



**UNICAMP**

**DEBORA PIGNATARI DRUCKER**

**A INTEGRAÇÃO DA INFORMAÇÃO SOBRE BIODIVERSIDADE E  
ECOSSISTEMAS PARA EMBASAR POLÍTICAS DE CONSERVAÇÃO: O  
PROJETO BIOTA GRADIENTE FUNCIONAL COMO ESTUDO DE CASO**

**CAMPINAS**

**2012**





Universidade Estadual de Campinas  
Instituto de Filosofia e Ciências Humanas

DEBORA PIGNATARI DRUCKER

A INTEGRAÇÃO DA INFORMAÇÃO SOBRE BIODIVERSIDADE E  
ECOSSISTEMAS PARA EMBASAR POLÍTICAS DE CONSERVAÇÃO: O  
PROJETO BIOTA GRADIENTE FUNCIONAL COMO ESTUDO DE CASO

ORIENTADOR: PROF. DR. CARLOS ALFREDO JOLY

CO-ORIENTADORA: PROFA. DRA. LEILA DA COSTA FERREIRA

Tese de Doutorado apresentada ao Instituto  
de Filosofia e Ciências Humanas da  
Universidade Estadual de Campinas  
(IFCH/UNICAMP) para obtenção do título de  
Doutora em Ambiente e Sociedade, área de  
concentração "Aspectos Biológicos de  
Sustentabilidade e Conservação".

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE  
DEFENDIDA PELA ALUNA DEBORA PIGNATARI DRUCKER E ORIENTADA PELO PROF.DR CARLOS ALFREDO JOLY  
Assinatura do Orientador

A handwritten signature in blue ink, appearing to read "C. Joly", is written over a faint circular stamp.

CAMPINAS  
2012

FICHA CATALOGRÁFICA ELABORADA POR  
CECÍLIA MARIA JORGE NICOLAU – CRB8/3387 – BIBLIOTECA DO IFCH  
UNICAMP

D84i	<p>Drucker, Debora Pignatari, 1978-</p> <p>A integração da informação sobre a biodiversidade e ecossistemas para embasar políticas de conservação: o projeto Biota Gradiente Funcional como estudo de caso / Debora Pignatari Drucker. -- Campinas, SP : [s. n.], 2012.</p> <p>Orientador: Carlos Alfredo Joly. Co-orientador: Leila da Costa Ferreira. Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Filosofia e Ciências Humanas.</p> <p>1. Biodiversidade – Conservação. 2. Ecologia. 3. Gestão do conhecimento. 4. Projeto Biota Gradiente Funcional. 5. Ecossistemas. 6. Mata Atlântica – Conservação. 7. Parque Estadual da Serra do Mar (SP) - Conservação. I. Joly, Carlos Alfredo, 1955- II. Ferreira, Leila da Costa, 1958- III. Universidade Estadual de Campinas. Instituto de Filosofia e Ciências Humanas. IV. Título.</p>
------	---

Informação para Biblioteca Digital

**Título em Inglês:** Integrating biodiversity and ecosystem information to support conservation policies: the Biota Functional Gradient project as a case study

**Palavras-chave em inglês:**

Biodiversity – Conservation

Ecology

Knowledge management

Project Biota Functional Gradient

Ecosystems

Atlantic Forest – Conservation

State Park of Serra do Mar (SP, Brazil) - Conservation

**Área de concentração:** Aspectos Biológicos de Sustentabilidade e Conservação

**Titulação:** Doutor em Ambiente e Sociedade

**Banca examinadora:**

Carlos Alfredo Joly [Orientador]

Antonio Mauro Saraiva

Michael Maier Keller

Paula Felício Drummond de Castro

Thomas Michael Lewinsohn

**Data da defesa:** 20-08-2012

**Programa de Pós-Graduação:** Ambiente e Sociedade – NEPAM

**UNIVERSIDADE ESTADUAL DE CAMPINAS**

**INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS**

**Debora Pignatari Drucker**

**A Integração da Informação sobre Biodiversidade e Ecossistemas para Embasar Políticas de Conservação: o Projeto Biota Gradiente Funcional como Estudo de Caso**

Tese de doutorado apresentada ao Programa de Doutorado em Ambiente e Sociedade do Instituto de Filosofia e Ciências Humanas e ao Núcleo de Estudos e Pesquisas Ambientais da Universidade Estadual de Campinas (IFCH/NEPAM/UNICAMP), Área de Concentração "Aspectos Biológicos da Sustentabilidade e Conservação", sob orientação do Prof. Dr. Carlos Alfredo Joly e co-orientação da Profa. Dra. Leila da Costa Ferreira.

Este exemplar corresponde à redação final da Tese defendida e aprovada pela

Comissão Julgadora em 20/08/2012.

BANCA EXAMINADORA:

Prof. Dr. Carlos Alfredo Joly (orientador)



Prof. Dr. Antonio Mauro Saraiva



Prof. Dr. Michael Maier Keller



Profa. Dra. Paula Felício Drummond de Castro



Prof. Dr. Thomas Michael Lewinsohn



Profa. Dra. Cristiana Simão Seixas (suplente)

Profa. Dra. Simone Aparecida Vieira (suplente)

Prof. Dr. Luciano Martins Verdade (suplente)

**Campinas, agosto de 2012**



## RESUMO

No Brasil, há numerosas instituições e cientistas que produzem conhecimento sobre nossos ecossistemas e muito já se sabe sobre nossa diversidade biológica. Novas descobertas e orientações para a tomada de decisão baseadas em conhecimento científico podem ser feitas pelo uso da informação que já existe, mas que em grande parte está dispersa, mal documentada e inacessível aos interessados. Uma integração é, portanto, necessária e pode ocorrer tanto por meio da consolidação de dados de estudos semelhantes, de forma a expandir escalas temporais e espaciais de análise, como pela associação de conhecimentos gerados por estudos com diferentes abordagens. Novas tecnologias computacionais que permitem lidar com informações de procedência diversa e características heterogêneas podem contribuir para representar o estado atual do conhecimento do sistema terrestre e para comunicar esse conhecimento entre os diversos atores interessados. Nessa direção, os objetivos deste estudo foram analisar o sistema de informação desenvolvido para o Projeto Temático Biota Gradiente Funcional, como estudo de caso para a integração de dados ecológicos e de biodiversidade, e também discutir as aplicações e limitações desse sistema para a conservação e gestão ambiental. Para isso, proponho um arcabouço para investigar uma infraestrutura de informação sobre biodiversidade e discuto o potencial dessa infraestrutura como objeto-ponte entre ciência e processos de tomada de decisão em conservação e gestão ambiental; analiso também as características de dados ecológicos, boas práticas para documentá-los e algumas ferramentas existentes para gerenciá-los, de forma a promover a integração do conhecimento ecológico, descrevo as etapas de desenvolvimento do sistema de informação para o Projeto Temático Biota e o analiso pelas lentes do arcabouço de investigação proposto. O sistema foi eficiente para gerenciar os dados e metadados do Projeto fornecidos pelos pesquisadores. Foi possível recuperar a informação por consultas integradas tanto ao catálogo de metadados, pelo uso de palavras-chave, quanto à base de dados, pela seleção de variáveis e de parâmetros temporais e espaciais. A possibilidade de consultar qualquer variável de interesse, independentemente da associação com uma entidade taxonômica, configura um sistema de informação sobre biodiversidade original e inovador. No sistema estudado, a base de dados é verdadeiramente genérica: a única restrição é a exigência da informação espacial. Os resultados permitiram concluir que avanços em ferramentas computacionais para gerenciar dados de biodiversidade e ecossistemas permitem análises que potencializam o valor de dados coletados em pesquisas individuais. Para perpetuar iniciativas de desenvolvimento tecnológico, como a do sistema aqui investigado, é necessário empreender mais esforços, no sentido de se promoverem os arranjos institucionais, legislativos e administrativos que atuam como alicerces da infraestrutura da informação e conferem sustentabilidade dos sistemas a médio e longo prazo. Além disso, é preciso capacitar recursos humanos no uso e desenvolvimento desses recursos tecnológicos e promover a sensibilização dos atores envolvidos, a fim de tornar o compartilhamento e a curadoria de dados uma prática amplamente realizada.

Palavras-chave: Biodiversidade. Ecologia, Gestão da Informação, Gestão do conhecimento, Ecoinformática, Infraestrutura de Informação.





## ABSTRACT

There are numerous scientists and institutions that produce knowledge about Brazilian ecosystems, and a lot is already known about the biological diversity therein. New findings and guidelines for decision making based on scientific knowledge can be made by using information that already exists, but is largely dispersed, poorly documented and inaccessible to users. It is therefore necessary to integrate biodiversity information, which may occur by consolidating data from similar studies in order to expand temporal and spatial analysis scales, or by associating knowledge generated by studies with different approaches. New computer technologies are continuously developed by information scientists, providing opportunities for the Biodiversity Informatics field to improve its capacity of making sense of the growing amount of heterogeneous data of diverse provenance, thus contributing to represent the current state of knowledge on the Earth's Systems and to communicate this knowledge among the various stakeholders. This study aimed at analyzing the information system developed for the Thematic Project Biota Functional Gradient as a case study on biodiversity and ecological data integration, and at discussing the applications and limitations of this system for conservation and environmental management. To this end, I propose a framework for investigating biodiversity information infrastructures and discuss their potential as boundary objects between science and decision making in conservation and environmental management; analyze the characteristics of ecological data, the best practices to document them and some existing tools to manage them in order to promote the integration of ecological knowledge; describe the stages of development of the information system for the Thematic Project Biota and discuss it through the lens of the proposed analytical framework. The information system was efficient to manage data and metadata provided by researchers. Information retrieval was possible by querying the metadata catalog based on keywords, and by querying the database system based on the selection of variables and parameters of space and time. The ability to query any variable of interest, regardless of association with a taxonomic entity, set up an original and innovative information system on biodiversity. In the studied system, the database is truly generic: the only restriction is the mandatory requirement for spatial information. The results showed that advances on computational tools for managing biodiversity and ecosystem data enable analyses that maximize the value of data collected in individual surveys. To perpetuate technological development initiatives such as the system investigated here, it is necessary to undertake further efforts in order to promote institutional arrangements, legislation and administration that act as foundations for information infrastructures and provide sustainability over time. Furthermore, it is necessary to train human resources in the use and development of technological tools, and to promote awareness and a collaborative culture between the actors involved in order to make data sharing and curation widely adopted practices.

Keywords: Biodiversity, Ecology, Information Management, Knowledge Management, Coinformatics, Information Infrastructure.



# SUMÁRIO

Resumo .....	vii
Abstract .....	ix
Lista de ilustrações .....	xiii
Lista de tabelas .....	xv
Agradecimentos .....	xix
Prefácio .....	xxiii
Introdução .....	29
Capítulo I: Análise de componentes necessários para compor uma infraestrutura de informação sobre biodiversidade e ecossistemas para mobilizar o conhecimento para apoiar processos de tomada de decisão .....	45
1.1. Introdução .....	45
1.2. Interface entre ciência e política .....	47
1.3. Infraestrutura de Informação sobre biodiversidade .....	50
Pessoas .....	55
Arcabouço Institucional .....	59
Dados .....	65
Tecnologia .....	72
Normas e Padrões .....	74
1.3. Discussão .....	76
1.3. Considerações Finais .....	80
Capítulo II: Avanços na integração e gerenciamento de dados ecológicos .....	83

Capítulo III: O sistema de informação do Projeto Temático Gradiente Funcional (BIOTA/FAPESP) como estudo de caso para a integração da informação sobre biodiversidade e ecossistemas .....	97
3.1 Introdução .....	97
3.2 Objeto de Estudo .....	98
3.3 Requisitos .....	100
3.4 Soluções .....	102
3.5 Teste do Sistema .....	111
3.6 Análise dos resultados e considerações finais .....	120
Conclusão.....	125
Referências .....	131
Anexo 1. Documentação da base de dados delineada para o Projeto Temático Biota Gradiente Funcional .....	151
Anexo 2. Lista de levantamentos biológicos catalogados no servidor do Projeto Temático Gradiente Funcional .....	161

## LISTA DE ILUSTRAÇÕES

Figura 1. Papéis interativos da ciência e de outros setores da sociedade na conservação da biodiversidade .....	37
Figura 1.1. Esquema ilustrativo de fatores que afetam ações políticas de governos municipais relacionadas com mudanças ambientais globais .....	48
Figura 1.2. Infraestrutura de informação em biodiversidade como objeto-ponte na interface entre o conhecimento científico e sistemas de tomada de decisão .....	52
Figura 1.3. Relação entre componentes de uma Infraestrutura de Informação .....	54
Figura 1.4. O ciclo de vida dos dados, desde o planejamento até a análise .....	66
Figura 1.5. Distribuição de 12.025 auxílios da NSF por valores, em 2007 .....	71
Figura 2.1. Exemplo de uma planilha de levantamento ecológico, sem anotações que permitam atribuir significado aos dados .....	86
Figura 2.2. A planilha da Figura 2.1, preparada para importação em banco de dados ou pacote estatístico .....	87
Figura 2.3. Metadados para a planilha das Figuras 2.1 e 2.2. ....	90
Figura 3.1. Modelo Entidade-Relacionamento da Base de Dados do Projeto Temático Biota Gradiente Funcional .....	103

Figura 3.2. Schema da tabela de variáveis com tipo número inteiro .....	104
Figura 3.3. Tabela de descrição de variáveis .....	107
Figura 3.4. Visualização do Atlas do Sistema de Informação desenvolvido para o Projeto Biota Gradiente Funcional pelo uso de um dispositivo móvel tablet .....	109
Figura 3.5. Exemplo de formulário de consulta .....	113
Figura 3.6. Exemplo de resultado de consulta .....	114
Figura 3.7. Localização das 14 parcelas de 1 hectare do Projeto Temático Gradiente Funcional .....	115
Figura 3.8. Visualização dos indivíduos arbóreos levantados nas Parcelas G e H .....	116
Figura 3.9. Formulário de cadastro de metadados de novos levantamentos .....	117
Figura 3.10. Página inicial do Repositório de Dados do Projeto Temático Biota Gradiente Funcional .....	118

## LISTA DE ABREVIATURAS E SIGLAS

BIOTA-FAPESP – Programa de Pesquisas em Caracterização, Conservação, Recuperação e Uso Sustentável da Biodiversidade do Estado de São Paulo

CDB – Convenção sobre Diversidade Biológica

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

CONCAR – Comissão Nacional de Cartografia

DataONE – *Data Observation Network for Earth*

DMP – *Data Management Plan*

EML – *Ecological Metadata Language*

ESA – *Ecological Society of America*

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo

FGDC – *Federal Geographic Data Committee*

GBIF - *Global Biodiversity Information Facility*

GEF – *Global Environment Facility*

GEO – *Group on Earth Observations*

GEO BON – *Biodiversity Observation Network*

GSDI – *Global Spatial Data Infrastructure*

IBP – *International Biological Program*

ICSU - *International Council for Science*

INDE – Infraestrutura Nacional de Dados Espaciais

IOC – *Intergovernmental Oceanographic Commission*

KNB – *Knowledge Network for Biocomplexity*

LBA – Programa de Grande Escala da Biosfera-Atmosfera na Amazônia

LTER – *Long Term Ecological Research*

NCEAS – *National Center for Ecological Analysis and Synthesis*

NEON – *National Ecological Observatory Network*

NERC - *Natural Environment Research Council*  
NSF – *National Science Foundation*  
OBIS – *Ocean Biogeographic Information System*  
OGC – *Open Geospatial Consortium*  
PELD - *Pesquisas Ecológicas de Longa Duração*  
PISCO – *Partnership for Interdisciplinary Studies of Coastal Oceans*  
PNB – *Política Nacional de Biodiversidade*  
PNUD – *Programa das Nações Unidas para o Desenvolvimento*  
PPBio – *Programa de Pesquisa em Biodiversidade do Ministério de Ciência, Tecnologia e Inovação*  
RAINFOR – *Amazonian Forest Inventory Network*  
SAEON – *South African Environmental Observatory Network*  
STS - *Science and Technology Studies*  
TDWG – *Biodiversity Information Standards*  
TFRI – *Taiwan Forestry Research Institute*  
UNESCO - *United Nations Educational, Scientific and Cultural Organization*  
XML – *eXtensible Markup Language*



*“A floresta tem muitas gentes”*

*“Dersu Uzala”, filme de Akira Kurosawa*



## AGRADECIMENTOS

Ao Carlos Joly, pela confiança e incentivo ao longo da orientação deste estudo e pelo exemplo de dedicação à conservação da biodiversidade.

À Leila da Costa Ferreira, por ampliar minha visão sobre ciência, sociedade e ambiente, pelas sugestões ao texto e pelo constante apoio.

Ao Thomas Lewinsohn, pelo apoio e sugestões ao longo do desenvolvimento deste estudo, pelas sugestões no exame de qualificação e na pré-banca, pelo empréstimo de livros e por compartilhar o fascínio pela Ciência Ecológica, bem como a vontade de fazê-la cruzar fronteiras e alcançar voos mais altos.

Ao Michael Keller, pelas sugestões na pré-banca e pelo empréstimo de livros.

À Myanna Lahsen, pelas sugestões no exame de qualificação e pela literatura disponibilizada.

Ao Daniel Hogan, *in memoriam*, pelos ensinamentos em Demografia e por proporcionar uma perspectiva abrangente das questões ambientais.

À Lúcia da Costa Ferreira, pelas sugestões ao longo da disciplina de Projetos e pelo apoio constante.

À Cristiana Seixas, pela amizade, pela ajuda na estruturação da tese, pelas sugestões ao conteúdo, pelos empréstimos de livros.

À Simone Vieira, pelo carinho, pelo apoio ao longo do desenvolvimento deste estudo e pelas sugestões ao texto final.

Ao Mateus Batistella, pelos ensinamentos em Ecologia de Paisagem e pelo apoio ao meu desenvolvimento profissional em Gestão da Informação e do Conhecimento.

Aos colegas do Projeto Temático Biota Gradiente Funcional, pela disposição em colaborar, compartilhando informações dos admiráveis trabalhos que desenvolveram.

Aos professores Flavio Santos, Marlies Sazima, Simey Thury Vieira Fisch e à Luciana Alves, pelo apoio desde a fase inicial do projeto.

Ao José Salim, pela parceria no desenvolvimento do sistema de informação deste estudo, pelos ensinamentos em Computação e por compartilhar o interesse pelas ferramentas de Ecoinformática.

Ao George Shepherd, pelas aulas de análise multivariada e pela valiosa oportunidade de trocar idéias sobre padrões e processos em ecossistemas.

Aos professores Leonardo Galletto, Mauro Almeida, Paulo Inácio Prado, Roberto Guimarães, Roberto Carmo, pelas aulas e ensinamentos. À Claudia Bauzer e ao Ricardo Torres, pelas sugestões no início do projeto de tese.

Ao Tiago Duque-Estrada, pelas afinidades profissionais e pela convivência em clima de colaboração. Torço para que possamos seguir colaborando um com o outro.

Ao William Magnusson, pelos ensinamentos em Ecologia e pelo incentivo profissional.

Ao Matt Jones, pelos ensinamentos em Ecoinformática e pelo entusiasmo em trabalhar de forma colaborativa.

Ao Rolf de By, pela colaboração no trabalho que originou o presente estudo, pelos ensinamentos e sugestões.

Ao Laurindo Campos, pelos ensinamentos desde os tempos do INPA.

Aos colegas do SIB-BR – Sistema de Informação sobre a Biodiversidade e Ecossistemas Brasileiros – Mercedes Bustamante, David Oren, Antonio Saraiva, Keila McFraden, Eduardo Dalcin, Pedro Correa, Nelson Simões, Celso Monteiro e Wagner Léio: aprendo muito com a oportunidade de debater com vocês temas tão relevantes para a conservação e gestão ambiental no Brasil.

Ao João Meidanis, pelas conversas e sugestões.

Às equipes do NEPAM, IFCH e IB: Maria de Fátima Rodrigues Moreira, Waldinei Araújo, Alcebíades Júnior, Maria Lúcia Mendonça Fernandes Pinto, Neusa Trevisan, Débora Matavelli, Renato Belinelo e Diego Carvalho por todo apoio dado para a realização deste estudo.

À Mercedes de Paula Ferreira, pela super eficiente revisão do português – praticamente em tempo real. Erros remanescentes são minha responsabilidade.

Ao Luciano Verdade, por me incentivar a cursar o doutorado no NEPAM.

Ao Paulo Dalgarrondo, pelos debates interessantes na monitoria da disciplina Biologia e Sociedade.

À Juliana Farinaci, pelo apoio e sugestões ao longo de toda a trajetória, e pelo SOS em emergências de uma forasteira em Barão Geraldo. Sua amizade foi uma das conquistas que essa trajetória me proporcionou.

À Joana Bezerra, Juliana Seidel e Zoraide Pessoa, colegas da Turma das “Lulus”, foi uma honra conviver com vocês durante o curso de doutorado.

Aos colegas do curso de doutorado em Ambiente e Sociedade, em especial ao Allan Yu, Bia Mendes, Carol Joly, Gabi di Giulio, Igor Carvalho, Jorge Calvimontes, Leonardo Teixeira, Luciana Gomes, Luziana Garuana, Marília Giesbrecht, Rafael Martins, Ramon Bicudo, Roberto Donato, Rodrigo Rodrigues e Satya Caldenhof. Ao Edu Viglio, Eduardo Marandola, Estevão Bosco, Leonardo Freire de Mello, Marcelo Fetz e Ricardo Ojima.

Aos colegas da Embrapa: Bibiana Teixeira, pelo apoio e compreensão; Vera Viana, super eficiente ao providenciar artigos e pela edição das referências bibliográficas – erros remanescentes são minha responsabilidade; Celina Takemura, pela ajuda com a mineração de texto; Cristina Criscuolo e Edson Bolfe, pelo apoio; Alessandro Tchani, Célia Grego, Cristina Rodrigues, Luciana Zenerato e Rosângela Pereira, pelo apoio, em especial nos momentos mais difíceis; Bianca Pedroni, Carlos Ronquim, Fábio Torresan, Gustavo Bayma, Janice Leivas, Luciana Spinelli, Ricardo Guimarães e Sandra Furlan pela amizade.

Ao Anselmo Nogueira, pela amizade e sugestões ao Capítulo 2.

Ao Leonardo Meireles e Flávia Natércia, não seria possível sonhar sem amigos como vocês por perto. À Nívea Dias, companheira de programas culturais, com direito a papos incríveis sobre briófitas. Aos colegas do IB, que ampliaram meu conhecimento em Biologia: Aline Lopes, André Rochelle, Arildo Dias, Bruno Rosado, Cristina Baldauf, Daniela Vinha, Felipe Amorim, Gustavo Shimizu, Jacira

Rabelo, Juliano Melis, Larissa Pereira, Leonardo Ré, Lorena Fonseca, Marcelo Freire, Marcos Scaranello, Maurício Sampaio, Pedro Cavalin, Rafael Costa.

À May, Oscar e família, Moema, Patrícia e Riza, pelas trocas e percepções sobre educação infantil. Ao Saturnino, pelo guaraná de Maués. Aos camaradas angoleiros, em especial à Vani, Mari, Mirna e Guga.

Ao Raul Vinhas Ribeiro, pelo apoio para eu me estabelecer em Campinas.

Aos amigos e colegas de trabalho, pela paciência com o meu comportamento antissocial nos últimos meses dessa trajetória. Em especial à Carolina Fincatti, Juliana Napolitano, Renata Zambello e Sylvia Mota, distantes no espaço, mas sempre próximas ao coração.

Aos meus pais, Sônia e José, pelo apoio em toda a trajetória, pela paciência e compreensão. Sem vocês, eu certamente não teria chegado até aqui.

Ao Claudio, Tania, Zezé, Altair, Esther, Nelson - meus familiares -, pelo apoio, paciência e compreensão.

Ao Gabriel, essa pessoa incrível com a qual tenho a sorte de conviver, por ser meu filho, e com quem aprendo e a quem ensino todos os dias.

Ao Núcleo de Pesquisas e Estudos Ambientais – NEPAM, pela formação que me proporcionou e pelo apoio institucional.

Ao Instituto de Biologia (IB) e ao Instituto de Filosofia e Ciências Humanas Embrapa (IFCH) da UNCIAMP e à EMBRAPA, pelo apoio institucional.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa concedida na fase inicial do curso.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pela bolsa concedida pelo Processo 2007/59409-4 e pelo financiamento ao projeto Biota Gradiente Funcional (03/12595-7). Ao FAEPEX/UNICAMP e ao Programa de Doutorado em Ambiente e Sociedade, pelo auxílio financeiro para participação em evento. Ao PPBio/MCT, INCT/CENBAM/CNPq/MCT - Estudos Integrados da Biodiversidade Amazônica e PELD – Pesquisas Ecológicas de Longa Duração – CNPq/MCT.

## PREFÁCIO

Em 1986, aos 8 anos, conheci o *slogan* e a bandeira da Fundação SOS Mata Atlântica – “Estão Tirando o Verde da Nossa Terra”. Aquela bandeira com parte do verde “comido” muito me impressionou. É a memória mais antiga que tenho da compreensão sobre a perda de florestas como consequência de nossas escolhas. Em 1992, veio a RIO-92 e eu, adolescente, me encontrava cercada de jornais e revistas, em busca de ampliar meu entendimento sobre nossas escolhas e suas consequências.

Não por acaso, escolhi Engenharia Florestal como profissão, para aprender sobre florestas e para contribuir para o enfrentamento do desafio da sustentabilidade. Quis conhecer a Amazônia – subi pela primeira vez em uma torre meteorológica na Reserva Cuieiras, Manaus, em 2000, e percebi que precisaria de mais tempo para sentir o que é aquele mundo. Fiz estágio no exterior pelo LBA, tive contato com pontos de vistas externos sobre o Brasil e sobre a Amazônia e, ao rodar modelos hidrológicos sobre um pedacinho de floresta que conheci por apenas um mês, quis saber o que acontecia na floresta quando chove, de perto, o ano inteiro. Saí de lá, pra continuar a vida e os estudos. Quando retornei, decidi exercer a profissão de Engenheira Florestal no Amazonas. Trabalhei na Reserva do Cuieiras, com intenso trabalho de campo, e vi igarapés pequenos de terra firme tornarem-se alagamentos incríveis por um ou dois dias, e depois voltarem a ter dois metros de largura. Trabalhei com a educação de técnicos florestais, viajei por uma pequena parte dos interiores do Amazonas, Pará, Acre e Roraima, visitei comunidades, tive uma vivência muito diferente daquela em São Paulo – na capital, onde nasci, e no interior, onde cursei a faculdade.

Venho de uma família de intelectuais, que valoriza muito o conhecimento e aprendi a ver o mundo pelas lentes das ciências. Minha mãe, bióloga, contava-me histórias sobre Darwin para dormir. Decidir cursar o mestrado no INPA foi consequência da maturidade que me fez compreender que a minha maneira de me relacionar com o mundo é pelos estudos. Optei por Ecologia, estudei a vegetação ripária, que ocorre à beira desses igarapés de terra firme com os quais tanto gostei de conviver e ali muito aprendi. Aprendi que a Ciência Ecológica é complexa e fascinante. Em florestas tropicais, um grande desafio. As vivências na Reserva Ducke e na Reserva do Cuieiras são experiências que, acredito, não esquecerei. Mas havia uma inquietação: meu estudo contribuiu para o entendimento de um grupo de plantas em um pequeno pontinho naquele mundo verde. Como esse conhecimento pode ser auxiliar às escolhas humanas sobre conservação? E se os resultados de meu estudo forem combinados com os de outros estudos? Meu mestrado foi em um Sítio de Pesquisa Ecológica de Longa Duração, a Reserva Ducke, uma das áreas mais estudadas no país. Outra coisa que me intrigava era: porque é tão difícil saber o que os outros pesquisadores encontraram em seus estudos?

Quando terminei o mestrado, em 2005, soube que os coordenadores do Programa de Pesquisa em Biodiversidade, que estava em seu início, estavam interessados em preservar os dados de pesquisa para o reuso em novos estudos. Passei a trabalhar com gestão da informação de levantamentos de campo do Programa. A gestão da informação me interessou porque entendi que essa é uma área de atuação na qual é possível aproximar conhecimento e ação. Procurei o doutorado em Ambiente e Sociedade, por entender que esse curso interdisciplinar me auxiliaria a ampliar o olhar e a compreensão. E isso realmente aconteceu. No Instituto de Biologia, cursei disciplinas, fiz monitoria, interagi com alunos e professores e percebi que o que eu havia aprendido sobre ecologia, no mestrado, era uma gota em um mar de conhecimento. No NEPAM, cursei disciplinas e o contato com colegas e professores me “abriu a cabeça” para perceber olhares diferentes do meu. Esse curso me ensinou muita coisa, e uma delas é que a



construção do conhecimento é feita de inúmeras maneiras, e que é possível propor novas abordagens para a prática científica.

Esta tese é uma contribuição para um campo pouco explorado. Aqui, a interdisciplinaridade está presente – interface entre Ecologia, Ciência da Informação e interpretação da ciência. A construção do conhecimento nesse campo está em estágios iniciais – não há método estabelecido e consagrado, os primeiros estudos são base para a construção de métodos. Junto-me aos que estão dando os primeiros passos. Torço para que os resultados inspirem outros estudos nessa área.

Na Embrapa, onde trabalho há cerca de um ano, vou liderar um projeto de gestão da informação geoespacial, recém-aprovado. Difundir o conhecimento científico, associá-lo com outros sistemas de conhecimento e utilizá-los para embasar nossas escolhas – essa é minha área de atuação. Na minha opinião, *estamos* tirando o verde da nossa Terra – a diminuição do verde é consequência de nossas escolhas. Está na hora de assumirmos nossas responsabilidades: compreender as consequências de nossas escolhas é o primeiro passo para fazermos escolhas melhores.



*“As coisas mudam no devagar depressa dos tempos”*

*Guimarães Rosa*



## INTRODUÇÃO

Este estudo visa a contribuir para a integração da informação sobre biodiversidade e ecossistemas, a fim de fomentar processos de decisão e gestão ambiental. Para delinear a problemática deste trabalho, contextualizarei o estado atual de reconhecimento da necessidade de mobilizar o conhecimento científico para tratar de ameaças ambientais globais; discorrerei brevemente sobre o histórico da correlação entre ciência ecológica e a solução de problemas ambientais desde a década de 1950 até a atualidade; descreverei como o conhecimento sobre a biodiversidade vem sendo representado em sistemas de informação, desde a década de 1990, quando esse campo passou a se solidificar. Para finalizar esta introdução, apresentarei, ainda, os objetivos e a estrutura dos demais Capítulos do presente trabalho.

Mudanças no uso da terra, no clima, na ciclagem de nutrientes e na distribuição de espécies estão criando novas combinações de condições físicas e biológicas cujas consequências para o funcionamento de ecossistemas precisam ser mais bem compreendidas (FLEISHMAN et al., 2011). Cientistas vêm documentando evidências de transições em sistemas ecológicos, em diferentes escalas e até mesmo em nível planetário (BARNOSKY et al., 2012). Espécies de gramíneas invasoras alteraram a dinâmica do fogo no oeste dos Estados Unidos (KNICK et al., 2005); há registros da perda de colônias de abelhas por sintomas desconhecidos, com consequências para a polinização e a agricultura<sup>1</sup>; a extinção de espécies de peixes, devido à sobrepesca, e sua relação com a degradação de ecossistemas costeiros podem estar levando-os a um colapso (JACKSON et al., 2001); há sinais de que os ciclos de energia e da água,

---

<sup>1</sup> A “Colony Collapse Disorder”, no Brasil conhecida como “desaparecimento de abelhas” tem sido reportada em todo o mundo, na última década *Colony Collapse Disorder Progress Report*, (2010).

em porções do bioma amazônico, estão se alterando em função das mudanças no uso da terra e no regime climático (DAVIDSON et al., 2012).

Essas evidências de que podemos estar próximos a um ponto de inflexão<sup>2</sup> em escala planetária reforça a importância de melhorarmos a capacidade de fazer previsões de transições críticas, em escalas globais e locais, pela detecção de sinais de alerta (BARNOSKY et al., 2012; SCHEFFER et al., 2009). Avançar na capacidade de realizar tais previsões tem aplicações em prevenção, adaptação e mitigação e na compreensão das atividades humanas que atuam como causas dessas transições (PETERS et al., 2008). A perda da biodiversidade, por exemplo, foi considerada um dos processos interligados do sistema terrestre que já teria ultrapassado os limites planetários (ROCKSTROM, 2009). Outros dois entre os nove limites identificados pelo mesmo estudo – mudanças climáticas e ciclo do nitrogênio – também teriam sido excedidos, e a interação entre esses subsistemas terrestres comprometidos poderia propagar efeitos para outros subsistemas ainda considerados dentro dos limites.

Tal reconhecimento da interconexão entre os limites planetários revela a importância de se estudarem as transições críticas sob a lente da interdisciplinaridade (WINOWIECKI, 2011). E mobilizar esse conhecimento para a solução de problemas motivou o estabelecimento de iniciativas com o “Belmont Forum” e o “Future Earth”. O primeiro é um grupo formado pelas principais agências financiadoras de pesquisas em mudanças globais, com o objetivo de fomentar pesquisas conjuntas com vistas às urgências dos problemas ambientais, e está contido na aliança mais ampla que é a segunda, a “Future Earth”, financiada pelo ICSU (“International Council for Science”) para integrar iniciativas em pesquisa sobre a sustentabilidade global.

A representação do conhecimento gerado por programas de pesquisa e iniciativas como as supracitadas, em sistemas que realcem a relevância, a credibilidade e a legitimidade das informações produzidas (CASH et al., 2003), aliada a novas ferramentas de comunicação (GROFFMAN et al., 2010), configura uma via interessante para mobilizar o conhecimento para apoiar sistemas de tomada de decisão.

---

<sup>2</sup> Tradução livre para *tipping point*, termo amplamente utilizado na literatura internacional.

Nossa compreensão do mundo, incluindo-se esses desafios ambientais, pode avançar de maneira significativa, com a aplicação efetiva de novas tecnologias que permitem lidar com grandes quantidades de dados de procedência diversa e características heterogêneas (LYNCH, 2009). Para compreender mudanças nos ecossistemas, por exemplo, ecólogos precisam incorporar dados de outras disciplinas, como a Física (PENNISI, 2005). As diferentes áreas do conhecimento estão se deparando com questões semelhantes: como documentar a procedência dos dados, como realizar sua curadoria, como preservá-los e, acima de tudo, como compreendê-los (PARSONS et al., 2011). Desafios tecnológicos cruciais incluem a necessidade de aumentar a capacidade de captar, processar, analisar e visualizar a informação. O “dilúvio” de dados traz à tona novos modos de transformar informação em conhecimento e, ao mesmo tempo, muda a maneira de relacionar o conhecimento com o estabelecimento de políticas e com ações práticas (HEY; TREFETHEN, 2003).

Na área da Biologia, a relação entre o avanço do conhecimento e a solução de problemas ambientais vem sendo debatida há décadas. No final da década de 1950, começou a ser concebido o IBP - Programa Biológico Internacional -, construído com base na representação da natureza como uma máquina que poderia ser controlada (KWA, 1987). O IBP, conduzido entre 1968 e 1974, foi fortemente norteado pela Ecologia de Sistemas, uma abordagem focada na construção de modelos matemáticos de ecossistemas, que forneceria o conhecimento científico para a solução de problemas ambientais (WORTHINGTON, 1975), em uma época na qual o interesse pelas implicações políticas e econômicas de problemas ambientais tornou-se mundial (SHACKLEY, 1997). O IBP baseava-se na proposta de que se poderia controlar o fluxo de energia e nutrientes em ecossistemas, e a elaboração de planos de gerenciamento de ecossistemas baseados nesse controle era mais valorizada do que a própria compreensão teórica (SHACKLEY, 1997). Tais modelos permitiriam o melhor manejo, predição e controle da natureza. Essa abordagem de macromodelagem foi característica da época e também constou no famoso relatório “Os Limites do Crescimento” (“The Limits to Growth” (MEADOWS et al., 1972)), que previu que os limites de recursos naturais seriam alcançados em cerca de um século e propôs ações políticas para diminuir os crescimentos econômicos e populacionais.

Contudo, caiu em descrédito a percepção da natureza como um sistema que poderia ser controlado, que deveria estar em equilíbrio e cujo bom funcionamento poderia aumentar a produtividade ou, ainda, que uma pane nesse sistema poderia levar a uma catástrofe (KWA, 1987). Até mesmo os idealizadores do IBP reconheceram que a abordagem foi ambiciosa (WORTHINGTON, 1975). Pesquisas em modelagem que focavam no controle em larga escala foram, então, abandonadas na Meteorologia, na Economia e na Ecologia, concomitantemente, ainda na década de 1970 (KWA, 1987), juntamente com a ideia de equilíbrio e estabilidade como características de ecossistemas naturais (SHACKLEY, 1997). E esse pensamento reducionista sobre ecossistemas – de que são equilibrados e estáveis, por exemplo - foi aos poucos sendo substituído pela visão de que eles, em realidade, estão constantemente mudando (MAURER, 1998). Como consequência, aumentou a complexidade de modelos matemáticos para representar ecossistemas sob essa perspectiva - o teste de hipóteses não pode ser conduzido como uma decisão probabilística simples entre duas alternativas claramente definidas a respeito de um único parâmetro (HILBORN; MANGEL, 1997). Ecólogos devem estar preparados para desenvolver explicações mais complicadas para fenômenos ecológicos, pela incorporação de processos que operam em múltiplas escalas.

Em 1987, a Comissão Mundial para o Meio Ambiente e Desenvolvimento (WCED – “World Commission for Environment and Development”) publicou o Relatório Brundtland – “Nosso Futuro Comum”. O Relatório tratou da necessidade de se buscar o desenvolvimento sustentável, concebido como “o desenvolvimento que satisfaz as necessidades presentes, sem comprometer a capacidade das gerações futuras de suprir suas próprias necessidades”. Com sua publicação e às vésperas da Conferência RIO 92, a discussão sobre a sustentabilidade ganhou força em diversas áreas do conhecimento, inclusive na Biologia, mais especificamente na Ecologia. Um importante debate sobre como o conhecimento ecológico pode contribuir para a sustentabilidade do manejo de recursos emergiu também, com a publicação do relatório “SBI – A Iniciativa Sustentável da Biosfera - Uma Agenda de Pesquisas Ecológicas” (LUBCHENCO et al., 1991). A SBI, realizada pela Sociedade Ecológica dos Estados Unidos (ESA), definiu prioridades para a pesquisa básica em ciência ecológica e



defendeu sua “comunicação, disseminação e utilização para embasar esforços no sentido de garantir a sustentabilidade da biosfera”.

No entanto, a ideia de que o conhecimento científico ecológico evitaria a degradação do recurso explorado foi interpretada por alguns autores como a sustentação do argumento de que a tecnologia poderia resolver qualquer problema. Uma das reações a isso foi de Ludwig et al. (1993), destacando que a razão pela qual os recursos foram historicamente superexplorados, frequentemente levando ao seu colapso e extinção, não estaria tão diretamente relacionada ao nível de conhecimento biológico, mas sim ao poder e à riqueza gerados pela exploração. Além disso, na opinião dos autores, seria impossível haver consenso científico quanto à resposta de sistemas ecológicos ao manejo de recursos naturais, e essas incertezas deveriam ser abertamente levadas em consideração, para embasar políticas e ações de manejo de recursos. Advertiram, ainda, que desconsiderar as incertezas poderia levar à manipulação de questões ambientais por interesses políticos e econômicos, mas COSTANZA (1993), por sua vez, destacou que políticos e empreendedores demandam certezas, ao lidarem com a exploração de recursos e, normalmente, a legislação ambiental está amparada por termos regulatórios descritos de forma determinística. Outros autores interpretaram alguns argumentos de LUDWIG et al. (1993) como um discurso contra a realização de pesquisas em Ecologia básica i.e. (MANGEL et al., 1993; MOONEY; SALA, 1993). Muitos dos aspectos da discussão sobre sustentabilidade levantados nesse debate permanecem atuais.

A visão tradicional de que a Ecologia está restrita ao que pode ser resolvido experimentalmente levou alguns, mesmo entre ecólogos, a afirmar que ela é uma ciência que tem pouco a contribuir para o manejo ambiental (HANSSON, 1993): Shrader-Frechette e McCoy (1993) discutiram as limitações da Ecologia para resolver problemas ambientais práticos e desenvolveram argumentação para sustentar sua proposição de que a metodologia de estudos de caso é a contribuição que a Ecologia pode dar para encontrar soluções; outros autores, nesse período, também enfocaram a necessidade de ampliação da escala em pesquisas ecológicas (TURNER et al., 1995; BROWN, 1995; MAURER, 1999), que nas décadas anteriores enfatizaram pesquisas manipulativas como experimentos sobre interações em escala local i.e. (TILMAN,

1982). Análises de estruturas de comunidades biológicas passaram a incluir, a partir de então, fatores históricos e biogeográficos.

No Brasil, o termo “desenvolvimento sustentável” passa a ser enfatizado nos diversos setores do Governo com a conferência RIO 92 (JOELS; CÂMARA, 2001). Dentre os compromissos assumidos pelos membros da CDB (Convenção sobre Diversidade Biológica), assinada por 192 países e a União Europeia, durante a Conferência das Nações Unidas sobre Meio Ambiente e Desenvolvimento – RIO 92, estão o desenvolvimento de estratégias, políticas, planos e programas nacionais de biodiversidade. A CDB define diversidade biológica como

a variabilidade de organismos vivos de qualquer origem, compreendendo, entre outros, ecossistemas terrestres, marinhos e outros ecossistemas aquáticos, e os complexos ecológicos de que fazem parte; isto inclui a diversidade dentro de espécies, entre espécies e de ecossistemas (Disponível em <http://www.cbd.int/>. Acesso em 19/08/2010).

A CDB foi ratificada pelo Brasil, em 1994 (Decreto Legislativo Nº 2, 3 de fevereiro de 1994), e a Política Nacional de Biodiversidade (PNB) foi decretada em 2002 (Decreto do Poder Executivo Nº 4.339, 22 de agosto 2002). As diretrizes para a PNB contemplam sete componentes principais:

- (1) Conhecimento da Biodiversidade,
- (2) Conservação da Biodiversidade,
- (3) Uso Sustentável dos Componentes da Biodiversidade,
- (4) Monitoramento, Avaliação e Mitigação dos Impactos sobre a Biodiversidade,
- (5) Acesso aos Recursos Genéticos e Conhecimentos Tradicionais Associados e Repartição de Benefícios,
- (6) Educação, Sensibilização Pública, Informação e Divulgação sobre Biodiversidade,
- (7) Fortalecimento Jurídico e Institucional para a Gestão da Biodiversidade.

Apesar da separação em itens, todas as diretrizes estão intimamente relacionadas, sendo que não podem ser pensadas separadamente. Na PNB está explícita a importância da promoção de pesquisas ecológicas e estudos sobre o papel

dos seres vivos na funcionalidade dos ecossistemas. A Ecologia é a ciência que estuda os organismos vivos, suas interações entre si e com o ambiente.

O Programa de Pesquisas em Caracterização, Conservação, Recuperação e Uso Sustentável da Biodiversidade do Estado de São Paulo (BIOTA-FAPESP), lançado em março de 1999, constituiu-se como uma resposta da comunidade científica do Estado de São Paulo à CDB (CASTRO 2011). Seu objetivo é conhecer, mapear e analisar a biodiversidade do Estado de São Paulo, incluindo a fauna, a flora e os microrganismos, e também avaliar as possibilidades de exploração sustentável de plantas ou de animais com potencial econômico, além de subsidiar a formulação de políticas de conservação dos remanescentes florestais.

O BIOTA-FAPESP foi denominado Instituto Virtual da Biodiversidade, por sua forma de organização, integrando, via internet, pesquisadores de várias instituições e estudantes. Assim, cientistas das principais universidades públicas paulistas, institutos de pesquisa e organizações não-governamentais participam de projetos para conhecer, mapear e analisar a biodiversidade, distribuída em ambientes terrestres, marinhos e em outros ecossistemas, bem como propor alternativas e políticas públicas para preservá-la. O BIOTA-FAPESP envolve mais de 1.200 profissionais.

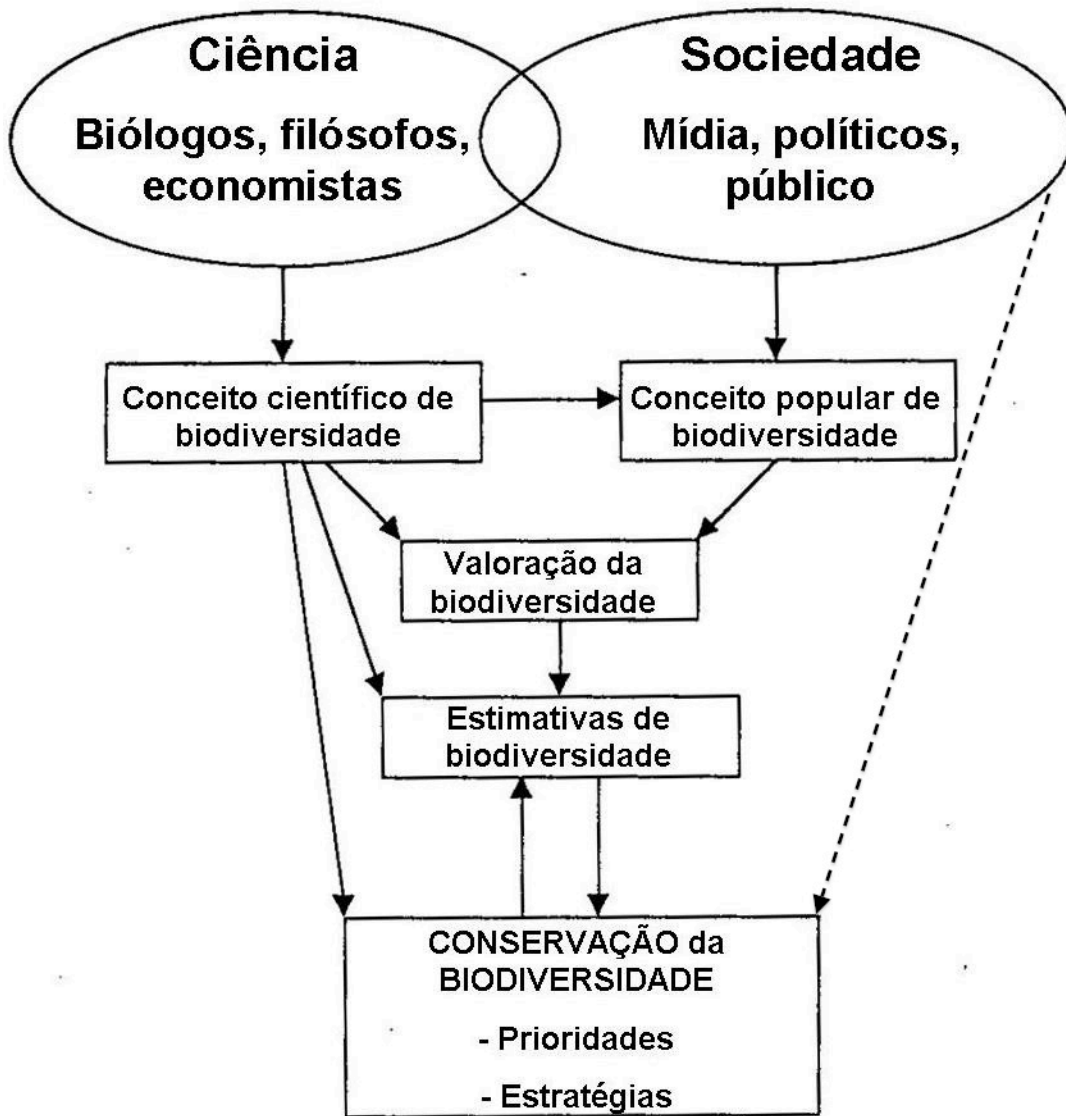
As informações taxonômicas produzidas pelo Programa BIOTA-FAPESP ([www.biota.org.br](http://www.biota.org.br)) estão em bancos de dados abertos à comunidade científica do Brasil e do exterior. A padronização das coletas permitiu a construção do Sistema de Informação Ambiental, SinBIOTA (<http://sinbiota.biota.org.br/>), que cadastra as coletas de organismos realizadas no Estado de São Paulo e as integra com coordenadas geográficas, possibilitando consultas a partir do nome científico dos organismos, do nome do coletor, da localidade ou da data de coleta.

O SinBIOTA está assentado sobre uma base cartográfica com os remanescentes de vegetação nativa, as áreas reflorestadas com espécies exóticas, as unidades de conservação, a rede de rios e de estradas e as áreas urbanas. Esta é a segunda base de dados: o Atlas do Programa BIOTA-FAPESP, que incorpora o Inventário Florestal de São Paulo, levantamento coordenado pelo Instituto Florestal - feito a partir de levantamentos de campo, de fotos aéreas e de imagens de satélite, o

Inventário monitora a área ocupada pelos remanescentes de vegetação nativa do Estado de São Paulo.

O SinBiota está integrado à rede "SpeciesLink" ("Um sistema distribuído de informação que integra em tempo real dados primários de coleções científicas", processo Fapesp 2001/02175-5): a base SinBiota/SpeciesLink vem permitindo estudos importantes para embasar a conservação, utilizando, por exemplo, ferramentas analíticas para modelagem de nicho (SIQUEIRA; DURIGAN, 2007); já um estudo de síntese amplo e baseado nas observações de espécies georreferenciadas recentemente resultou em contribuição para a formulação de políticas públicas. O Programa BIOTA-FAPESP produziu, junto com a Secretaria Estadual do Meio Ambiente, uma série de mapas que constituem o suporte científico para orientar as estratégias de conservação, preservação e restauração da biodiversidade nativa do Estado de São Paulo (RODRIGUES et al., 2008). O mapa da biodiversidade paulista (RODRIGUES; BONONI, 2008), que utilizou principalmente a base de dados SinBiota/SpeciesLink, reuniu esforços de pesquisadores especialistas em diversos grupos taxonômicos do Programa Biota. A relação entre o BIOTA-FAPESP e a formulação de políticas de conservação no Estado de São Paulo foi apontada internacionalmente como modelo de interface entre ciência e política (JOLY et al., 2010).

O termo biodiversidade é uma construção social e política, e o desenvolvimento de estratégias e prioridades de conservação de biodiversidade é produto da interação entre a comunidade científica e outros setores da sociedade (KORICHEVA; SIIPI, 2004). A Figura 1 é um esquema ilustrativo dessa interação, na qual o conceito científico de biodiversidade influencia o conceito popular de biodiversidade, e esses dois conceitos determinam formas de atribuição de valor a ela. Sua valoração é o porquê de se preocupar com a conservação da biodiversidade, seja pelo seu valor intrínseco, pelos serviços ambientais que ela promove ou pelos fármacos que pode fornecer, dentre inúmeros valores possíveis.



**Figura 1.** Papéis interativos da ciência e de outros setores da sociedade, no desenvolvimento do conceito de biodiversidade e na conservação da biodiversidade. Modificado de Koricheva e SIPI (2004).

O estabelecimento de estratégias e prioridades de conservação da biodiversidade influencia e é influenciado pelas estimativas de biodiversidade, ou seja, pelo que e como a biodiversidade é medida. Essas estimativas, por sua vez, estão relacionadas com os conceitos e com a valoração da biodiversidade.

Até por volta de 1995, a Ciência da Biodiversidade orientou-se principalmente pela taxonomia e sistemática e para a compreensão de como as espécies estão distribuídas e como evoluem (LARIGAUDERIE et al., 2012). No entanto, listas de espécies não fornecem informações sobre funcionalidades. Mas, após o “Millenium Ecosystem Assessment” (2005), a importância da compreensão da relação entre a perda de biodiversidade e a funcionalidade de ecossistemas vem crescendo (CARDINALE et al., 2007; HOOPER et al., 2012), aproximando a Ciência da Biodiversidade e a Ciência Ecológica. Serviços ecossistêmicos e o ser humano como parte de sistemas socioecológicos passam a ser conceitos-chave na Ciência da Biodiversidade (LARIGAUDERIE et al., 2012), aproximando-a também das Ciências Sociais e Econômicas. E a evolução da Ciência da Biodiversidade tem implicações para o estabelecimento de estratégias de conservação: por exemplo, um estudo em 27 localidades na Mata Atlântica revelou que os modos de reprodução e a diversidade filogenética de anfíbios são afetados por variáveis que representam a umidade, refletindo diferenças na suscetibilidade à dessecação (da SILVA et al., 2012). Esses resultados demonstram a importância de incorporar o avanço do conhecimento sobre estimativas de biodiversidade em modelos preditivos que auxiliem a compreensão dos mecanismos que regulam a distribuição de espécies.

Para que o conceito científico de biodiversidade contemple a definição da CDB, no estabelecimento de estratégias e prioridades de conservação da biodiversidade (levando em consideração a biodiversidade para além do número de espécies, como aspectos de diversidade funcional e interações entre espécies), é necessário avançar em modelos que representem e descrevam dados ecológicos e que possam ser usados em conjunto com sistemas de informação taxonômica, em coleções científicas como o SinBiota/SpeciesLink.

Em 2009, foi realizado o *workshop* “BIOTA + 10, definindo metas para 2020”, no qual se fez um balanço dos avanços do Programa e uma discussão sobre as

prioridades para os dez anos seguintes: houve consenso em apontar o SinBIOTA como uma das principais conquistas do Programa. Ao mesmo tempo, uma série de demandas foi elencada como prioritária para uma nova versão do sistema de informação, o SinBIOTA 2.0 (“Science Plan” (JOLY et al., 2010)), em consonância com os avanços da Ciência da Biodiversidade em todo o mundo, que entende que a representação do conceito científico da biodiversidade vai além da ocorrência das espécies *per se*. No mesmo ano, foi aprovado pelo edital FAPESP & Microsoft Research o projeto “SinBIOTA 2.0 - Sistema de Informações do Programa BIOTA/FAPESP: planejando os próximos 10 anos” (Processo FAPESP 09/53151-0). O Projeto objetivou incorporar novas funcionalidades ao Sistema de Informação Ambiental do Programa de Pesquisa BIOTA-FAPESP e desenvolveu um protótipo do novo sistema (MIRA et al., 2011). O presente estudo é complementar aos avanços alcançados pelo projeto SinBIOTA 2.0.

Enquanto algumas áreas da ciência já têm a tradição de lidar com ferramentas e repositórios para lidar com grandes quantidades de dados, outras áreas, como as baseadas em levantamentos de campo, ainda não possuem um quadro de trabalho para gerenciar dados gerados pelas novas formas de instrumentação que a tecnologia viabilizou (BORGMAN et al., 2006). No caso da Ciência Ecológica, tradicionalmente conduzida por grupos de pesquisa pequenos que geram dados por observações de campo locais, os quais podem ser integrados no tempo e no espaço para avançar significativamente a compreensão de processos, as práticas de gerenciamento de dados foram pouco estudadas (BORGMAN et al., 2006). A Ciência Ecológica passou a se beneficiar da aquisição de dados por sensoriamento remoto nos anos 1990 (KWA, 2005) e, mais recentemente, vem também utilizando instrumentos de campo que geram maior volume de dados, com a distribuição de sensores (PORTER et al., 2012) para estudar, por exemplo, o crescimento de plantas, comportamento de aves e variáveis micrometeorológicas.

As novas fontes de informação tornaram o desafio de estudar o gerenciamento de dados ecológicos ainda mais complexo: agora, além de avançar nos estudos sobre processos e padrões de documentação e gerenciamento de dados ecológicos gerados por observações de campo e associá-los com informações taxonômicas, é necessário

considerar as particularidades do tratamento da informação obtida por sensores de campo e remotos. Além disso, é preciso estudar conceitos e técnicas para integrar essas fontes de dados em arcabouços adaptados às questões de pesquisa de interesse (JONES et al., 2006). É central para a compreensão de padrões, processos e tendências do mundo natural a transição de um cenário no qual dados ecológicos estão guardados em gavetas ou em bancos de dados desconexos, destinados a gerar enorme variedade de conclusões de estudos locais, o que dificulta o julgamento de quão gerais ou idiossincráticas essas descobertas científicas podem ser (MICHENER; JONES, 2012), para um cenário em que os dados podem ser integrados para responder perguntas em escalas temporais e espaciais mais amplas.

Esse contexto configura como um atraente tema de estudo a representação de dados ecológicos gerados em levantamentos de campo e sua integração com aqueles obtidos por sensores automáticos no campo e remotos. Essa temática torna-se ainda mais desafiadora com o enfoque em um *hotspot* de biodiversidade, a Floresta Atlântica, bioma caracterizado por alta taxa de endemismo e pelo alto grau de ameaça (MYERS et al., 2000). A Floresta Atlântica foi uma das maiores florestas tropicais das Américas, com cerca de 150 milhões de hectares em condições ambientais altamente heterogêneas, porém estimativas recentes indicam que hoje apenas 11,4% a 16%<sup>3</sup> da área original são cobertos por remanescentes da Floresta Atlântica, sendo que 9% estão protegidos em unidades de conservação (RIBEIRO et al., 2009). Ao mesmo tempo, a região da Mata Atlântica é a de maior densidade populacional do Brasil (HOGAN, 2000). A associação de alta relevância para a conservação com o alto grau de ameaça tornou essa região a pioneira a motivar esforços voltados à conservação e restauração no país (RODRIGUES et al., 2009). A tendência a associar a biodiversidade e os ecossistemas com a economia depende também da informação disponível sobre os mesmos, enfatizando a importância de estudar e difundir conhecimento sobre estudos em ecossistemas (RING et al., 2010).

---

<sup>3</sup> Estimativas variam dependendo da inclusão ou não de remanescentes de floresta em regeneração e de fragmentos menores do que 100 ha.



O Parque Estadual da Serra do Mar é um dos maiores remanescentes de Floresta Ombrófila Densa Atlântica no Brasil e desempenha importante papel para sua conservação (JOLY et al., 2008). Considerando que a maior parte dos remanescentes florestais desse bioma ocorre em áreas de baixa aptidão agrícola, notadamente áreas íngremes, um fator importante para a compreensão de sua estrutura e funcionamento é o estudo de como os padrões e processos ocorrem ao longo de gradientes altitudinais (JOLY et al., 2012).

O Projeto Temático do Programa BIOTA/FAPESP Gradiente Funcional ("Composição florística, estrutura e funcionamento da Floresta Ombrófila Densa dos Núcleos Picinguaba e Santa Virgínia do Parque Estadual da Serra do Mar") foi concebido nesse contexto de demanda por estudos que contribuam para a compreensão da estrutura e funcionamento da Mata Atlântica. O principal objetivo do projeto foi investigar quais são as características intrínsecas das espécies que determinam a composição florística, estrutura e funcionamento das fitofisionomias da Floresta Ombrófila Densa, ao longo de um gradiente de 1000m de altitude (JOLY; MARTINELLI, 2004).

O Projeto Temático Biota Gradiente Funcional, que enfoca diferentes aspectos dos estudos sobre biodiversidade e ecossistemas, configurando-se como interdisciplinar, associado aos desafios de representar as diferentes dimensões do conhecimento ecológico e sobre a biodiversidade em sistemas de informação e de comunicar esse conhecimento para informar processos de tomada de decisão motivaram a formulação do presente estudo.

Nesse contexto, a pergunta norteadora deste trabalho é: como ferramentas computacionais podem contribuir para a integração da informação sobre biodiversidade e ecossistemas e para a mobilização do conhecimento científico para apoiar processos de tomada de decisões sobre a conservação e a gestão ambiental?

O objetivo principal deste estudo foi, assim, delinear e implementar um sistema de informação para integrar dados ecológicos e de biodiversidade do Projeto Temático Biota Gradiente Funcional e discutir as aplicações e limitações desse sistema para a conservação e gestão ambiental. Os objetivos específicos foram:

1. Investigar os elementos que compõem uma infraestrutura de informação sobre biodiversidade e discutir o potencial dessa infraestrutura para apoiar processos de tomada de decisão em conservação e gestão ambiental;
2. analisar as características de dados ecológicos, boas práticas para documentá-los e algumas ferramentas existentes para gerenciá-los;
3. analisar o sistema de informação desenvolvido para o Projeto Temático Biota Gradiente Funcional como estudo de caso para a integração da informação sobre biodiversidade e ecossistemas;
4. propor recomendações para a preservação, compartilhamento e integração de dados ecológicos para agências de fomento e instituições de pesquisa.

Nesse sentido, a estrutura da tese é a seguinte:

No Capítulo 1, investigo componentes de uma infraestrutura de informação sobre biodiversidade e discuto o potencial dessa infraestrutura como objeto-ponte entre ciência e processos de tomada de decisão em conservação e gestão ambiental. No Capítulo 2, discuto as características de dados ecológicos, boas práticas para documentá-los e algumas ferramentas existentes para gerenciá-los, para promover a integração do conhecimento ecológico. No Capítulo 3, apresento o sistema de informação desenvolvido para o Projeto Temático Biota Gradiente Funcional, como estudo de caso para a integração da informação sobre biodiversidade e ecossistemas. Por fim, com base nas considerações dos Capítulos 1 e 2 sobre as aplicações e técnicas computacionais para integrar a informação sobre biodiversidade e ecossistemas, utilizadas para delinear o estudo de caso do Capítulo 3, e a partir da análise do sistema de informação do Projeto Temático Biota Gradiente Funcional, considerando suas limitações, proponho recomendações para a FAPESP e outras agências de fomento e instituições de pesquisa em relação a preservação, compartilhamento e integração de dados ecológicos de futuros projetos.

É reconhecido que outros sistemas de conhecimento, como o conhecimento de populações locais (BERKES, 2009; HUNTINGTON, 2011) e os provenientes de iniciativas para acomodar informações de cidadãos-cientistas (HOCHACHKA et al., 2012) muito têm a contribuir para monitorar e compreender mudanças nos

ecossistemas. Contudo, neste trabalho, o termo conhecimento é utilizado para denotar apenas o conhecimento científico.

É importante, ainda, registrar que, neste trabalho, adoto o termo gestão como tradução para o termo *management*, que pode ser traduzido como manejo, gestão, gerenciamento ou administração, em português do Brasil (SEIXAS et al., 2011).



# **CAPÍTULO I. ANÁLISE DE COMPONENTES NECESSÁRIOS PARA COMPOR UMA INFRAESTRUTURA DE INFORMAÇÃO SOBRE BIODIVERSIDADE E ECOSSISTEMAS PARA MOBILIZAR O CONHECIMENTO PARA APOIAR PROCESSOS DE TOMADA DE DECISÃO.**

## **1.1. INTRODUÇÃO**

Alterações no ciclo do nitrogênio (VITOUSEK et al., 1997a, 1997b), chuva ácida (LIKENS, 1991) e escassez de água potável (POSTEL et al., 1996) são alguns exemplos de mudanças ambientais resultantes da ação humana. Apesar de aparentemente desconexas, essas questões estão interconectadas e são consequências das escolhas humanas em relação ao sistema terrestre (JASANOFF; WYNNE, 1998). O sistema terrestre compreende a união de processos, componentes e interações físicas, químicas, biológicas e sociais que, em conjunto, determinam o estado e a dinâmica do planeta (LEEMANS et al., 2001).

Historicamente, diferentes disciplinas dividem a complexidade em sistemas fechados, com fronteiras bem definidas e que podem ser estudados de um ponto de vista externo (TAYLOR, 2005). No entanto, as mudanças ambientais globais requerem avanços na compreensão do sistema terrestre de forma integrada e configuram para a comunidade científica o desafio de compor a ciência do sistema terrestre (LEEMANS et al., 2001). Mas, mesmo assim, as especificidades disciplinares devem ser reconhecidas e utilizadas para analisar os objetos relacionados com os problemas ambientais com a amplitude necessária para compreendê-los (FERREIRA, 2004).

A perda de diversidade biológica é um dos componentes das mudanças ambientais globais (ROCKSTORM, 2009), e há crescentes evidências de que essas extinções estão alterando o funcionamento de ecossistemas (HOOPER et al., 2012). Para compreender as causas e consequências desses processos, é necessário

integrar o conhecimento de inúmeras disciplinas, tanto das ciências naturais como sociais (LARIGAUDERIE et al., 2012). Essa integração depende de uma profunda compreensão das especificidades da Ciência Biológica e suas disciplinas – principalmente Taxonomia, Ecologia, Genética e Evolução (BOWKER, 2000). A partir dessa ótica de limites bem definidos entre disciplinas, é possível representar e comunicar o conhecimento biológico para integrá-lo com as outras áreas do conhecimento necessárias para interpretar o fenômeno da perda de diversidade, compreender suas consequências para o bem-estar humano, fazer previsões e propor ações para minimizar os danos. Analogamente, a interpretação de outros componentes das mudanças ambientais globais, como as mudanças climáticas, demandam conhecimento biológico para permitir a ampliação da compreensão de suas causas e consequências.

A capacidade de acompanhar as alterações na diversidade e na distribuição das espécies e sua relação com fatores ambientais é um instrumento determinante para definir estratégias e mecanismos para a conservação (DÍAS et al., 2006). Essa capacidade é também essencial para compreender e prever efeitos de mudanças do uso da terra e do clima na biodiversidade (BELLARD et al., 2012), bem como o raciocínio inverso – compreender o efeito da perda de diversidade no funcionamento do sistema terrestre (HOOPER et al., 2012). Para viabilizar as análises que permitirão esse avanço do conhecimento são necessários recursos de informação sobre biodiversidade.

Mas, além de ampliar as possibilidades de compreensão da dinâmica das mudanças ambientais globais, é preciso avançar na interpretação da ciência e na comunicação do conhecimento científico com sistemas de tomada de decisão (CLARK et al., 2001; COOK et al., 2010; LIKENS, 2010; TAYLOR, 2005). E, embora o papel da ciência seja reconhecido por muitos como crucial para lidar com os desafios colocados pela necessidade de conciliar desenvolvimento e conservação ambiental (KATES et al., 2001), ainda há poucos estudos sobre como criar sistemas que efetivamente mobilizem o conhecimento científico para apoiar decisões sobre a gestão ambiental (CASH et al., 2003). A organização da produção do conhecimento em bases de dados e a utilização

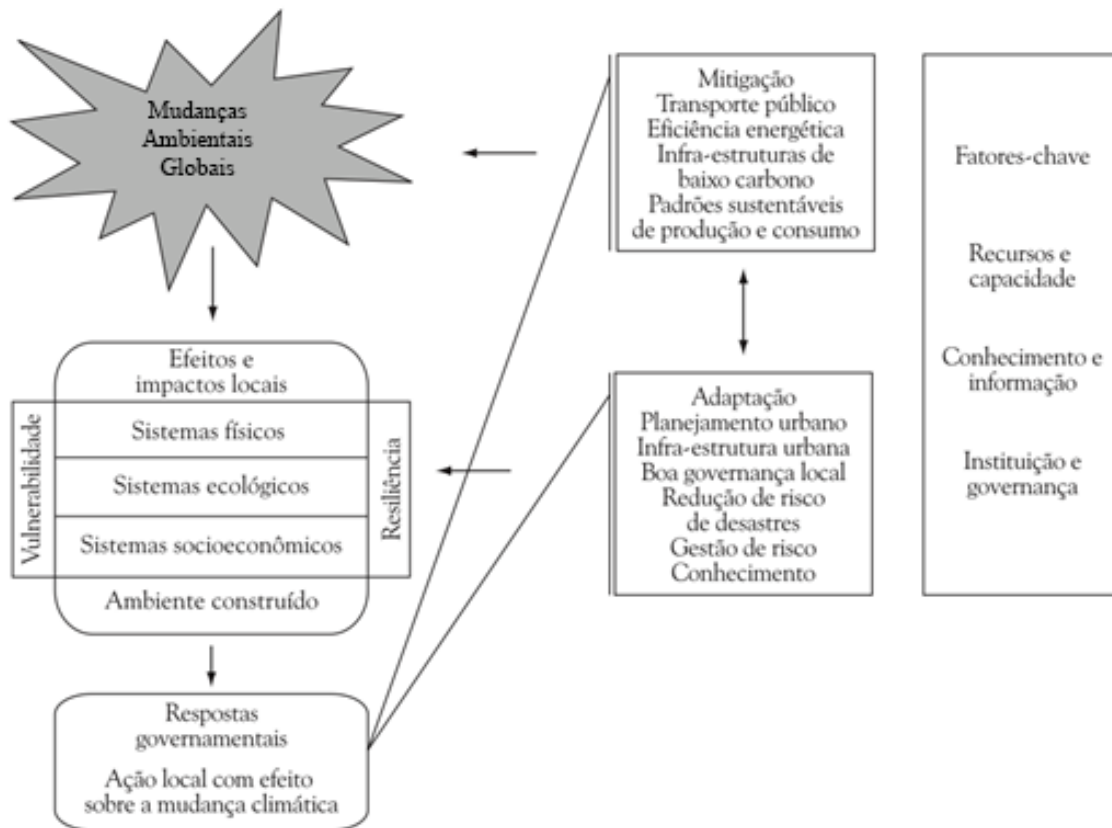
de sistemas de informação como instrumentos para comunicar resultados é uma alternativa para enfrentar esse desafio (GROFFMAN et al., 2010; HINE, 2006).

Nesse contexto, o objetivos do presente Capítulo são investigar os elementos que compõem uma infraestrutura de informação sobre biodiversidade e discutir o potencial dessa infraestrutura como objeto-ponte entre ciência e processos de tomada de decisão em conservação e gestão ambiental.

## **1.2. INTERFACE ENTRE CIÊNCIA E POLÍTICA**

Na última década, a literatura sobre a interface entre ciência e política na temática das mudanças climáticas vem se constituindo de forma sólida (JASANOFF; WYNNE, 1998; LAHSEN; NOBRE, 2007; MARTINS; FERREIRA, 2011). Apesar de a crise ambiental englobar, entre outros aspectos, tanto a questão climática como a perda da biodiversidade e degradação de ecossistemas, inclusive a interação entre os componentes da biosfera e da atmosfera - por exemplo efeitos de mudanças climáticas na distribuição de espécies - (BELLARD et al., 2012), o componente biológico das mudanças ambientais globais vem sendo menos documentado na literatura sobre a interface entre política e ciência. Entretanto, os desafios ambientais globais requerem ações que não podem ser compartimentadas, e a ciência para a sustentabilidade (KATES et al., 2001) deve contemplar componentes climáticos, ecológicos, econômicos e sociais. Essa complexidade é ilustrada na Figura 1.1.

O esquema da Figura 1.1. ilustra a relação imbricada entre sistemas ecológicos, físicos e socioeconômicos com ações locais para enfrentar as mudanças globais. Martins e Ferreira (2011) originalmente propuseram o esquema da Figura 1.1. para ilustrar os fatores que afetam ações políticas ao nível local – ou seja, nível de governos e instituições municipais – para o enfrentamento das mudanças climáticas. Proponho que os fatores que afetam as ações locais podem ser considerados no enfrentamento das mudanças ambientais globais em um sentido mais amplo, que englobe também a perda de biodiversidade e a degradação de ecossistemas.



**Figura 1.1. Esquema ilustrativo de fatores que afetam ações políticas de governos municipais relacionadas com mudanças ambientais globais. Modificada de Martins e Ferreira (2011).**

Dentre os fatores-chave apresentados na Figura 1.1., estão o conhecimento e a informação disponível para conceber e implementar estratégias de longo prazo em torno das mudanças globais em nível local. Porém, apesar da existência de inúmeros estudos em temáticas relacionadas às mudanças ambientais globais, o conhecimento não está integrado e disponível para informar decisões (OSTROM, 2009). Martins e Ferreira (2011) fornecem análises detalhadas dos demais componentes do esquema.

A representação do conhecimento em arcabouços que permitam sua integração configura-se de crucial importância para o esforço de mobilizar o conhecimento para embasar tomadas de decisão (CLARK et al., 2001). Para isso, são necessários sistemas de informação representativos das disciplinas necessárias para compreender as consequências das decisões atuais, ressaltando as pesquisas que indicam



dificuldades de tomada de decisões estratégicas sob grande incerteza por exemplo, (BRADSHAW; BORCHERS, 2000). Nesse sentido, é importante destacar que, ainda que seja possível consolidar o conhecimento e aprender fatos sobre o mundo real, esse conhecimento sempre é acompanhado de algum componente de incerteza (KING, 2011). Em estudos sobre mudanças climáticas, há uma crescente literatura discutindo como tratar a relação entre o reconhecimento da incerteza e os sistemas de tomada de decisão (SLUIJS, VAN DER et al., 2010).

A existência de fronteiras claras entre ciência e política é um tema controverso (GUSTON, 2009; LATOUR, 1994; SCHOR, 2008), sendo cada vez mais reconhecida a importância de se criarem mecanismos que facilitem a comunicação entre as fronteiras do conhecimento e da ação (GUSTON, 2009). Na mesma direção, estudos sobre a interface entre a ciência e a política vêm discutindo crescentemente a importância de organizações de contorno (*boundary*) ou organizações-ponte (*bridging*), na tomada de decisão sobre problemas ambientais (CASH et al., 2003; GUSTON, 2009; MCNIE, 2007; SHACKLEY; WYNNE, 1996; STAR; GRIESEMER, 1989). Essas organizações atuam como intermediárias entre as arenas da ciência e da política, dispensando a necessidade de delimitar fronteiras definidas, por isso podem ser especificamente concebidas para atuar nessa intermediação ou podem estar presentes em organizações com responsabilidades e atribuições mais amplas (GUSTON, 2009).

Cientistas e tomadores de decisão frequentemente diferem em expectativas tanto sobre o que são evidências sólidas e argumentos convincentes, como na caracterização da incerteza (MCNIE, 2007), e essas diferenças também ocorrem dentro da própria esfera científica, mesmo entre profissionais de uma mesma área de conhecimento – por exemplo, mudanças climáticas (SLUIJS, VAN DER et al., 2010). É exatamente nesse sentido que a efetividade de organizações intermediárias entre sistemas que ligam conhecimento à ação está relacionada com suas funções: mediação, comunicação e tradução do conhecimento nessa interface, não havendo necessidade de existirem formalmente ou de se seguir um modelo único de estruturação (CASH et al., 2003). E os aspectos da relevância, transparência e legitimidade da informação utilizada para mobilizar o conhecimento científico para o manejo de recursos são apontados como fundamentais nessa intermediação (CASH et

al., 2003; CUTTS et al., 2011). Essas características e funções de organizações- ponte as tornam adaptáveis a diferentes pontos de vista (STAR; GRIESEMER, 1989), por trazerem as áreas de *expertise* de diferentes grupos para a mesa: isso torna o acesso à informação e ao processo de produção da informação mais transparente aos diferentes interessados (CASH et al., 2003).

Exemplos de organizações-ponte nos Estados Unidos são as iniciativas financiadas pelo programa da National Science Foundation (NSF) intitulado “Tomada de Decisão sob Incerteza” – “National Science Foundation Program Decision Making Under Uncertainty” – DMUU: Center for Research on Environmental Decisions (Earth Institute, Universidade de Columbia), Decision Center for a Desert City (Universidade do Estado do Arizona), Climate Decision Making Center (Universidade Carnegie Mellon), Center for Robust Decision Making on Climate and Energy Policy (Universidade de Chicago).

Uma função essencial de organizações-ponte é facilitar a criação de objetos-ponte. “Obras-ponte” ou “objetos-ponte” foram as primeiras sugestões de representação das interações entre ciência e política, pela literatura da sociologia da ciência (GIERYN, 1983; STAR; GRIESEMER, 1989). Objetos-ponte podem ser adotados e interpretados por múltiplos atores e instituições (STAR; GRIESEMER, 1989) e podem ser, por exemplo, cenários, modelos preditivos (CASH et al., 2003) e sistemas de informação geográfica (CUTTS et al., 2011).

### **1.3. INFRAESTRUTURA DE INFORMAÇÃO SOBRE BIODIVERSIDADE**

Concomitantemente ao avanço na base teórica sobre a interface entre ciência e política; a Ciência da Informação e a Biblioteconomia – que são os campos tradicionalmente incumbidos de representar o conhecimento – vêm sofrendo transformações significativas nos últimos 50 anos, devido ao desenvolvimento de novas ferramentas computacionais (BORGMAN et al., 2006). O desenvolvimento de

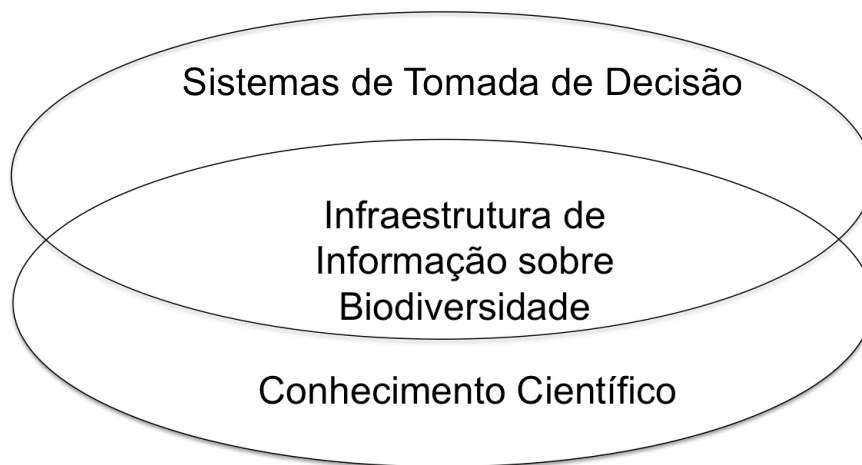
repositórios institucionais emergiu como uma nova estratégia, que permite a universidades e outras organizações acelerarem a comunicação científica, transformando-se para além de seu histórico papel de passivamente dar suporte às editoras (LYNCH, 2003). Dessa forma, infraestruturas de informação têm fornecido novas ferramentas e serviços para a prática científica e para a representação do conhecimento. Essas novas possibilidades de troca de informação entre cientistas e entre cidadãos e cientistas têm incentivado mudanças na comunicação científica – movimentos “Open Science”, “Open Data” e “Open Knowledge” vêm ganhando adeptos - por exemplo Woelfle et al. (2011). Editoras de revistas científicas vêm sendo pressionadas a mudar seus modelos de negócios de forma a torná-los menos restritivos e mais favoráveis à disseminação do conhecimento, como no caso do boicote à Elsevier (<http://thecostofknowledge.com/>).

A informação sobre biodiversidade é produzida por inúmeras organizações no Brasil – na maioria dos casos por instituições públicas, em todas as esferas de governo. Apesar de sua importância – tanto para o avanço do conhecimento quanto para embasar decisões sobre a gestão ambiental – essa informação está dispersa e não se sabe quantas e quais instituições detêm recursos de informação sobre o tema (LEWINSOHN; PRADO, 2002). Nesse contexto, ainda é difícil para os usuários saberem o quê, onde e como essa informação pode ser acessada e quem é responsável pelo seu armazenamento e manutenção.

Essa dificuldade de acesso aos recursos de informação em biodiversidade no Brasil evidencia que eles não são bem gerenciados: em todos os níveis da administração, há duplicação de esforços e de gastos, o que dificulta a tomada de decisão para o planejamento, gestão e resposta aos problemas ambientais da comunidade em geral. Isso se deve fortemente ao fato de não haver uma estrutura disponível para que usuários, geradores e gestores de recursos de informação sobre biodiversidade colaborem eficientemente para construir uma Infraestrutura Nacional de Dados sobre Biodiversidade. Uma Infraestrutura de Informação sobre Biodiversidade compõe o arcabouço básico que deve ser eficientemente coordenado e gerenciado para os interesses da sociedade (GURALNICK et al., 2007).

É preciso, no entanto, registrar que, atualmente, o crescente reconhecimento da importância da gestão da informação sobre biodiversidade como interesse nacional vem impulsionando setores do Governo Federal, aliado ao PNUD (Programa das Nações Unidas para o Desenvolvimento – “Global Environment Facility” (JORDAN, 1994)), a estimular a colaboração de instituições de relevância para a temática da biodiversidade (International Bank for Reconstruction and Development (2008); UNEP (2010)), com o intuito de desenvolver uma estratégia para um sistema de informação sobre biodiversidade e ecossistemas.

A fim de que se forme a base de conhecimento necessária para informar decisões, é fundamental a consolidação de conceitos e práticas de infraestrutura de informação para dados de biodiversidade (GURALNICK et al., 2007). Dessa forma, uma infraestrutura de informação sobre biodiversidade pode ser considerada um objeto-ponte entre a ciência e processos de decisão em conservação e gestão ambiental, conforme o esquema teórico ilustrado na Figura 1.2.



**Figura 1.2.** Infraestrutura de informação em biodiversidade como objeto-ponte na interface entre o conhecimento científico e sistemas de tomada de decisão.

Infraestrutura da informação é um tema tratado em profundidade em Estudos sobre Ciência e Tecnologia (“STS - Science and Technology Studies” –, uma visão abrangente é apresentada em Bowker et al., 2010). Tais estudos evidenciam o caráter dinâmico desse campo, dificultando sua definição pela adoção de conceitos fechados. Neste estudo, adotei a definição de infraestrutura de informação do “Global Spatial Data Infrastructure” (GSDI): “Conjunto de tecnologias, políticas e arranjos institucionais que facilitam a disponibilidade e o acesso aos dados” (GSDI, 2009).

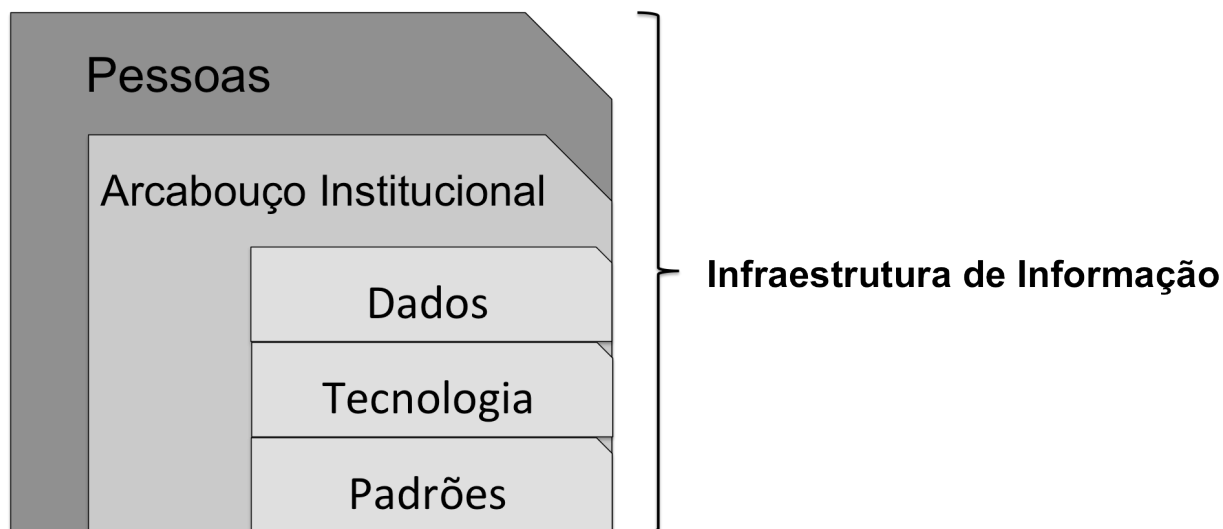
Esse organismo global – o GSDI – foi uma das fontes utilizadas pela CONCAR (Comissão Nacional de Cartografia), para propor a INDE (Infraestrutura Nacional de Dados Espaciais), instituída pelo Decreto N° 6.666, de 27/11/2008. Nesse instrumento, a INDE é definida como

um conjunto integrado de tecnologias, políticas, padrões, mecanismos de coordenação, monitoramento e acordos necessários para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento, a disseminação e o uso dos dados geoespaciais de origem federal, estadual, distrital e municipal. (COMITÊ DE PLANEJAMENTO DA INFRAESTRUTURA NACIONAL DE DADOS ESPACIAIS, 2010)

E, mesmo considerando que a INDE esteja inserida no universo da geoinformação, tais conceitos fornecem elementos para explorar características que compõem uma infraestrutura de informação de uma forma geral. É importante ressaltar que dados sobre biodiversidade são, em grande parte, passíveis de representação geográfica – em especial os levantamentos de campo enfocados nesse estudo.

No presente estudo, baseio-me na subdivisão em cinco componentes elaborada por Warnest (2005) e adotada no Plano de Ação da INDE (COMITÊ DE PLANEJAMENTO DA INFRAESTRUTURA NACIONAL DE DADOS ESPACIAIS, 2010), para analisar uma **INFRAESTRUTURA DE INFORMAÇÃO SOBRE BIODIVERSIDADE**. O primeiro componente, que permeia todos os demais, são as **Pessoas** ou **Atores** envolvidos com o planejamento, aquisição, tratamento, gerenciamento, análise e uso da informação. O segundo é o **Arcabouço Institucional**, ou seja, o contexto de governança e coordenação que garante a implementação e longevidade da infraestrutura, por meio de acordos, articulações e regras. O terceiro

são os **Dados** propriamente ditos; no caso de levantamentos de campo de biodiversidade, comumente são representados em termos de conjuntos de dados heterogêneos e seus respectivos metadados, glossários, dicionários de dados, ontologias. O quarto componente é a **Tecnologia**, ou seja, as ferramentas computacionais que viabilizam o acesso, aquisição, armazenamento e distribuição da informação, o que inclui elementos de desenvolvimento de *software* e capacidade de armazenamento, processamento e conectividade. Por último, o quinto componente são as **Normas e Padrões** que asseguram a integração da informação: modelos de dados, padrões de metadados e de interoperabilidade. O esquema da Figura 1.3. ilustra a relação entre os elementos – o componente **Pessoas** permeia todos os demais.; o componente **Arcabouço Institucional** determina a configuração dos outros três elementos.



**Figura 1.3. Relação entre componentes de uma Infraestrutura de Informação: o elemento Pessoas permeia todos os demais.**

Nas próximas seções, discorrerei sobre os cinco componentes da Infraestrutura de Informação sobre Biodiversidade.

### 1.3.1 Pessoas

As **Pessoas** ou **Atores** envolvidos com o planejamento, aquisição, tratamento, gerenciamento, análise e uso da informação sobre biodiversidade são o componente-chave que permeia todos os demais. Em primeiro lugar, um aspecto a destacar é o fato de que, no universo da ciência da biodiversidade e ecossistemas, pesquisadores têm o hábito de planejar cuidadosamente seus experimentos e amostragem de campo e, crescentemente, têm se preocupado antecipadamente em planejar as análises estatísticas que serão realizadas (ZUUR et al., 2010). Contudo, os mesmos pesquisadores não têm o hábito de planejar como os dados coletados serão preservados, como será feita sua documentação, como se dará o compartilhamento desses dados com colegas. Assim, é comum acontecer, até mesmo dentro de um mesmo grupo de pesquisa, que uma nova questão científica provoque um movimento de “garimpo” de dados, e isso ocorre pelo contato com ex-alunos, ou em buscas no computador antigo do pesquisador líder - máquina já encostada no canto do laboratório aguardando ser doada - ou, até mesmo, em *back-ups* em mídias obsoletas e inadequadas para a preservação de informação ao longo prazo.

Adicionalmente, mesmo quando informações de interesse são encontradas, elas frequentemente não estão documentadas com metadados que permitam sua compreensão, e é comum que até mesmo o responsável pelo levantamento desses dados tenha esquecido detalhes importantes para sua interpretação e reutilização. Essa ausência de cultura de planejar e executar a preservação dos dados resulta em perda de informações valiosas (ZIMMERMAN, 2008). E, nesse cenário, ainda é preciso lembrar que o arquivamento de dados de pesquisas em repositórios é oneroso, em termos de recursos financeiros necessários e tempo investido, o que provavelmente contribuiu para apenas recentemente agências financiadoras e revistas científicas passarem a incentivar o arquivamento de dados em repositórios.

De todo modo, a importância do fator humano em infraestruturas de informação, a colaboração e compartilhamento de dados são tópicos que vêm sendo cada vez mais estudados e documentados (BORGMAN et al., 2006). E, além do já citado, há ainda a resistência por parte dos pesquisadores para tornar dados disponíveis, por inúmeros

motivos, dentre os quais a falta de tempo para se dedicar ao preenchimento de formulários de documentação de dados (NELSON, 2009). Mas também existem os pesquisadores que se preocupam com a possibilidade de seus dados serem mal interpretados ou utilizados de forma inapropriada (COSTELLO, 2009), ou o fato de que um mesmo conjunto de dados pode ser usado para mais de um estudo pelo mesmo pesquisador ou grupo de pesquisa que se esforçou para levantá-lo, e os pesquisadores querem usufruir da segurança de explorar seus próprios dados antes de expô-los à comunidade. Por outro lado, estudos mostraram que a exposição de dados à comunidade científica pelo compartilhamento em repositórios é benéfica também ao pesquisador/grupo de pesquisa, pelo aumento de citações (PIWOWAR et al., 2007) e do número de publicações geradas pelo reuso de dados arquivados, conferindo transparência e amplitude à colaboração (PIWOWAR, 2011). Sem falar no fato de que a exposição de dados brutos aumenta a chance de livrar os pesquisadores de erros, pois a comunidade científica que lida com informações similares pode detectar inconsistências que tenham passado despercebidas pelos geradores desses dados (NELSON, 2009).

Uma via promissora para incentivar que mais dados tornem-se disponíveis é a prática de publicá-los, ao invés de simplesmente compartilhá-los (COSTELLO, 2009). Um exemplo é a iniciativa da ESA ("Ecological Society of America") de publicar "Data Papers" vinculados ao repositório "Ecological Archives", que são compilações e sínteses de bases de dados e metadados associados. Esses periódicos publicam as bases de dados com a citação, resumo e outras informações associadas, analogamente à publicação de artigos, de forma que os créditos são claramente dos responsáveis pelos levantamentos de dados, contornando a maior parte dos problemas apontados pelos pesquisadores relutantes em compartilhar dados. As revistas "American Naturalist", "Evolution" e "Heredita" passaram a requerer que os dados utilizados nos artigos por eles publicados sejam livremente disponibilizados em repositórios digitais (BUTLIN, 2010; RAUSHER et al., 2010; WHITLOCK et al., 2010). A "Biotropica" estuda fazer o mesmo (BRUNA, 2010).

Essa prática de publicar dados passou recentemente a ser incentivada também pelo GBIF ("Global Biodiversity Information Facility") (COSTELLO et al., 2012). Os



autores argumentam que, além de a publicação de dados ser uma prática que incentiva pesquisadores a compartilharem informações, como já argumentado nas iniciativas citadas anteriormente, a qualidade de dados pode ser verificada por mecanismos de revisão pelos pares, analogamente à publicação de trabalhos completos em revistas científicas. De fato, mecanismos para a aferição de qualidade de dados de biodiversidade são necessários. Um estudo mostrou que a interpretação de dados ecológicos, para que sejam úteis para o reuso, requer o detalhamento de circunstâncias que frequentemente são deixadas para trás, quando os dados são inseridos em repositórios públicos, comprometendo-se sua qualidade (ZIMMERMAN, 2008).

Sem dúvida, o esforço de produzir um conjunto de dados bem descrito e cuidadosamente validado deve ser recompensado. A publicação de dados, de forma análoga à publicação de artigos, é uma tendência que reforça o valor dos dados e deve produzir um efeito benéfico ao seu compartilhamento. Contudo, essa prática tem desvantagens: o acúmulo de conjuntos de dados desconexos em repositórios tende a dificultar a obtenção de dados de interesse por consultas e o reuso demandará esforços em etapas anteriores à geração de resultados em novas análises. E as métricas de reuso devem ser capazes de captar o valor dos dados disponibilizados, e não seu volume. Além disso, a ciência do século XXI traz à tona desafios que vão além do correto arquivamento de dados pela publicação: é necessário elaborar mecanismos que permitam o reuso de forma interativa, colaborativa e dinâmica (PARSONS; FOX, 2012)<sup>4</sup>.

É esperado que exemplos de reuso de dados incentivem pesquisadores a compartilharem, contudo a investigação sobre como dados disponibilizados ao público estão sendo reutilizados em novos estudos ainda é recente. Um estudo sobre o reuso de dados de expressão gênica demonstrou que, mesmo em uma área do

---

<sup>4</sup> Artigo submetido ao “Data Science Journal”. Os autores expuseram o manuscrito à comunidade da ESIP (“Earth System Information Partners”) e ficaram surpresos com as 50 páginas de comentários e sugestões que receberam. Uma evidência de como a prática científica está se tornando cada vez mais colaborativa.

conhecimento na qual repositórios, padrões e políticas já estão maduros, muito da informação ainda está inacessível (PIWOWAR 2011).

As pessoas que gerenciam agências de fomento estão criando incentivos para que os dados de pesquisas sejam disponibilizados. Em 2011, a NSF (“National Science Foundation”) e o NERC (“Natural Environment Research Council”) passaram a exigir que as propostas de novos auxílios viessem acompanhadas por planos de gerenciamento de dados (DMP – “Data Management Plan”). Um DMP descreve os dados que serão criados e como serão gerenciados e disponibilizados ao longo do tempo. O conteúdo de um DMP deve incluir: os tipos de dados que serão gerados, os padrões que serão aplicados, provisões para o arquivamento, preservação e acesso, previsão do término da coleta de dados. Há de se imaginar que, em sua fase inicial, esses incentivos de revistas e agências de fomento devem causar um impacto nas práticas dos pesquisadores, uma vez que eles terão que se deparar com essas questões com as quais não estão habituados.

Além da cultura de disponibilizar dados, outras etapas do processo científico podem se tornar mais transparentes. Fluxos de trabalhos científicos, ou *scientific workflows*, são representações de processos envolvidos na realização de uma análise científica (JONES et al., 2006): combinam-se dados e procedimentos computacionais em uma sequência estruturada de etapas que implementam uma solução para uma questão científica (ALTINTAS et al., 2004). Muitas vezes, uma análise científica envolve vários processos independentes entre a aquisição dos dados brutos e os resultados que serão publicados – o uso de *workflows* permite que um processo analítico seja repetido inúmeras vezes e possibilita até mesmo monitoramento de parâmetros e indicadores pré-definidos. Além disso, *workflows* podem ser arquivados e compartilhados, permitindo que o processo de analítico que resulta na geração de conhecimento científico seja comunicado de forma transparente.

Essa possibilidade expõe as decisões que o cientista toma ao longo do tratamento dos dados aos seus pares, permitindo que as análises sejam refeitas. Usuários de dados e de *workflows* podem ser outros cientistas. Além disso, essa transparência aproxima a comunidade científica de outro grupo de usuários, os tomadores de decisão. Por exemplo, no Parque Nacional Kruger, África do Sul,

cientistas e gestores desenvolveram uma solução para informar decisões sobre o manejo do Parque, baseada em um *workflow*. A densidade de elefantes, a densidade de árvores e a dinâmica da população de búfalos foram utilizadas como indicadores ambientais do Parque; dados sobre esses indicadores são coletados continuamente ali e, a partir do conhecimento sobre os padrões de variação dos valores obtidos, foram estabelecidos limites mínimos e máximos de valores aceitáveis. Um *workflow* é utilizado para analisar periodicamente esses dados e monitorar como os indicadores estão se comportando ao longo do tempo. Para ser eficaz, uma infraestrutura de informação sobre biodiversidade precisa acomodar as demandas dos usuários, como esses gestores do caso do Parque Kruger.

Além de indivíduos envolvidos com planejamento, aquisição, análise e uso de informação sobre biodiversidade, o componente **Pessoas** também está presente em outras etapas do procedimento científico, como, por exemplo, na figura de gestores da informação em biodiversidade, personagens cruciais para se viabilizarem os processos necessários para representar o conhecimento científico. A análise dos demais componentes da infraestrutura aprofundará o tema de atribuições e necessidades desses personagens.

### ***1.3.2 Arcabouço Institucional***

Entende-se por **Arcabouço Institucional** a rede de acordos e articulações com vistas a parcerias e cooperações, para fins de interesse comum. O fortalecimento institucional, os marcos legais, a captação de recursos e a gestão/coordenação, bem como regras formais e não formais estão contidos nesse Arcabouço.

Sob o ponto de vista legislativo, os marcos legais de destaque para a informação científica em biodiversidade e ecossistemas, no Brasil, são a Infraestrutura Nacional de Dados Espaciais (INDE, Decreto Nº 6.666, de 27 de novembro de 2008), a Lei de Acesso à Informação (Lei Nº 12.527, de novembro de 2011) e a Lei de Acesso Público aos Dados e Informações Existentes nos Órgãos e Entidades Integrantes do SISNAMA – Sistema Nacional do Meio Ambiente (Lei Nº 10.650, de 16 de abril de 2003).

A Infraestrutura Nacional de Dados Espaciais (INDE, Decreto Nº 6.666, de 27 de novembro de 2008 (<http://www.inde.gov.br/>)) foi instituída com o objetivo de integrar os dados geoespaciais existentes nas diversas instituições do Governo brasileiro, harmonizando-os, disseminando-os e proporcionando seu uso efetivo. É definida como o conjunto integrado de tecnologias; políticas; mecanismos e procedimentos de coordenação e monitoramento; padrões e acordos, necessários para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento, a disseminação e o uso dos dados geoespaciais de origem federal, estadual, distrital e municipal. A Comissão Nacional de Cartografia (CONCAR), por meio do trabalho dos seus comitês especializados, tem produzido os padrões e normas a serem adotados na produção dos dados e informações geoespaciais pelos órgãos públicos.

Na mesma linha, anos depois, promulgou-se a Lei de Acesso à Informação (Lei Nº 12.527, de novembro de 2011), que se aplica às

entidades privadas sem fins lucrativos que recebam, para realização de ações de interesse público, recursos públicos diretamente do orçamento ou mediante subvenções sociais, contrato de gestão, termo de parceria, convênios, acordos, ajustes ou outros instrumentos congêneres. (Lei Nº 12.527, 2011).

No caso de informações de pesquisas, a mesma Lei declara que esse acesso à informação “não compreende as informações referentes a projetos de pesquisa e desenvolvimento científicos ou tecnológicos cujo sigilo seja imprescindível à segurança da sociedade e do estado”. Há uma seção intitulada “Da Classificação da Informação quanto ao Grau e Prazos de Sigilo”, que dispõe sobre as informações consideradas “imprescindíveis à segurança da sociedade ou do Estado”, na qual constam informações cuja divulgação ou acesso irrestrito possam “prejudicar ou causar risco a projetos de pesquisa e desenvolvimento científico ou tecnológico, assim como a sistemas, bens, instalações ou áreas de interesse estratégico nacional [...]”. Essas informações podem ser consideradas ultrassecretas, secretas ou reservadas, com prazo máximo de restrição de acesso à informação de 25, 15 e 5 anos, respectivamente.

Por outro lado, a Lei Nº 10.650, de 16 de abril de 2003, já dispunha sobre o acesso público aos dados e informações existentes nos órgãos e entidades integrantes do SISNAMA (Sistema Nacional do Meio Ambiente), determinando que

os órgãos e entidades da Administração Pública, direta, indireta e fundacional, integrantes do Sisnama, ficam obrigados a permitir o acesso público aos documentos, expedientes e processos administrativos que tratem de matéria ambiental e a fornecer todas as informações ambientais que estejam sob sua guarda, em meio escrito, visual, sonoro ou eletrônico. (Lei Nº 10.650, 2003).

Do ponto de vista de dados gerados por cientistas de universidades e instituições de pesquisa, alguns projetos e programas de pesquisa nacionais e internacionais instituíram suas políticas de compartilhamento de dados. De uma forma geral, tratam-se de regras sobre prazos e procedimentos para o compartilhamento de dados em diferentes níveis – dentro dos grupos de pesquisa envolvidos com sua geração, dentro do programa de pesquisa guarda-chuva, para o público externo. Alguns documentos fazem alusão explícita a regras que regulam como deve ser acordada a coautoria, no caso de uso de conjuntos de dados que gerem publicações. Esse é o caso da política de dados do LBA (Programa de Grande Escala da Biosfera-Atmosfera na Amazônia (<http://lba.inpa.gov.br>)), do Código de Ética, Compartilhamento de Dados e Política de Publicação para participantes da RAINFOR (“Amazonian Forest Inventory Network” ([http://www.geog.leeds.ac.uk/projects/rainfor/manuals/RAINFOR\\_Policy\\_Portuguese.pdf](http://www.geog.leeds.ac.uk/projects/rainfor/manuals/RAINFOR_Policy_Portuguese.pdf))) e da Política de Dados e Publicações do Programa FAPESP de Pesquisa sobre Mudanças Climáticas Globais (<http://www.fapesp.br/4483>).

A política de dados do PPBio (Programa de Pesquisa em Biodiversidade do Ministério de Ciência, Tecnologia e Inovação ([http://ppbio.inpa.gov.br/Port/docsinternos/politica\\_dou.pdf](http://ppbio.inpa.gov.br/Port/docsinternos/politica_dou.pdf))), determina prazos de disponibilização de metadados e de dados e institui a categoria de dados sensíveis, a saber:

Dados sensíveis: são dados preliminares ou consolidados que, se liberados ao acesso público, possam resultar em efeito adverso ao local e/ou às comunidades de origem da mesma e, por isso, passíveis de restrição. Podem ser considerados dados sensíveis: (a) a localização de espécies que estejam na lista de espécies ameaçadas de extinção; (b) dados de espécie que possa ser roubada ou traficada por sua raridade ou valor econômico (considerando sua potencialidade: como fornecedora de produtos que venham a ser utilizados na indústria farmacêutica ou química; como agente de controle biológico; entre outras); (c) a localização de habitats e sítios arqueológicos, culturais ou históricos cujo acesso possa ameaçar sua integridade; (d) informações utilizadas em decisões de Política de Estado que possam vir a interferir no alcance das metas e objetivos da mesma. (Política de dados do PPBio 2009).

A restrição de acesso a dados de espécies ameaçadas foi considerada não aderente à legislação brasileira por pelo menos um estudo (GONZALES, 2010), evidenciando-se a necessidade de promover o debate sobre a disponibilização de dados sobre a biodiversidade e ecossistemas pela comunidade científica no Brasil. O prazo de disponibilização após 2 anos da coleta do dado, estabelecido na política do PPBio, é inspirado principalmente nos prazos estabelecidos na política de dados do LTER (“Long Term Ecological Research” (<http://www.lternet.edu/data/netpolicy.html>)). No entanto, no caso dessa rede norte-americana, a maior parte das informações disponíveis são apenas os metadados, enquanto o acesso aos dados ainda é mais restrito. O NCEAS (“National Center for Ecological Analysis and Synthesis” (<http://nceas.ucsb.edu/datapolicy>)) tem uma política de dados sintética e direta: um dos objetivos do Centro é promover a disponibilização de dados para a comunidade científica no sentido amplo, então a política determina que os dados devem ser depositados no repositório do centro. O diferencial é também solicitar o depósito da informação sobre as ferramentas analíticas usadas. Mais sintética ainda é a política de dados do PISCO (“Partnership for Interdisciplinary Studies of Coastal Oceans” (<http://piscoweb.org/data/data-sharing-policy>)), definida em apenas três itens e uma cláusula de exceções.

Já a política de dados da NSF (“National Science Foundation” (<http://www.nsf.gov/bfa/dias/policy/dmp.jsp>)) declara que “é esperado que pesquisadores compartilhem dados com outros cientistas”, porém não explicita detalhes. Recentemente (há cerca de um ano), a NSF passou a exigir planos de gerenciamento de dados em todas as propostas submetidas para financiamento.

Uma série de iniciativas internacionais de redes de sistemas de informação sobre o sistema terrestre estão em andamento:

(a) GBIF

O GBIF (“Global Biodiversity Information Facility”), estabelecido em 2001, é uma iniciativa multilateral estabelecida por acordos intergovernamentais via Memorando de Entendimento, no qual participantes conduzem cooperação técnico-científica para ampliar o acesso à informação sobre biodiversidade. A cada cinco anos, os signatários expressam a intenção de manter sua participação no GBIF ou de tornarem-se novos membros. O GBIF é um organismo de coordenação criado com o objetivo de promover os esforços técnicos e científicos para desenvolver e manter uma unidade de informação global para compartilhar informação digital sobre biodiversidade. Para a entidade, dados sobre biodiversidade referem-se a dados científicos, principalmente sobre espécies biológicas e sobre espécimes ou observações de organismos individuais, e ela tem focado em dados de ocorrências de espécies e listas de espécies depositadas em museus ou compiladas para determinadas localidades. O GBIF tem reforçado intensamente o mecanismo de publicar dados como principal incentivo a que os provedores de dados compartilhem os mesmos. O GBIF tem uma política de dados” (<http://www.gbif.org/participation/being-a-part-of-gbif/policies/>), a qual coloca as atividades dessa iniciativa global como não infratora de propriedade intelectual.

A entidade trabalha em íntima colaboração com o TDWG (“Biodiversity Information Standards”<sup>5</sup>, conhecido anteriormente por “Taxonomic Databases Working Group”), uma associação científica e educacional afiliada à União Internacional de Ciências Biológicas (“International Union of Biological Sciences”). O TDWG foi formado para estabelecer uma colaboração internacional entre projetos de bases de dados de

---

<sup>5</sup> Em tradução livre, Padrões de Informação sobre Biodiversidade.

biodiversidade pelo desenvolvimento, adoção e promoção de padrões e diretrizes para registrar e compartilhar dados sobre organismos biológicos. Historicamente, o TDWG destaca-se pela atuante participação na evolução e disseminação do padrão para registros de ocorrências “DarwinCore” (WIECZOREK et al., 2012). Atualmente, além de manter o trabalho relacionado ao padrão “DarwinCore”, o grupo está também focado no desenvolvimento de padrões para o intercâmbio de dados de biodiversidade, em especial o Tapir (“Access Protocol for Information Retrieval”).

#### (b) GEO BON

O “Group on Earth Observations” (GEO) é uma parceria voluntária entre governos e organizações internacionais, lançada em 2005 com o objetivo de estabelecer um sistema de observação global em nove áreas consideradas de “benefício social”: desastres, saúde, energia, clima, água, tempo, ecossistemas, agricultura e biodiversidade. O GEO BON (“Biodiversity Observation Network”) é o componente do GEO associado ao tema da biodiversidade (“Geo Bon Strategy e website [www.earthobservations.org/geobon.shtml](http://www.earthobservations.org/geobon.shtml))).

#### (c) OBIS

O OBIS (“Ocean Biogeographic Information System”) é um repositório de bases de dados sobre espécies marinhas, criado pelo Censo da Vida Marinha, uma parceria internacional para o estudo da vida marinha, e a partir de 2010, passou a ser gerenciado pela IOC (“Intergovernmental Oceanographic Commission”), da UNESCO (IOC of UNESCO, 2012)

#### (d) DataONE –

O DataONE (“Data Observation Network for Earth”) é um projeto financiado pela NSF (“National Science Foundation”), dentro do programa “DataNet”, o qual é focado em ciberinfraestrutura (ALLARD, 2012). Sua missão é possibilitar inovar a geração de conhecimento pelo acesso universal a dados sobre a vida na Terra e sobre o ambiente que a sustenta. Suas ações têm focado principalmente no desenvolvimento colaborativo de ferramentas que permitam a integração de diferentes sistemas de



informação, na comunicação com o público e na capacitação de recursos humanos. Recentemente, a rede DataONE lançou um portal que integra diferentes repositórios de dados, como do LTER (“Long Term Ecological Research”) e do LBA, em um só sistema de busca, sinalizando que a integração da informação por meio de metadados é um caminho para promover estudos interdisciplinares.

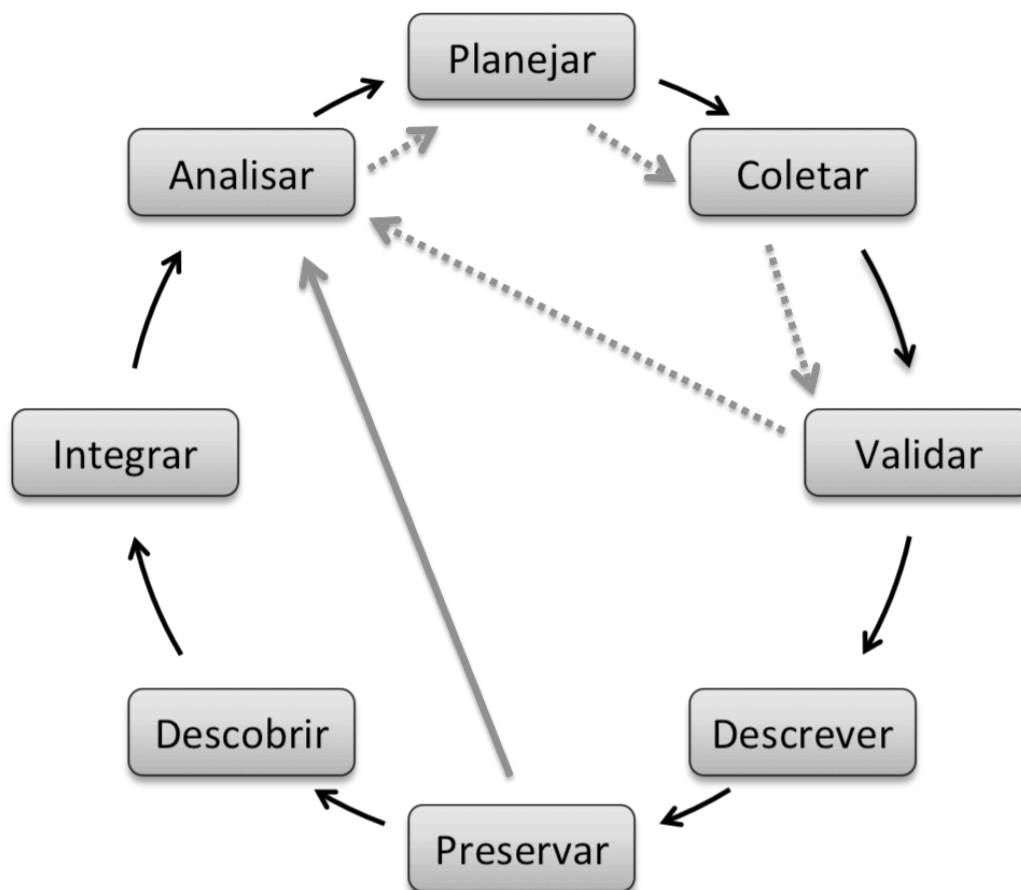
Mesmo assim, apesar de haver algumas iniciativas no sentido de estabelecer governança, por meio de acordos e articulações, para integrar e viabilizar a manutenção de diferentes sistemas de informação sobre biodiversidade existentes, esse é o aspecto menos contemplado da infraestrutura da informação sobre biodiversidade até o momento no país. Além da sua importância para estabelecer regras e normas de disponibilização e reuso de dados, esse componente é fundamental para viabilizar a interface com os sistemas de tomada de decisão. Articulações nessa interface são necessárias para que a infraestrutura da informação em biodiversidade atue de fato como objeto-ponte, permitindo que as demandas dos usuários sejam reconhecidas e que arranjos e soluções sejam desenvolvidos para atendê-las.

### ***1.3.3 Dados***

No presente estudo, dados são a representação da informação obtida por estudos científicos (por exemplo, por observações de campo ou experimentos). Dados são acrescidos de metadados, os quais descrevem as propriedades dos dados e, assim, permitem sua compreensão e interpretação (NATIONAL SCIENCE FOUNDATION CYBERINFRASTRUCTURE COUNCIL, 2007).

O conhecimento científico é derivado da informação representada pelos dados e que é incorporada ao acervo de fatos, princípios e teorias científicas (MICHENER; JONES, 2012). Strasser et al. (2011) propuseram um esquema para descrever o “ciclo de vida dos dados”, um contínuo de etapas de curadoria necessárias para transformá-los em novo conhecimento. Em uma perspectiva ampla, curadoria de dados é entendida como o conjunto de atividades e processos realizados para gerenciar, manter, validar e armazenar dados de pesquisa de forma a preservá-los e permitir seu

reuso (TRELOAR et al., 2007). O esquema do ciclo de vida dos dados elaborado por Strasser et al. (2011), ilustrado na Figura 1.4., foi concebido a partir do ciclo de vida de dados proposto em trabalhos anteriores, de diretrizes de melhores práticas de curadoria de dados ecológicos e ambientais (GIOIA, 2009; WHITLOCK, 2011).



**Figura 1.4.** O ciclo de vida dos dados, desde o planejamento até a análise modificado de Strasser et al. (2011). As setas cinza pontilhadas representam o fluxo mais comum em pesquisas ecológicas, o qual desconsidera etapas importantes para a longevidade e reuso dos dados. A seta cinza contínua representa as etapas que ocorrem em um cenário de pesquisa no qual não ocorre o reuso de dados de outros levantamentos.

As etapas do ciclo de vida dos dados não são necessariamente exclusivas e os passos não precisam ser sequenciais (MICHENER; JONES, 2012). Normalmente, em pesquisas ecológicas promovidas por um ou um pequeno grupo de pesquisadores, há crescente preocupação com o planejamento (ZUUR, 2010) com a coleta, validação e análise, porém as etapas de descrição, preservação, descoberta e integração, importantes para garantir a longevidade e reuso dos dados, são frequentemente desconsideradas. Há também projetos de pesquisa que promovem o levantamento de novos dados e são delineados com a preocupação de preservar a informação. Esses, normalmente, passam pelas etapas de planejar, coletar, validar, descrever, preservar e, depois, passam direto para a etapa de analisar. As etapas de descoberta e integração são realizadas em pesquisas nas quais há o reuso de dados. Pesquisadores envolvidos em um esforço de estudo de síntese provavelmente se preocuparão com a etapa de validar, depois daquela de descobrir e antes daquela de integrar. Há ferramentas computacionais de ecoinformática associadas com cada etapa do ciclo de vida dos dados (MICHENER; JONES, 2012).

Além dos dados e dos metadados associados, há elementos adicionais necessários para compor a representação formal da informação e do conhecimento. Trata-se de recursos da Ciência da Informação, relacionados à definição formal de termos (BOWKER, 2000; FUGAZZA et al., 2010). Um deles é a construção de ontologias, que são modelos de representação de conceitos e das relações entre eles (JONES et al., 2006; MADIN et al., 2008): os componentes de representação formal capturam o significado dos termos, reduzem a ambiguidade e facilitam a busca pela informação. Com as perspectivas de um cenário de crescente disponibilização de dados e metadados, esses recursos de informação tornam-se cada vez mais importantes.

Dados de biodiversidade comumente são considerados como aqueles associados à informação taxonômica (espécie, gênero, família) para cada espécime ou grupo de espécimes, bem como dados sobre a coleta e localidade (coletor ou observador, data, metodologia, características do sítio, coordenadas geográficas, nome das localidades e das unidades políticas) (COLWELL, 1996). Ainda que registros sobre ocorrências de espécies estejam sendo crescentemente disponibilizados, é difícil

encontrar dados representativos de distribuição de espécies que não de grupos bastante conhecidos, em locais bem estudados. Frequentemente, somos incapazes de responder a questões simples como listar quais espécies já foram registradas em determinada localidade (GURALNICK et al., 2007). As soluções computacionais para responder a essas perguntas existem, porém há problemas difíceis de tratar, em termos de qualidade dos dados, especialmente quanto à nomenclatura, uma vez que a evolução da ciência da Taxonomia leva a mudanças hierárquicas e alterações de nomes de taxa, de forma dinâmica (HUANG; QIAO, 2011; SANTOS; BRANCO, 2011). Além disso, há barreiras culturais que fazem com que ainda muito poucos pesquisadores e instituições compartilhem informações de ocorrências de espécies (CURRY; HUMPHRIES, 2007).

A comunidade científica mundial em informática para a biodiversidade tem se debruçado nos desafios de se catalogarem espécies e há aplicações importantes para o sucesso desse objetivo. Por exemplo, a estratégia mundial para a conservação de plantas é uma das ações para atingir as metas de Aichi (<http://www.cbd.int/sp/targets/>) propostas em consonância com a CDB, e uma das metas é, em 2020, ter estabelecido o catálogo mundial *on-line* de plantas.

No entanto, dados de biodiversidade são mais amplos do que ocorrências de espécies em determinado local e em determinada data. Muitas vezes, levantamentos de biodiversidade estão relacionados com a ocorrência de entidades taxonômicas. Contudo, especialmente em ecossistemas megadiversos, essas entidades taxonômicas frequentemente não são conhecidas no momento de sua observação (de BY et al., 2008), comumente, pode levar anos até que o espécime coletado seja associado a um nome de espécie – no caso de bactérias, isso pode demorar décadas. É preciso representar a informação sobre biodiversidade sem necessariamente associá-la a uma entidade taxonômica.

Listas de espécies também são insuficientes para fornecer informação sobre redes de interação entre espécies. Um exemplo é o esforço monumental feito para identificar as espécies de herbívoros e de plantas, em uma comunidade de alta diversidade, em Papua-Nova Guiné, recentemente reportado por Novotny et al. (2010): em mais de 15 anos, aproximadamente 7.000 relações alimentares foram traçados

entre mais de 200 espécies de plantas e 1490 espécies de insetos, oriundas de 11 guildas alimentares distintas. As comparações evidenciaram padrões distintos de especialização e diversificação nas diferentes guildas, restringindo a possibilidade de usar uma única guilda como *surrogate* para a comunidade inteira. Esses números impressionantes podem representar apenas 15% de toda a riqueza de herbívoros e de interações existentes em uma floresta tropical (LEWINSOHN, 2010). Listas de espécies também são insuficientes para analisar a contribuição de uma espécie individual para a conservação da comunidade biológica como um todo, temática que pode ser estudada por diferentes abordagens e que requer dados de interações entre espécies.

No presente estudo, dados ecológicos também são considerados dados sobre biodiversidade, mesmo que a entidade biológica não seja necessariamente representada taxonomicamente: por exemplo, dados de respiração do solo estão relacionados a um amplo conjunto de organismos e os métodos mais comumente utilizados para a estimativa desse componente do balanço de carbono em ecossistemas não dissociam a contribuição de diferentes entidades taxonômicas.

Fica claro, então, que a complexidade de gerenciar dados ecológicos é desafiadora, devido à grande heterogeneidade em sua sintaxe, semântica e *schema*<sup>6</sup> (JONES et al., 2006). Dados ecológicos são obtidos de variadas maneiras: observações de campo, as quais podem ser anotadas manualmente em papel, gravadas, imageadas, obtidas por sensores automáticos locais e gravadas em *data-loggers*; dados obtidos por sensores remotos; dados obtidos por simulação em modelos. As informações são representadas por medidas, pontos, coordenadas, frequências, categorizações. O avanço do conhecimento leva a mudanças nas representações dos conceitos científicos, ao mesmo tempo em que o avanço tecnológico permite medir novos aspectos dos organismos e do ambiente. Assim, não é possível prever quais variáveis serão medidas em novos estudos, o que torna conceitos clássicos de modelagem de bases de dados insuficientes para gerenciar essa informação (de BY et al., 2008). As bases de dados mais amplamente utilizadas são relacionais, e a maioria sustenta apenas um modelo de dados (JONES et al.,

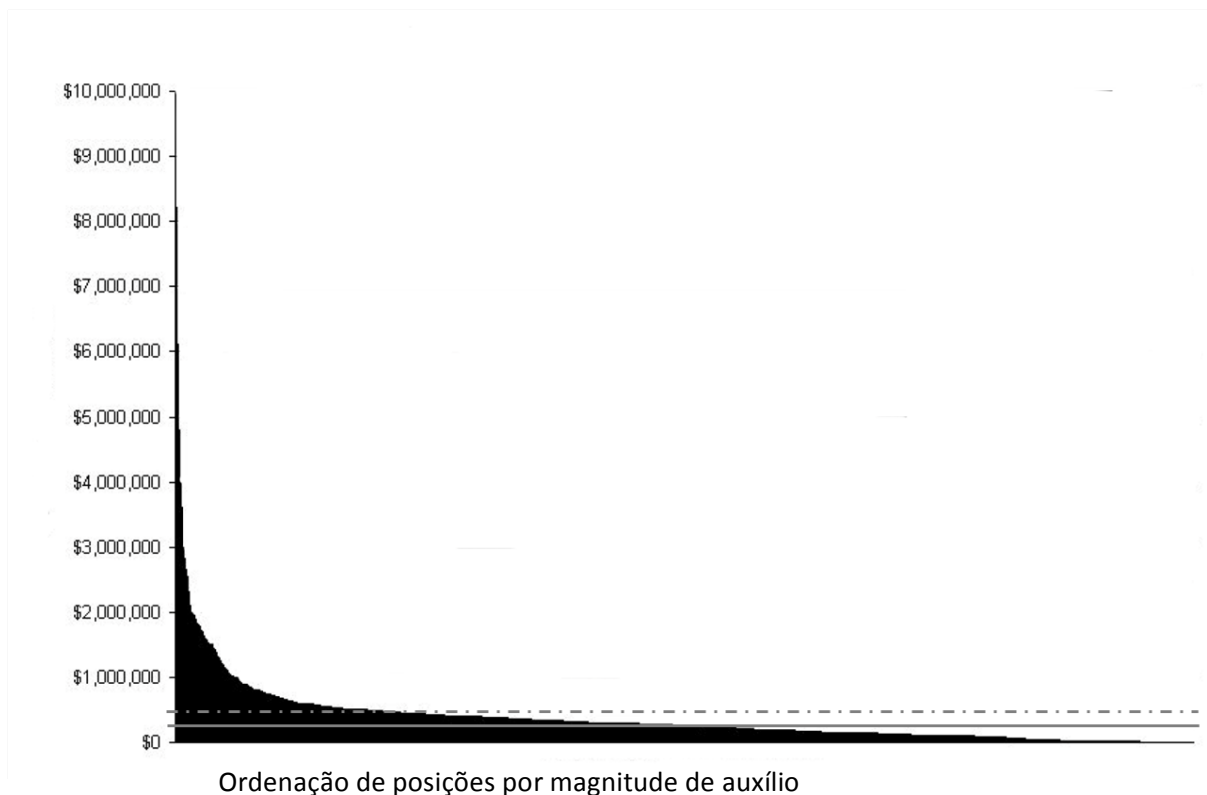
---

<sup>6</sup> *Schema* é a estrutura de uma tabela ou de um modelo de relacionamento de tabelas. O termo em inglês é frequentemente utilizado.

2006). Isso torna sua manutenção laboriosa e custosa, pois os modelos de dados precisarão ser alterados quando for preciso acomodar uma nova variável não prevista anteriormente, configurando um retrabalho minucioso, para evitar perda de dados.

Alguns projetos de grande porte em florestas tropicais, como o TEAM (“Tropical Ecology Assessment and Monitoring”) ou RAINFOR (“An international network to monitor the structure, composition and dynamics of Amazonian forests”, MALHI et al., 2002)) estabeleceram planilhas de campo pré-estruturadas e desenvolveram sistemas de informação com modelos relacionais, para gerenciar essa informação (PEACOCK et al., 2007). Essas soluções são úteis porque atendem demandas específicas de perguntas propostas por projetos e podem compor quadros mais completos de informação, se documentados em padrões conhecidos. Essas bases de dados relacionais são eficazes para gerenciar grandes volumes de dados, coletados em inúmeras localidades para representar parâmetros importantes para a compreensão de ecossistemas, como é o caso da biomassa. Mas, apesar do sucesso de casos como esses, nos quais algumas variáveis e muitos levantamentos são as principais características, a integração da informação sobre biodiversidade requer a representação da “diversidade de dados de biodiversidade”, que podem variar muito de um levantamento para outro (BOWKER, 2000). Para decifrar a complexidade da trajetória da vida no planeta é necessário conduzir uma grande variedade de pesquisas “no varejo” – de pequena escala e com delineamentos pouco padronizados, que também devem ser acomodados e preservados em infraestruturas de informação.

Ainda que muitos esforços estejam sendo empenhados para se preservarem dados de biodiversidade, a maior parte dos dados já coletados e que continuam sendo gerados por projetos de levantamentos de campo provavelmente será guardada em gavetas e computadores pessoais e não será disponibilizada. Esses dados estão representados na cauda do gráfico da Figura 1.5. e foram descritos por Heidorn (2008).



*Figura 1.5.* Distribuição de 12.025 auxílios da NSF por valores, em 2007 – valor total de US\$ 2.9 bilhões. Poucos auxílios concentram a maior parte dos investimentos, representados à esquerda no eixo X; enquanto auxílios menores estão à direita do eixo X – a linha cinza pontilhada marca o valor de corte de US\$ 500.000,00 por auxílio, enquanto a linha cinza contínua marca o valor de corte de US\$ 250.000,00. A cauda longa são os projetos “no varejo”: menor aporte de recursos por projeto, menor infraestrutura de informação frequentemente geram dados dependentes da observação e anotação humana, mais trabalhosos de se obter do que projetos que têm mais recursos para obter dados por instrumentos automáticos. Esses dados gerados por projetos da cauda longa são mais difíceis de preservar, encontrar e reusar. (Gráfico modificado de Heidorn, 2008).

Enquanto para a maioria dos dados gerados por projetos de grande porte (“Big Data”, edição especial da “Nature”, 2008) existe uma infraestrutura de informação planejada, a infraestrutura necessária para armazenar e permitir o reuso de dados de projetos menores é ainda muito tímida. Esses dados da “cauda longa” podem conter tipos de informação similares que, quando passíveis de cruzamento, têm o potencial de permitir novas descobertas. Algumas delas podem não ser previsíveis (HEIDORN, 2008).

A maior parte dos estudos em Ecologia gerados por grupos de pesquisa fornece informações restritas no espaço e no tempo, dificultando a compreensão das mudanças que ocorrem nos sistemas ecológicos (SHUGART, 1998). A necessidade de mais dados para servir de base de estudos para compreender como os diferentes componentes da biodiversidade e dos ecossistemas estão conectados motivou o estabelecimento de uma rede de sítios de pesquisa de longa duração (HOBBIE, 2003) - iniciada nos Estados Unidos, atualmente há sítios de pesquisa em todo o mundo – que se mostrou valiosa, já que a compreensão das mudanças como componentes fundamentais de ecossistemas e o acompanhamento a longo prazo fornecem informações sobre como os ecossistemas funcionam (SHUGART, 1998). No Brasil, há a rede PELD (BARBOSA et al., 2004), iniciada em 2000, com 12 sítios; em 2010, a rede foi expandida para 26 sítios e incluiu uma chamada para gestão da informação, evidenciando a preocupação da agência financiadora (CNPq) com a preservação dos dados a longo prazo.

#### ***1.3.4 Tecnologia***

O componente Tecnologia é composto pelas ferramentas computacionais que viabilizam o acesso, aquisição, armazenamento e distribuição da informação, o que inclui elementos de desenvolvimento de *software* (mecanismos de busca, análise e visualização da informação) e capacidade de armazenamento, processamento e conectividade. Ferramentas de análise e modelagem são fundamentais para sintetizar dados e gerar informação para o avanço do conhecimento e embasamento de decisões.

Modelos são simplificações de realidades mais complexas, concebidos a partir do conhecimento teórico e aplicados a um conjunto de dados reais ou imaginários, e essas simplificações são representações apropriadas para determinadas escalas temporais e espaciais. Modelos ecológicos são expressões matemáticas desenvolvidas para serem análogas ao ecossistema de interesse. O principal intuito de seu uso é integrar informação e produzir predições das respostas dos ecossistemas a mudanças, por isso estão desempenhando papel crescente no desenvolvimento de teorias



ecológicas, em diversas escalas. Um importante papel de modelos é o de explorar fenômenos que ocorrem em escalas temporais e espaciais, nas quais observações extensivas e diretas são proibitivas, senão impossíveis. Ecólogos têm obtido sucesso considerável para prever respostas de ecossistemas em uma amplitude de escalas. O desafio é determinar até que ponto essa predicabilidade pode ser generalizada ou aplicada a outras condições (SHUGART, 1998).

Modelagem de Nicho, também conhecida como Modelos de Distribuição de Espécies, é uma ferramenta que vem se consagrando como útil para compreender a distribuição de espécies e como elas podem ser afetadas pelo aquecimento global (COLOMBO; JOLY, 2010), dentre outras aplicações (PEREIRA et al., 2010). Esse tipo de análise é baseada em associações entre ocorrências conhecidas de espécies e variáveis ambientais, para determinar condições sob as quais essas espécies têm maior probabilidade de manter populações viáveis. Essa correlação é conhecida como a estimativa do “envelope ambiental” – obtido por variáveis climáticas, hidrológicas, pedológicas – para o qual determinada espécie tem maior probabilidade de ocorrer (SOBERÓN; NAKAMURA, 2009).

Entretanto, muitas outras análises são realizadas para compreender a biodiversidade, e a transparência dos modelos científicos têm muito a contribuir com a interface ciência-política. Ferramentas com o pacote R e o Kepler (ALTINTAS et al., 2004) documentam fluxos de trabalho (*workflows*, mencionados na seção **1.3.1, Pessoas**), o que permite não apenas compreender como determinada pesquisa chegou a determinado resultado, mas também reproduzir as análises e certificar que os resultados estão corretos, ou aplicá-las para outros conjuntos de dados com características similares.

Pesquisas ecológicas e levantamentos de biodiversidade são tradicionalmente realizados em escala local, permitindo a compreensão de fenômenos pontuais e desconexos. A disponibilização de dados permite que novos tipos de questões sejam respondidas – ampliando escalas temporais e espaciais de análise e permitindo compreender padrões e processos relevantes para a tomada de decisão estratégica. E, mesmo sabendo que muitas dessas variáveis coletadas na escala local são fortes indicadoras de como mecanismos ecológicos operam, a capacidade de fazer previsões

em escalas maiores é limitada com dados locais. Novas propostas de redes de monitoramento, como o NEON (“National Ecological Observatory Network”, (KELLER et al., 2008)), estão mudando o cenário para viabilizar análises preditivas em Ecologia, pelo acesso a bases de dados mais representativas do tempo e do espaço.

Como mencionado na seção **1.3.2. Arcabouço Institucional**, o exemplo do portal do “DataONE” para a integração da informação de diferentes disciplinas por meio de metadados configura-se como uma via interessante para permitir a interdisciplinaridade. Metadados descrevem a estrutura dos dados e a semântica da informação contida nos mesmos (HARRIS et al., 2006). Ferramentas gerenciadoras de catálogos de metadados modelam a estrutura dos arquivos XML, que descrevem os dados ao invés da estrutura do modelo de dados, tornando-os independentes dos modelos de dados para gerenciar os metadados e, por consequência, os dados associados (BERKLEY, 2001). Identificadores dos metadados fornecem descrição da identidade e da versão do conjunto de dados, requisitos que permitem informar sobre a procedência da informação (JONES et al., 2006).

#### ***1.3.4 Normas e Padrões***

Normas e padrões permitem a descoberta, o intercâmbio, a integração e a usabilidade da informação, assim sistemas de referência e padrões de metadados e de interoperabilidade são fundamentais nesse item. Trata-se de convenções que descrevem o contexto da informação contida em um registro ou conjunto de registro (LYNCH, 2003; BORGMAN, 2006). No caso da informação sobre biodiversidade, os padrões de metadados mais amplamente adotados atualmente são:

##### (a) “Dublin Core”

O “Dublin Core” é um padrão genérico para descrever recursos de informação: pode ser utilizado tanto para descrever documentos como livros, artigos de jornal e outras publicações textuais, como imagens, filmes e arquivos de sons (<http://dublincore.org/documents/dces>)

(b) “Darwin Core”

O “Darwin Core” é um conjunto de padrões e inclui um glossário para descrever a ocorrência de espécies, documentada por observações, espécimes em coleções ou amostras (WIECZOREK et al., 2012).

(c) EML

O EML (“Ecological Metadata Language”) é uma especificação de metadados para descrever conjuntos de informações (comumente representadas em tabelas). É modular e por isso adapta-se às necessidades dos usuários (FEGRAUS et al., 2005). Esse padrão será descrito em mais detalhes no Capítulo 2.

(d) FGDC

O FGDC (“Federal Geographic Data Committee”) foi concebido para a infraestrutura nacional de dados espaciais dos Estados Unidos. Contempla um perfil de dados biológicos (“FGDC Biological Data Working Group” and “USGS Biological Resources Division”, 1999). Está sendo adaptado para o ISO 19115.

(e) ISO 19115

O ISO 19115 é um padrão da Organização Internacional para a Padronização (“International Organization for Standardization”), para descrever a informação geográfica. É um padrão complexo, com mais de 400 elementos, sendo que 20 deles são obrigatórios.

(f) Perfil de Metadados Geoespaciais do Brasil

O Perfil de Metadados Geoespaciais do Brasil é uma das normas definidas pela INDE (mencionada no item **1.3.2**) e é baseado no padrão ISO 19115:2003. O Perfil MGB contém o conjunto básico e necessário de elementos que retratam as características dos produtos geoespaciais, para permitir sua identificação, avaliação e utilização. Esse conjunto básico é proposto como o núcleo comum a todos os tipos de produtos geoespaciais. Como mencionado em **1.3.2**, grande parte dos dados de biodiversidade são espacializáveis, portanto o perfil MGB é importante também para representar espacialmente os dados de biodiversidade.

## 1.4. DISCUSSÃO

Há o entendimento de que a ação humana está ultrapassando limites do sistema terrestre de forma irreversível (ROCKSTORM, 2009). Um desses limites é o da perda de diversidade biológica (DÍAS et al., 2006). Há também esforços para mobilizar o conhecimento científico para apoiar decisões. CASH et al. (2003) propuseram que sistemas de conhecimento para apoiar processos de tomada de decisão sobre o sistema terrestre devem ser transparentes, legítimos e relevantes. Expor os dados e fluxos de tratamento e análise por meio de tecnologias computacionais tem um papel importante para contemplar esses aspectos e atuar como objeto-ponte entre ciência e processos de tomada de decisão. Para isso, uma infraestrutura da informação sobre biodiversidade e ecossistemas, contemplando os componentes Pessoas, Arcabouço Institucional, Dados, Tecnologias e Normas e Padrões, foi analisada.

A análise do componente 'Pessoas' permitiu identificar necessidades no planejamento, aquisição, tratamento, gerenciamento, análise e uso da informação sobre biodiversidade. Além do planejamento da pesquisa pelo cientista, pessoas que atuam em órgãos de fomento estão crescentemente atribuindo valor aos dados como um produto do empreendimento científico e, além das agências de fomento, pessoas em revistas científicas estão criando mecanismos de incentivo para que os produtores de dados se preocupem com sua documentação também na aquisição. A cultura da documentação de metadados também está sendo incentivada e é necessário que pessoas capacitadas atuem na validação da qualidade dos dados e metadados depositados em repositórios. Pessoas precisam ser capacitadas para que atuem na interface entre a Computação e a Biologia, para que cooperem para gerar elementos como ontologias e vocabulários controlados, ferramentas de visualização e comunicação com usuário. E, para que uma infraestrutura de informação sobre biodiversidade e ecossistemas possa ser considerada um objeto-ponte entre a ciência e processos de tomada de decisão em conservação e gestão ambiental, ela deve acomodar as demandas de outro grupo de usuários: os tomadores de decisão. Essas pessoas podem ser gestores de parques, por exemplo, interessados no monitoramento

da biodiversidade: considerando que as mudanças são uma característica dos ecossistemas, é tarefa complexa compreender os ciclos dos ecossistemas, os efeitos das escolhas humanas na biodiversidade e como as mudanças globais afetam o bem-estar humano (PRESSEY et al., 2007); ao mesmo tempo, os gestores de parques precisam tomar decisões urgentes e com base em pouca informação (COOK et al., 2010).

Holling (1973) propôs que o ciclo de experimentação, aprendizado e adaptação seria uma maneira de manejar ecossistemas, nos quais as mudanças são parte do sistema. Nesse sistema, conhecido como manejo adaptativo, a decisão de, por exemplo, não extrair determinado recurso em determinado local, levaria um ciclo de aprendizado sobre os efeitos dessa prática no recurso de interesse, fornecendo subsídio para a decisão de como utilizar aquele recurso. A infraestrutura de informação deve permitir que o usuário defina parâmetros, tome decisões com base nesses parâmetros, monitore o que aconteceu, aprenda com a experiência e redefina parâmetros.

Mais recentemente, vem crescendo o reconhecimento de que a governança de muitos recursos como a pesca, produtos florestais e bacias hidrográficas, por exemplo, é muito complexa para ser tratada por uma só agência (em geral, governamental). O co-manejo é o compartilhamento de poderes e responsabilidades entre governos locais e usuários de recursos (BERKES, 2009). A prática tem demonstrado que, em sistemas socioecológicos nos quais os sistemas de trabalho estão bem desenvolvidos, o manejo adaptativo tende a incorporar elementos do co-manejo e vice-versa. As pessoas aprendem fazendo, através dos arranjos e regras locais, em um sistema que envolve também alguma experimentação (BERKES, 2009). Através de ciclos de aprendizado, redes de conhecimento podem incorporar novos elementos para lidar com problemas que podem aumentar em escala (CASH et al., 2006). Uma infraestrutura de informação que acomode informações provenientes de sistemas de conhecimento que não apenas o científico é a forma mais equitativa de atuar como objeto-ponte – ou seja, mediando interesses e conhecimento entre os diferentes atores envolvidos com a gestão e conservação ambiental. Assim, apesar de o presente trabalho não tratar diretamente de outros sistemas de conhecimento que não o científico, os elementos de infraestrutura

discutidos aqui se aplicam também a arcabouços mais complexos, como a do co-manejo adaptativo. E é interessante lembrar que há iniciativas promissoras de incorporação de conhecimento de cidadãos-cientistas em sistemas de informação sobre biodiversidade (COOPER et al., 2007; CUTTS et al., 2011; HOCHACHKA et al., 2012).

Entretanto, do ponto de vista do componente 'Arcabouço Institucional', apesar de algumas iniciativas estarem em andamento para estabelecer redes de parceria e cooperação em nível nacional, isso vem ocorrendo apenas recentemente e é o aspecto menos contemplado da infraestrutura da informação sobre biodiversidade no país, até o momento. É importante que o fortalecimento institucional dos múltiplos atores ocorra de forma equilibrada, para viabilizar resultados de longo prazo. Esse componente é importante não apenas para estabelecer regras e normas de disponibilização e reuso de dados pelos cientistas, mas também para intermediar os interesses dos atores envolvidos com a tomada de decisão. Essa interface é necessária para que a infraestrutura da informação em biodiversidade atue de fato como objeto-ponte, permitindo que as demandas dos usuários sejam reconhecidas e que arranjos e soluções sejam desenvolvidos para atendê-las.

A análise do componente 'Dados' evidenciou a complexidade dos dados de biodiversidade e ecossistemas, notadamente presentes na cauda longa (HEIDORN, 2008). Esses dados vão além de informações que possam ser associadas a uma entidade taxonômica. As práticas dos pesquisadores nas diferentes etapas do ciclo de vida dos dados (STRASSER et al., 2011) tendem a desconsiderar aspectos de preservação e documentação. Além disso, a Ecologia como ciência integradora, que possa ser analisada em escalas temporais e espaciais mais amplas e caminhe para auxiliar mais em modelos preditivos, ainda é uma visão de futuro e não uma prática cotidiana em grupos de pesquisa.

O potencial de análises em larga escala baseadas em modelagem computacional, com uso de dados depositados em repositórios públicos, foi demonstrado recentemente em um estudo sobre descoberta de fármacos (LOUNKINE et al., 2012). Fármacos podem atuar em diversas proteínas e centenas delas têm implicado efeitos colaterais. Identificar os mecanismos que atuam em efeitos adversos

de fármacos é oneroso por métodos empíricos convencionais, então métodos que possam identificar efeitos colaterais no início do processo de descoberta de um fármaco são ferramentas altamente desejadas por toxicologistas (KOLAJA, 2012). O modelo de LOUNKINE et al. (2012), baseado em dados de 656 fármacos comerciais depositados em repositórios públicos, previu 246 novas associações entre fármacos e efeitos colaterais, evidenciando mecanismos de efeitos colaterais pouco considerados previamente. Esse método abre nova maneira de avaliar a segurança de novos medicamentos e pode ser usado desde o início do processo de detecção de novos fármacos. Da mesma forma, a possibilidade de uso de dados ecológicos e de biodiversidade disponibilizados em repositórios públicos, associado ao uso de técnicas computacionais avançadas, pode promover descobertas com aplicações importantes.

A análise do componente 'Tecnologia' evidenciou que colabora com o requisito de transparência de objetos-ponte a transição da prática científica, em Ecologia e Biodiversidade, de uma cultura na qual os métodos são minimamente reportados em artigos para uma cultura de modelos abertos de análise e *workflow* científico.

A apresentação do componente 'Normas e Padrões' indicou que há padrões amplamente adotados para descrever dados de biodiversidade. O importante é que algum dos padrões seja adotado para documentar os dados, o que promoverá sua descoberta, intercâmbio, integração e usabilidade.

## 1.5. CONSIDERAÇÕES FINAIS

As mudanças ambientais globais demandam conhecimento científico, para permitir a compreensão da dinâmica de sistemas naturais. É necessário entender tanto a natureza das mudanças como a resposta de sistemas naturais às mudanças, ao mesmo tempo em que as pressões das atividades humanas alteram o ambiente nas escalas local, regional e global. Diferentes disciplinas utilizam conceitos e linguagens diferentes para descrever e explicar os processos que levam à deterioração ambiental (OSTROM, 2009). A alocação de esforços para implementar infraestruturas de informação, considerando-se os aspectos institucionais, as características dos dados, a tecnologia, normas e padrões, é uma via para organizar resultados em sistemas de conhecimento (CASH et al., 2003) e mobilizar o conhecimento científico para apoiar sistemas de tomada de decisão.

Apesar da compreensão de que ecossistemas estão sempre em mudança, as ideias de equilíbrio da natureza, natureza selvagem e floresta intocada são conceitos que ainda influenciam as práticas conservacionistas (SCOONES, 1999; SHUGART, 1998), bem como o conceito de biodiversidade como riqueza de espécies ainda influencia a conservação biológica.

A ciência da Biodiversidade passou de um foco na sistemática e taxonomia, nos anos 1980, para uma visão mais dinâmica do papel da biodiversidade no funcionamento dos ecossistemas e, atualmente, está inserida no contexto de valoração de serviços ambientais para o bem-estar humano (LARIGAUDERIE; MOONEY, 2010).

Analogamente, a ciência da Informática para a Biodiversidade (“Biodiversity Informatics”) teve uma rápida ascensão nos anos 2000 (BISBY, 2000), essencialmente com enfoque na digitalização de dados de espécimes. A Informática para a Biodiversidade é um campo do conhecimento relativamente recente, que pode ser definido como a criação, curadoria, análise e interpretação da informação sobre biodiversidade (SOBERON; PETERSON, 2004). Esse campo nasceu no gerenciamento de dados da comunidade que estuda a sistemática e a taxonomia e, apesar de essa informação vir sendo acumulada há séculos, só passou a ser satisfatoriamente gerenciada nas últimas duas décadas.



No contexto atual, é necessário avançar nos modelos de representação do conhecimento sobre a biodiversidade, para contemplar dados sobre ecossistemas – não necessariamente vinculados a uma entidade taxonômica; dados sobre a diversidade de interações entre espécies e sua relação com o funcionamento de ecossistemas; dados de espécimes cuja identificação levará anos ou décadas para ser obtida. O uso de tecnologias computacionais para representar o conhecimento sobre a biodiversidade nessa perspectiva ampla deve ser considerado de forma a acomodar dados historicamente menos enfatizados em sistemas de informação sobre a biodiversidade.

Os cinco elementos de infraestrutura de informação analisados devem ser considerados em todas as escalas: desde dentro de um grupo de pesquisa, passando por equipes multidisciplinares, até agências de fomento ou instituições de ensino e pesquisa. Também em termos de infraestrutura de informação, a diversidade de instituições proposta por OSTROM (2009), em um sentido mais amplo, é necessária para lidar com a complexidade de sistemas socioecológicos.

Em relação ao Arcabouço Institucional, há algumas iniciativas no sentido de estabelecer governança para integrar e viabilizar a manutenção de diferentes sistemas de informação sobre biodiversidade existentes. Contudo, esse é o aspecto menos contemplado da infraestrutura da informação sobre biodiversidade até o momento no país. O reconhecimento das demandas de diferentes atores quanto aos recursos de informação sobre biodiversidade está fortemente ligado à esse componente.

Na última década, houve avanços em alguns aspectos da infraestrutura de informação necessária para integrar o conhecimento científico sobre biodiversidade e ecossistemas, a fim de subsidiar decisões práticas sobre o gerenciamento de recursos: além do avanço no desenvolvimento de *softwares*, evoluiu também o estabelecimento de normas e padrões para documentar a informação e permitir a interoperabilidade de sistema. No entanto, para que essas normas, padrões e ferramentas sejam efetivamente adotados de forma a viabilizar a integração da informação, é necessário empreender mais esforços no sentido de promover os arranjos institucionais, legislativos e administrativos que atuam como alicerces da infraestrutura da informação e conferem sustentabilidade dos sistemas, a médio e longo prazo. Além disso, é

preciso capacitar recursos humanos no uso desses recursos tecnológicos e promover a aculturação dos atores envolvidos, no sentido de tornar o compartilhamento de dados uma prática amplamente realizada.

## **CAPÍTULO II, AVANÇOS NA INTEGRAÇÃO E GERENCIAMENTO DE DADOS ECOLÓGICOS<sup>7</sup>**

Apesar dos enormes avanços recentes de conhecimento da vida na Terra, nas últimas décadas, há ainda muitas lacunas de informação sobre a distribuição, abundância e interação das espécies entre si e com o ambiente. O desconhecimento é ainda maior em ecossistemas megadiversos, como florestas tropicais ou recifes de corais. Ao mesmo tempo, considerando os cenários atuais de perda de espécies e de habitat e as mudanças globais, a necessidade de informação para a compreensão de padrões e processos ecossistêmicos e de manutenção da biodiversidade cresceu vertiginosamente.

No Brasil, há numerosas instituições e cientistas que produzem conhecimento sobre nossos ecossistemas e muito já se sabe sobre nossa diversidade biológica. Novas descobertas e orientações para a tomada de decisão baseadas em conhecimento científico podem ser feitas pelo uso da informação que já existe, mas que em grande parte está dispersa, mal documentada e inacessível aos interessados (LEWINSOHN; PRADO, 2002; REICHMAN et al., 2011). Essa integração pode se dar tanto por meio da consolidação de dados de estudos semelhantes, de forma a expandir escalas temporais e espaciais de análise, como pela associação de conhecimentos gerados por estudos com diferentes abordagens.

A Ecologia como ciência permite uma grande amplitude de abordagens, desde o nível de genes ao estudo da paisagem e mesmo de continentes ou oceanos inteiros. Por isso, não é uma ciência que pode ser facilmente delimitada por metodologias ou procedimentos específicos (BERRY, 1989). Historicamente, de acordo com o foco de cada pesquisa, a evolução de estudos ecológicos se deu por meio do desenvolvimento de subdisciplinas, tais como Autoecologia, Ecologia de Populações, Ecologia de

---

<sup>7</sup> Artigo publicado em “Natureza & Conservação”, em julho de 2011. DRUCKER, D.P. 2011. Avanços na Integração e Gerenciamento de Dados Ecológicos. *Natureza & Conservação* 9(1): 115-120. doi: 10.4322/natcon.2011.016.

Comunidades, Ecologia de Ecossistemas ou, mais recentemente, a Ecologia de Paisagens.

Ainda que fragmentado, o conhecimento gerado pelas subdisciplinas vem contribuindo para o avanço da compreensão de padrões e processos naturais. Contudo, à medida que a pesquisa é focada em determinado ramo da Ecologia, as práticas científicas passam a divergir ao longo do tempo, e o estabelecimento de conexões e a produção de sínteses do conhecimento tornam-se mais complexos. Estudos de síntese do conhecimento ecológico atual são necessários para estabelecer pontes entre a Ecologia e disciplinas que tratam da interação entre sociedade e ambiente. Esse conhecimento vem sendo cada vez mais utilizado por disciplinas como Política Ambiental, Biologia da Conservação, Manejo de Bacias Hidrográficas e Estudos das Mudanças Climáticas Globais.

Para viabilizar estudos de síntese do conhecimento ecológico é necessária a colaboração entre cientistas que atuam em diferentes linhas de pesquisa, o que depende do acesso a dados ecológicos e de outras disciplinas correlatas. Um conjunto de dados normalmente é coletado para responder a determinada pergunta de um pesquisador ou grupo de pesquisa. No entanto, não há como prever quais dados serão importantes para estudos futuros, e novas análises poderão ser realizadas utilizando o mesmo conjunto de dados. Conjuntos de dados representativos de diferentes grupos taxonômicos e situações ecológicas são necessários, por exemplo, para entender os padrões de distribuição de abundâncias de espécies observados na natureza (PRADO, 2009), entretanto, ao contrário de pesquisadores de outras áreas de conhecimento como a Física, pesquisadores em Ecologia ou Taxonomia não têm ainda a cultura de disponibilizar os dados primários que foram analisados e publicados em revistas científicas. Vários aspectos relacionados aos dados primários não estão contidos nos artigos, mas podem ser clarificados pela disponibilização dos mesmos, acompanhados por anotações sobre procedimentos de coleta.

Dados ecológicos podem ser obtidos por observações diretas no campo, uma prática que provém da origem da ciência ecológica e que volta a ser valorizada atualmente, em paralelo a estudos baseados em experimentos (SAGARIN; PAUCHARD, 2010). Um dos motivos é a quantidade crescente de publicações que

evidenciaram efeitos de mudanças climáticas por estudos de monitoramento em sistemas naturais vide meta-análises, em Parmesan e Yohe, (2003), Root et al., (2003). Além disso, o acúmulo de dados observacionais permite a expansão das escalas espaciais e temporais de análise e possibilita a detecção de padrões insuspeitos. Essas diferentes abordagens em Ecologia, aliadas às tradições de pesquisa distintas tanto em suas subdisciplinas como em áreas afins, levam à produção de dados altamente heterogêneos, que podem ser representados por contagens de indivíduos, medidas de variáveis ambientais (por exemplo, temperatura ou precipitação) ou representações de processos como competição ou herbivoria, ou de seus efeitos (por exemplo, grau de abertura de dossel ou quantidade de dano em folhas). As terminologias utilizadas também variam de acordo com a linha de pesquisa, bem como a forma de estruturar os dados digitalmente (JONES et al., 2006).

Essas características tornam uma tabela de dados ecológicos avulsa muito difícil de interpretar, especialmente por alguém que não participou de sua coleta e digitalização. Ecólogos comumente armazenam dados em planilhas digitais e frequentemente as utilizam simultaneamente como um bloco de anotações, ao invés de tratá-las como tabelas que serão usadas em um banco de dados. Por exemplo, é comum pesquisadores colorirem uma ou outra célula para destacar determinado valor ou característica, ou mesclar duas ou mais células de uma linha ou coluna em uma só, dentre outros procedimentos que os auxiliam a explorar seu conjunto de dados. Infelizmente, raramente são agregadas explicações que levem outra pessoa a compreender por que determinada célula está colorida de amarelo ou de vermelho, como no exemplo da Figura 2.1.

Parcela	Espécie	Qde	Altitude
LO2 3500	<i>Rapatea paludosa</i>	133	55
	<i>Monotagma spicatum</i>	60	
	<i>Pepinia sprucei</i>	473	
	<i>Ischnosiphon arouma</i>	3	
LO4 0000	<i>Rapatea paludosa</i>	0	45
	<i>Monotagma spicatum</i>	19	
	<i>Pepinia sprucei</i>	120	
	<i>Ischnosiphon arouma</i>		

*Figura 2.1.* Exemplo de uma planilha de levantamento ecológico, sem anotações que permitam atribuir significado aos dados. Não é possível interpretar o que significam as células coloridas ou as células vazias; note-se que, além delas, há uma com valor 0. As siglas, abreviações e unidades não estão explicadas. Também não há identificação de localidade, data ou de responsável pelos dados. Esses dados provêm de um estudo real, realizado pela autora. Para mais explicações, veja Figuras 2.2 e 2.3.

Além disso, a maioria das análises realizadas em pacotes estatísticos requer que a tabela de dados tenha o mesmo tipo de dado em cada coluna, o que também é necessário em bancos de dados, para permitir a associação entre diferentes tabelas. Células mescladas dificultam a importação em bancos de dados ou pacotes estatísticos - por exemplo, em uma coluna com dados em números reais, não pode haver uma célula com texto. Além disso, células vazias não são autoexplicativas: não sabemos se aquele item foi medido e foi nulo, se não foi coletado ou se significa que o valor estava abaixo da capacidade de detecção do instrumento utilizado no levantamento de dados. Assim, ecólogos frequentemente copiam seus dados em uma segunda tabela “limpa”, que atende os requisitos do programa estatístico para analisar os dados, como na Figura 2.2. Esse processo é redundante, trabalhoso e vulnerável a multiplicação de erros.

Data	Parcela	Espécie	Contagem	Altitude (m)
26/04/04	LO2 3500	<i>Rapatea paludosa</i>	133	55
26/04/04	LO2 3500	<i>Monotagma spicatum</i>	60	55
26/04/04	LO2 3500	<i>Pepinia sprucei</i>	473	55
26/04/04	LO2 3500	<i>Ischnosiphon arouma</i>	3	55
28/04/04	LO4 0000	<i>Rapatea paludosa</i>	0	45
28/04/04	LO4 0000	<i>Monotagma spicatum</i>	19	45
28/04/04	LO4 0000	<i>Pepinia sprucei</i>	120	45
28/04/04	LO4 0000	<i>Ischnosiphon arouma</i>	*	45

Figura 2.2. A planilha da Figura 2.1, preparada para importação em banco de dados ou pacote estatístico. Note que as células vazias anteriores agora têm valores, distinguindo-se da que anteriormente tinha 0. É importante ter um valor ou marcador que assinala dados omissos (que não foram obtidos), distinguindo-os daqueles que foram medidos com valor zero.

A combinação da heterogeneidade de dados ecológicos, devido às múltiplas abordagens de pesquisa possíveis, com a cultura dos ecólogos de armazenar dados em planilhas soltas, torna a integração de dados ecológicos um desafio. Um procedimento que vem sendo adotado com sucesso em iniciativas nacionais e internacionais para tornar esse caminho menos árduo é a documentação do conjunto de dados (MICHENER et al., 1997). Essa documentação, ou descrição de conjuntos de dados, é conhecida por *metadados*, os quais fornecem informação sobre que entidades, processos ou medidas são representados nos dados; além disso, documentam como os dados foram coletados, com quais objetivos, quando, onde e por quem, informações fundamentais para que possam ser utilizados em novas análises. Por exemplo, um pesquisador pode registrar dados de serapilheira (camada de folhas e material orgânico depositados diretamente acima do solo em florestas) para medir a produtividade da floresta, enquanto os mesmos dados podem ser de interesse para alguém que estuda a ciclagem de um determinado nutriente, ou sua correlação com a variação de abundância de uma planta ou bactéria. A *anotação* dos procedimentos de coleta de dados pode tornar a mesma série de dados úteis para a investigação dessas diferentes questões (MADIN et al., 2008). Esse procedimento de documentação amplia a utilidade e colabora com a longevidade do conjunto de dados, pois os próprios pesquisadores que levantaram os dados tendem, com o passar do tempo, a esquecer

detalhes dos procedimentos de coleta relevantes para sua interpretação, ou mesmo as abreviações usadas ao anotá-los.

Um exemplo de especificação para a documentação de dados é o EML (“Ecological Metadata Language” (FEGRAUS et al., 2005)), implementado em módulos estruturados em linguagem XML (“eXtensible Markup Language”), uma linguagem de marcação que facilita a recuperação da informação por mecanismos de busca. Cada módulo descreve um aspecto da documentação de um conjunto de dados, como o módulo de descrição de métodos de coleta ou o de abrangência geográfica, taxonômica e temporal do conjunto de dados. A Tabela 2.1 contém os módulos e a explicação do que descrevem.

**Tabela 2.1. Módulos da especificação de metadados EML e suas descrições (MICHENER et al., 2011).**

Módulo	Conteúdo
EML-resource	Informação geral sobre os metadados, como quem os criou, título e palavras-chave
EML-party	Informação detalhada sobre pessoas e organizações envolvidas
EML-coverage	Abrangência temporal, geográfica e espacial do conjunto de dados
EML-project	Contexto da pesquisa que originou o conjunto de dados, por exemplo, objetivos e financiamento
EML-methods	Procedimentos de campo e de laboratório, usados para criar o conjunto de dados e procedimentos de controle de qualidade dos dados
EML-attribute	Nome, definição, unidade e domínio das variáveis
EML-access	Níveis de acesso permitido para usuários ou grupos de usuários



Essa especificação vem sendo adotada por diversas redes de grupos de pesquisa, como o LTER norte-americano (“Long Term Ecological Research” (BRUNT, 2006; MICHENER et al., 2011)), o SAEON (“South African Environmental Observatory Network”) e o TEAM (“Tropical Ecosystem Assessment and Monitoring”), da “Conservation International”, e de instituições como o NCEAS (“National Center for Ecological Analysis and Synthesis” (JONES et al., 2006) e o TFRI (“Taiwan Forestry Research Institute”). No Brasil, adotaram o protocolo EML para documentar os dados coletados o Projeto Temático Gradiente Funcional, do Programa BIOTA-FAPESP (“Composição florística, estrutura e funcionamento da Floresta Ombrófila Densa dos Núcleos Picinguaba e Santa Virgínia, do Parque Estadual da Serra do Mar” (JOLY et al., 2008)), o PPBio Amazônia Ocidental (Programa de Pesquisa em Biodiversidade, MCT, COSTA; MAGNUSSON, 2010)) e o PELD (Programa de Pesquisas Ecológicas de Longa Duração, CNPq). Um dos módulos do EML é a descrição da tabela de dados, o que permite a interpretação de tabelas avulsas, como a da Figura 2.2. A Figura 2.3 ilustra alguns dos itens contidos no módulo de atributos da especificação EML.

O editor de metadados Morpho (HIGGINS et al., 2002) pode ser usado para criar arquivos de metadados EML. Trata-se de um programa que pode ser instalado gratuitamente em qualquer computador pessoal e que cria documentação de metadados por meio de etapas simples. O programa possui uma interface em português, dentre outros idiomas, e, além de criar metadados, também serve para buscar metadados armazenados em um computador local, ou disponíveis na rede mundial de computadores, por meio de palavras-chave.

Attribute(s) Info:	Attribute Description (label and definition)	Measurement Type and Domain
"Data"	Data - Dia, mês e ano que foi realizada a coleta	date Time Format DD-MM-YYYY Precision 0 Enumerated Domain nominal
"Parcela"	Parcela - Identificador da parcela de 2x100 metros onde os dados foram coletados; baseado nas trilhas padronizadas da Reserva Ducke.	Order Code Definition "LO2.3500" LO = Trilha Leste-Oeste, 3500 = parcela instalada na metragem 3500 m ao longo da trilha, Source Code Definition "LO4.0000" LO = Trilha Leste-Oeste, 0000 = parcela instalada no início da trilha
"Espécie"	Nome da Espécie - Nome da espécie herbácea estudada	Enumerated Domain nominal Order Code Definition "Rapatea paludosa" Rapateaceae, Rapatea paludosa Aubl. Source Code Definition "Monotagma spicatum" Marantaceae, Monotagma spicatum (Aubl.) J.F. Macbr. Source Order Definition "Pepinia sprucei" Bromeliaceae, Pepinia sprucei (Baker) G.S. Varad. & Chihmartin Source Code Definition "Ischnosiphon acoums" Marantaceae, Ischnosiphon acoums (Aubl.) Koern. Source
"Contagem"	Número de Indivíduos - Número de indivíduos	Unit dimensionless Type natural ratio
"Altitude"	Altitude - altitude	Unit meter Type natural ratio
Data Set Owner(s): Indivíduos: <b>Drucker</b>		

Figura 2.3. Metadados para a planilha das Figuras 2.1 e 2.2.; módulo de descrição da tabela de atributos. À esquerda, a descrição das variáveis representadas em cada coluna; à direita, explicação sobre as unidades e domínios das medidas representadas na tabela de dados.

Outra ferramenta que permite a elaboração e recuperação de metadados em EML e conjuntos de dados associados é o Metacat, um Catálogo de Metadados (BERKLEY et al., 2004) desenvolvido pelo KNB (“Knowledge Network for Biocomplexity”). O Metacat é um repositório flexível de metadados e dados para uso em servidores e fornece um formulário de registro de metadados, podendo ser personalizado para atender às necessidades de uma instituição ou projeto. Tanto as consultas como os itens mais comumente utilizados pelos usuários podem ser automatizados: por exemplo, localidades geográficas revisitadas repetidamente para levantamentos de dados podem ser gravadas para que não seja necessário reinserir a mesma informação. Os projetos brasileiros citados acima (Gradiente Funcional do BIOTA-FAPESP, PPBio e PELD) criaram interfaces de busca em português. Além disso, esse repositório de metadados permite diferentes níveis de acesso a grupos de usuários distintos. A equipe responsável pela coleta dos dados pode ter acesso irrestrito a todo o acervo de dados; técnicos governamentais responsáveis pelo gerenciamento de unidades de conservação podem acessar metadados e dados de localidades contidas em unidades de conservação ou suas áreas de amortecimento circundantes; o público em geral pode acessar apenas os metadados e resultados de pesquisas já publicadas.

Essas ferramentas computacionais viabilizam a documentação de metadados, o que permite conectar os dados existentes sem ser necessária a centralização em um único sistema. Isso facilita a recuperação de dados isolados em “ilhas de dados” de cada subdisciplina (JONES et al., 2006) ou projeto e instituição. Essa possibilidade permite não só o embasamento sólido em dados empíricos para a formulação de hipóteses, mas também o teste de hipóteses em escalas espaciais e temporais apropriadas. Em alguns modelos preditivos de efeitos de mudanças climáticas, por exemplo, há uma grande dificuldade em estimar determinados parâmetros que representam os ecossistemas florestais. Embora de fato existam regiões pouco estudadas, a omissão de informação sobre as áreas nas quais já foi feito pelo menos um levantamento de campo configura uma perda de oportunidade de se avançar o conhecimento e melhorar os modelos e suas previsões.

Ecólogos estudam sistemas complexos nos quais fatores históricos e redes de interações não lineares têm papéis importantes. O compartilhamento de dados ecológicos e de áreas correlatas tem o potencial de ampliar o conhecimento atual sobre a interação de espécies entre si e com o ambiente. O avanço da compreensão de sistemas naturais fornece embasamento para a tomada de decisões quanto a inúmeros desafios da sociedade, incluindo desastres naturais, saúde pública, saneamento básico, abastecimento de água e produção de alimentos, dentre outros.

A documentação de metadados acompanhados por conjuntos de dados representam um importante avanço quanto ao compartilhamento de dados ecológicos – Hackett et al. (2008) descrevem casos bem-sucedidos do NCEAS, um centro criado nos Estados Unidos para compilar e integrar dados ecológicos existentes e promover estudos de síntese em ecologia, junto com o desenvolvimento de novos modelos teóricos. Um exemplo de síntese conduzida no NCEAS é o estudo de Halpern e Floeter (2008) sobre a relação entre riqueza de espécies e diversidade funcional, em comunidades de peixes de recife, em 36 localidades no Oceano Atlântico, baseados em dados coletados pelos autores e dados compilados da literatura. Em outro estudo, Jackson et al., (2001) examinaram tendências históricas de sobrepesca ao longo do tempo.

Contudo, ainda há desafios tecnológicos a serem enfrentados para permitir a análise integrada de dados tão heterogêneos e gerar informação relevante para embasar decisões sobre o gerenciamento de recursos naturais. Um grande desafio é a recuperação de dados quantitativos por meio de consultas às bases de dados disponibilizadas. Pelo repositório Metacat, é possível recuperar os metadados e, através deles, acessar conjuntos de dados de interesse, mesmo que dados e metadados estejam distribuídos em diferentes servidores. Porém, essa solução não permite consultar mais de uma tabela ao mesmo tempo; por exemplo, realizar uma consulta “mostre-me todos os registros obtidos no município de Ubatuba de árvores com diâmetro, à altura do peito, maior do que 30 cm”. Um banco de dados relacional pode ser modelado para determinado projeto de pesquisa com variáveis definidas *a priori* e resolver esse problema de forma pontual; porém, é difícil antever quais dados serão coletados em estudos futuros e que precisarão ser acomodados nos sistemas de

informação *a posteriori* (de BY et al., 2008). Outro desafio é promover a integração de diferentes repositórios como o do KNB, com outros de características semelhantes como o “Mercury” (no Brasil, adotado com o nome de “Beija-Flor”, pelo programa LBA – Programa de Larga Escala da Biosfera-Atmosfera na Amazônia (<http://lba.inpa.gov.br/lba/>)). Nesse sentido de integração de catálogos de metadados, a rede DataONE é um recente esforço para fornecer um padrão comum para que essas iniciativas possam se comunicar e serem consultadas conjuntamente e, assim, potencializar as iniciativas de compartilhamento de dados existentes (REICHMAN et al., 2011). Isso tem a vantagem de evitar a replicação desnecessária de conjuntos de dados, por meio do estabelecimento de identificadores unívocos.

Como a cultura de colaboração e de desenvolvimento de pesquisas de síntese são recentes na prática científica, especialmente no caso da Ciência Ecológica, há dificuldades, resistências e hábitos refratários ao compartilhamento de dados ecológicos. A primeira é de caráter institucional, pois instituições de pesquisa geralmente não dispõem de recursos para manter bases de dados com registros. Em geral, há um desequilíbrio entre o número de pesquisadores responsáveis pela aquisição de dados em campo ou laboratório e o número de pessoas trabalhando com a infraestrutura em tecnologia de informação (LYNCH, 2008). Mais recentemente, na última década, agências financiadoras no Brasil e no exterior começaram a demandar que projetos submetidos contenham um plano de gerenciamento de dados a serem obtidos na pesquisa. A política de dados formalizada pelo PPBio<sup>8</sup> demonstra que o debate sobre o compartilhamento de dados no âmbito nacional já vem acontecendo e envolvendo pesquisadores de variadas instituições em diferentes estados e regiões do país. Com o avanço da Ciência da Informação, há a crescente necessidade de capacitação de recursos humanos em gerenciamento de dados, em cursos de graduação em Ciências Biológicas e áreas afins. Somado a isso, é preciso agregar mais pessoal capacitado em Ciência da Informação, para trabalhar em colaboração com pesquisadores em Ecologia e Ambiente. Note-se que, embora haja novos programas e cursos para a área de Bioinformática, de modo geral estes são voltados

---

<sup>8</sup> [http://ppbio.inpa.gov.br/Port/docsinternos/politica\\_dou.pdf](http://ppbio.inpa.gov.br/Port/docsinternos/politica_dou.pdf).

para a Genômica e Proteômica, sem abranger a informação voltada para Biodiversidade e Ciências Ambientais.

Além da falta de tempo e de estrutura dos pesquisadores, outro fator que limita o compartilhamento de dados é o receio dos pesquisadores de que seus dados sejam usados de forma inapropriada, por ignorância sobre o contexto dos estudos ou por má fé. Os responsáveis pelos levantamentos tendem a desconfiar que outros não respeitarão os direitos de uso dos dados, ao publicarem estudos sem consultar os responsáveis ou negociar coautorias.

A ideia de que os benefícios de compartilhar dados ecológicos superam amplamente os riscos, pelas oportunidades de novas parcerias e estudos colaborativos que fomentam, ainda não é clara para a maioria dos pesquisadores. Uma via para incentivar que mais pesquisadores disponibilizem seus dados é a publicação dos dados em si (COSTELLO, 2009). Um exemplo importante é a iniciativa da ESA (“Ecological Society of America”) de criar o repositório “Ecological Archives”, que são compilações e sínteses de bases de dados e metadados referentes a artigos publicados em seus periódicos (“Ecology”, “Ecological Applications” e outros), em EML. Além disso, tais periódicos passaram a publicar “Data Papers”, que documentam bases de dados com a citação, resumo e outras informações associadas, no formato de artigos normais. A consequência disso é que dados utilizados para análises em novos estudos serão citados em referências bibliográficas como qualquer outro artigo, o que pode ser contabilizado como produção para os cientistas responsáveis pelos levantamentos de dados. Os periódicos “Heredity”, “Evolution” e “American Naturalist” recentemente anunciaram em conjunto que passarão a exigir a disponibilização de dados de pesquisas como pré-requisito para publicá-las (WHITLOCK et al., 2010). A melhor razão, porém, para pesquisadores aceitarem e investirem em disponibilizar adequadamente seus dados é que já há evidências de que pesquisas cujos dados brutos estão acessíveis tendem a ser mais citadas do que pesquisas que não compartilham dados (PIWOWAR et al., 2007). Apesar de muitos estudos serem necessários, a informação já existente nos permite fazer certas inferências e previsões para o futuro. O desenvolvimento de ferramentas para o gerenciamento de dados e

metadados ecológicos é um passo essencial no avanço de estudos de síntese do conhecimento ecológico.





# **CAPÍTULO III. O SISTEMA DE INFORMAÇÃO DO PROJETO TEMÁTICO GRADIENTE FUNCIONAL (BIOTA/FAPESP) COMO ESTUDO DE CASO PARA A INTEGRAÇÃO DA INFORMAÇÃO SOBRE BIODIVERSIDADE E ECOSSISTEMAS<sup>9</sup>**

## **3.1. INTRODUÇÃO**

A ciência da biodiversidade é baseada em um grande número de disciplinas, para permitir o avanço de nossa compreensão sobre a trajetória da vida na Terra (BOWKER, 2000). Grandes volumes de dados primários, que variam de genômica a dados ambientais, são crescentemente produzidos e arquivados em inúmeras bases de dados desconexas, trazendo desafios para sua manutenção, curadoria, interoperabilidade, análise e interpretação (JONES et al., 2006). Em paralelo, cientistas da computação vêm desenvolvendo novas tecnologias continuamente, e o campo da informática para a biodiversidade pode se beneficiar de tais avanços para melhorar a capacidade de tratar essa enorme quantidade de dados de forma a produzir informação que faça sentido (BISBY, 2000).

A representação de espécimes depositadas em coleções biológicas e os desafios técnicos de informatizar acervos foram os primeiros passos na direção de tornar acessível a informação sobre biodiversidade (CURRY; HUMPHRIES, 2007). Sequentemente, soluções para representar registros de ocorrências de espécies associadas a coordenadas geográficas passaram a receber maior atenção, ao mesmo tempo em que ferramentas analíticas para modelagem de nicho também passaram a ser mais consideradas (CANHOS et al. 2004).

No contexto atual, o reconhecimento da comunidade científica em relação à importância de compreender a relação entre biodiversidade e funcionalidade, o que requer a abordagem da biodiversidade como além do número de espécies (ROMERO,

---

<sup>9</sup> Artigo em preparação por Drucker, D.P., Salim, J.A. e Joly, C.A.

2008), demanda ferramentas computacionais que permitam a representação de recursos de informação não necessariamente relacionados a entidades taxonômicas, ou seja, a evolução do conceito de biodiversidade configura novos desafios tecnológicos para sua representação em sistemas de informação.

Assim, o presente Capítulo tem o objetivo de contribuir para o desenvolvimento de soluções tecnológicas que permitam a representação da informação sobre biodiversidade e ecossistemas para além de ocorrências de espécies. Para isso, analiso o sistema de informação desenvolvido para o Projeto Temático do Programa BIOTA/FAPESP Gradiente Funcional ("Composição florística, estrutura e funcionamento da Floresta Ombrófila Densa dos Núcleos Picinguaba e Santa Virgínia do Parque Estadual da Serra do Mar", FAPESP 03/12595-7), como estudo de caso para a integração da informação sobre biodiversidade e ecossistemas e o analiso com base no arcabouço de investigação apresentado no Capítulo 1. Trata-se de um esforço para contribuir para o desafio de reforçar o valor dos dados existentes, de forma a torná-los disponíveis para pesquisas futuras, bem como para informar decisões sobre a conservação do Parque Estadual da Serra do Mar.

### **3.2. OBJETO DE ESTUDO**

O Projeto Temático Gradiente Funcional tem o objetivo principal de investigar a estrutura e o funcionamento da Floresta Ombrófila Densa Atlântica. Com abordagem multidisciplinar, o Projeto Temático vem levantando informações em fitossociologia, biologia da reprodução, ecofisiologia da germinação, ecofisiologia da fotossíntese, biomassa e dinâmica do carbono, entre diversas outras áreas de fundamental importância para a compreensão dos mecanismos reguladores da biodiversidade nessas florestas e para o estabelecimento de ações pertinentes para sua conservação, manejo e restauração.

O Projeto é um estudo de caso, representativo de estudos em biodiversidade e ecossistemas, nos quais as questões cruzam escalas geográficas e temporais, enquanto muito das investigações individuais de pesquisadores são realizadas em

escalas locais (BROWN et al., 2001). Há nele pesquisas sobre variação de biomassa (ALVES et al., 2010), estoques de carbono e nitrogênio (VIEIRA et al., 2011), tolerância à inundação em áreas de restinga (OLIVEIRA; JOLY, 2009), ecofisiologia de espécies arbóreas (ROSADO, 2012), polinização por beija-flores, para citar alguns. Além disso, foi realizado na Mata Atlântica, um *hot-spot* de biodiversidade (MYERS, 2000). O sistema de informação foi concebido para integrar dados e metadados dessa grande variedade de informação.

Os coordenadores do Projeto Temático já possuíam experiência anterior em programas de pesquisa que têm tradição em preservar e compartilhar dados ao longo prazo (BIOTA-FAPESP e LBA – “Experimento de Grande Escala da Biosfera-Atmosfera”) e conceberam o desenho amostral com vistas a essa mesma finalidade. Foram instaladas 14 parcelas quadradas de 1ha, subdivididas em 100 subparcelas de 10m × 10m. As parcelas foram delimitadas por equipe especializada em Topografia, utilizando-se instrumentos de alta precisão, como teodolito e nível digitais, altímetro e GPS. O limite externo das parcelas foi definido colocando-se estacas de 1,2m de tubo de PVC 3/4, a cada 10m. A cada 50m, foi colocada uma estaca de tubo de PVC de 5 polegadas, georreferenciada de forma a permitir a plotagem da parcela em imagens de satélite e ortofotos. A partir dessa delimitação externa, as parcelas foram subdivididas em subparcelas de 10m × 10m, tendo uma estaca em cada um dos vértices. Em cada ponto correspondente à posição das estacas, foi registrado nível altimétrico (JOLY, 2012), o que facilitou o posicionamento das observações de campo, e as diferentes equipes puderam utilizar um identificador comum às subparcelas. Como estas são quadradas, valores X e Y de um plano cartesiano puderam ser utilizados para determinar pontos de observação, quando foi o caso. A localização geográfica foi identificada como o elemento de integração dos diferentes levantamentos do Projeto, sendo que a maioria deles foi realizada dentro das parcelas instaladas .

A equipe de levantamento arbóreo coletou dados de todos os indivíduos de espécies de plantas com perímetro à altura do peito maior do que 15 cm, em todas as 14 parcelas instaladas. Isso resultou em 22.991 registros de indivíduos arbóreos no primeiro censo, realizado em 2005.

### 3.3. REQUISITOS

Frequentemente, a identificação do material coletado em inventários biológicos acontece meses ou anos após a coleta, dificultando a acomodação em bases de dados fundamentadas em ocorrência de espécies, como é o caso do GBIF. (“Global Biodiversity Information Facility”). É necessário, então, um sistema de informação para levantamentos de biodiversidade que contemple dados de espécimes ainda não determinadas, ou seja, no qual o código de coleta seja suficiente para referenciar a informação na base de dados. Os sistemas para dados taxonômicos não são estáticos, pois a ciência que estuda a evolução realiza descobertas que são refletidas em alterações nos sistemas de referência. Assim, um sistema de informação para levantamentos de biodiversidade não precisa incorporar um subsistema de dados taxonômicos, e sim trabalhar em conjunto com os sistemas de informação em coleções já existentes, para acomodar automaticamente as alterações nas informações taxonômicas.

Além disso, muitas vezes os dados de levantamentos de campo são observacionais, ou seja, não necessariamente relacionados a espécimes físicos coletados e depositados em coleções biológicas. Por exemplo, dados sobre populações de espécies são fundamentais para embasar pesquisas em biodiversidade e ambiente, bem como políticas públicas sobre uso e manejo de recursos. Dados sobre biodiversidade não necessariamente estão relacionados com entidades taxonômicas, como é o caso de muitos estudos sobre o fluxo de nutrientes. Dados sobre o meio físico em escala local, como características físicas e químicas do solo e abertura do dossel são necessários para compreender relações ecológicas, estratégias adaptativas e distribuição de espécies. Dados sobre estrutura da floresta e biomassa são imprescindíveis para estimar estoques e dinâmica do carbono. É necessário, portanto, um sistema que comporte informações ecológicas e ambientais que possam ser relacionadas com as bases de dados taxonômicos já existentes. Esses dados normalmente são levantados em conjuntos e representados por tabelas; assim, um levantamento de campo, em geral, tem como produto uma tabela com um conjunto de observações de, por exemplo, diâmetro de espécies arbóreas de inúmeros indivíduos.

Assim, um sistema que acomode esse tipo de informação deve ser desenhado com base na acomodação de tabelas e seus *schemas*, os quais definem o domínio dos dados em cada coluna.

Outro desafio para o gerenciamento de levantamentos biológicos de campo é que não é possível prever novos projetos nem os dados que serão coletados no futuro. Tradicionalmente, em bancos de dados relacionais, as colunas das tabelas são definidas previamente e de forma estática, e as tabelas que compõem o banco de dados são relacionadas, por meio de colunas, com variáveis análogas pelo modelo entidade-relacionamento. Isso significa que a acomodação de uma nova variável, não prevista no início de um projeto, demanda a definição de um novo relacionamento dentro do banco. Esse processo não é escalável e, em levantamentos biológicos, o surgimento de variáveis não previstas *a priori* é a regra e não a exceção. Para tratar essa questão, de By et al., (2008) propuseram um sistema de banco de dados de inventários genérico, que acomoda dados de todo e qualquer levantamento, inclusive aqueles não previstos.

Outro requisito importante é o posicionamento dos levantamentos no tempo e no espaço: o monitoramento de variáveis que representem as diferentes formas de medir a biodiversidade ao longo do tempo é crucial para permitir a compreensão da natureza dinâmica dos ecossistemas (SHUGART 1998). Do ponto de vista da mobilização do conhecimento científico para informar decisões sobre conservação, em práticas como o manejo adaptativo (BERKES 2008), que é baseado no trabalho em ciclos, a dimensão temporal é fundamental para compreender processos. Por outro lado, em termos geográficos, muito da Biologia da Conservação é baseada no espaço, e análises espaciais têm atraído crescentemente ecólogos e conservacionistas (LANDEIRO; MAGNUSSON, 2011).

Do ponto de vista do usuário, é necessário que o sistema permita consultas baseadas em nomes de variáveis – por exemplo, “mostre-me todos os levantamentos que mediram temperatura”. É desejável também que seja possível realizar consultas baseadas em operadores com as variáveis – por exemplo, “mostre-me todos os indivíduos arbóreos com valores de perímetro, à altura do peito, maiores do que 30 centímetros”. Ou, ainda, como a distribuição de espécies permanece um assunto

importante para a Biologia, é necessário poder consultar o sistema com base em nomes científicos – por exemplo, “posicione as ocorrências de *Euterpe edulis* nas parcelas estudadas”. É necessário acessar também os metadados, ou seja, possibilitar a consulta sobre os responsáveis pelos levantamentos, objetivos dos estudos, método de coleta, direitos de uso. Metadados permitem a usuários que não tenham participado do levantamento compreenderem como e por que os dados foram levantados e, até mesmo, replicarem o mesmo método, em situação semelhante.

### **3.4. SOLUÇÕES**

O desafio de desenvolver um sistema de informação para levantamentos biológicos que seja genérico e que possa acomodar dados de todo e qualquer levantamento, inclusive aqueles não previstos, foi abordado com o desenvolvimento de uma técnica de fragmentação vertical de tabelas simples de dados, a qual foi baseada no delineamento proposto por de By et al. (2008). Essa técnica associa toda e qualquer observação com o local de coleta georreferenciado, e as tabelas são fragmentadas e recompostas de acordo com as necessidades do usuário. Essas características também contemplam o requisito de incorporar informações espaciais ao sistema. O modelo da base de dados está esquematizado na Figura 3.1.

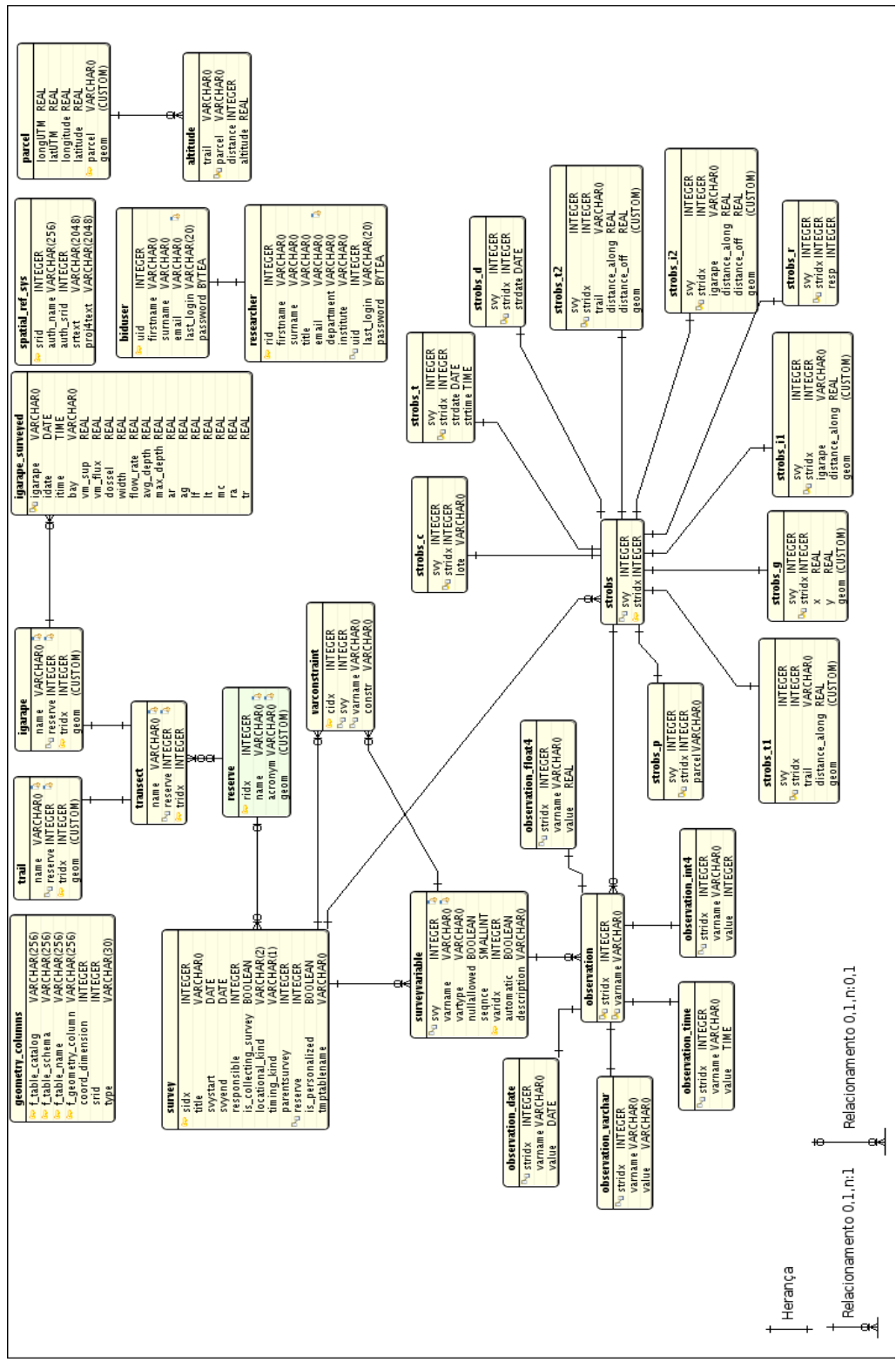


Figura 3.1. Modelo Entidade-Relacionamento da Base de Dados do Projeto Temático Biota Gradiente Funcional. No Anexo 1 são apresentados detalhes desse modelo.

O sistema prevê cinco tipos de dados computacionais: números inteiros (*integer4*), números reais (*float*), datas (*date*), horários (*time*), textos (*character varying*). O usuário informa ao sistema quais e de que tipos são as métricas utilizadas na pesquisa. Assim, uma tabela guarda informações sobre as variáveis definidas por um estudo – nome e tipo da variável, informação sobre a possibilidade de aceitar valores nulos e uma descrição. Os valores das variáveis são armazenados em outras cinco tabelas, de acordo com seu tipo – números inteiros, números reais, datas, horários e textos, respectivamente. A Figura 3.2. ilustra o *schema* da tabela de variáveis com tipo número inteiro.

Column	Type	Modifiers
stridx	integer	not null
varname	character varying	not null
value	integer	

**Figura 3.2. Schema da tabela de variáveis com tipo número inteiro. O atributo “stridx” é uma anotação da observação: pode ser uma anotação espacial, que define qual tipo de informação geográfica está associada ao levantamento, ou outra anotação sobre o levantamento, como nome do responsável pela coleta. O atributo “varname” é o nome da variável e “value” é o valor observado.**

A técnica utilizada para possibilitar o armazenamento de dados genéricos foi a fragmentação de dados, que consiste em dividir um banco de dados em unidades lógicas menores, chamadas fragmentos. Foi utilizada a fragmentação vertical para dividir a relação por atributos e armazenar as informações das colunas das variáveis, nas tabelas definidas por tipo de variável. Assim, quando o usuário define a variável “perímetro à altura do peito”, especificando que o tipo de dados será números reais, a respectiva coluna identificada pelo próprio usuário será fragmentada e inserida na tabela de armazenamento de variáveis reais.



O sistema desenhado por de By et al. (2008) foi baseado no desenho de trilhas instalado na Reserva Ducke, Manaus, AM (BACCARO et al., 2008). Foi realizado um redesenho do sistema com base nas trilhas instaladas nas 14 parcelas do Temático Biotá Funcional, baseado no trabalho de Salim (2009). Para isso, foi necessário incorporar as informações fornecidas pela equipe de topografia responsável pela alocação das parcelas na base de dados em Postgresql, com o complemento para armazenar dados geográficos Postgis. O sistema foi projetado para receber 4 tipos de informações geográficas: distância ao longo de uma trilha em metros; distância para uma trilha em metros; latitude e longitude de um ponto, em UTM zona 23 Sul e Datum WGS 84; um polígono que delimita uma parcela, definida pelo número da mesma. Assim, qualquer dado inserido no sistema deve necessariamente ter uma referência geográfica, garantindo a integração. O usuário necessita apenas de uma coordenada geográfica ou indicar em qual parcela foi realizado o levantamento. E pode indicar, ainda, em qual ponto das trilhas que delimitam as subparcelas trabalhou (distância ao longo de uma trilha) ou a que distância das mesmas trilhas realizou o levantamento de dados (distância da trilha – neste caso, precisa indicar os valores X e Y, com base nos vértices da parcela, para determinar pontos de observação).

Alguns processos automatizados de verificação da qualidade e consistência dos dados foram implantados. Assim, ao entrar com uma tabela de dados no sistema, é preciso informar o tipo de dado que ela contém (e.g. inteiro, real, alfa-numérico, data, data e hora), a unidade (e.g. m/s, cm, dd/mm/aa) e o intervalo (ou “range”), ou seja, os valores mínimo e máximo contidos no conjunto de dados. Assim, se o valor máximo de data informado for 01/10/2009, o sistema acusará um erro ao tentar entrar com uma data posterior, como 30/11/2009. Ao entrar com a tabela de dados, é preciso também fazer uma descrição dos dados de cada atributo.

O fluxo de informação no sistema de informação é o seguinte:

1. Criação de um novo levantamento: nessa etapa, são inseridos o título, o nome do pesquisador responsável e o tipo de localização (distância ao longo de uma trilha, distância de uma trilha, latitude e longitude de um ponto ou subparcela).

2. Definição de variáveis: as variáveis definidas pelo usuário devem ser cadastradas individualmente, indicando um nome, o tipo (inteiro, real, texto, data ou hora), a aceitação ou não de valores nulos e uma descrição.

3. Acesso à tabela gerada pelo sistema: o sistema cria uma tabela que contém as variáveis definidas pelo usuário, adicionada de uma ou duas colunas geradas automaticamente e que se referem à dimensão espacial. As colunas automáticas variam de acordo com o tipo de informação espacial gerada pelo usuário. Por exemplo, se o usuário optar pelo tipo de informação geográfica “distância da trilha”, serão geradas 3 colunas: identificador da trilha (texto), valor X e valor Y. A tabela gerada pelo sistema não contém dados, apenas as definições das variáveis.

4. A tabela criada em 3 é povoada com os dados das variáveis selecionadas pelo usuário e inserida no sistema pelo serviço de *upload* do arquivo. Nessa etapa, o usuário seleciona o levantamento cujos dados vai inserir pelo título e insere a tabela de dados. Ao fazê-lo, o sistema realiza a fragmentação vertical em estruturas lógicas menores, preenchendo as tabelas de variáveis de acordo com o tipo especificado na tabela de descrição de variáveis, ilustrada na Figura 3.3.

5 – Após a entrada de dados através de arquivos em formato CSV (*comma separated values*), as localizações geográficas podem ser calculadas de acordo com o tipo de localização escolhido pelo usuário. Por exemplo, se o usuário escolheu a distância ao longo de uma trilha para fornecer tais informações, o sistema utiliza o valor de distância fornecido na tabela de dados e o identificador da trilha para calcular a localização geográfica em WGS 84 das observações contidas no arquivo.

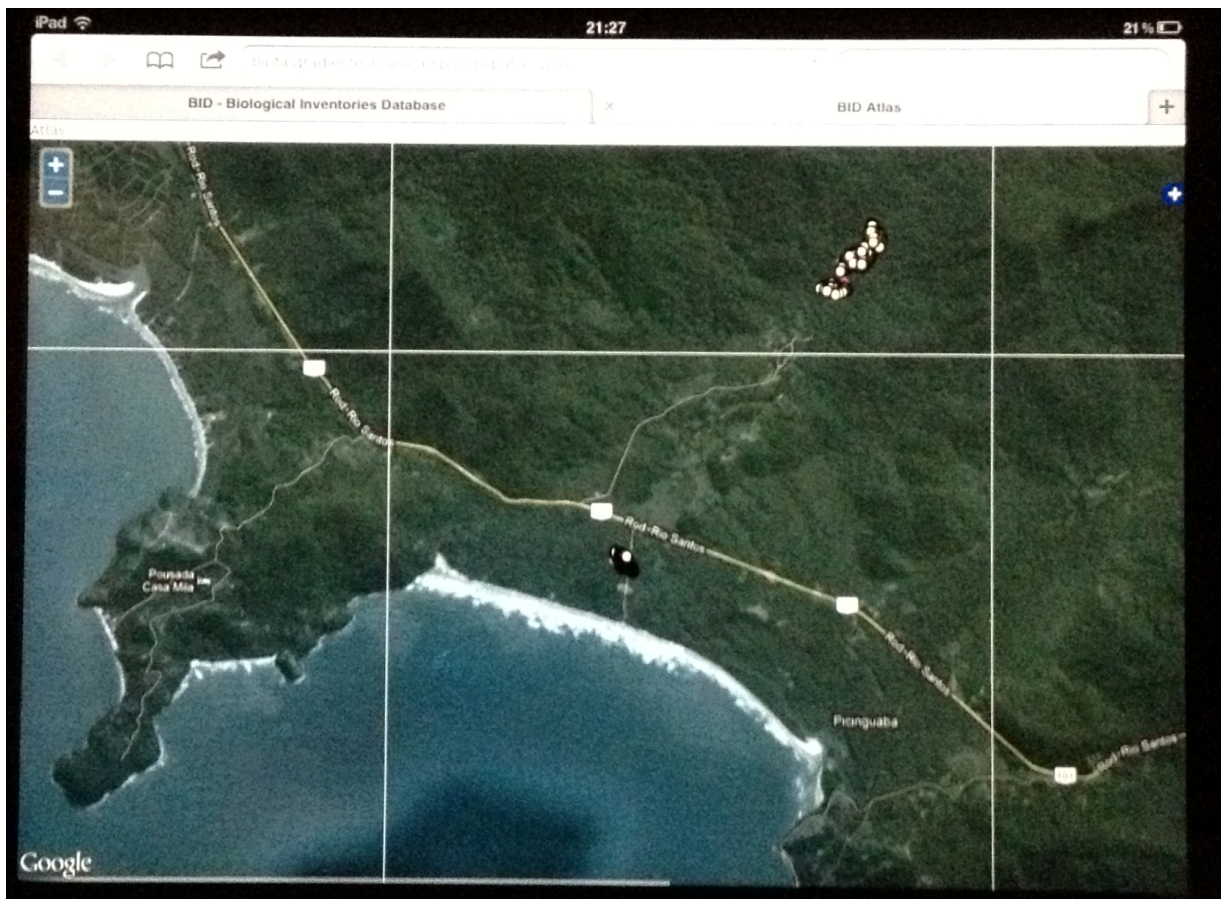
svy	varname	vartype	nullallowed	seqnce	varidx	automatic	description
3	trail	vchar	f	1	4	t	
3	distance_along	float4	f	2	5	t	
3	distance_off	float4	f	3	6	t	
3	CodParcel	int4	t	4	7	f	Código da parcela
3	Parcel	vchar	t	5	8	f	Identificador da Parcela (letra)
3	Face	int4	t	6	9	f	Face de exposição
3	County	vchar	t	7	10	f	Indefinido
3	Longitude	float4	t	8	11	f	Longitude
3	Latitude	float4	t	9	12	f	Latitude
3	SubParcel	int4	t	10	13	f	Código identificador da subparcela
3	Plate_old	vchar	t	11	14	f	Placa antiga
3	plate	vchar	t	12	15	f	Placa atual
3	X	float4	t	13	16	f	Distância fora da trilha
3	Y	float4	t	14	17	f	Distância ao longo da trilha
3	Dead	int4	t	15	18	f	Indivíduo morto
3	Date_Dead	date	t	16	19	f	Data da morte do indivíduo
3	GD	vchar	t	17	20	f	Indefinido
3	CodSpecies	int4	t	18	21	f	Código da espécie
3	Family	vchar	t	19	22	f	Taxonomia: Família
3	Genus	vchar	t	20	23	f	Genêro
3	Species	vchar	t	21	24	f	opposita
3	pap	float4	t	22	25	f	Perímetro na altura do peito
3	pap_a	float4	t	23	26	f	PAP - Ramificação 1
3	pap_b	float4	t	24	27	f	PAP - Ramificação 2
3	pap_c	float4	t	25	28	f	PAP - Ramificação 3
3	pap_e	float4	t	26	29	f	PAP - Ramificação 5
3	h_pap	float4	t	27	30	f	Indefinido
3	copa	float4	t	28	31	f	Indefinido
3	h_total	float4	t	29	32	f	Indefinido
3	h_fuste	float4	t	30	33	f	Indefinido
3	pap_d	float4	t	31	34	f	PAP - Ramificação 4
4	trail	vchar	f	1	35	t	
4	distance_along	float4	f	2	36	t	
4	distance_off	float4	f	3	37	t	
4	sub	int4	t	4	38	f	Subparcela
4	n.o. trep	vchar	t	5	39	f	Número
4	ff 1	vchar	t	6	40	f	Indefinido
4	ff 2	vchar	t	7	41	f	Indefinido
4	ff 3	vchar	t	8	42	f	Indefinido
4	ff 4	vchar	t	9	43	f	Indefinido
4	Met. sc.	vchar	t	10	44	f	Indefinido
4	Familia	vchar	t	11	45	f	Família
4	SP	vchar	t	12	46	f	Espécie
4	autor	vchar	t	13	47	f	Autor taxonomia

**Figura 3.3. Tabela de descrição de variáveis. A coluna “vartype” indica o tipo de variável , o qual definirá em qual tabela de variável os valores serão inseridos no processo de fragmentação vertical.**

Foi também desenvolvida uma interface *web* para a entrada e saída de dados: criação de novos levantamentos, inserção de arquivos de dados, realização de consultas e visualização dos dados espacializados em um atlas ligado a um servidor de mapas (Mapserver e Openlayers). Essa interface de visualização permite o acesso aos dados de todas as variáveis definidas pelo usuário que inseriu dados daquele levantamento. Isso significa que a seleção de um ponto de observação no mapa, por exemplo, gerará uma janela com a lista das variáveis e valores existentes para aquele ponto. Além da visualização da informação no atlas, é possível visualizar as tabelas

originais de cada levantamento e interagir com o sistema por meio de consultas avançadas. Nesse caso, o usuário define a consulta por meio de operadores simples, como “mostre-me todas os indivíduos de árvores da família Myrtaceae”: o usuário escolhe a variável “Família”, em uma lista de variáveis disponíveis, e informa a restrição “Myrtaceae”.

Para a implementação do sistema, foi utilizado o Sistema de Gerenciamento de Banco de Dados Objeto Relacional (SGBDOR) PostgreSQL, um SGBD de código aberto e a extensão espacial PostGIS. A fragmentação vertical e cálculos sobre as localizações espaciais foram implementados com a linguagem procedural PL/pgSQL embutida no banco de dados como gatilhos (triggers). Para a interface foram utilizadas as linguagens PHP, HTML4 e Javascript aplicando folhas de estilo CSS (*Cascade Style Sheet*). Tais tecnologias são livres e acessíveis em qualquer navegador web, inclusive dispositivos móveis. A Figura 3.4. ilustra o acesso ao Atlas do sistema pelo uso de um dispositivo móvel *tablet*.



**Figura 3.4. Visualização do Atlas do Sistema de Informação desenvolvido para o Projeto Biota Gradiente Funcional pelo uso de um dispositivo móvel tablet. É possível visualizar as observações de árvores nas parcelas de restinga e terras baixas, a BR-101 (Rio-Santos) e a sede do Núcleo Pinguaba do Parque Estadual da Serra do Mar.**

Além da solução encontrada para o gerenciamento de dados quantitativos, toda a informação foi documentada em EML (“Ecological Metadata Language” (FEGRAUS et al., 2005)), uma especificação de metadados amplamente adotada no mundo todo para descrever dados ecológicos de forma estruturada. Metadados são cruciais para fornecer informação sobre pessoas e instituições envolvidas com o levantamento, métodos adotados, coberturas espaciais, temporais e taxonômicas (MICHENER et al., 2006). Os metadados foram acomodados em um servidor Metacat instalado no Instituto de Biologia da Universidade de Campinas, em São Paulo (Unicamp). O Metacat é um

catálogo de metadados que permite a acomodação, consulta e recuperação de dados e metadados em XML. Essas escolhas foram feitas porque os dois componentes mostraram-se eficientes para gerenciar conjuntos de dados, em outras iniciativas nacionais e internacionais de gestão da informação de pesquisas ecológicas (MICHENER; JONES 2012). O Metacat é facilmente customizável, e o EML é versátil por exigir algumas informações mínimas e permitir que as demais sejam opcionais.

Para obter os metadados das pesquisas do Projeto Temático, foram feitas duas solicitações pela lista de discussão da equipe e algumas conversas pessoais, para esclarecer dúvidas. Para catalogar os metadados, foi utilizado o editor de metadados Morpho, aplicativo desenvolvido especialmente para gerar arquivos XML no padrão EML. O Morpho é de fácil manuseio, não requer conhecimento avançado em linguagem de programação e adere facilmente ao Metacat.

Os dados de lianas e arbóreas foram fornecidos pelos responsáveis pelos levantamentos. No caso dos dados de arbóreas, as tabelas foram fornecidas após o esforço de associar dados coletados pela equipe de estrutura da vegetação com dados coletados pela equipe de botânica. A estrutura da vegetação foi medida antes do que os levantamentos botânicos, e a relação entre os levantamentos se dá pelo número da placa colocada em cada indivíduo medido. Considerando que o número de indivíduos levantados variou entre 1230 e 2071 por parcela, a possibilidade de ocorrerem erros ao associar dados de botânica com os de estrutura da vegetação é grande. Os pesquisadores relataram que, algumas vezes, a equipe da botânica não encontrou algumas árvores medidas pela equipe de estrutura da vegetação. Assumiu-se que as árvores morreram. Em outros casos, a equipe de botânica encontrou indivíduos arbóreos que não foram plaqueados pela equipe de estrutura da vegetação, e inseriram novos registros com números de placas que não ocorreram na medida da estrutura da vegetação. Outra consideração sobre os dados de arbóreas é o fato de a planilha fornecida pelos responsáveis possuir algumas células em amarelo, “que geralmente indica que há algum dado faltando para o registro do indivíduo”, conforme relatado. Além disso, “células em rosa indicam que os indivíduos devem ser medidos com uma escada”. Esse tipo de prática no gerenciamento de dados por ecólogos foi discutido no Capítulo 2 desse trabalho.

### 3.5. TESTE DO SISTEMA

O sistema de informação é composto pelas ferramentas (discutidas no Capítulo 2) para gerenciamento de metadados, já existentes e adotadas internacionalmente, acrescidas de uma base de dados genérica e habilitada espacialmente, o que permite a acomodação de qualquer novo levantamento, mesmo com variáveis não previstas antes, desde que com a informação da localização geográfica. Essa base de dados genérica foi proposta por de By et al. (2008) para a Reserva Ducke, Manaus, AM. No presente trabalho, a concepção teórica dessa base foi adaptada ao caso do Projeto Biotá Gradiente Funcional e implementada pela primeira vez, configurando um trabalho original. É possível realizar consultas com o componente de visualização dos dados espacialmente, consultar diretamente as tabelas de dados e, ainda, consultar os levantamentos por meio de consultas aos metadados.

A base de dados genérica foi testada com o levantamento de espécies arbóreas, realizado em todas as 14 parcelas do Projeto Temático Biotá Gradiente Funcional (ALVES et al., 2010; JOLY et al., 2008; JOLY et al., 2012) e com os dados do levantamento de lianas realizado na parcela J (ALVES et al. 2012). Uma série de consultas aos dados inseridos na base foi testada. Um exemplo de formulário de consulta avançada pode ser visualizado na Figura 3.5., enquanto um exemplo de resultado da consulta está ilustrado na Figura 3.6.

Um exemplo de visualização espacial, ou seja, consultas baseadas nos dados de localização das observações em campo, está ilustrado na Figura 3.7. Nesse exemplo, foi desenhada no mapa a posição X e Y dos indivíduos arbóreos com PAP (Perímetro à Altura do Peito) > 15 cm na Parcela E. As linhas que dividem as parcelas de 1ha em 100 subparcelas de 10m x 10m servem como base para o georreferenciamento de qualquer observação realizada dentro de uma parcela ou fora dela, desde que fornecida a distância para uma das linhas de uma das parcelas e o ângulo em relação à linha. Na Figura 3.8., é possível visualizar todos os indivíduos arbóreos levantados nas Parcelas G e H.

A Figura 3.9. ilustra a página de cadastro de novos levantamentos. A Figura 3.10. ilustra a página inicial do Repositório de Dados do Biota Funcional, pela qual é possível fazer buscas por metadados e dados do Projeto Temático Biota Funcional por meio de palavras-chave. Quando um novo levantamento é inserido, pela página de cadastro ou pelo programa Morpho, os itens do padrão EML são preenchidos e existe a opção de entrar ou não com a tabela de dados. Os metadados, acompanhados ou não da tabela de dados, formam um pacote de dados. É possível, também, definir níveis de permissões de acesso - por exemplo, permitir que usuários externos visualizem os metadados, porém não tenham acesso aos dados. Mesmo com a ausência de acordos formais prévios sobre a disponibilização de dados e metadados dos levantamentos, a colaboração dos pesquisadores permitiu que metadados de 30 levantamentos conduzidos pelo Projeto Temático Gradiente Funcional fossem documentados com o programa Morpho, que possibilita a entrada de mais itens de metadados do que a catalogação de metadados pelo formulário *on-line* de cadastro.<sup>10</sup> A recuperação das informações de metadados foi testada com sucesso no sistema – por exemplo, a busca pela palavra-chave “carbono” retornou 5 pacotes de dados.

A interação com os pesquisadores que se propuseram a colaborar foi bastante variável: enquanto alguns se entusiasmaram com a proposta, fornecendo também as tabelas de dados para teste do sistema e respondendo prontamente às dúvidas sobre os dados e metadados fornecidos, outros enviaram os metadados e não responderam às dúvidas. Houve também aqueles que não se manifestaram, ou seja, não forneceram dados e metadados. A descrição de dados ecológicos é complexa e a documentação de metadados deve ser cuidadosa para assegurar sua qualidade e permitir o efetivo reuso em análises futuras. Um estudo mostrou que a habilidade de um pesquisador para compreender dados ecológicos foi determinante para seu reuso (Zimmerman 2008). Isso ilustra como a colaboração em Ecologia ainda está em estágios iniciais e a necessidade de acumular experiência em curadoria e reuso de dados para que os erros mais comuns sejam sistematizados e que se consolide uma base para desenvolver ferramentas automáticas para detecção de erros.

---

<sup>10</sup> A lista dos levantamentos catalogados encontra-se no Anexo 2.



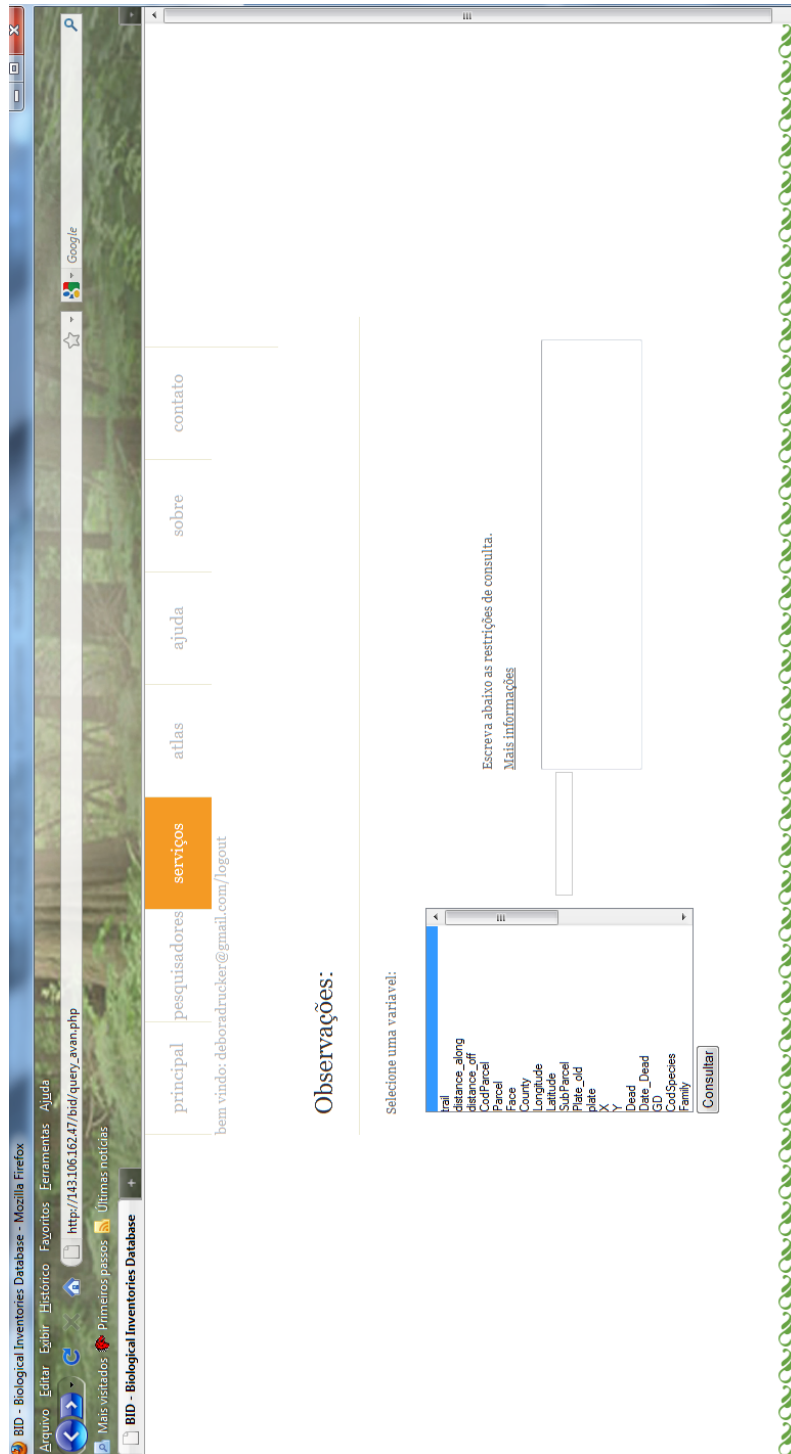


Figura 3.5. Exemplo de formulário de consulta.

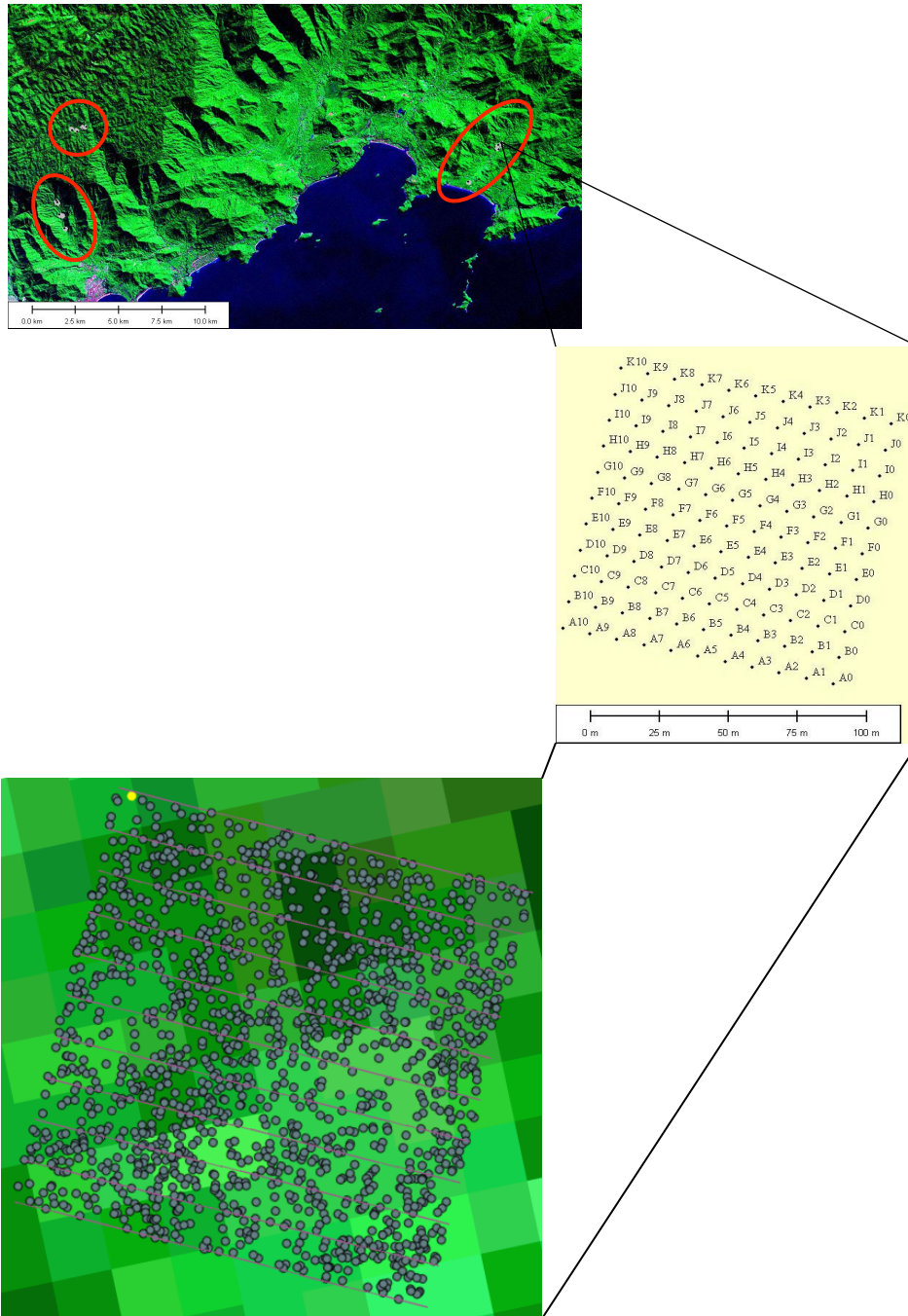
http://142.106.162.47/isd/include/get\_obs.php?id=1  
 Show 50 entries

Search:

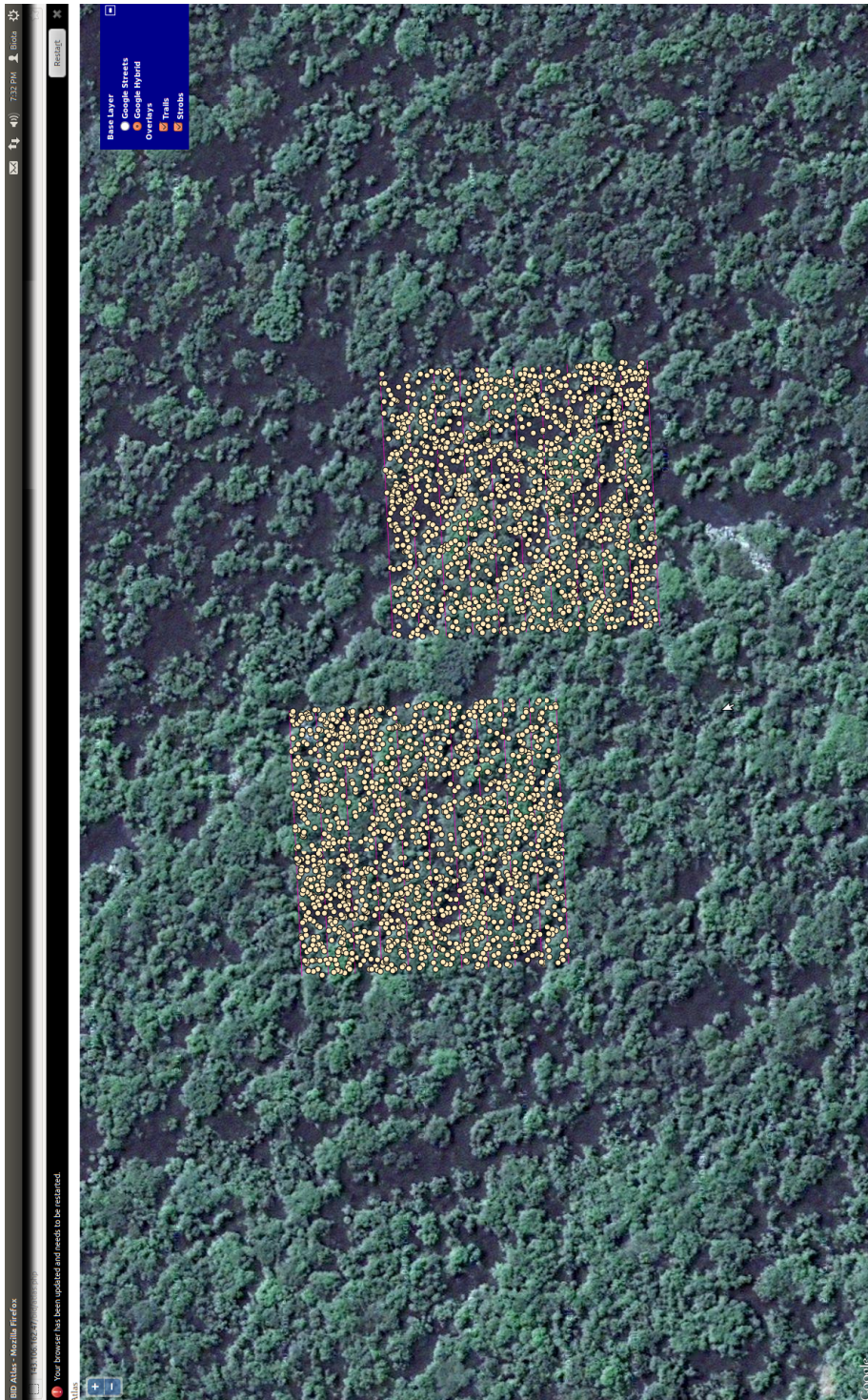
CodParcel	CodSpecies	copa	County	Date_Death	Dead	Face	Family	GD	Genus	h_fuste	h_pap	h_total	Latitude	Longitude	pap	pap_a	pap_b	pap_c
22543	1	420	Picinguaba	0	0	0	Nyctaginaceae		Guapira	6	1.3	10	0	0	44			
22544	1	371	Picinguaba	0	0	0	Euphorbiaceae		Pera	5	1.3	7	0	0	15			
22545	1	357	Picinguaba	0	0	0	Araceae		Euterpe	8	1.3	10	0	0	28			
22546	1	378	Picinguaba	0	0	0	Fabaceae		Sclerobolium	6	1.3	12	0	0	130			
22547	1	390	Picinguaba	0	0	0	Melastomataceae		Miconia	8	1.3	10	0	0	25.5			
22548	1	420	Picinguaba	0	0	0	Nyctaginaceae		Guapira	7	1.3	8	0	0	23.5			
22549	1	407	Picinguaba	0	0	0	Myrtaceae		Eugenia	4	1.3	7	0	0	28.1			
22550	1	365	Picinguaba	0	0	0	Clusiaceae		Kielmeyera	6	1.3	7	0	0	28			
22551	1	371	Picinguaba	0	0	0	Euphorbiaceae		Pera	5	1.3	6	0	0	20.2			
22552	1	371	Picinguaba	0	0	0	Euphorbiaceae		Pera	4	1.3	10	0	0	35.3			
22553	1	371	Picinguaba	0	0	0	Euphorbiaceae		Pera	7	1.3	8	0	0	30.4			
22554	1	371	Picinguaba	0	0	0	Euphorbiaceae		Pera	3	1.3	6	0	0	20.5			
22555	1	351	Picinguaba	0	0	0	Aquifoliaceae		Ilex	3	1.3	6	0	0	15.8			
22556	1	371	Picinguaba	0	0	0	Euphorbiaceae		Pera	6	1.3	7	0	0	23			
22557	1	468	Picinguaba	0	0	0	Myrtaceae		Gomidesia	5	1.3	9	0	0	78	45.9		
22558	1	347	Picinguaba	0	0	0	Anacardiaceae		Tapirira	5	1.3	10	0	0	66			
22559	1	416	Picinguaba	0	0	0	Myrtaceae		Myrcia	6	1.3	9	0	0	41			
22560	1	416	Picinguaba	0	0	0	Myrtaceae		Myrcia	4	1.3	7	0	0	19.3			
22561	1	357	Picinguaba	0	0	0	Araceae		Euterpe	7	1.3	8	0	0	23.8			
22562	1	396	Picinguaba	0	0	0	Myrsinaceae		Rapanea	5	1.3	7	0	0	16.7			
22563	1	389	Picinguaba	0	0	0	Melastomataceae		Miconia	3	1.3	5	0	0	19.2			
22564	1	371	Picinguaba	0	0	0	Euphorbiaceae		Pera	6	1.3	7	0	0	19			
22565	1	390	Picinguaba	0	0	0	Melastomataceae		Miconia	5	1.3	9	0	0	23.8			
22566	1	358	Picinguaba	0	0	0	Bigoniaceae		Jacaranda	7	1.3	10	0	0	77			
22567	1	371	Picinguaba	0	0	0	Euphorbiaceae		Pera	6	1.3	7	0	0	25			
22568	1	368	Picinguaba	0	0	0	Euphorbiaceae		Alchornea	6	1.3	7	0	0	21.1			
22569	1	371	Picinguaba	0	0	0	Euphorbiaceae		Pera	6	1.3	7	0	0	15.4			
22570	1	368	Picinguaba	0	0	0	Euphorbiaceae		Alchornea	7	1.3	14	0	0	101.3			
22571	1	354	Picinguaba	0	0	0	Araliaceae		Schefflera	6	1.3	7	0	0	16.1			
22572	1	407	Picinguaba	0	0	0	Myrtaceae		Eugenia	7	1.3	10	0	0	54.3			
22573	1	416	Picinguaba	0	0	0	Myrtaceae		Myrcia	7	1.3	10	0	0	39.6			
22574	1	390	Picinguaba	0	0	0	Melastomataceae		Miconia	5	1.3	7	0	0	18.6			
22575	1	358	Picinguaba	0	0	0	Bigoniaceae		Jacaranda	8	1.3	11	0	0	56.2			
22576	1	416	Picinguaba	0	0	0	Myrtaceae		Myrcia	4	1.3	6	0	0	33			
22577	1	371	Picinguaba	0	0	0	Euphorbiaceae		Pera	6	1.3	7	0	0	20.9			

Concluido

Figura 3.6. Exemplo de resultado de consulta



**Figura 3.7. Localização das 14 parcelas de 1 hectare do Projeto Temático Gradiente Funcional.** Destacadas pelas elipses, detalhe da divisão da Parcela E em subparcelas e exemplo de consulta da base de dados habilitada em Sistema de Informação Geográfica. A consulta foi: “retorne a posição X e Y dos indivíduos arbóreos com PAP (Perímetro à Altura do Peito) > 15 cm na Parcela E”.



**Figura 3.8.** Visualização dos indivíduos arbóreos levantados nas Parcelas G e H.



Composição, estrutura  
e funcionamento da  
Floresta Ombrófila  
Densa do Parque Estadual  
da Serra do Mar




Principal	Repositório	Registro	Logout
-----------	-------------	----------	--------

Use esse formulário para inserir uma nova descrição de dados .  
Se você tem dúvidas, comentários ou sugestões sobre esse formulário, por favor contate [gradientefuncional@gmail.com](mailto:gradientefuncional@gmail.com) .  
**\*Indica um campo de preenchimento obrigatório.**


Nome da pessoa responsável pelo preenchimento do formulário (O que é isso?)		Esconder
*Primeiro Nome	<input type="text"/>	
*Último Nome	<input type="text"/>	
Informação Básica (O que é isso?)		Esconder
*Título do Conjunto de Dados	<input type="text"/>	
Principal Responsável pelo Conjunto de Dados (O que é isso?)		Esconder
*Primeiro Nome	<input type="text"/>	
*Último Nome	<input type="text"/>	
Nome da Organização	<input type="text"/>	
E-Mail	<input type="text"/>	
Telefone	<input type="text"/>	
FAX	<input type="text"/>	
Endereço	<input type="text"/>	
Cidade	<input type="text"/>	
Estado	<input type="text"/>	
CEP	<input type="text"/>	
País	<input type="text"/>	
Colaboradores (O que é isso?)		Esconder
Primeiro Nome	<input type="text"/>	
Último Nome	<input type="text"/>	
Papel	<input type="text" value="Co-responsável"/>	
<a href="#">Adicionar Colaborador</a>		
Resumo (O que é isso?)		Esconder
*Resumo (max. 350 palavras)	<input type="text"/>	
Palavras-Chave (O que é isso?)		Esconder

Figura 3.9. Seção de formulário de cadastro de metadados de novos levantamentos.



**Biota**  
FAPESP

Composição, estrutura  
e funcionamento da  
Floresta Ombrófila  
Densa do Parque Estadual  
da Serra do Mar



**gradiente.funcional**

Principal
Repositório
Cadastro
Logout

### Repositório de Dados do Biota Gradiente Funcional

Bem vindo ao Repositório de Dados do Biota Gradiente Funcional. Esse repositório contém dados de levantamentos realizados no âmbito do Projeto Temático. Os dados armazenados nesse repositório estão conectados à rede KNE ([Knowledge Network for Biodiversity](#)), um repositório internacional de dados. Algumas das séries de dados contidas aqui foram geradas por pesquisadores em esforços individuais, enquanto outras são resultado de esforços conjuntos por grupos de pesquisadores. As descrições de cada conjunto de dados contêm mais informações sobre pessoas e instituições envolvidas. Para perguntas, comentários e sugestões, por favor contate: [biota@uniscamp.br](mailto:biota@uniscamp.br).

**Busca por Dados**

Buscar somente nos campos "Título", "Resumo", "Palavras-Chave", "Pessoas Envolvidas" (mais rápido)

Buscar todos os campos (mais lento)

Esta ferramenta permite a busca por conjuntos de dados de interesse. Ao inserir um texto no quadradinho e clicar no botão "Buscar", a busca será conduzida apenas nos campos "Título", "Resumo", "Palavras-Chave" e "Pessoas Envolvidas". Ao optar pela opção "Buscar todos os campos", a busca ocorrerá em todos os campos (isso fará com que a busca leve mais tempo).

O caractere "%" pode ser usado como um "coringa" (ou "wildcard") nas buscas (por exemplo, "biodiversidade%" localizará qualquer frase que contenha a palavra biodiversidade).

[Procurar dados cadastrados no Repositório do Biota Gradiente Funcional](#)

[Ver Mapa Interativo](#)


**Cadastre informação sobre um novo conjunto de dados do Biota Gradiente Funcional**

[Cadastre Informação Online](#) - Esta é uma maneira simples de fornecer informação (metadados) sobre um conjunto de dados. A página de cadastro serve para entrar com informação sobre um novo conjunto de dados associado com as pesquisas conduzidas no âmbito do Biota Funcional.


OU

Use o software [Morpho](#) para cadastrar e documentar conjuntos de dados, bem como fazer "upload" de tabelas de dados.


Morpho é uma ferramenta de gerenciamento de dados para ecólogos. Foi criado para fornecer fácil acesso e manipulação de metadados (e.g. documentação) e de dados (tanto localmente como parte de uma rede). Morpho permite que ecólogos criem seus metadados (ou seja, que descrevem seus dados em um formato padronizado), e cria um catálogo de metadados que pode ser usado para consultas, edição ou visualização. Além disso, Morpho viabiliza o acesso aos servidores da rede, de modo que possam também ser usados em consultas e visualização de dados acessíveis publicamente.




Esse repositório é uma iniciativa do Programa Biota-FAPESP e da Universidade Estadual de Campinas e é baseado no software desenvolvido pela rede [Knowledge Network for Biodiversity \(KNE\)](#), acomodando metadados que são compatíveis com a linguagem de metadados [Biological Metadata Language \(BML\)](#).



FAPESP



IB  
Instituto de Biologia



UNICAMP

DEPARTAMENTO DE BIOLOGIA VEGETAL  
Rua Monteiro Lobato 970 Cidade Universitária "Zelmaro Vaz" - São Geraldo - Campinas - SP - Caixa Postal 6109 Cep 13.083-970  
Fones: 0XX19 3521-6165 Fax: 0XX19 3521-6165  
Copyright © Universidade Estadual de Campinas. Todos os Direitos Reservados  
Unicamp  
Contato • Termos de Uso • Acessibilidade

**Figura 3.10.** Página inicial do Repositório de Dados do Projeto Temático Biota Gradiente Funcional

Na tentativa de contribuir para a padronização de termos, foi feito um exercício de mineração de texto dos termos utilizados nos metadados para verificar os termos mais utilizados e detectar ambiguidades.<sup>11</sup> Esse exercício teve o objetivo de criar um corpus e detectar os termos mais frequentes, o que é o primeiro passo para criar um vocabulário controlado. Os termos mais frequentes nas palavras-chave foram “floresta” e “espacial” (unigramas), “atlântica gradiente” e “funcional mata” (bigramas) – neste caso, termos que, agrupados, não têm sentido. Considerando todos os módulos dos arquivos de metadados, os termos mais frequentes foram “estadual”, “serra”, “parque”, “projeto” (unigramas) e “parque estadual” e “projeto temático”. É certo que a quantidade de pacotes de dados utilizados compõem um universo restrito para a elaboração de um dicionário de dados, porém mesmo assim já é claro que, para as especificidades desse Projeto, alguns termos podem ser padronizados, para evitar ambiguidades ou erros de grafia – como, por exemplo, a variável “polinização”, levantada por duas pesquisas diferentes e representada por parâmetros distintos. Comparando-se com o repositório de dados do PPBio (Programa de Pesquisa em Biodiversidade – MCTI), também com o Metacat, percebe-se que os termos mais frequentes nas palavras-chave foram “Amazônia”(unigrama) e “parcelas permanentes” (bigrama). Nota-se ainda um regionalismo, com a alta frequência do termo “igarapé”, próprio da Região Norte do Brasil para se referir a pequenos riachos: se incorporada em uma ontologia, para facilitar buscas por termos de interesse, essa palavra seria mais facilmente encontrada em mecanismos de busca.

De todo modo, a solução para o gerenciamento de dados de levantamentos biológicos apresentada aqui permite que a informação seja acessada por meio de consultas qualitativas (por palavras-chave), quantitativas (por retomada de valores numéricos) ou espaciais (por representações do espaço geográfico).

É importante registrar que uma limitação do sistema é o fato de o usuário precisar inserir a mesma informação duas vezes – no catálogo de metadados e no sistema de informação genérico. Isso ocorre para título, nome do responsável pelo

---

<sup>11</sup> Além dos 30 metadados catalogados pelo Projeto Gradiente Funcional, foram utilizados 279 pacotes de dados catalogados no repositório do PPBio e 59, no repositório do PELD – Amazônia Ocidental.

levantamento e na descrição das variáveis da tabela de dados. Para evitar essa duplicação de esforços, é possível desenvolver uma funcionalidade adicional ao programa Morpho, por meio de um *plug-in*, para que se permita o povoamento automático da base de dados genérica, pela inserção de dados e metadados pelo Morpho. Essa etapa está em desenvolvimento, porém ainda não funcional.

Outra limitação é o fato de a genericidade do sistema permitir que usuários adotem nomes diferentes para a mesma variável, ou adotem nomes iguais para representar variáveis diferentes, o que dificulta consultas comparativas de dados entre diferentes levantamentos, uma vez que as relações entre as variáveis são desconhecidas. Uma forma de contornar esse problema é o desenvolvimento de um dicionário de variáveis.

### 3.6. ANÁLISE DOS RESULTADOS E CONSIDERAÇÕES FINAIS

Com base no arcabouço proposto no Capítulo 1, analiso os componentes **Pessoas**, **Arcabouço Institucional**, **Dados**, **Tecnologia** e **Padrões** do sistema de informação desenvolvido para Projeto Temático BIOTA Gradiente Funcional. O componente **Pessoas** foi o maior limitante: o sistema foi idealizado e implementado por uma aluna de pós-graduação e um aluno de graduação da UNICAMP. Apesar das tentativas de recrutar bolsistas ou realizar acordos de cooperação, a grande deficiência de profissionais habilitados para trabalhar a gestão da informação em biodiversidade no mercado impediu a consolidação de uma equipe. Por outro lado, os dois alunos envolvidos com o desenvolvimento do sistema foram capacitados para trabalhar nessa área. Além de pessoas para desenvolver o sistema, é necessário haver pessoas para sua manutenção, o que é frequentemente negligenciado em projetos na área – há um entendimento frequente de que a contratação temporária de um serviço terceirizado será o suficiente para tratar as demandas de, por exemplo, um grupo de pesquisa ou projeto. Contudo, esse tipo de serviço temporário só faz sentido quando inserido em um contexto mais amplo de infraestrutura de informação que possa manter os



benefícios ao longo do tempo. Outro grupo de pessoas, os usuários, sejam eles pesquisadores ou gestores, foi pouco contemplado no presente sistema, que enfocou mais aspectos técnicos.

Em relação ao **Arcabouço Institucional**, não havia acordos e regras formais previamente estabelecidos no contexto do Projeto Temático. As interações com os pesquisadores e a colaboração se deu por meio de acordos informais. Editais recentes da FAPESP e do CNPq vêm colocando a obrigatoriedade de depósito de dados e metadados em repositórios públicos, evidenciando que há avanços em curso para a formalização de regras para o compartilhamento de dados e metadados. O caminho da informalidade das regras para sua institucionalização foi observado no caso do PPBio Amazônia Ocidental: nos primeiros quatro anos, o compartilhamento de dados e metadados também se deu por acordos informais e amadureceu para a institucionalização de uma política de dados em 2008. Os **Dados** do Projeto Temático foram representativos da grande diversidade de abordagens em pesquisas biológicas de campo, variando de dados de sensores meteorológicos até dados genéticos.

Em relação à **Tecnologia**, o sistema de informação foi eficiente para gerenciar dados e metadados do Projeto Temático. Foi possível realizar consultas integradas tanto ao catálogo de metadados, pelo uso de palavras-chave, quanto à base de dados, pela seleção de variáveis e de parâmetros temporais e espaciais. A possibilidade de consultar qualquer variável de interesse, independentemente da associação com uma entidade taxonômica, configura um sistema de informação sobre biodiversidade original e inovador. No sistema estudado, a base de dados é verdadeiramente genérica: a única restrição é a exigência da informação espacial. Como a localização pode ser apenas um par de coordenadas ou mesmo a definição de uma das parcelas de 1 ha do projeto, trata-se de um requisito de simples fornecimento. O exemplo de visualização das observações pela funcionalidade do atlas em dispositivos móveis sinaliza que não estamos longe de utilizar tecnologias da informação como a apresentada nesse estudo em tempo real no campo. Hoje, verificar os valores observados de perímetro de espécies arbóreas da Mata Atlântica já é possível com um telefone celular com acesso à internet. Ainda que a conectividade seja um fator limitante, a tendência de desenvolvimento tecnológico pode mudar esse cenário em um curto espaço de tempo.

Foram adotadas **Normas e Padrões** internacionalmente aceitas, como aquelas recomendadas pela OGC<sup>12</sup>. A área do conhecimento de Biodiversidade e Ecossistemas ainda proporciona muitas oportunidades para se consolidar padrões. Por exemplo, a ambiguidade do uso dos termos para as consultas integradas mostrou-se problemática, pois a falta de termos controlados dificulta o discernimento entre termos iguais, utilizados para representar informações de natureza diferentes, ou o uso de termos diferentes para representar a mesma entidade. O uso de ontologias, *web-semântica* e vocabulários controlados são alternativas para solucionar esses problemas.

É possível avançar em ferramentas computacionais pontualmente, mas sua sustentabilidade, a longo prazo, depende dos elementos de infraestrutura discutidos no Capítulo 1. Apesar de acordos sobre o compartilhamento de dados e metadados, nesse estudo de caso, não terem sido firmados antes do desenvolvimento desse sistema de informação, 30 pesquisadores concordaram em preencher o formulário de metadados. Ainda assim, o fato de o sistema de informação ter sido proposto à equipe com o Projeto Temático já em andamento pode ter levado muitos pesquisadores a não se interessarem em compartilhar, ou não terem reservado tempo para isso. Os avanços para a integração da informação obtidos com o desenho e implementação do sistema foram mais significativos nos aspectos de tecnologia, dados, normas e padrões. O enfoque maior em provedores e consumidores de dados e em acordos e arranjos para promover o treinamento e o incentivo ao compartilhamento de dados pode conferir maior sustentabilidade ao sistema aqui apresentado. Aliado à isso, é necessário também a costura institucional que viabilize mecanismos de financiamento, treinamento e fixação de pessoal de desenvolvimento de sistemas.

Esse sistema de informação apresentado é um esforço para contribuir para o desafio de reforçar o valor dos dados existentes, de forma a torná-los disponíveis para pesquisas futuras. Inúmeras etapas e funcionalidades podem ser adicionadas para avançar na integração de dados de biodiversidade. O tratamento dos dados e das ferramentas, em conjunto com a abordagem de aspectos institucionais e culturais – elementos fundamentais de uma infraestrutura de informação, como discutido no

---

<sup>12</sup> *Open Geospatial Consortium* - <http://www.opengeospatial.org/>.

Capítulo 1 - podem colaborar com a longevidade e a utilidade do esforço aqui apresentado.

Adicionar ao sistema funcionalidades que acomodem demandas de usuários que atuem na gestão ambiental ou cidadãos cientistas é um tema interessante a ser explorado por futuros sistemas. A concepção de novas ferramentas de visualização, a possibilidade de acomodar dados gerados por cidadãos cientistas e adicionados por dispositivos móveis e o uso de workflows para apoiar o co-manejo adaptativo são tendências da evolução de sistemas de informação como objetos-ponte entre a conhecimento científico e sistemas de tomada de decisão.



## CONCLUSÃO

Evidências de transições críticas em sistemas ecológicos em diferentes escalas (BARNOSKY et al., 2012) demandam o estudo do sistema terrestre sob a ótica de diferentes disciplinas (FERREIRA, 2004; LARIGAUDERIE et al., 2012). O papel da ciência é reconhecidamente importante para auxiliar o enfrentamento dos desafios das mudanças ambientais globais (KATES et al., 2001), contudo como mobilizar o conhecimento científico para informar decisões sobre a gestão ambiental ainda é uma questão pouco amadurecida nesse debate (CASH et al., 2003). Novas tecnologias computacionais que permitem lidar com informações de procedência diversa e características heterogêneas podem contribuir para representar o estado atual do conhecimento do sistema terrestre sob diferentes óticas e para comunicar esse conhecimento entre grupos com culturas científicas distintas.

A prática científica está mudando com o acesso a novas ferramentas tecnológicas e com a grande quantidade de dados passíveis de análise (HEY et al., 2009). As pesquisas estão se configurando cada vez mais como esforços colaborativos e a comunicação científica está mais dinâmica. Compreender esses dados é um grande desafio, e práticas de curadoria passam a se tornar relevantes nas diferentes áreas do conhecimento (PARSONS et al., 2011), pois, ao mesmo tempo em que pesquisadores de diferentes formações deparam-se com cada vez mais dados para interpretar e transformar em novo conhecimento, a maneira de comunicar esse conhecimento para os diferentes setores da sociedade também passa por transformações.

Assim, esse cenário de permeabilidade do conhecimento entre diferentes grupos de pesquisa e entre a ciência e outros setores da sociedade configura novas possibilidades de mobilizar o conhecimento científico para informar decisões sobre as escolhas humanas sobre a gestão ambiental. Mais especificamente, avanços computacionais viabilizam a representação do conhecimento em sistemas que realçam a relevância, a credibilidade e a legitimidade das informações produzidas (CASH et al.,

2003), uma via interessante a fim de mobilizar o conhecimento para apoiar sistemas de tomada de decisão.

A perda de biodiversidade é um dos limites considerados irreversíveis em estudos sobre as transições críticas que o sistema terrestre está enfrentando (DÍAS et al., 2006; ROCKSTORM, 2009). Neste estudo, investiguei como ferramentas computacionais podem contribuir para a integração da informação sobre biodiversidade e ecossistemas e para a mobilização do conhecimento científico de maneira a apoiar processos de tomada de decisões sobre a conservação e a gestão ambiental. O Projeto Temático Biota Gradiente Funcional foi escolhido como estudo de caso por ser representativo de projetos de pesquisa multi-institucionais e multidisciplinares na temática da biodiversidade.

Para tanto, o presente trabalho foi dividido em quatro Capítulos. Na Introdução, delineei a problemática que motivou este estudo: nossas estratégias para a conservação e gestão ambiental dependem das limitações de nosso conhecimento atual sobre a biodiversidade (KORICHEVA; SIIPI, 2004). A representação do conhecimento sobre biodiversidade em sistemas de informação é fortemente aliada às pesquisas em taxonomia e sistemática, com enfoque em dados sobre espécimes em coleções biológicas e sobre observações de ocorrências de espécies associadas. Neste trabalho, estudei a representação do conhecimento sobre biodiversidade e ecossistemas, de forma a contemplar pesquisas sobre os mecanismos que levam as ocorrências observadas, bem como representar espécies não identificadas e a diversidade de interações e informações relevantes para a biodiversidade que não estão associadas a alguma entidade taxonômica, como, por exemplo, dados de carbono no solo.

No Capítulo 1, investiguei componentes de uma infraestrutura de informação sobre biodiversidade e discuti o potencial dessa infraestrutura como objeto-ponte entre ciência e processos de tomada de decisão em conservação e gestão ambiental. A análise da alocação de esforços para implementar infraestruturas de informação, considerando-se os recursos humanos em todas as etapas do ciclo de vida dos dados, aspectos institucionais, as características dos dados, a tecnologia, normas e padrões, revelou-se uma via para organizar resultados em sistemas de conhecimento (CASH et

al., 2003) e mobilizar o conhecimento científico para apoiar sistemas de tomada de decisão. A transparência de sistemas de conhecimento é considerada fundamental na intermediação entre ciência e sociedade, e a infraestrutura de informação em biodiversidade mostrou-se uma via interessante para tornar algumas etapas do processo científico mais transparentes - é o caso da exposição de dados e métodos de trabalho, ou fluxos de trabalhos científicos, que são representações de processos envolvidos na realização de uma análise científica (JONES et al., 2006). Para que essa transparência seja viabilizada em infraestruturas de informação, é necessário considerar os cinco elementos analisados, em todas as escalas: desde dentro de um grupo de pesquisa, passando por equipes multidisciplinares, até agências de fomento ou instituições de ensino e pesquisa.

Ainda no Capítulo 1, detectei que, na última década, houve avanços em alguns aspectos da infraestrutura de informação necessária para integrar o conhecimento científico sobre biodiversidade e ecossistemas, a fim de subsidiar decisões práticas sobre o gerenciamento de recursos: além do avanço no desenvolvimento de *softwares*, evoluiu também o estabelecimento de normas e padrões para documentar a informação e permitir a interoperabilidade de sistemas. No entanto, para que essas normas, padrões e ferramentas sejam efetivamente adotados de forma a viabilizar a integração da informação, é necessário empreender mais esforços no sentido de promover os arranjos institucionais, legislativos e administrativos que atuam como alicerces da infraestrutura da informação e conferem a sustentabilidade dos sistemas, a médio e longo prazo. Além disso, é preciso capacitar recursos humanos no uso desses recursos tecnológicos e promover a aculturação dos atores envolvidos, no sentido de tornar o compartilhamento de dados uma prática amplamente realizada.

No Capítulo 2, discuti as características de dados ecológicos, especialmente sua heterogeneidade; boas práticas para documentá-los, como a importância de documentar procedimentos de coleta; e algumas ferramentas existentes para gerenciá-los, como a recuperação da informação em catálogos de metadados, para promover a integração do conhecimento ecológico.

No Capítulo 3, analisei o sistema de informação desenvolvido para o Projeto Temático Biota Gradiente Funcional, como estudo de caso para a integração da

informação sobre biodiversidade e ecossistemas. Apresentei os requisitos e o processo decisório para as escolhas dos componentes do sistema. O sistema de informação foi eficiente para gerenciar dados e metadados do Projeto Temático. Foi possível realizar consultas integradas tanto ao catálogo de metadados, pelo uso de palavras-chave, quanto à base de dados, pela seleção de variáveis e de parâmetros temporais e espaciais. A possibilidade de consultar qualquer variável de interesse, independentemente da associação com uma entidade taxonômica, configura um sistema de informação sobre biodiversidade original e inovador. No sistema estudado, a base de dados é verdadeiramente genérica: a única restrição é a exigência da informação espacial. O exemplo de visualização das observações pela funcionalidade do atlas em dispositivos móveis sinaliza que não estamos longe de utilizar tecnologias da informação como a apresentada nesse estudo em tempo real no campo.

Apesar dos resultados positivos, houve também pontos que precisam ser melhorados. A ambiguidade do uso dos termos para as consultas integradas mostrou-se problemática, pois a falta de termos controlados dificulta o discernimento entre termos iguais, utilizados para representar informações de natureza diferentes, ou o uso de termos diferentes para representar a mesma entidade. O uso de ontologias, *web-semântica* e vocabulários controlados são alternativas para solucionar esses problemas. Ficou clara também a necessidade de desenvolver funcionalidades que acomodem demandas de usuários que atuem na gestão ambiental ou cidadãos cientistas. Novas ferramentas de visualização, a possibilidade de acomodar dados gerados por cidadãos cientistas e adicionados por dispositivos móveis e o uso de *workflows* para apoiar o co-manejo adaptativo são componentes que aproximam os diversos atores envolvidos com a gestão da informação em biodiversidade e configuram infraestruturas de informação de biodiversidade como objetos-ponte entre a geração de conhecimento e sistemas de tomada de decisão.

A incorporação do conhecimento local e de observações de cidadãos-cientistas podem auxiliar nossa compreensão do sistema terrestre. Sistemas de observação que forneçam subsídio ao manejo adaptativo já estão funcionando, como no exemplo do Parque Kruger, na África do Sul. Além disso, o DataONE integrou dados de diferentes disciplinas do sistema terrestre via integração de metadados, indicando que a



interdisciplinaridade para enfrentar os desafios ambientais pode se beneficiar de ferramentas computacionais.

A investigação revelou que, apesar de as tecnologias e o estabelecimento de normas estarem avançando rapidamente, pouco esforço vem sendo colocado nos aspectos institucionais e administrativos, comprometendo a sustentabilidade dos avanços obtidos. A formação de recursos humanos em gestão da informação é também um gargalo que demanda mais atenção.

A partir da análise do sistema de informação do Projeto Temático Biota Gradiente Funcional, considerando suas limitações, proponho recomendações para a FAPESP e outras agências de fomento e instituições de pesquisa em relação a preservação, compartilhamento e integração de dados ecológicos de futuros projetos.

Para promover a integração de informações sobre biodiversidade e ecossistemas, recomendo que:

1. agências financiadoras promovam uma consulta à comunidade científica, para elaborar uma Política de Dados levantados com recursos de Projetos do Programa BIOTA e correlatos. Nessa política, deverá estar requisitado o uso de padrões internacionalmente aceitos;
2. as propostas venham acompanhadas de um plano de gestão dos dados, tal como a NSF vem fazendo há dois anos. Nesse plano, os acordos sobre o compartilhamento de dados e metadados estarão previstos antes do início do projeto;
3. agências financiadoras aloquem recursos para implementar repositórios de dados e metadados institucionais, para acomodar as informações de projetos que não possuam infraestrutura da informação para fazê-lo;
4. agências financiadoras mantenham recursos humanos e financeiros para a devida gestão desses repositórios;
5. a capacitação de pesquisadores, alunos e técnicos no uso de ferramentas para documentação, preservação, análise e visualização de dados, bem como para as demais etapas do ciclo de vida de dados;

6. a disponibilização de dados bem documentados e de acordo com a política estabelecida seja reconhecida como produtividade de pesquisadores, grupos de pesquisa e programas de pós-graduação.

## REFERÊNCIAS

ALLARD, S. DataONE: Facilitating eScience through Collaboration, **Journal of eScience Librarianship**, v. 1, n. 1, Article 3. 2012.

ALTINTAS, I; BERKLEY, C; JAEGER, E; JONES, M; LUDÄSCHER, B; MOCK, S. Kepler: An Extensible System for Design and Execution of Scientific Workflows. **Proceedings of the The Future of Grid Data Environments**, Global Grid Forum 10. 2004.

ALVES, L. F.; VIEIRA, S. A.; SCARANELLO, M. A. et al. Forest structure and live aboveground biomass variation along an elevational gradient of tropical Atlantic moist forest (Brazil). **Forest Ecology and Management**, v. 260, n. 5, p. 679-691, jul. 2010.

BACCARO, F. B.; DRUCKER, D.; VALE, J. D.; OLIVEIRA, M. L.; MAGALHÃES, C. U.; LEPSCH-CUNHA, N.; MAGNUSSON, W. E. A Reserva Ducke. In: OLIVEIRA, M. L.; BACCARO, F. B.; BRAGA-NETO, R.; MAGNUSSON, W. E. (Org.). **Reserva Ducke: a biodiversidade amazônica através de uma grade**. Manaus: Attema, 2008. p. 11-20.

BARBOSA, F. A. R.; SCARANO, F. R.; SABARÁ, M G.; ESTEVES, F. A. Brazilian LTER: ecosystem and biodiversity information in support of decision-making. **Environmental Monitoring and Assessment**, v. 90, n. 1/3, p. 121-133, 2004.

BARNOSKY, A. D.; HADLY, E. A.; BASCOMPTE, J. et al. Approaching a state shift in Earth's biosphere. **Nature**, v. 486, n. 7401, p. 52-58, 2012.

BELLARD, C.; BERTELSMEIER, C.; LEADLEY, P.; THUILLER, W.; COURCHAMP, F. Impacts of climate change on the future of biodiversity. **Ecology letters**, p. 365-377, jan. 2012.

BERKES, F. Evolution of co-management: role of knowledge generation, bridging organizations and social learning. **Journal of environmental management**, v. 90, n. 5, p. 1692-702, abr. 2009.

BERKLEY, C. et al. Metacat: a schema-independent XML database system. In: INTERNATIONAL CONFERENCE ON SCIENTIFIC AND STATISTICAL DATABASE MANAGEMENT (SSDBM '01), 13., 2001, Fairfax. **Proceedings...** Fairfax, VA: SSDBM, 2001. p. 171-179.

BERRY, R. J. Ecology: where genes and geography meet. **Journal of Animal Ecology**, v. 58, p. 733–759, 1989.

BIG DATA. **Nature**, 455, n. 7209, p. 1–136, 2008.

BISBY, F. A. The Quiet Revolution: Biodiversity Informatics and the Internet. **Science**, v. 289, n. 2309, 2000.

BORGMAN, C. L.; WALLIS, J. C.; ENYEDY, N. **Little Science Confronts the Data Deluge: Habitat Ecology, Embedded Sensor Networks, and Digital Libraries**. Los Angeles: Papers, Center for Embedded Network Sensing, 2006. p. 1-20.

BOWKER, G. C. Biodiversity Datadiversity. **Social Studies of Science**, v. 30, n. 5, p. 643-683, out. 2000.

BRADSHAW, G. A.; BORCHERS, J. G. Uncertainty as Information: Narrowing the Science-policy Gap. **Conservation Ecology**, v. 4, n. 7, p. 1-10, 2000.

BROWN, J.H.. **Macroecology**. Chicago: University of Chicago Press, 1995.

BROWN, J. H.; WHITHAM, T. G.; MORGAN ERNEST, S. K.; GEHRING, C. A. Complex species interactions and the dynamics of ecological systems: long-term experiments. **Science**, v. 293, n. 5530, p. 643-50, jul. 2001.

BRUNA, E. M. Scientific journals can advance tropical biology and conservation by requiring data archiving. **Biotropica**, v. 42, n. 399–401, 2010.

BRUNDTLAND, G. **Our Common Future**: from one earth to one world. Nova York: Oxford University Press, 1987.

BRUNT, J. W. LTER advances Ecological Informatics. **The LTER Network News**, v. 19, n. 2, p. 16-17, 2006.

BUTLIN, R. Data archiving. **Heredity**, v. 106, n. 709, 2010.

de BY, R. A.; DRUCKER, D. P.; SANTOS, L. C. Base de dados para inventários de biodiversidade. In: OLIVEIRA, M. L.; BACCARO, F. B.; BRAGA-NETO, R.; MAGNUSSON, W. E. (Org.). Reserva Ducke: a biodiversidade amazônica através de uma grade. Manaus: Attema Design Editorial, 2008. p. 145-160. Disponível em: <[http://ppbio.inpa.gov.br/Port/public/LivroRFAD\\_ebook.pdf/](http://ppbio.inpa.gov.br/Port/public/LivroRFAD_ebook.pdf/)>. Acesso em: 19 set. 2009.

CANHOS, V.P.; S. SOUZA, S.; GIOVANNI, R.; CANHOS, D.A.L. Global biodiversity informatics: setting the scene for a “new world” of ecological modeling. **Biodiversity Informatics**, v. 1, p. 1-13, 2004.

CARDINALE, B. J.; WRIGHT, J. P.; CADOTTE, M. W. et al. Impacts of plant diversity on biomass production increase through time because of species complementarity. **Science**, v. 104, n. 46, 2007.

CASH, D. W.; ADGER, W. N.; BERKES, F. et al. Scale and Cross-Scale Dynamics : governance and Information in a multilevel world. **Ecology and Society**, v. 11, n. 2, 2006.

CASH, D. W.; CLARK, W. C.; ALCOCK, F., DICKSON NM, ECKLEY N., Guston, D.H.; JÄGER, J. MITCHELL, R.B. Knowledge systems for sustainable development. **Proc. Natl. Acad. Sci. USA**, v. 100 p. 8086–8091, 2003.

CASTRO, P. F. D. **Avaliação de impacto de programas de pesquisa em biodiversidade**. 2011. 231 f.Tese (Doutorado). Universidade Estadual de Campinas,Campinas, SP.

CLARK, J. S.; CARPENTER, S. R.; BARBER, M. et al. Ecological Forecasts: an emerging imperative. **Science**, v. 293, n. 5530, p. 657-660, 2001.

COLOMBO, A.; JOLY, C. A. Brazilian Atlantic Forest lato sensu: the most ancient Brazilian forest , and a biodiversity hotspot, is highly threatened by climate change. **Brazilian Journal of Biology**, v. 70, n. 3, p. 697-708, 2010.

COLUMBIA, B.; ODUM, E. P. **Best Practices for Preparing Ecological Data Sets to Share**. Columbia: Columbia University, 1998. p. 1997-1998.

COLWELL, R. K. **Biota: the biodiversity database manager**. Sunderland, Massachusetts: Sinauer Associates, Inc., Publishers, 1996.

COMITÊ DE PLANEJAMENTO DA INFRAESTRUTURA NACIONAL DE DADOS ESPACIAIS. **Plano de ação para implantação da infraestrutura nacional de dados espaciais** .Rio de Janeiro: Comissão Nacional de Cartografia, 2010. 203 p.

COOK, C. N.; HOCKINGS, M.; CARTER, R. (BILL). Conservation in the dark? The information used to support management decisions. **Frontiers in Ecology and the Environment**, v. 8, n. 4, p. 181-186, maio, 2010.

COOPER, C. B.; DICKINSON, J.; PHILLIPS, T.; BONNEY, R.. Citizen science as a tool for conservation in residential ecosystems. **Ecology and Society**, v. 12, n. 2 p. 11, 2007.

COSTA, F. R. C.; MAGNUSSON, W. E. The need for large-scale, integrated studies of biodiversity – the experience of the Program for Biodiversity Research in Brazilian Amazonia. **Natureza & Conservação**, v. 8, n. 1, p. 3-12, 2010.

COSTANZA, R. Developing ecological research that is relevant for achieving sustainability. **Ecological Applications**, v. 3, n. 4, p. 579-581, 1993.

COSTELLO, M. J. Motivating Online Publication of Data. **American Institute of Biological Sciences**, v. 59, n. 5, p. 418-427, 2009.

CURRY, G.; HUMPHRIES, C. (Ed.). Biodiversity Databases. **Systematics**, v. 2007, n. 1270, abr. 2007.

CUTTS, B. B.; WHITE, D. D.; KINZIG, A. P. Participatory geographic information systems for the co-production of science and policy in an emerging boundary organization. **Environmental Science & Policy**, v. 14, n. 8, p. 977–985, 2011.

USDA. **Colony Collapse Disorder Progress Report**. Washington, jun., 2010.

DAVIDSON, E. A.; ARAÚJO, A. C. DE; ARTAXO, P. et al. The Amazon basin in transition. **Nature**, v. 481, n. 7381, p. 321-328, jan. 2012.

DÍAZ, S.; FARGIONE, J.; CHAPIN, F. S.; TILMAN, D. Biodiversity loss threatens human well-being. **PLoS biology**, v. 4, n. 8, e277, 2006.

FEGRAUS, E. H.; ANDELMAN, S.; JONES, M. B.; SCHILDHAUER, M. 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**, v. 86, p. 158–168, 2005..

FERREIRA, L.C. Idéias para uma Sociologia da questão ambiental - teoria social , sociologia ambiental e interdisciplinaridade. **Desenvolvimento e Meio Ambiente** (UFPR), Curitiba, v. 10, p. 77-89, 2004.

FLEISHMAN, E.; BLOCKSTEIN, D. E.; HALL, J. A. et al. Top 40 Priorities for Science to Inform US Conservation and Management Policy. **BioScience**, v. 61, n. 4, p. 290-300, abr. 2011.

FUGAZZA, C.; DUPKE, S.; VACCARI, L. Matching SKOS Thesauri for Spatial Data Infrastructures. **Communications in Computer and Information Science**, v. 108, p. 211-221, 2010.

GIERYN, T. F. Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists **American Sociological Review**, v. 48, n. 6. p. 781-795. 1983.

GIOIA, P. Managing biodiversity data within the context of climate change: towards best practice. **Austral Ecology**, v. 35, n. 4, p. 392-405, nov. 2009.

GONZALEZ, M. Análise das restrições de acesso a dados de espécies ameaçadas, previstas em políticas de coleções biológicas científicas brasileiras, à luz do direito ambiental e da ciência da informação. **Ciência da Informação**, Brasília, DF, n. 39, nov. 2010. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1705>>. Acesso em: 07 ago. 2012.

GROFFMAN, P. M.; STYLINSKI, C.; NISBET, M. C. et al. Restarting the conversation: challenges at the interface between ecology and society. **Frontiers in Ecology and the Environment**, v. 8, n. 6, p. 284-291, 2010.

GURALNICK, R. P.; HILL, A. W.; LANE, M. Towards a collaborative, global infrastructure for biodiversity assessment. **Ecology letters**, v. 10, n. 8, p. 663-72, ago. 2007.

GUSTON, D. H. Boundary Organizations in Environmental Policy and Science: An Introduction. **Science, Technology & Human Values**, v. 26, n. 4, p. 399-408, 2009.

HACKETT, E. et al. Ecology transformed: the National Center for Ecological Analysis and Synthesis and the changing patterns of ecological research. In: OLSON, G.; ZIMMERMAN, A.; BOS, N. (Ed.). **Scientific collaboration on the internet**. Boston, MA: Massachusetts Institute of Technology, 2008. p. 277-296.



HALPERN, B. S.; FLOETER, S. R. Functional diversity responses to changing species richness in reef fish communities. **Marine Ecology Progress Series**, v. 364, p. 147-156, 2008.

HANSSON, L. Why ecology fails at application: should we consider variability more than regularity? **Oikos**, v. 100 p. 624-627, 2003.

HARRIS, D.; LATIFUR, K.; RAYMOND, P.; BHAVANI, T. Standards for secure data sharing across organizations. **Computer Standards & Interfaces**, v. 29 n. 1, p. 86-96, 2007.

HEIDORN, P. B. Shedding Light on the Dark Data in the Long Tail of Science. **Library Trends**, v. 57, n. 2, p. 280-299, 2008.

HEY, T.; TREFETHEN, A. The Data Deluge: an e-science perspective. **Grid Computing**. [S.l.]: John Wiley & Sons, 2003. p. 809-824.

HIGGINS, D.; BERKLEY, C.; JONES, M. B. Managing heterogeneous ecological data using Morpho. In: INTERNATIONAL CONFERENCE ON SCIENTIFIC AND STATISTICAL DATABASE MANAGEMENT, 14., 2002, Edinburgh. **Proceedings...** Edinburgh: SSDBM, 2002.

HILBORN, R.; MANGEL, M. **The ecological detective: confronting models with data**. Princeton:Princeton University Press, 1997.

HINE, C. Databases as scientific instruments and their role in the ordering of scientific work. **Social Studies Of Science**, v. 38, n. 2, p. 269-298, apr. 2006.

HOBBIE, J. E. Scientific accomplishments of the Long Term Ecological Research program: An introduction. **BioScience** v. 53, p. 17–20. 2003.

HOCHACHKA, W. M.; FINK, D.; HUTCHINSON, R. A. et al. Data-intensive science applied to broad-scale citizen science. **Trends in ecology & evolution**, v. 27, n. 2, p. 130-7, fev. 2012.

HOGAN, D. J. Demographic dynamics and Environmental Change in Brazil. **Ambiente & sociedade**, v. 9, p. 1-30, 2000.

HOLLING, C.S. Resilience and Stability of Ecological Systems. **Annual Review of Ecology and Systematics** v. 4, p. 1-23, 1973.

HOOPER, D. U.; ADAIR, E. C.; CARDINALE, B. J. et al. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. **Nature**, v. 486, n. 7401, p. 105-108, maio 2012.

HUANG, X.; QIAO, G. Biodiversity data sharing is not just about species names: response to Santos and Branco. **Trends in Ecology & Evolution**, v. 26, p. 377-378, 2011.

HUNTINGTON, H. P. The local perspective. **Nature**, v. 478, n. 182, p. 7-8, 2011.

INTERNATIONAL BANK FOR RECONSTRUCTION AND DEVELOPMENT. **Global environment facility trust fund grant agreement: national biodiversity mainstreaming and institutional consolidation project**. Caixa Economica Federal; Fundo Brasileiro Para Biodiversidade – FUNBIO, mar. 2008. Acting as an Implementing Agency of the Global Environment Facility.

JACKSON, J. B.; KIRBY, M. X.; BERGER, W. H. et al. Historical overfishing and the recent collapse of coastal ecosystems. **Science**, v. 293, n. 5530, p. 629-37, jul. 2001.

JASANOFF, S.; WYNNE, B. Science and decisionmaking. In: RAYNER, S.; MALONE, E. L. (Ed.). **Human Choice and Climate Change Volume One**. Columbus: Batelle Press, 1998. p. 1–88.

JOELS, L. C.; CÂMARA, G. O workshop “Modelos e cenários para a Amazônia: o papel da ciência”. **Parcerias Estratégicas**, v. 12, n. p. 29-34. set. 2001.

JOLY, C. A.; RODRIGUES, R. R.; METZGER, J. P.; VERDADE, L. M.; OLIVEIRA, M. C.; BOLZANI, V. S. Biodiversity Conservation Research, Training, and Policy in São Paulo. **Science**, v. 328, n. 5984, p. 1358-1359. 2010.

JOLY, C. A.; HADDAD, C. F. B.; VERDADE, L. M.; OLIVEIRA, M. C.; BOLZANI, V. S. **BIOTA**: Science plan and strategies for the next decade, document resulting from the Workshop BIOTA + 10: setting agenda and priorities for 2020, 2009. Disponível em: <<http://www.biota.org.br/SciencePlan.pdf>>. Acesso em: 12 nov. 2011.

JOLY, C. A.; MARTINELLI, L. A.; ALVES, L. F.; TAMASHIRO, J. Y.; AIDAR, M. P. M.; CAMARGO, P. B.; ASSIS, M. A.; BERNACCI, L. C.; DURIGAN, G. As parcelas permanentes do Projeto Temático BIOTA Gradiente Funcional: Composição florística, estrutura e funcionamento da Floresta Ombrófila Densa dos Núcleos Picinguaba e Santa Virgínia do Parque Estadual da Serra do Mar, São Paulo, Brasil. In: SANQUETTA, C. A. (Ed.). **Experiências de monitoramento no bioma Mata Atlântica com uso de parcelas permanentes**. Curitiba: Programa PELD-CNPq & Multigraph, 2008. p. 109-148.

JOLY, C. A.; ASSIS, M. A.; BERNACCI, L. C.; TAMASHIRO, J. Y.; CAMPOS, M. C. R.; GOMES, J. A. M. A.; LACERDA, M. S.; SANTOS, F. A. M.; PEDRONI, F.; PEREIRA, L. S.; PADGURSCHI, M. C.; PRATA, E. M. B.; RAMOS, E.; TORRES, R. B.; ROCHELLE, A.; MARTINS, F. R.; ALVES, L. F.; VIEIRA, S. A. MARTINELLI, L. A.; CAMARGO, P. B.; AIDAR, M. P. M.; EISENLOHR, P. V.; SIMÕES, E.; VILLANI, J. P.; BELINELLO, R. Florística e fitossociologia em parcelas permanentes da Mata Atlântica do sudeste do Brasil ao longo de um gradiente altitudinal. **Biota Neotropica**, v. 12, n. 1. Disponível em: <<http://www.biotaneotropica.org.br/v12n1/pt/abstract?article+bn0181201201>>. Acesso em: 12 mar. 2012.

JONES, M. B.; SCHILDHAUER, M. P.; REICHMAN, O. J.; BOWERS, S. The New Bioinformatics: integrating ecological data from the gene to the biosphere. **Annual Review of Ecology, Evolution, and Systematics**, v. 37, n. 1, p. 519-544, dez. 2006.

JORDAN, A. Institutions for global environmental change. **Global Environmental Change**, v. 4, n. 3, p. 265-267, 1994.

KATES, R. W.; CLARK, W. C.; CORELL, R. et al. Sustainability science. **Science**, v. 292, n. 5517, p. 641-642, abr. 2001.

KING, G. Ensuring the data-rich future of the social sciences. **Science**, v. 331, n. 6018, p. 719-21, fev. 2011.

KNICK, S. T.; HOLMES, A. L.; MILLER, R. F. **The role of fire in structuring sagebrush habitats and bird communities Conference Paper: Studies in Avian Biology**. 2005

KOLAJA, K. Drug discovery: Computer model predicts side effects. **Nature** v. 486, p. 326–327, 2012

KORICHEVA, J.; SIIPI, H. The phenomenon of biodiversity. In: OKSANEN, M.; PIETARINEN, J. (Ed.). **Philosophy and Biodiversity**. Cambridge: Cambridge University Press, 2004, p. 27-53.

KWA, C. Representations of Nature Mediating between Ecology and Science Policy: the case of the international biological programme. **Social Studies of Science**, v. 17, n. 3, p. 413-442, 1987.

KWA, C. Local ecologies and global science, discourses and strategies of the international geosphere–biosphere programme. **Social Studies of Science**, v. 35, p. 923–950, 2005.

LAHSEN, M.; NOBRE, C. A. Challenges of connecting international science and local level sustainability efforts: the case of the Large-Scale Biosphere – Atmosphere Experiment in Amazonia. **Environmental Science and Policy**, v. 10, p. 62-74, 2007.

LANDEIRO, V. L.; MAGNUSSON, W. E. The geometry of spatial analyses: implications for conservation biologists. **Natureza & Conservação**, v. 9, n. 1, p. 7-20, 2011.

LARIGAUDERIE, A.; PRIEUR-RICHARD, A.-H.; MACE, G. M. et al. Biodiversity and ecosystem services science for a sustainable planet: the DIVERSITAS vision for 2012–20. **Current Opinion in Environmental Sustainability**, v. 4, n. 1, p. 101-105, fev. 2012.

LATOUR, B. **Jamais fomos modernos**: ensaios de antropologia simétrica. Rio de Janeiro: Ed. 34, 1994. 149 p.

LEEMANS, R.; ASRAR, G.; BUSALACCHI, A. et al. Developing a common strategy for integrative global environmental change research and outreach: the Earth System Science Partnership ( ESSP ) Strategy paper. **Current Opinion in Environmental Sustainability**, v. 1, n. 2, oct. p. 4-13, 2001.

LEWINSOHN, T. M.; PRADO, P. I. **Biodiversidade brasileira**: síntese do estado atual do conhecimento. São Paulo: Editora Contexto, 2002.

LEWINSOHN, T. M. A large trophic quilt. **Journal of Animal Ecology**, v. 79, p. 1143–1145, 2010.

LIKENS, G. E. Human-accelerated environmental change. **BioScience**, v. 41, p. 130, 1991.

LIKENS, G. E. The role of science in decision making: does evidence-based science drive environmental policy? **Frontiers in Ecology and the Environment**, v. 8, n. 6, 2010.

LOUNKINE, E.; KEISER, M. J.; WHITEBREAD, S.; MIKHAILOV, D.; HAMON, J.; JENKINS, J. L.; LAVAN, P.; WEBER, E.; DOAK, A. K.; CÔTÉ, S.; SHOICHET, B. K.; URBAN, L. Large-scale prediction and testing of drug activity on side-effect targets. **Nature**, v. 486, p. 361–367, 2012.

LUBCHENCO, J. et al. The Sustainable Biosphere Initiative: an ecological research agenda. **Ecology**, v. 72, p. 371-412, 1991.

LUDWIG, D. Environmental sustainability: magic, science, and religion in natural resource management. **Ecological Applications**, v. 3, n. 4 p. 555-558, 1993.

LYNCH, C. Jim Gray's fourth paradigm and the construction of the scientific record. **The Fourth Paradigm**, 2009. p. 177-183. Scholarly Communication.

LYNCH, C. How do your data grow. **Nature**, v. 455, p. 28-29, 2008.

LYNCH, C. A. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. **ARL Bimonthly Report**, v. 226, p. 1-7, 2003.

MADIN, J. S.; BOWERS, S.; SCHILDHAUER, M. P.; JONES, M. B. Advancing ecological research with ontologies. **Trends in ecology & evolution**, v. 23, n. 3, p. 159-68, mar. 2008.

MALHI, Y., PHILLIPS, O.L., LLOYD, J., BAKER, T., WRIGHT, J., ALMEIDA, S., ARROYO, L., FREDERIKSEN, T., GRACE, J., HIGUCHI, N., KILLEEN, T., LAURANCE, W.F., LEAÑO, C., LEWIS, S., MEIR, P., MONTEAGUDO, A., NEILL, D., NÚÑEZ VARGAS, P., PANFIL, S.N., PATIÑO, S., PITMAN, N., QUESADA, C.A., RUDAS-LL., A., SALOMÃO, R., SALESKA, S., SILVA, N., SILVEIRA, M., SOMBROEK, W.G., VALENCIA, R., VÁSQUEZ MARTÍNEZ, R., VIEIRA, I.C.G. AND VINCETI, B. An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR). **Journal of Vegetation Science**, v. 13 p. 439–450, 2002.

MANGEL, M.; HOFMAN, R. J.; NORSE, E. A.; TWISS JR, J. R. Sustainability and ecological research. **Ecological Applications**, v. 3, n. 4, p. 573-575, 1993.

MARTINS, R. D. A.; FERREIRA, C. Opportunities and constraints for local and subnational climate change policy in urban areas: insights from diverse contexts. **International Journal of Global Environmental Issues**, v. 11, n. 1, p. 37-53, 2011.

MAURER, B. A. Ecological Science and Statistical Paradigms: at the threshold. **Science**, v. 279, n. 5350, p. 502-503, jan. 1998.

MAURER, B. A. **Untangling ecological complexity**: the macroscopic perspective. Chicago: University of Chicago Press, 1999.

MCNIE, E. C. Reconciling the supply of scientific information with user demands : an analysis of the problem and review of the literature. **Environmental Science and Policy**, v. 10, p. 17-38, 2007.

MICHENER, W. Meta-information concepts for ecological data management. **Ecological Informatics**, v. 1, n. 1, p. 3-7, 2006.

MEADOWS, D. L. **The Limits to Growth**: a report for the club of rome project on the predicament of mankind. New York: Universe Books, 1972.

MICHENER, W. K.; JONES, M. B. Ecoinformatics: supporting ecology as a data-intensive science. **Trends in ecology & evolution**, v. 27, n. 2, p. 85-93, fev. 2012.

MICHENER, W. K. Meta-information concepts for ecological data management. **Ecological Informatics**, v. 1, p. 3-7. 2006.

MICHENER, W. K. et al. Nongeospatial metadata for the ecological sciences. **Ecological Applications**, v. 7, p. 330–342, 1997.

MICHENER, W. K. et al. Long term ecological research and information management, **Ecological Informatics**, v. 6, n. 1, p. 13-24, 2011.

MIRA, C.; FEIJAO, P.; DUQUE-SETRADA, T.; NEIDANIS, J.; JOLY, C. **The SinBiota 2.0 Biodiversity Information System**. IEEE e-Science Conference. p.142-149, 2011.

MOONEY, H. A.; SALA, O. E. Science and sustainable use. **Ecological Applications**, v. 3, n. 4, p. 564-566, 1993.

MYERS, N.; MITTERMEIER, R. A.; MITTERMEIER, C. G.; FONSECA, G. A. B.; KENT, J. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, p. 853–858, 2000.

NATIONAL SCIENCE FOUNDATION. **Cyberinfrastructure Vision for 21st Century Discovery**. Arlington, mar. 2007. 58 p..

NELSON, B. Data sharing: empty archives. **Nature**, v. 461, p. 160-163, 2009..

NOVOTNY, V.; MILLER, S. E.; BAJE, L.; BALAGAWI, S.; BASSET, Y.; CIZEK, L. Guild-specific patterns of species richness and host specialization in plant–herbivore food webs from a tropical forest. **Journal of Animal Ecology**, v. 79, p. 1193–1203, 2010.

OLIVEIRA, V. C.; JOLY, C. A. Flooding tolerance of *Calophyllum brasiliense* Camb. (Clusiaceae): morphological, physiological and growth responses. **Trees**, v. 24, n. 1, p. 185-193, nov. 2009.

OSTROM, E. A general framework for analyzing sustainability of social-ecological systems. **Science**, v. 325, n. 5939, p. 419-22, jul. 2009.

PARMESAN, C.; YOHE, G. A globally coherent fingerprint of climate change impacts across natural systems. **Nature**, v. 421, p. 37–42, 2003.

PARSONS, M.A.; FOX, P.A. Is data publication the right metaphor? **Data Science Journal** (submetido). 8 p. Disponível em: <[http://dl.dropbox.com/u/546900/parsons\\_fox\\_metaphor\\_dsj\\_open.docx](http://dl.dropbox.com/u/546900/parsons_fox_metaphor_dsj_open.docx)>. Acesso em: 05 maio 2012.

PARSONS, M. A.; GODOY, O.; LEDREW, E. et al. A conceptual framework for managing very diverse data for complex, interdisciplinary science. **Journal of Information Science**, v. 19, out. 2011.

PEACOCK, J.; BAKER, T. R.; LEWIS, S. L.; LOPEZ-GONZALEZ, G.; PHILLIPS, O. L. The RAINFOR database: monitoring forest biomass and dynamics. **Journal of Vegetation Science**, v. 18, n. 4, p. 535-542, aug. 2007.

PENNISI, E. How will big pictures emerge from a sea of biological data? **Science**, v. 309, n. 5731, p. 94, jul. 2005.



PEREIRA, H. M.; LEADLEY, P. W.; PROENÇA, V. et al. Scenarios for global biodiversity in the 21st century. **Science**, v. 330, n. 6010, p. 1496-501, dez. 2010.

PETERS, D. P.; GROFFMAN, P. M.; NADELHOFFER, K. J. et al. Living in an increasingly connected world: a framework for continental-scale environmental science. **Frontiers in Ecology and the Environment**, v. 6, n. 5, p. 229-237, jun. 2008.

PIWOWAR, H. A.; DAY, R. S.; FRIDSMA, D. B. Sharing detailed research data is associated with increased citation rate. **PloS one**, v. 2, n. 3, p. e308, jan. 2007.

PIWOWAR, H. A. Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. **PLoS One**, v. 6, n. 7, 2011.

PORTER, J. H.; HANSON, P. C.; LIN, C.-C. Staying afloat in the sensor data deluge. **Trends in ecology & evolution**, v. 27, n. 2, p. 121-9, fev. 2012.

PRADO, P. I. Distribuições de abundâncias de espécies: avanços analíticos para entender um padrão básico em ecologia. **Ciência e Ambiente**, v. 39, p. 121-136, 2009.

POSTEL, S. L.; DAILY, G. C.; EHRLICH, P. R. Human appropriation of renewable fresh water. **Science**, v. 271, p. 785–788, 1996.

PETERSON, G. et al. Uncertainty, climate change, and adaptive management. **Conservation Ecology**, v. 1, n. 2, p. 4, 1997.

RAUSHER, M. D. et al. Data archiving. **Evolution**, v. 64, p. 603–604, 2010.

REICHMAN, O. J.; JONES, M. B.; SCHILDHAUER, M. P. Challenges and opportunities of open data in ecology. **Science**, v. 331, p. 703-705, 2011.

RIBEIRO, M. C.; METZGER, J. P.; MARTENSEN, A. C.; PONZONI, F. J.; HIROTA, M. M. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. **Biological Conservation**, v. 142, n. 6, p. 1141-1153, jun. 2009.

RING, I.; HANSJÜRGENS, B.; ELMQVIST, T.; WITTMER, H.; SUKHDEV, P. Challenges in framing the economics of ecosystems and biodiversity: the TEEB initiative. **Current Opinion in Environmental Sustainability**, v. 2, n. 1/2, p. 15-26, maio 2010.

ROCKSTRÖM, J.; STEFFEN, W.; NOONE, K.; PERSSON, Å.; CHAPIN, F. S.; LAMBIN, E. F.; LENTON, T. M.; FOLEY, J. A. A safe operating space for humanity. **Nature**, 461, p. 7263, p. 472-475, 2009.

RODRIGUES, R. R.; BONONI, V. L. R. (Org.). **Diretrizes para conservação e restauração da biodiversidade no Estado de São Paulo**. São Paulo: Instituto de Botânica, 2008. 246 p.

RODRIGUES, R. R.; LIMA, R. A. F.; GANDOLFI, S.; NAVE, A. G. On the restoration of high diversity forests: 30 years of experience in the Brazilian Atlantic Forest. **Biological Conservation**, v. 142, n. 6, p. 1242-1251, jun. 2009.

ROOT, T. L. et al. Fingerprints of global warming on wild animals and plants. **Nature**, v. 421, p. 57–60, 2003.

ROMERO, T. **Diversidade de interação**. Agência FAPESP. Disponível em: <<http://agencia.fapesp.br/9737>>. Acessado em: 27 nov. 2011.

ROSADO, B. H. P.; OLIVEIRA, R. S.; JOLY, C. A.; AIDAR, M. P. M.; BURGESS, S. S. O. Diversity in nighttime transpiration behavior of woody species of the Atlantic Rain Forest, Brazil. **Agricultural and Forest Meteorology**, v. 158, p. 13-20, jun. 2012.

SAGARIN, R.; PAUCHARD, A. Observational approaches in ecology open new ground in a changing world. **Frontiers in Ecology and the Environment**, v. 8, p. 379–386, 2010.

SALIM, J. A. Um repositório espacialmente habilitado para dados ecológicos genéricos. 2009. 43f. Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) - Instituto de Computação, Universidade Estadual de Campinas.

SANTOS, A. M.; BRANCO, M. The quality of name-based species records in databases. **Trends in Ecology & Evolution**, v. 26, p. 377-378, nov. 2011.

SCHEFFER, M.; BASCOMPTE, J.; BROCK, W. A. et al. Early-warning signals for critical transitions. **Nature**, v. 461, n. 7260, p. 53-59, 2009.

SCHOR, T. **Ciência e Tecnologia**: o caso do experimento de grande escala da biosfera- atmosfera na Amazônia (LBA). São Paulo: AnnaBlume/Fapesp, 2008. 147 p.

SCOONES, I. New ecology and the social sciences: what prospects for a fruitful engagement? **Annu. Rev. Anthropol.** v. 28, p. 479-507, 1999.

SEIXAS, C. S.; KALIKOSKI, D. C.; ALMUDI, T. et al. Gestão compartilhada do uso de recursos pesqueiros no Brasil: elementos para um programa nacional. **Ambiente & Sociedade**, v. 14, n. 1, p. 23-44, 2011.

SHACKLEY, S. Trust in models? The mediating and transformative role of computer models in the environmental discourse. In: REDCLIFT, M.; WOODGATE, G. (Ed.). **The International Handbook of Environmental Sociology**. Cheltenham: Edward Elgar, 1997. p. 237-260.

SHRADER-FRECHETTE, K. S.; MCCOY, E. D. **Method in ecology**: strategies for conservation. Cambridge: Cambridge University Press, 1993.

SHUGART, H. H. **Terrestrial Ecosystems in Changing Environments**. Cambridge University Press, 1998. 552 p.

DA SILVA, F. R., ALMEIDA-NETO, M., DO PRADO, V. H. M., HADDAD, C. F. B. AND DE CERQUEIRA ROSSA-FERES, D. Humidity levels drive reproductive modes and phylogenetic diversity of amphibians in the Brazilian Atlantic Forest. **Journal of Biogeography**, v. 39, n. 9, p. 1720–1732, 2012.

SIQUEIRA, M. F.; DURIGAN, G. Modelagem da distribuição geográfica de espécies lenhosas de cerrado no Estado de São Paulo. **Revista Brasileira de Botânica**, v. 30, p. 239-249, 2007.

SLUIJS, J. P. VAN DER; EST, R. VAN; RIPHAGEN, M. Beyond consensus: reflections from a democratic perspective on the interaction between climate politics and science. **Current Opinion in Environmental Sustainability**, v. 2, n. 5/6, p. 409-415, dez. 2010.

SOBERÓN, J.; NAKAMURA, M.. Niches and distributional areas: Concepts, methods and assumptions **Proceedings of the National Academy of Sciences USA**. v. 106, suppl. 2, p. 19644-19650, 2009.

STAR, S.; GRIESEMER, J. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39". **Social Studies of Science**, v. 19, n. 3 p. 387–420, 1989.

STRASSER, C.; COOK, R.; MICHENER, W.; BUDDEN, A.; KOSKELA, R. DataONE promoting data stewardship through best practices. In: ENVIRONMENTAL INFORMATION MANAGEMENT CONFERENCE, 1., 2011, California. **Proceedings...** California: University of California, 2011. p. 126-131.

TAYLOR, P. J. **Unruly complexity**: ecology, interpretation, engagement. Chicago: University of Chicago Press, 2005.

TILMAN, D. Resource competition and community structure. **Monogr. Pop. Biol.** v. 17, Princeton, N.J.: Princeton University Press, 1982. p. 296.

TRELOAR, A.; GROENEWEGEN, D.; HARBOE-REE, C. The Data Curation Continuum: managing data objects in institutional repositories. **D-Lib Magazine**, v. 13, n.9/10, p. 1-13, 2007.

TURNER, M. G.; GARDNER, R. H.; O'NEILL, R. V.. Ecological dynamics at broad scales; ecosystems and landscapes. **BioScience**, Special Supplement v. 45, n. 6, p. 29–33, 1995.

UNEP. **Gerenciamento e uso de informações para ampliar a capacidade brasileira em conservar e utilizar a biodiversidade**. [S.l.], [2010]. Documento do Projeto.

VIEIRA, S. A.; ALVES, L. F.; DUARTE-NETO, P. J. et al. Stocks of carbon and nitrogen and partitioning between above- and belowground pools in the Brazilian coastal Atlantic Forest elevation range. **Ecology and evolution**, v. 1, n. 3, p. 421-34, nov. 2011.

VITOUSEK, P. M.; ABER, J. D.; HOWARTH, R. W. Human alteration of the global nitrogen cycle: sources and consequences. **Ecol Appl**, v. 7, p. 737–50, 1997a.

VITOUSEK, P. M.; MOONEY, H. A.; LUBCHENCO, J.; MELILLO, J. M. Human domination of Earth's ecosystems. **Science**, v. 277, p. 494–499, 1997b.

WARNEST, M. **A collaboration model for national spatial data infrastructure in federated countries**. 2005. 279 f. Dissertation (Ph.D. in Geomatics) – Department of Geomatics, University of Melbourne, Australia.

WHITLOCK, M. C. et al. Data archiving. **American Naturalist**, v. 175, p. 145–146, 2010.

WHITLOCK, M. C. Data archiving in ecology and evolution: best practices. **Trends in ecology & evolution**, v. 26, n. 2, p. 61-5, fev. 2011.

WIECZOREK, J.; BLOOM, D.; GURALNICK, R. et al. Darwin Core: an evolving community-developed biodiversity data standard. **PloS one**, v. 7, n. 1, p. e29715, jan. 2012.

WINOWIECKI, L.; SMUKLER, S.; SHIRLEY, K.; REMANS, R.; PELTIER, G.; LOTHES, E. KING, E.; COMITA, L.; BAPTISTA, S. ALKEMA, L. Tools for enhancing interdisciplinary communication. **Sustainability: Science, Practice & Policy**, v. 7, n. 1, p. 74-80.

WOELFLE, M., OLLIARO, P., TODD, MATTHEW H. Open science is a research accelerator. **Nature Chemistry** v. 3, p.745–748, 2011.

WORTHINGTON, E. B. (Ed.). **The Evolution of IBP**. New York: Cambridge University Press, 1975. 268 p.

ZIMMERMAN, A. S. New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. **Science, Technology & Human Values**, 2008.

ZUUR, A. F.; IENO, E. N.; ELPHICK, C. S. A protocol for data exploration to avoid common statistical problems. **Methods in Ecology and Evolution**, v. 1, n. 1, p. 3-14, 2010.

# ANEXO 1. DOCUMENTAÇÃO DA BASE DE DADOS DELINEADA PARA O PROJETO TEMÁTICO GRADIENTE FUNCIONAL<sup>13</sup>

## A1 Database Schema

To provide a generic repository for biological inventories, BID (Biological Inventories Database) has a vertical fragmentation of data which allow personalized surveys to be insert in database, thus no previous knowledge of a survey must be known in system design.

BID is divided in a schema that easily support any kind of survey to be insert in. This schema is show in appendix A in a graphical manner.

### A.1.1 Tables

- survey(sidx, title, startdate, enddate, responsible, is\_collection\_survey, locational\_kind, timing\_kind, supersurvey, reserve, is\_personalized, tmptablename):
  - *sidx*: unique identification of a survey. Type: integer
  - *title*: title of survey. Type: character varying.
  - *startdate*: start date of the survey. Type: date.
  - *enddate*: end data of the survey. Type: date.
  - *responsible*: identification of a researcher whose is responsible for the survey. Type: integer.
  - *is\_collection\_survey*: Indicates whether this survey will collect materials for inclusion in collection. Type: boolean.
  - *location\_kind*: specifies the type of locational data which this survey provided(for now, null values are allowed, which signal no locations are provided). Supported types are:
    - “g”: mentions an x and y geographic coordinate in any datum, but the intern datum used is WGS84, so a conversion is necessary.
    - “p”: this means that any location is a priori known as a parcel (parcel) for the reserve (reserve), and exists in the “parcel” table. This is useful for data which has not a specific location in a parcel (e.g. , animals).
    - “t1” (or “i1”): should provide locational information using two attributes: the trail name, and a distance along the trail, from start, in meters. (Similarly, “i1” means an igarape plus a distance.).

---

<sup>13</sup> SALIM, J. A. **Um repositório espacialmente habilitado para dados ecológicos genéricos**. 2009. 43p. Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) - Instituto de Computação, Universidade Estadual de Campinas.

- "t2" (or "i2"): should provide a locational information using three attributes: trail (igarape), distance along the trail (igarape) from start, distance off the trail, negative for larboard and positive for starboard, in meters.
- *timing\_king*: Indicates whether this surveys are timed or not. Support types are:
  - "n": this means that survey have no further timing information.
  - "d": this means that every information has a calendar date associated with it.
  - "h": this means that every information has a calendar date and time-of-day associated with it.
- *reserve*: identification of a reserve where the survey will be made. Type: integer.
- *is\_personalized*: indicates if a survey will store information about who was responsible for the collection of a sample. Type: boolean
- *tmptablename*: name of the temporary table which will store survey before the fragmentation process. Type: character varying.

surveyvariable(**varidx**,svy,varname,vartype,nullallowed,seqnce,automatic):

- *varidx*: unique identification of a variable created automatic or by system's users. Type: integer.
- *svy*: identification of what survey this variable is inhered. Type: integer
- *varname*: variable's name. Type: character varying.
- *vartype*: variable's type. Support types are: int4, float8, date, time and varchar. Type: character varying.
- *nullallowed*: specific if a null value is allowed by the variable. Type: boolean.
- *seqnce*: a sequence number of variables indicates the sequence of variable in a CSV(where data is import) file.
- *automatic*: indicates whether a variable is automatic, creates by the system (e.g., distance\_along) or by user. Type: boolean.

altitude(trail, **parcel** ,distance , altitude):

- *trail*: indicates the trail which a altitude is specified. Type: character varying.
- *parcel*: indicates what parcel a trail is associated. Type: character varying.
- *distance*: indicates the distance along the trail where altitude is specified. Type: integer.
- *altitude*: altitude of a specific location on trail. Type: real.

parcel( longUTM, latUTM, longitude, latitude, **parcel**, geom):

- *longUTM*: indicates the longitude in UTM projection, this value is obtained from longitude after conversion. Type: real.
- *latUTM*: indicates the latitude in UTM projection, this value is obtained from latitude after conversion. Type: real.
- *longitude*: longitude in WGS84.
- *latitude*: latitude in WGS84.
- *parcel*: unique identification of a parcel, name of parcel. Type: character varying.



- *geom*: geometry column with spatial information about a parcel. Type: geometry.
- researcher(**rid** , *firstname* , *surname*, *title*, *email*, *department*, *institute*):
  - *rid*: unique identification of a researcher. Type: integer.
  - *firstname*: first name of the researcher. Type: character varying.
  - *surname*: family name of the researcher. Type: character varying.
  - *title*: academic titles. Type: character varying.
  - *email*: email address. Type: character varying.
  - *department*: department within institute, if any. Type: character varying.
  - *institute*: affiliation of the researcher. Type: character varying.
- reserve(**ridx**, *name*, *acronym*, *geom*):
  - *ridx*: unique identification of a reserve. Type: integer
  - *name*: name of reserve. Type: character varying.
  - *acronym*: short name of reserve. Type: character varying.
  - *geom*: the area of the reserve as a “MULTIPOLYGON”, using WGS84. Type: geometry.
- transect(*name*,reserve,**tridx**):
  - *name*: name of trail or igarape. Type: character varying.
  - *reserve*: identification of a reserve which transect is associated. Type: integer.
  - *tridx*: unique identification of a trail or igarape. Type: integer.
- trail(*geom*): inherits transect:
  - *geom*: the trail as a “LINESTRING”, using WGS84. Type: geometry.
- igarape(*geom*): inherits transect:
  - *geom*: the igarape as a “LINESTRING”, using WGS84. Type: geometry.
- igarape\_survey( *igarape*, *date*, *time*, *bay* , *vm\_sup*, *vm\_flux*, *dossel*, *width*, *vazao*, *prof\_med*, *prof\_max*, *ar* , *ag*, *lf*, *lt*, *mc*, *ra*, *tr*):
- observation(**stridx**, **varname**):
  - *stridx*: unique identification of a strobs<sup>1</sup>. Type: integer.
  - *varname*: name of the variable which this observation represent. Type: character varying.
- observation\_date(*value*):
  - *value*: a dated observation. Type: date.
- observation\_time(*value*):
  - *value*: a timed observation. Type: time.

- `observation_int4(value)`:
  - *value*: a integer observation. Type: integer.
- `observation_float8(value)`:
  - *value*: a real observation. Type: real.
- `observation_varchar(value)`:
  - *value*: a character varying observation. Type: character varying.
- `strops(svy, stridx)`:
  - *svy*: a foreign key of a survey. Type: integer.
  - *stridx*: unique identification of a strob. Type: integer.
- `strops_c(lote)`: inherits strops:
  - *lote*: indicates a collecting strops. Type: character varying.
- `strops_d(datum)`: inherits strops:
  - *datum*: date which strops was collect. Type: date.
- `strops_g(x,y,geom)`: inherits strops:
  - *x*: longitude coordinate of a observation. Type: real.
  - *y*: latitude coordinate of a observation. Type: real.
  - *geom*: the point of observation as “point”, using WGS84.
- `strops_h(date,time)`: inherits strops:
  - *datum*: date when observation was made. Type: date.
  - *time*: time when observation was made. Type: time without time zone.
- `strops_i1(igarape,distance_along,geom)`: inherits strops:
  - *igarape*: identification of the igarape where this strob was collected. Type: character varying.
  - *distance\_along*: distance the igarape, from start, in meters. Type: real.
  - *geom*: location in igarape where a collect was made as a “point”, using WGS84.
- `strops_i2(igarape,distance_along,distance_off,geom)`: inherits strops:
  - *igarape*: identification of the igarape where this strob was collected. Type: character varying.
  - *distance\_along*: distance along the igarape, from start, in meters. Type: real.
  - *distance\_off*: distance off the igarape in meters. Type: real.
  - *geom*: location in igarape where a collect was made as a “point”, using WGS84.
- `strops_t1(trail,distance_along,geom)`: inherits strops:
  - *trail*: identification of the trail where this strob was collected. Type: character

- varying.
  - *distance\_along*: distance the trail, from start, in meters. Type: real.
  - *geom*: location in trail where a collect was made as a “point”, using WGS84.
- *strops\_t2*(trail,distance\_along,distance\_off,geom): inherits *strops*:
  - *trail*: identification of the trail where this strob was collected. Type: character varying.
  - *distance\_along*: distance along the trail, from start, in meters. Type: real.
  - *distance\_off*: distance off the trail in meters. Type: real.
  - *geom*: location in trail where a collect was made as a “point”, using WGS84.
- *strops\_p*(parcel): inherits *strops*:
  - *parcel*: parcel where collect was made. Type: character varying.
- *strops\_r*(res\_resp): inherits *strops*:
  - *res\_resp*: responsible researcher of this collect. Type: integer.

## A2 Vertical Fragmentation

The vertical fragmentation is performed in insert of data. First of all, the researcher needs to personalize a survey, specifying what types of locational and timing which will be used in collects, also can specify other parameters explain above in survey table. After that, the system automatic creates the appropriate variables for each type of survey (e.g., “t1”: *distance\_along*) and insert this variable in *surveyvariable* table setting the field *automatic* to true. Then, researcher can create new variables which is relevant for his research, specifying the name and the type of this new variables which is insert in *surveyvariable* table obeying the sequence number of automatic variables.

At this point, the researcher can obtain from the system a spreadsheet with all variables and fields which he has to collect to complete his research. With this spreadsheet filled with information collect in field (reserve), it is possible easily insert it in database. When a complete survey is insert into the system in a CSV file, the system perform a several steps to fragment the data in this file:

- 1) A function reads the CSV file into a temporary table which has the same schema that the CSV file.
- 2) With a temporary table filled, the system now can create for each row in this table a *strops* which will store a spatial information about the location where the researcher collects it, doing the appropriate calculations for obtain the spatial data.
- 3) At this point, the system can fill observations for each variable which was previous created in the survey specification. So, for each variable listed in *surveyvariable* table that is not automatic (*field automatic* is false) the system reads information from the temporary table for appropriate observation table looking for variable type.
- 4) Here, the process is done, now the temporary table can be delete

to preserve disk space, and each observation has a strob associated with it, so queries can easily be done for a variable value either its location.

## A2.1 Functions and Triggers

To provide the requirements of BID, the system has several functions written in PL/PGSQL procedural language. Above is present all functions and their specifications.

### A2.1.1 Functions

- `bid_insert_survey`(*t* character varying, *inidate* date, *trmdate* date, *rspesq* integer, *collects* boolean, *lockind* character varying, *timkind* character varying, *personal* boolean, *rsrv* integer, *superlev* integer):  
**Description:** creates a new survey specified accords with given parameters.
  - *t*: title of survey.
  - *inidate*: start date of the survey.
  - *trmdate*: end date of the survey.
  - *rspesq*: responsible researcher for survey.
  - *collects*: indicates if survey is a collecting type.
  - *lockind*: indicates the locational kind of the survey, in set: p,t1, t2, i1, i2.
  - *timkind*: indicates the timing kind of the survey, in set: n,d,h;
  - *personal*: indicates if survey is personalized.
  - *rsrv*: indicates what reserve the survey is from.
  - *superlev*: indicates what is the parent survey of this new survey. This value can be null.
  
- `bid_insert_svyvar`(*svynr* integer, *seq* smallint, *vname* character varying, *vtype* character varying, *nulled* boolean):  
**Description:** creates a new variable for a specific survey.
  - *svynr*: survey identification which variable is owned.
  - *seq*: sequence number of the new variable obeying the previous ones that was created.
  - *vname*: name of the new variables.
  - *vtype*: type of the new variables. Support types are: date, time, float4, int4 and varchar.
  - *nulled*: indicates if the new variables can allow null values.
  
- `bid__insert_svyvar`(*svynr* integer, *vname* character varying, *vtype* character varying, *nulled* boolean):  
**Description:** auxiliary function that in fact creates new variable. This function should not be called direct, only called by `bid_insert_levantvar` function.(Note de double '\_' instead of one in `bid_insert_levantvar` function).
  - *svynr*: survey identification which variable is owned.
  - *vname*: name of the new variables.

- *vtype*: type of the new variables. Support types are: date, time, float4, int4 and varchar.
  - *nulled*: indicates if the new variables can allow null values.
- `bid_insert_researcher`(*fname* character varying, *lname* character varying, *ttle* character varying, *eml* character varying, *dept* character varying, *inst* character varying):
 

**Description:** inserts a new researcher .

    - *fname*: first name of researcher.
    - *lname*: last name of researcher.
    - *ttle*: academic title.
    - *eml*: email address.
    - *dept*: department within institute, if any.
    - *inst*: affiliation of researcher.
  - `bid_insert_igarape`(*name* character varying, *res* character varying):
 

**Description:** inserts a new igarape.

    - *name*: name of igarape.
    - *res*: reserve where igarape is localized.
  - `bid_insert_reserve`(*fullname* character varying, *shortname* character varying):
 

**Description:** inserts a new reserve.

    - *fullname*: full name of reserve.
    - *shortname*: a acronym for reserve.
  - `bid_insert_trail`(*name* character varying, *res* character varying):
 

**Description:** inserts a new trail.

    - *name*: trail's name.
    - *res*: reserve where trail is localized.
  - `bid_insert_user`(*fname* character varying, *lname* character varying, *ulogin* character varying, *upassword* character varying):
 

**Description:** inserts a new user of BID.

    - *fname*: first name of user.
    - *lname*: family name of user.
    - *ulogin*: user login name.
    - *upassword*: password for user.
  - `bid_createtable_for_csvfile`(*svynr* integer):
 

Description: creates a new temporary table for accommodate data from CSV file, searching in surveyvariable table for variables.

    - *svynr*: survey which will receive data.
  - `bid_datasheet`(*svynr* integer):
 

**Description:** creates a spreadsheet to be filled with data collect in field.

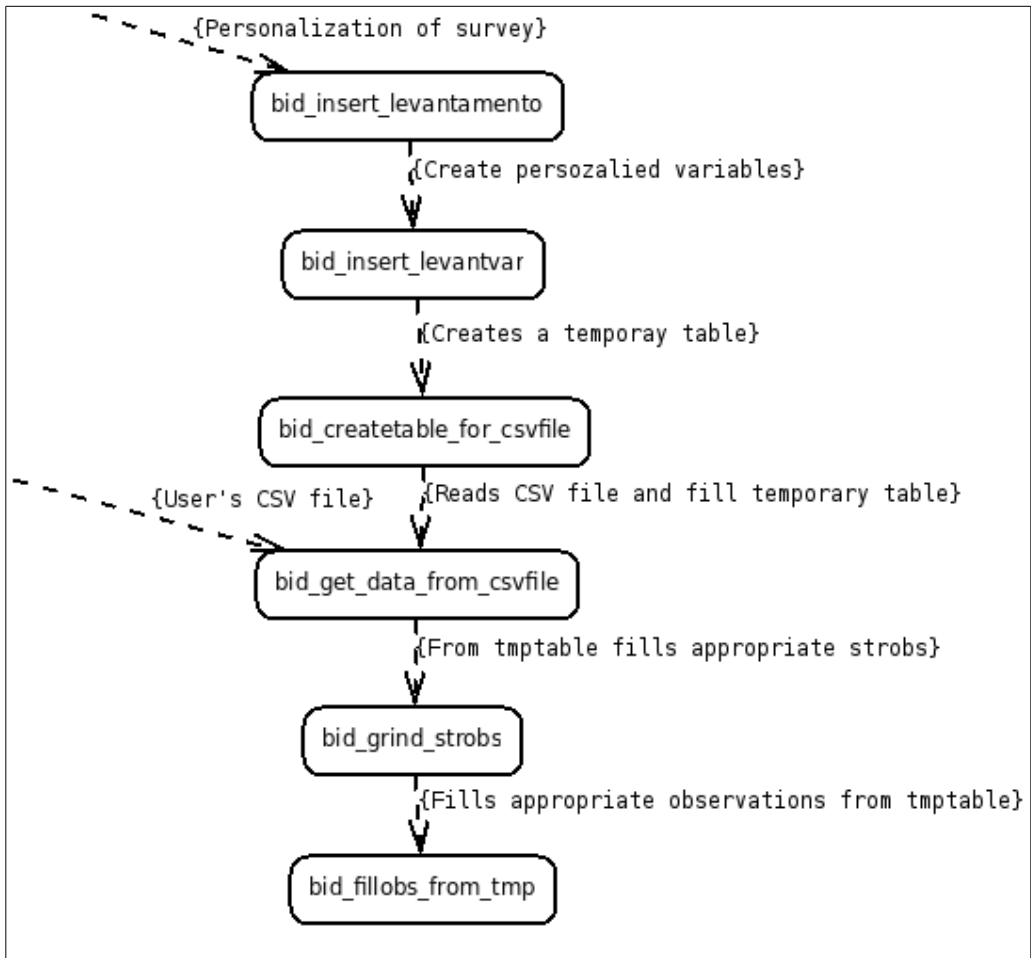
- Svynr: survey which spreadsheet must be generate.
- bid\_empty\_tmptable(svynr integer):  
**Description:** erase data from temporary table of given survey.
  - *svynr*: survey which temporary table will be erase.
- bid\_get\_data\_from\_csvfile(svynr integer, csvpath character varying):  
**Description:** reads CSV file and filled temporary table of given survey.
  - *svynr*: survey which data will be filled.
  - *csvpath*: full path to CSV file.
- bid\_get\_researchers():  
**Description:** return all researchers registered in system.
- bid\_remove\_svyvar(svynr integer, vname character varying):  
**Description:** removes a variable from surveyvariable table.
  - *svynr*: survey which variable is associated.
  - *vname*: name of variable to be removed.
- bid\_remove\_tmptable(svynr integer):  
**Description:** removes the temporary table of given survey.
  - *svynr*: survey which temporary table to be removed.
- bid\_fillobs\_from\_tmp(svynr integer):  
**Description:** fills observations with data stored in temporary table of given survey.
  - *svynr*: survey which temporary table must be read to fill observations.
- bid\_grind\_strobs(svynr integer):  
**Description:** fills strobs with data stored in temporary table of given survey and calculates the spatial location of each strob.
  - *svynr*: survey which strobs must be create.
- bid\_\_insert\_strobs(svynr integer, tablename character varying, subtype character varying, attlist character varying):  
**Description:** auxiliary function which perform insertion into appropriate strobs table, one in set: p, i1, i2, t1, t2, g, r, d, h and c. This function should not be called direct, only called for bid\_insert\_strobs. (Note the double “\_” instead of one in bid\_insert\_strobs function.
  - *svynr*: survey which strobs is owned.
  - *tablename*: name of the temporary table of given survey.
  - *subtype*: subtype of strobs, one in set: p, i1, i2, t1, t2, g, r, d, h and c.
  - *attlist*: list of attributes for each subtype. This is make dynamic for bid\_insert\_strobs function.
- bid\_get\_fullstrobs(svynr integer):

- Description:** gets all observations from given survey.
- *svynr*: survey where data is stored.

### A2.1.2 Triggers

- `strobs_trigger()`:  
**Description:** trigger called before a insert or update in strobs table for types i1, i2, t1, t2, and g. Calculates the geometry information of each strobs and store it in geom column of appropriate strobs table.

### A3 Schematics of BID vertical fragmentation



The picture above shows how BID perform the vertical fragmentation. The flowchart presented has simply informations about the role process, just for better understanding of its functionality.



## **ANEXO 2. LISTA DE LEVANTAMENTOS BIOLÓGICOS DO PROJETO TEMÁTICO GRADIENTE FUNCIONAL, CATALOGADOS NO SERVIDOR METACAT, INSTALADO NO INSTITUTO DE BIOLOGIA, UNICAMP-SP.**

---

### **1. Juliano van Melis**

“Lianas: biomassa em florestas neotropicais e relação riqueza e biomassa em um trecho de Floresta Ombrófila Densa Atlântica.”

### **2. Bruno Henrique Pimentel Rosado**

“O papel de características morfofisiológicas e arquitetura do sistema radicular de espécies arbóreas, nos fluxos de água e nutrientes ao longo de um gradiente altitudinal na Mata Atlântica.”

### **3. Lorena Coutinho Nery da Fonseca**

“Cargas de pólen transportadas por beija-flores em diferentes estratos de vegetação, em Floresta Atlântica, sudeste do Brasil.”

### **4. Ludmila Mickeliunas Pansarin**

“Biologia floral e reprodutiva e anatomia do labelo de *Cyrtopodium polyphyllum* Vell. (Orchidaceae, Cyrtopodiinae).”

### **5. Ludmila Mickeliunas Pansarin**

“Biologia floral e reprodutiva do gênero *Cirrhaea* Lindl. (Orchidaceae) e evolução dos sistemas de polinização em Stanhopeinae.”

### **6. Valéria Forni Martins**

“Chuva de sementes e padrão espacial de espécies arbóreas ornitocóricas, em Floresta Ombrófila Densa de Terras Baixas.”

### **7. Máira de Campos Gorgulho Padgurschi**

“Floresta Ombrófila Densa Montana com Taquaras: caracterização estrutural de um trecho do Núcleo Santa Virgínia/PESM, São Paulo.”

### **8. Sabrina Ribeiro Latansio-Aidar**

“Ecofisiologia do uso de Nitrogênio, em espécies de Floresta Ombrófila Densa Montana e Cerrado *sensu stricto*, SP.”

### **9. André Luis Casarin Rochelle**

“Heterogeneidade ambiental e diversidade de um trecho de Floresta Ombrófila Densa Atlântica, no Parque Estadual da Serra do Mar, em Ubatuba, São Paulo”

### **10. Carolina Bernucci Virillo**

“Dinâmica e estrutura populacional de quatro espécies arbóreas, no Núcleo Picinguaba – Parque Estadual da Serra do Mar, SP.”

### **11. Pedro Ortman Cavalin**

“Atributos funcionais de indivíduos arbóreos jovens, em um gradiente altitudinal na Floresta Ombrófila Densa do litoral do estado de São Paulo”.

### **12. Viviane Camila de Oliveira**

“Estratégias adaptativas, crescimento e assimilação de carbono de seis espécies arbóreas tropicais submetidas à saturação hídrica do solo.”

### **13. Eduardo Magalhaes Borges Prata**

“Composição florística, fitossociologia e distribuição de espécies arbóreas, numa área de Floresta Ombrófila Densa Atlântica em Picinguaba, Ubatuba-SP”

### **14. Aline Lopes e Lima**

“Papel ecológico da absorção foliar de água em *Drimys brasiliensis* Miers. (Winteraceae).”

**15. Christiane Erondina Corrêa**

“Estrutura de populações de *Psychotria nuda* (Rubiaceae), em Floresta Ombrófila Densa.”

**16. Rafael Carvalho da Costa**

“Padrão espacial e dinâmica populacional de duas espécies arbóreas da Mata Atlântica do litoral de São Paulo.”

**17. Vinícius Lourenço Garcia de Brito**

“Biologia da polinização e reprodução de *Tibouchina pulchra* Cogn. (Melastomataceae) em gradiente altitudinal no sudeste do Brasil.”

**18. Luciana F. Alves, Marco A. Assis, Simone A. Vieira, Luiz A. Martinelli, Fernando R. Martins, Carlos A. Joly**

“Variação na abundância e biomassa de lianas ao longo de um gradiente altitudinal de Floresta Ombrófila Densa”

**19. Simone A. Vieira, Luciana F. Alves, Flavio A. M. Santos, Luiz A. Martinelli, Carlos A. Joly**

“Variação do estoque de madeira morta ao longo de um gradiente altitudinal da Floresta Ombrófila Densa”

**20. Luciana F. Alves, Simone A. Vieira, Luiz A. Martinelli, Plínio B. B. Camargo, Flavio A. M. Santos, Carlos A. Joly**

“Dinâmica da comunidade arbórea ao longo de um gradiente altitudinal de Floresta Ombrófila Densa”

**21. Luciana F. Alves, Simone A. Vieira, Luiz A. Martinelli, Plínio B. B. Camargo, Flavio A. M. Santos, Carlos A. Joly**

“Variação da estrutura e biomassa da Floresta Ombrófila Densa ao longo de um gradiente altitudinal”

**22. Ewerton Caltran Manarin**

“Estudo de 2 fatores responsáveis por fixação de Nitrogênio, na Floresta Ombrófila Densa do Estado de São Paulo – Ubatuba: estimativa de fixação biológica de Nitrogênio por cianobactérias e comparação da fixação por leguminosas em diferentes altitudes.”

**23. Adriana Cristina Rosa Saraiva, Simey Thury Vieira Fisch**

“Influência das variações microambientais na comunidade de palmeiras da Floresta Ombrófila Densa Montana, Núcleo Santa Virgínia, Parque Estadual da Serra do Mar – SP”

**24. Simey Thury Vieira Fisch**

“Distribuição da comunidade de palmeiras no gradiente altitudinal da Floresta Ombrófila Densa, dos Núcleos Picinguaba e Santa Virgínia do Parque Estadual da Serra do Mar”

**25. Bianca Netto Rodrigues, Simey Thury Vieira Fisch**

“Análise morfológica e cromossômica de populações da palmeira clonal *Geonoma elegans* Mart, em diferentes altitudes no núcleo Picinguaba, Parque Estadual da Serra do Mar – SP”

**26. Carlos Eduardo Pereira Nunes, Marlies Sazima**

“Biologia da reprodução de *Elleanthus* (Orchidaceae), no Parque Estadual da Serra do Mar, SP.”

**27. Nivea Dias dos Santos**

“Zonação altitudinal de briófitas na Mata Atlântica do sudeste do Brasil”

**28. Larissa Giorgeti Veiga, Carlos Alfredo Joly, Simone Aparecida Vieira, Flavio Antonio Maës dos Santos e Luciana Alves**

“Estoque e produção de madeira morta, em diferentes altitudes de Mata Atlântica, no Parque Estadual da Serra do Mar, SP.”

**29. Gabriel de Castro Vasconcellos Saenz**

“Estratégias de uso de água de três espécies arbóreas, em uma restinga florestal em Ubatuba, SP.”

**30. Andrea Tavares**

“Sistemas de polinização em uma área de Floresta Atlântica submontana do litoral de SP”.