada ferramenta apresenta os modelos gerados e disponibilizam estes para análise. Nesse contexto também, um entendimento dos padrões de metadados existentes para a área de biodiversidade se fez necessário para verificar a usabilidade deles para a MDE e quais soluções seriam possíveis para uma interoperabilidade entre essas ferramentas.

O objetivo principal desta tese é apresentar um meio pelo qual seja possível realizar uma interoperabilidade para os modelos gerados pelas ferramentas de modelagem de distribuição de espécies e espera-se que o perfil de aplicação proposto por esta pesquisa venha a ser utilizado pela comunidade científica da área para o desenvolvimento ou adaptação das ferramentas existentes, proporcionando uma interoperabilidade entre as ferramentas de MDE.

O perfil de aplicação proposto, chamado de AP-SDM (sigla para *Application Profile – Species Distribution Modeling*) foi desenvolvido baseado nos estudos apresentados nesta pesquisa sobre os padrões de metadados existentes para área de Informática para Biodiversidade, mas que não atendiam por completo a padronização dos modelos gerados pelas ferramentas de MDE.

Por intermédio dos estudos de caso de duas ferramentas de modelagem, openModeller e MaxEnt, realizando todo o processo de modelagem de distribuição de espécies com exemplos factíveis de reprodução, foi possível verificar a complexidade no processo de modelagem, na preparação dos dados, na escolha dos algoritmos, na análise dos dados e também na publicação dos resultados.

Baseado nesse contexto dos estudos de caso, foi desenvolvido um ambiente de testes chamado SIME (Sistema de Modelagem de Espécies) para simular a utilização do perfil de aplicação AP-SDM e da arquitetura computacional proposta nesta pesquisa para proporcionar interoperabilidade para os modelos gerados pelas ferramentas nos estudos de caso apresentados.

Foi possível verificar, por meio dos estudos de caso, que utilizando-se de uma forma de padronização, como o perfil de aplicação AP-SDM ou de padrões de metadados (forma limitada, dependendo do modelo), é possível promover uma interoperabilidade entre as ferramentas de modelagem de distribuição de espécies, pois a padronização permite que elas tenham um modelo pelo qual podem nortear a entrada e a saída de dados.

Conclui-se também que é necessário um forte incentivo ao uso de plataformas e soluções computacionais para a publicação e disponibilização de dados científicos de MDE, pois isso proporciona a toda comunidade uma grande

quantidade de informações já processadas e disponíveis para análise e reuso, por isso a importância da padronização desses dados processados.

Entende-se, portanto, que esta tese proporcionou soluções originais para as lacunas da área da engenharia da computação, aplicadas à informática para biodiversidade. Sendo assim, procurou-se apresentar de maneira clara o conteúdo conceitual atualizado, apresentando soluções computacionais inovadoras e realizando estudos de caso reproduzíveis para a comunidade científica, tendo a preocupação em disponibilizar todas as informações necessárias para a continuidade da pesquisa e o avanço da ciência.

REFERÊNCIAS

ADHIKARI, D.; BARIK, S. K.; UPADHAYA, K. **Habitat distribution modelling for reintroduction of Ilex khasiana Purk., a critically endangered tree species of northeastern India** *Ecological Engineering*. 40: 37-43 p. 2012.

AGOSTINHO, S. et al. Developing a learning object metadata application profile based on LOM suitable for the Australian higher education context. **Australasian Journal of Educational Technology**, v. 20, n. 2, p. 191-208, 2004.

ALEIXO, D. V. B. S.; CATARINO, M. E.; BAPTISTA, A. A. R. P. Joa Archival Description Application Profile. International Conference on Dublin Core and Metadata ApplicationsDC-2015, 2015, São Paulo.

AMADOR, V. B. et al. **Extension of IEEE LOM Standard for Describing Educational Interactive Application: An Accessibility approach**. *IEEE Latin America Transactions*: IEEE. 14 2016.

ANACLETO, T. C. S.; OLIVEIRA, G. **Métodos para indicação de áreas para conservação: uma análise a partir da modelagem de nicho de tatus, no estado do Mato Grosso**. *Caminhos de Geografia*. 15: 108-117 p. 2014.

ARAÚJO, M. B. Bioensembles. 2015. Disponível em: <<http://www.maraujolab.com/resources/software/>>. Acesso em: Feb-15-2015.

ARIÑO, A. H.; CHAVAN, V.; KING, N. The Biodiversity Informatics Potential Index. **BMC Bioinformatics**, v. 12, n. Suppl 15, p. 1-17, 2011. ISSN 14712105.

AUSTIN, M. P. **Spatial prediction of species distribution: an interface between ecological theory and statistical modelling**. *Ecological Modelling*: Elsevier: 101-118 p. 2002.

BAIR, S. A.; STEUER, S. M. B. Developing a Premodern Manuscript Applicatino Profile Using Dublin Core. **Journal of Library Metadata**, v. 13, n. 1, p. 1-16, 2013.

BERENDSOHN, W. G. et al. Biodiversity information platforms: From standards to interoperability. **ZooKeys**, v. 150, p. 71-87, 2011. ISSN 13132989.

BIOMOD. Software - The BIOMOD Package. 2015. Disponível em: <<http://www.will.chez-alice.fr/Software.html>>. Acesso em: Jan-15-2015.

BLAIR, G. S. et al. **The role of ontologies in emergent middleware: supporting interoperability in complex distributed systems.** Proceedings of the 12th International Middleware Conference. Lisbon, Portugal: International Federation for Information Processing 2011.

BLONDER, B. et al. **The *n*-dimensional hypervolume** ~~The *n*-dimensional hypervolume~~. Global Ecology and Biogeography. 23: 595-609 p. 2014.

BOMAN, C. et al. **Map and Geospatial Information round table - Best practices when ugin ALAIR.** MAGIRT - American Library Association. 1 2016.

BOOCH, G.; RUMBAUGH, J.; JACOBSON, I. **UML Guia do Usuário.** 2nd. Rio de Janeiro: CAMPUS, 2006. 474.

BOOTH, T. H. et al. **bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies.** **Diversity and Distributions**, v. 20, n. 1, p. 1-9, 2014.

BORBA, C. Study case of interoperability between Species Distribution Modeling Tools. DOI: [doi:10.7910/DVN/MC76QC](https://doi.org/10.7910/DVN/MC76QC), Harvard Dataverse, 2017.

BORBA, C.; CORRÊA, L. P. L. Dublin Core to Ensure Interoperability between Models Generated by Tools of Species Distribution Modeling. International Conference on Dublin Core and Metadata Applications, 2014, Austin, TX. p.179-180.

_____. Use of Metadata for Interoperability between Models of Species Distribution Modeling. International Conference on Dublin Core and Metadata Applications, 2015, São Paulo - Brazil.

BOTKIN, D. B. et al. **Changing the Global Environment.** London: Academic Press Inc., 1989.

BOUNTOURI, L. et al. Metadata interoperability in public sectors information. **Journal of Information Science**, v. 35, n. 2, p. 204-231, 2009.

BREIMAN, L. **Random Forests.** Machine Learning. 45: 5-32 p. 2001.

BROENNIMANN, O. et al. **Measuring ecological niche overlap from occurrence and spatial environmental data.** Global Ecology and Biogeography. 21: 481-497 p. 2012.

BSI. **Interoperability between metadata systems used for learning, education and training. Code of practice for the development of application profiles**: BSI 2005.

BUONAZIA, I.; MASCI, M. E. **Un Dublin Core Application Profile per il Portale della Cultura Italiana**. Il PICO Application Profile: Scuola Normale Superiore di Pisa 2007.

CARPENTER, G.; GILLISON, A. N.; WINTER, J. **DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals**. Biodiversity & Conservation. 2: 667-680 p. 1993.

CARRIER, S. W. **The Dryad Repository Application Profile: Process, Development, and Refinement**. 2008. 69 (Master of Science in Information Science). School of Information and Library Science, University of North Carolina at Chapel Hill

CASTRO, F. F.; SANTOS, P. L. V. A. D. C. **Os Metadados como instrumentos tecnológicos na padronização e potencialização dos recursos informacionais no âmbito das bibliotecas digitais na era da web semântica**. Inf. & Soc.: Est. João Pessoa. 17: 13-19 p. 2007.

CASTRO, J. A.; RIBEIRO, C.; SILVA, J. R. Designing Application Profile Using Qualified Dublin Core: A case study with fracture mechanics datasets. International Conference on Dublin Core and Metadata Applications, 2013.

CEN, E. C. F. S. **Guidelines and support for building application profiles in e-learning** 2006.

CHEN, Y.-N.; CHEN, S.-J. **Metadata Lifecycle Model and Metadata Interoperability**. Glasgow, UK 2005.

CHEN, Y.-N. et al. Functional Requirements of Metadata System: From User Needs Perspective., International Conference on Dublin Core and Metadata Applications, 2003. p.91-98.

CIASULLO, G.; LODI, G.; ROTUNDO, A. **Core Public Service Vocabulary: The Italian Application Profile** 2013.

CONVERTINO, M. et al. **Untangling drivers of species distributions: Global sensitivity and uncertainty analyses of MAXENT** Environmental Modelling & Software. 51: 296-309 p. 2014.

CORRÊA, L. P. L. et al. **Computational techniques for biologic species distribution modeling**: 308-325 p. 2011.

COYLE, K.; BAKER, T. **Guidelines for Dublin Core Application Profiles** 2009.

CURADO MALTA, M.; BAPTISTA, A. A. State of the Art on Methodologies for the Development of a Metadata Application Profile. In: DODERO, J. M.; PALOMO-DUARTE, M., *et al* (Ed.). **Metadata and Semantics Research: 6th Research Conference, MTSR 2012, Cádiz, Spain, November 28-30, 2012. Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p.61-73. ISBN 978-3-642-35233-1.

_____. **A Method for the Development of Dublin Core Application Profiles (Me4DCAP V0.1): A Description**. International Conference on Dublin Core and Metadata Applications: 90-103 p. 2013.

CURRIE, M. et al. Visualizing Interoperability: ARH, Aggregation, Rationalisation and Harmonisation. Int. Conference on Dublin Core and Metadata for e-Communities, 2002, Firenze University Press. p.177-183.

CUTLER, D. R. et al. **Random Forest for Classification in Ecology**. *Ecology*. 88 2007.

DALE, V. H. **Ecological Modeling for Resource Management**. 1. United States of America: Springer-Verlag New York, Inc, 2003. 328.

DATAONE, O. Data Life Cycle. 2016. Disponível em: <<https://www.dataone.org/data-life-cycle>>. Acesso em: 12/12/2016.

DCMI. **Dublin Core education application profile** 2017.

DE LA PASSADIERE, B.; JARRAUD, P. **ManUeL, un profil d'application de LOM pour CampusSciences**. Sciences et Technologies de l'Information et de la Communication pour. 11 2004.

DELATORRE, M. et al. **Implementing new northernmost records to modelling the distribution of *Hypsiboas caingua* (Anura: Hylidae) in South America**. *Brazilian Journal of Biology*, v. 74, n. 4, 2014.

DINIZ-FILHO, J. A. F. et al. **Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change**. *Ecography*. 32 2009.

DPLP. **Dicionário Priberam da Língua Portuguesa** 2013.

DUBOIS, G. et al. **The Digital Observatory for Protected Areas (DOPA) Explorer 1.0.** 2015.

DUBOIS, G. et al. **eHabitat, a multi-purpose Web Processing Service for ecological modeling.** *Environmental Modelling & Software*: Elsevier. 41: 123-133 p. 2013.

DUVAL, E. et al. **Metadata Principles and Practicalities.** *D-Lib Magazine*. 8 2002.

EADIE, M. **Towards an Application Profile for Images.** *ARIADNE, Web Magazine for Information Professionals*. <http://www.ariadne.ac.uk/issue55/eadie> 2008.

EESV. **Process and Methodology for Core Vocabularies** 2012.

EHABITAT. eHabitat - Ecological Forecasting for Policy Making - See more at: <http://ehabitat.jrc.ec.europa.eu/> - sthash.D8viksQc.dpuf. 2015. Disponível em: <<http://ehabitat.jrc.ec.europa.eu/>>. Acesso em: 10-Mar-2015.

ELITH, J. et al. **Novel methods improve prediction of species' distributions from occurrence data.** *Ecography*. 29: 129-151 p. 2006.

ELITH, J. et al. **A statistical explanation of maxent for ecologists.** *Diversity and Distributions*. 17: 43-57 p. 2011.

FATH, S. E. J. B. D. **Fundamentals of Ecological Modelling - Applications in Environmental Management and Research.** 4th. US: Elsevier B.V., 2011. 399 ISBN 978-0-444-53567-2.

FEGRAUS, E. H. et al. **Maximizing the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation.** *Bulletin of the Ecological Society of America*: 158-168 p. 2005.

FORDHAM, D. A. et al. **Plant extinction risk under climate change: are forecast range shifts alone a good indicator of species vulnerability to global warming?** . *Global Change Biology*: Blackwell Publishing Ltd 2012.

FRIESEN, N.; MASON, J.; WARD, N. Building Educational Metadata Profiles. Int. Conference on Dublin Core and Metadata for e-Communities, 2002, Firenze University Press. p.63-69.

GÁRCIA-QUISMONDO, M. A. M.; PRADO, J. C.; CERVERÓ, A. C. **Desarrollo de un Esquema de Metadatos para la Descripción de Recursos Educativos: El Perfil de Aplicación MIMETA**. *Revista Española de Documentación Científica*. 29: 551-571 p. 2006.

GBIF. **Darwin Core Archives – How-to Guide**. Copenhagen: Global Biodiversity Information Facility 2010.

_____. **Global Biodiversity Information Facility: fee and open access to biodiversity data** Global Biodiversity Information Facility: fee and open access to biodiversity data. <http://www.gbif.org/>, 2015. Disponível em: <<http://www.gbif.org/>>. Acesso em: Jan/15/2015.

GIOVANELLI, J. G. R. et al. **Modeling a spatially restricted distribution in the Neotropics: How the size of calibration area affects the performance of five presence-only methods**. *Ecological Modeling*. 221: 215-224 p. 2010.

GOETHALS, P. L. M.; CHON, T. S. Special issue of the 7th International Conference on Ecological Informatics, 13-16 December 2010, Ghent, Belgium: 'Unravelling complexity and supporting sustainability'. **Ecological Informatics**, v. 17, p. 1-2, Sep 2013. ISSN 1574-9541. Disponível em: <<Go to ISI>://WOS:000325907300001 >.

GUEDES, G. T. A. **UML Uma Abordagem Prática**. Novatec, 2008. 336.

GUISAN, A. et al. **Unifying niche shift studies: insights from biological invasions**. *Trends in Ecology & Evolution*: Elsevier. 29: 260-269 p. 2014.

GUO, Q.; KELLY, M.; GRAHAM, C. H. **Support vector machines for predicting distribution of Sudden Oak Death in California**. *Ecological Modeling*: Elsevier. 182: 75-90 p. 2005.

HEERY, R.; PATEL, M. **Application Profiles: Mixing and Matching Metadata Schemas**. *Web Magazine for Information Professionals*. Loughborough University Library 2000.

HIJMANS, R. J.; GUARINO, L.; MATHUR, P. **DIVA-GIS, version 7.5 - Manual** 2012.

HILLMANN, D. Using Dublin Core. 2005. Disponível em: <<http://www.dublincore.org/documents/usageguide/>>. Acesso em: Mar/04/2015.

HOFMEISTER, C. et al. A general model of software architecture design derived from five industrial approaches. **The Journal of Systems and Software**, v. 80, p. 106-126, 2007.

HORTAL, J. et al. **Understanding (insect) species distributions across spatial scales**. *Ecography*. 33: 51-53 p. 2010.

HUTCHINSON, G. E. **A treatise on limnology**. New York: John Wiley, 1957.

IABIN. **Inter-American Biodiversity Information Network** Inter-American Biodiversity Information Network. 2015. Disponível em: <<http://www.oas.org/en/sedi/dsd/iabin/>>. Acesso em: Fev/02/2015.

IEC, I. E. C. **TC65/290/DC: Industrial Process Measurement and Control**: IEC, International Electrotechnical Commission 2002.

IEEE. **The authoritative dictionary of IEEE standards terms**: Standards Information Network IEEE Press 2000.

IHLOW, F. et al. **On the brink of extinction? How climate change may affect global chelonian species richness and distribution** *Global Change Biology*. 18: 1520–1530 p. 2012.

IMS, G. L. C. **IMS Application Profile Guidelines Overview** 2005.

JORGENSEN, S. E. **Fundamentals of Ecological Modelling**. 2nd. Amsterdam - The Netherlands: Elsevier Science B.V., 1994. 628 ISBN 0-444-81572-4.

KHANUM, R.; MUMTAZ, A. S.; KUMAR, S. **Predicting impacts of climate change on medicinal asclepiads of Pakistan using Maxent modeling** *Acta Oecologica*. 49: 23-31 p. 2013.

KNB. **Ecological Metadata Language (EML) Specification**: The Knowledge Network for Biocomplexity 2015.

KRAUSE, E. M. et al. Evolution of an Application Profile: Advancing Metadata Best Practices through the Dryad Data Repository. International Conference on Dublin Core and Metadata Applications

DC-2015, 2015, São Paulo.

KREBS, C. J. **Ecology: The Experimental Analysis of Distribution and Abundance**. New York: Harper & Row, Publishers Inc., 1978.

LIMA-RIBEIRO, M. S.; DINIZ-FILHO, J. A. F. **Modelando a distribuição geográfica das espécies no passado: Uma abordagem promissora em paleontologia**. *Revista Brasileira de Paleontologia*. 15: 371-385 p. 2012.

Manual Global de Ecologia. 4ª. Augustus Editora, 2002. ISBN 85-85497-12-2.

MARCO JUNIOR, P. D.; SIQUEIRA, M. F. D. **Como determinar a distribuição potencial de espécies sob uma abordagem conservacionista?** *Megadiversidade*. 5 2009.

MARSHALL, C. E.; GLEGG, G. A.; HOWELL, K. L. Species distribution modelling to support marine conservation planning: The next steps. **Marine Policy**, v. 45, p. 330-332, 2014. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0308597X13002005>.

MASON, J.; GALATIS, H. **Theory and Practice of Application Profile Development**. *International Conference on Dublin Core and Metadata Applications*: 43-52 p. 2007.

MCINERNY, G. J.; ETIENNE, R. S. Ditch the niche - is the niche a useful concept in ecology or species distribution modelling? **Journal of Biogeography**, v. 39, p. 2096-2102, 2012a.

_____. Pitch the niche – taking responsibility for the concepts we use in ecology and species distribution modelling. **Journal of Biogeography**, v. 39, p. 2112–2118, 2012b.

_____. **Stitch the niche – a practical philosophy and visual schematic for the niche concept**. *Journal of Biogeography*. 39: 2103-2111 p. 2012c.

_____. **Stitch the niche – a practical philosophy and visual schematic for the niche concept** **Stitch the niche – a practical philosophy and visual schematic for the niche concept**. *Journal of Biogeography*. 39: 2103-2111 p. 2012d.

MICHENER, W. K. et al. **NonGeospatial Metadata for the Ecological Sciences**. *Ecological Applications*. 7: 330-342 p. 1997.

MICHENER, W. K.; JONES, M. B. **Ecoinformatics: supporting ecology as a data-intensive science**. *Trends in Ecology & Evolution*: Cell Press. 27: 85-93 p. 2012.

NATIVI, S.; MAZZETTI, P.; GELLER, G. N. **Environmental model access and interoperability: The GEO Model Web initiative**. *Environmental Modelling & Software*: Elsevier. 39: 214-228 p. 2013.

NBN. Metadata within the NBN. 2011. Disponível em: <<http://www.nbn.org.uk/Share-Data/Providing-Data/NBN-Metadata-Standard.aspx>>. Acesso em: Mar/12/2015.

NILSSON, M. **Description Set Profiles: A constraint language for Dublin Core Application Profiles**: Dublin Core Metadata Initiative 2008.

NILSSON, M.; BAKER, T.; JOHNSTON, P. The Singapore Framework for Dublin Core Application Profiles. 2008. Disponível em: <<http://dublincore.org/documents/singapore-framework/>>. Acesso em: Dez/15/2016.

ODUM, E., P. **Ecologia**. São Paulo: Guanabara Koogan, 2012. ISBN 9788527700610.

ODUM, E. P.; BARRET, G. W. **Fundamentos de Ecologia**. São Paulo: Cengage Learning, 2011. ISBN 978852210540.

OLIVEIRA, H. R.; CASSEMIRO, F. A. S. **Potenciais efeitos das mudanças climáticas futuras sobre a distribuição de um anuro da Caatinga *Rhinella granulosa* (Anura, Bufonidae)**. *Iheringia, Série Zoologia*. Porto Alegre. 103 2013.

ONYANCHIA, I.; KEIZER, J.; KATZ, S. A Dublin Core Application Profile in the Agricultural Domain. International Conference on Dublin Core and Metadata Applications, 2001, Tokyo, Japan.

PADONOU, E. A. et al. **Using species distribution models to select species resistant to climate change for ecological restoration of bowé in West Africa**. *African Journal of Ecology*, v. 53, n. 1, p. 83-92, 2015.

PALAVITSINIS, N.; MANOUSELIS, N.; ALONSO, S. S. Evaluation of a Metadata Application Profile for Learning Resources on Organic Agriculture. In: SPRINGER-VERLAG, MTSR, 2009, Springer-Verlar. p.270-281.

PANZACCHI, M. et al. **Searching for the fundamental niche using individual-based habitat selection modelling across populations** Searching for the

fundamental niche using individual-based habitat selection modelling across populations. *Ecography* 2014.

PEREIRA, I. M. **Modelagem de Distribuição e Conservação de Davilla Vand. (Dilleniaceae Salisb.).** 2014. (Ph.D.). Departamento de Biologia, Universidade de São Paulo

PETERSON, A. T. et al. The big questions for biodiversity informatics. **Systematics & Biodiversity**, v. 8, n. 2, p. 159-168, 2010. ISSN 14772000. Disponível em: <<http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=51743256&site=ehost-live>

<http://www.tandfonline.com/doi/abs/10.1080/14772001003739369>>.

PETERSON, A. T.; SOBERÓN, J. Integrating fundamental concepts of ecology, biogeography, and sampling into effective ecological niche modeling and species distribution modeling. **Plant Biosystems**, v. 146, n. 4, p. 789-796, 2012. ISSN 11263504. Disponível em: <<http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=83380598&site=ehost-live>

<http://www.tandfonline.com/doi/abs/10.1080/11263504.2012.740083>>.

PETERSON, A. T. et al. **Ecological Niches and Geographic Distributions.** United Kingdom: Princeton University Press, 2011. 328 ISBN 978-0-691-13686-8.

PETERSON, R. W. **A College of Integrated Studies: Education for the Professional Generalist.** *L&S Magazine*. University of Wisconsin: Spring 1988.

PHILIPS, S. J. et al. **Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data.** *Ecological Applications*. 19: 181-197 p. 2009.

PHILIPS, S. J.; DUDÍK, M.; SCHAPIRE, R. **A maximum entropy approach to species distribution modeling** A maximum entropy approach to species distribution modeling. ICML International Conference on Machine Learning, 2004, ACM New York. p.83.

_____. **Maximum entropy modeling of species geographic distributions.** *Ecological Modelling*: Elsevier. 190: 231-259 p. 2006.

PINAYA, J. L. D. **Processo de Pré-Análise para a Modelagem de Distribuição de Espécies** 2013. 112 (Master). Departamento de Engenharia de Computação da Escola Politécnica, Universidade de São Paulo

RECKNAGEL, F. **Ecological informatics: understanding ecology by biologically-inspired computation**. 2nd. Berlim - Germany: Springer, 2003. 496 ISBN 3-540-43455-0.

_____. Ecological informatics: A discipline in the making. **Ecological Informatics**, v. 6, n. 1, p. 1-3, Jan 2011. ISSN 1574-9541. Disponível em: <<Go to ISI>://WOS:000288361800001 >.

REITZ, J. M. **Online Dictionary for Library and Information Science** 2004.

ROBERTSON, T. et al. **The GBIF Integrated Publishing Toolkit: Facilitating the Eddicient Publishing of Biodiversity Data on the Internet**. PLOS One. 9 2014.

RODDER, D. et al. **A novel method to calculate climatic niche similarity among species with restricted ranges - the case of terrestrial Lycian salamanders**. Organisms Diversity & Evolution. 11: 409-423 p. 2011.

RODRIGUES, E. S. C. **Teoria da Informação e Adaptatividade na Modelagem de Distribuição de Espécies**. 2012. Escola Politécnica - Departamento de Engenharia Elétrica, Universidade de São Paulo

RODRIGUES, F. A. **Um método de referência para análise de desempenho preditivo de algoritmos de modelagem de distribuição de espécies**. 2012. (Ph.D.). Departamento de Engenharia de Computação da Escola Politécnica, Universidade de São Paulo

RUSSELL, S.; NORVING, P. **Inteligência Artificial**. Rio de Janeiro: Elsevier, 2004.

SALOKHE, G.; PESCE, V.; LIESTHOUT, J. **Organization Metadata Application Profile** 2008.

SAYÃO, L. F.; MARCONDES, C. H. **O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais**. TransInformação. 20: 133-148 p. 2008.

SCACHETTI-PEREIRA, R. **Desktop GARP**. Word wide web electronic publication: University of Kansas 2002.

SHAW, M.; GARLAN, D. **Software Architecture: Pespectives on an Emerging Discipline**. Prentice Hall, 1996.

SILVA, A.; BORBA, C.; CORRÊA, L. P. L. Metadata Standard for Models Generated by openModeller. International Conference on Dublin Core and Metadata Applications, 2015, São Paulo - Brazil.

SILVA, J. R. **Usage-driven Application Profile Generation Using Ontologies**. 2016. (Ph.D.). Faculdade de Engenharia, Universidade do Porto

SOBEK-SWANT, S. et al. **Potential distribution of emerald ash borer: What can we learn from ecological niche models using Maxent and GARP?**. Forest Ecology and Management. 281: 23-31 p. 2012.

SOBERÓN, J. **Commentary on Ditch, Stitch and Pitch: the niche is here to stay**. Journal of Biogeography, v. 41, n. 2, p. 414-417, 2014.

SOBERÓN, J.; NAKAMURA, M. Niches and distributional areas: concepts, methods, and assumptions. PNAS Proceedings of the National Academy of Sciences of the United States of America, 2009. p.19644-19650.

SOBERÓN, J.; PETERSON, A. T. **Interpretation of models of fundamental ecological niches and species' distributional areas**. Biodiversity Informatics. 2: 1-10 p. 2005.

SOUZA MUÑOZ, M. et al. openModeller: a generic approach to species' potential distribution modelling. **Geoinformatica**, v. 15, n. 1, p. 111-135, 2011. ISSN 13846175. Disponível em: <<http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=57407346&site=ehost-live>>.

http://download.springer.com/static/pdf/280/art%3A10.1007%2Fs10707-009-0090-7.pdf?auth66=1403799632_61a1d208c839e0822cd371326d52d637&ext=.pdf.

SPECIESLINK. SpeciesLink. 2015. Disponível em: <<http://splink.cria.org.br/>>. Acesso em: Mar/02/2015.

STOCKWELL, D. R. B.; NOBLE, I. R. **Induction of sets of rules from animal distribution data: A robust and informative method of data analysis**. Mathematics and Computers in Simulation: Elsevier. 33: 385-390 p. 1992.

STOCKWELL, D. R. B.; PETERSON, T. A. **Effects of sample size on accuracy of species distribution models**. Ecological Modeling: Elsevier. 148: 1-13 p. 2002.

TORRES, N. M. et al. **Can species distribution modelling provide estimates of population densities? A case study with jaguars in the Neotropics.** *Diversity and Distributions*, v. 18, n. 6, p. 615-627, 2012.

VANDERWAL, J. et al. **Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know?** *Ecological Modelling*. 220: 589–594 p. 2009.

VAPNIK, V. **The Nature of Statistical Learning Theory**. 1. New York: Springer, 1995. 188.

VERNADAT, F. B. **Enterprise modelling and integration: principles and applicatiks.** London: Chapman & Hall, 1996.

WCED, W. C. O. E. A. D. **Our Common Future**. UN, United Nations. 1987

WEIBEL, S. L.; LAGOZE, C. An element set to support resource discovery. *International Journal on Digital Libraries*, v. 1, n. 2, p. 176-186, 1997.

WIECZOREK, J. et al. **Darwin Core: An Evolving Community-Developed Biodiversity Data Standard.** *PLoS ONE* 7 2012.

WIECZOREK, J. et al. Darwin Core. 2014-11-08 2015. Disponível em: <<http://rs.tdwg.org/dwc/index.htm>>. Acesso em: Abril/15/2015.

WILSON, C. D.; ROBERTS, D.; REID, N. **Applying species distribution modelling to identify areas of high conservation value for endangered species: A case study using *Margaritifera margaritifera* (L.)** *Biological Conservation*. 144: 821-829 p. 2011.

WILSON, K. et al. Development of a metadata application profile at the State Library of New South Wales. 13th Information Online Conference, 2007, Sydney, Australia.

WISZ, M. S. et al. **Effects of sample size on the performance of species distribution models.** *Diversity and Distributions*. 14: 763–773 p. 2008.

YANG, X.-Q. et al. **Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills** *Ecological Engineering*. 51: 83-87 p. 2013.

YANG, Z. F. Ecological informatics for ecosystem conservation in view of environmental risk assessment and management. **Stochastic Environmental**

Research and Risk Assessment, v. 25, n. 5, p. 641-642, Jul 2011. ISSN 1436-3240. Disponível em: <<Go to ISI>://WOS:000292021500001 >.

ZSCHOCKE, T. et al. The LOM application profile for agricultural learning resources of the CGIAR. **International Journal Metadata, Semantics and Ontologies**, p. 13-23, 2009.