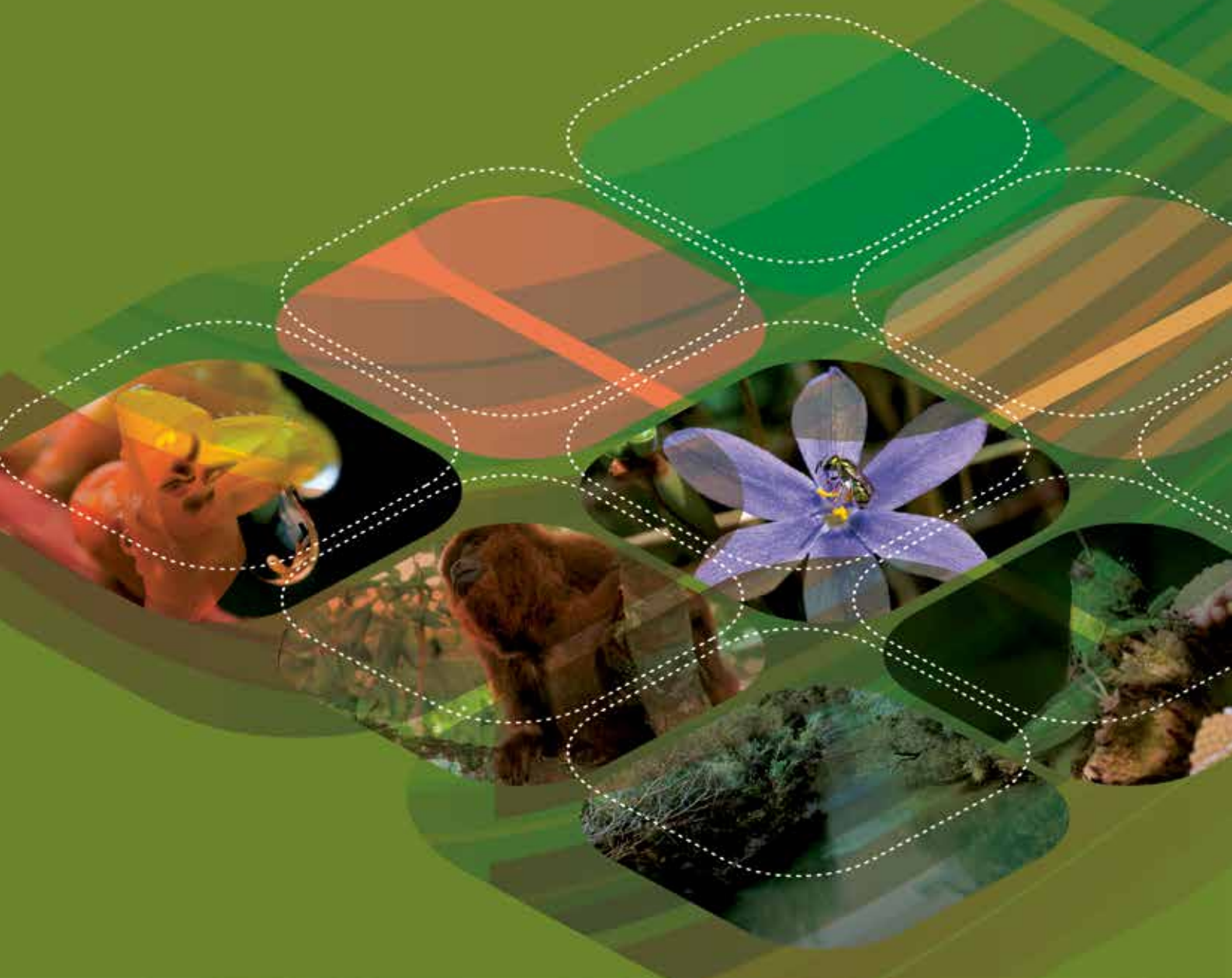


Diretrizes para a Integração de Dados de **Biodiversidade**



REPÚBLICA FEDERATIVA DO BRASIL

Presidenta

Dilma Rousseff

Vice-Presidente

Michel Temer

MINISTÉRIO DO MEIO AMBIENTE

Ministra

Izabella Mônica Teixeira

Secretária de Biodiversidade e Florestas

Ana Cristina Fialho de Barros

INSTITUTO CHICO MENDES DE CONSERVAÇÃO DA BIODIVERSIDADE

Presidente

Cláudio Carrera Maretti

Diretor de Pesquisa, Avaliação e Monitoramento da Biodiversidade

Marcelo Marcelino de Oliveira

Coordenadora Geral de Pesquisa e Monitoramento

Kátia Torres Ribeiro

Coordenadora de Monitoramento da Conservação da Biodiversidade

Tathiana Chaves de Souza

Flona da Restinga de Cabedelo (PB)

Foto: Wigold B. Schäffer

Ministério do Meio Ambiente
Secretaria de Biodiversidade e Florestas

Diretrizes para a Integração de Dados de **Biodiversidade**

MMA
Brasília
2015

INSTITUTO CHICO MENDES DE CONSERVAÇÃO DA BIODIVERSIDADE

Diretoria de Pesquisa, Avaliação e Monitoramento da Biodiversidade Coordenação Geral de Pesquisa e Monitoramento

EQSW 103/104 – Centro Administrativo Setor Sudoeste – Bloco D – 2º andar

Tel: (61) 2028-9090 | fax: (61) 2028-9068

CEP: 70670-350 | Brasília/DF

www.icmbio.gov.br/monitoramento

Catálogo na Fonte

Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis

D597 Diretrizes para a Integração de Dados de Biodiversidade / Daniel Lins da Silva... [et al].
 - Brasília: MMA, 2015.
 100 p.

ISBN 978-85-7738-195-1

1. Guia. 2. Banco de dados (Biodiversidade). 3. Biodiversidade. I. Silva, Daniel Lins da. II. Corrêa, Pedro Luiz Pizzigatti. III. Juarez, Keila Macfadem. IV. Fonseca, Rafael Luís. V. Ministério do Meio Ambiente. VI. Secretaria de Biodiversidade e Florestas. VII. Título.

CDU(2.ed.)504.7(036)

Referência para citar a publicação:

SILVA, D.L.; CORRÊA, P.L.P.; JUAREZ, K.M.; FONSECA, R.L. **Diretrizes para a Integração de Dados de Biodiversidade**. Brasília: MMA, 2015. 100 p.

A reprodução desta obra é permitida desde que citada a fonte. Essa permissão não se aplica às fotos e ilustrações, que foram cedidas exclusivamente para esta publicação.

Esta obra não pode ser comercializada.

Diretrizes para a Integração de Dados de **Biodiversidade**

COORDENAÇÃO EDITORIAL

Keila Macfadem Juarez e Rafael Luís Fonseca.

AUTORES

Daniel Lins da Silva, Pedro Luiz Pizzigatti Corrêa, Keila Macfadem Juarez e Rafael Luís Fonseca.

COLABORAÇÃO NOS TEXTOS E ACOMPANHAMENTO DO ESTUDO

Eduardo Dalcin e Jan Kleine Büning.

REVISÃO GRAMATICAL

Renata Canêdo de Oliveira

PROJETO GRÁFICO

Ângela Ester Magalhães Duarte / AM Design

FICHA CATALOGRÁFICA

Helionídia Oliveira

FOTOS

As fotos publicadas neste livro foram cedidas sem custos pelos autores.

Agradecemos gentilmente aos fotógrafos que as cederam para compor este livro.

CAPA

Composição com fotos de Maurício Mercadante e Wigold B. Schäffer.

REALIZAÇÃO

Esta publicação foi realizada pelo Projeto “Monitoramento da Biodiversidade com Relevância para o Clima em nível de UC, considerando medidas de adaptação e mitigação”. É um projeto do governo brasileiro, coordenado pelo Ministério do Meio Ambiente (MMA) e o Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio), no contexto da Cooperação Brasil-Alemanha, no âmbito da Iniciativa Internacional de Proteção ao Clima (IKI), do Ministério Federal do Meio Ambiente, Proteção da Natureza, Construção e Segurança Nuclear da Alemanha (BMUB). Prevê apoio técnico da Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.



Eugenia puniceifolia (Kunth) DC. MYRTACEAE
Jardim Botânico de Brasília (DF)
Foto: Maurício Mercadante

Sumário

Apresentação.....	9
1. Introdução.....	11
1.1 Objetivos do Livro.....	11
1.2 GT de Integração de Dados do Ministério do Meio Ambiente.....	13
1.3 Referências.....	13
2. Infraestruturas para gestão de dados de biodiversidade.....	15
2.1 Iniciativas para gestão de dados de biodiversidade.....	15
2.1.1 GBIF.....	15
2.1.2 NBII.....	16
2.1.3 DATAONE.....	16
2.1.4 IABIN PTN.....	16
2.2 Sistema de Informação de Biodiversidade do MMA.....	17
2.3 Arquitetura Básica de Interoperabilidade.....	20
2.3.1 Provedor de Dados.....	22
2.3.2 Agregador de Dados.....	22
2.3.3 Publicador de Dados.....	23
2.3.4 Portal Web.....	23
2.3.5 Serviços WEB.....	24
2.4 Definição de Conceitos e Nomenclaturas.....	25
2.5 Metadados.....	26
2.5.1 Importância dos Metadados.....	28
2.5.2 O que é um padrão de metadados?.....	29
2.5.3 Estrutura de um padrão de metadados.....	29
2.5.4 Vocabulários Controlados.....	31
2.5.5 Padrões sugeridos pelo GT de Integração de Dados da MMA.....	32
2.6 Tipos de dados.....	32
2.6.1 Dados Taxonômicos.....	32
2.6.2 Dados de Ocorrências.....	33
2.6.3 Coleções e Provedores de Dados.....	34
2.6.4 Dados Geoespaciais.....	35
2.6.5 Dados Bibliográficos.....	35
2.6.6 Multimídia.....	36
2.7 Referências.....	36

3. Padrões de Dados e Metadados.....	41
3.1 Darwin Core.....	41
3.1.1 Apresentação.....	41
3.1.2 Versões do Darwin Core.....	41
3.1.3 Objetivos.....	42
3.1.4 Detalhamento da Estrutura.....	43
3.1.5 Definição dos termos.....	44
3.1.6 Exemplo de registro Darwin Core.....	45
3.1.7 Simple Darwin Core.....	46
3.1.8 Darwin Core Archive (DwC-A).....	47
3.1.9 Referências.....	50
3.2 Dublin Core.....	51
3.2.1 Apresentação.....	51
3.2.2 Objetivos.....	52
3.2.3 Detalhamento da Estrutura.....	52
3.2.4 Definição dos termos.....	53
3.2.5 Exemplo de registro Dublin Core.....	54
3.2.6 Referências.....	55
3.3 EML.....	56
3.3.1 Apresentação.....	56
3.3.2 Objetivos.....	57
3.3.3 Detalhamento da Estrutura.....	58
3.3.4 Definição dos elementos.....	60
3.3.5 Exemplo de registro EML.....	63
3.3.6 Referências.....	63
3.4 Audubon Core.....	65
3.4.1 Apresentação.....	65
3.4.2 Objetivos.....	65
3.4.3 Detalhamento da Estrutura.....	66
3.4.4 Definição dos termos.....	67
3.4.5 Exemplo de registro Audubon Core.....	68
3.4.6 Referências.....	68
3.5 Plinian Core.....	68
3.5.1 Apresentação.....	68
3.5.2 Objetivos.....	69

3.5.3 Detalhamento da Estrutura.....	69
3.5.4 Definição dos Termos.....	70
3.5.5 Exemplo de registro Plinian Core.....	71
3.5.6 Referências.....	72
3.6 Padrões OGC.....	72
3.6.1 Apresentação.....	72
3.6.2 Objetivos.....	74
3.6.3 Detalhamento da Estrutura.....	74
3.6.4 Referências.....	75

4. Ferramentas e Protocolos para Manipulação de Dados de Biodiversidade.....	77
4.1 TAPIR.....	77
4.1.1 Apresentação.....	77
4.1.2 Funcionamento.....	78
4.1.3 TAPIR Wrappers.....	81
4.1.4 TapirLink.....	82
4.1.5 Referências.....	82
4.2 IPT.....	83
4.2.1 Apresentação.....	83
4.2.2 Funcionamento.....	83
4.2.3 Referências.....	84
4.3 HIT.....	85
4.3.1 Apresentação.....	85
4.3.2 Funcionamento.....	85
4.3.3 Referências.....	87

5. Ferramentas de apoio.....	89
------------------------------	----



Fêmea de Saí-azul alimentando-se de frutos da Magnólia-amarela (*Michelia champaca*).
Foto: Maurício Mercadante



Apresentação

A necessidade de integrar dados e gerar informações de qualidade para rápida tomada de decisão é fundamental para a implementação de estratégias efetivas de conservação da biodiversidade. Diversos temas relacionados a biodiversidade, como por exemplo, definição de áreas prioritárias para conservação ou elaboração de planos de ação para as espécies ameaçadas ou gestão de sistemas de áreas protegidas, necessitam de um grande esforço prévio de compilação de dados e qualificação da informação.

Infelizmente, nem sempre é possível acessar de forma simples e direta os dados e as informações necessárias que dão suporte a esses processos. Embora existam uma quantidade grande de dados e informações sobre a biodiversidade brasileira, em sua maioria, encontram-se espalhadas em diversas instituições e armazenadas em estruturas não padronizadas e em sistemas de informação independentes. Em muitos casos, há redundância de sistemas, falta de qualificação de dados e também de esforços para levantamento e análises de dados primários.

No caso específico de instituições governamentais, as informações precisam ser compartilhadas de forma transparente, seja internamente entre os componentes de sua estrutura, seja externamente com a sociedade. Esse fluxo de informação propicia maior integração entre os processos de gestão, melhoria na tomada de decisão nas estruturas de governança, bem como aumento da qualidade e confiabilidade dos resultados e impactos esperados.

A proposta deste livro é apresentar diretrizes técnicas para armazenamento, compartilhamento e publicação de dados e informações de biodiversidade. É evidente que as tecnologias e ferramentas computacionais são bastantes dinâmicas e se atualizam com grande velocidade. Porém, o conhecimento básico e conceitual sobre infraestrutura de dados, padrões de protocolos e metadados, e ferramentas computacionais aplicadas à gestão de dados, é fundamental para que gestores pensem em melhores práticas e políticas para o armazenamento, compartilhamento e gestão de informações, seja no âmbito de suas organizações, seja no ambiente de redes inteorganizacionais.

A publicação também é resultado de um longo e profundo processo de discussão com gestores e especialistas do tema, realizados pelo Grupo de Trabalho de Integração de Dados do Ministério do Meio Ambiente ao longo dos anos de 2011 até 2013. Todas as recomendações e diretrizes foram referendadas por esse grupo e, certamente, serão de grande valia para todos que trabalham com o uso de dados e informações para as políticas de biodiversidade.

Este livro é um passo importante na direção da integração das bases de dados sobre a biodiversidade brasileira, seja no âmbito do MMA e suas vinculadas, seja entre instituições públicas ou privadas.

Carlos Alberto de Mattos Scaramuzza

Dr. em Ecologia

Diretor do Departamento de Conservação da Biodiversidade

Secretaria de Biodiversidade e Florestas (SBF)

Ministério do Meio Ambiente (MMA)



Dalechampia linearis Baill. EUPHORBIACEAE
Parque Nacional da Chapada dos Veadeiros (GO)
Foto: Maurício Mercadante

1. Introdução

1.1 Objetivos do Livro

O objetivo desta publicação é apresentar um guia com os padrões e protocolos sugeridos pelo Grupo Técnico de Integração de Dados de Biodiversidade do Ministério do Meio Ambiente (GT-MMA) para a implementação de uma arquitetura para a interoperabilidade de dados de biodiversidade. Este guia foi criado a partir da compilação de informações de diversas fontes técnicas, como livros, manuais e sites de organizações tecnicamente relevantes na área da biodiversidade. Entre estas organizações destacam-se o *Global Biodiversity Information Facility* (GBIF)¹, o *Data Observation Network for Earth* (DATAONE)², o *Biodiversity Information Standards* (TDWG)³ e o *Atlas of Living Australia* (ALA)⁴.

As informações foram estruturadas sob a forma de um guia técnico da área de interoperabilidade de sistemas de biodiversidade, incluindo um catálogo de referências para as principais fontes de dados técnicos relacionados aos assuntos abordados.

Este guia pretende atender a diferentes necessidades, conforme o perfil do leitor:

1. Um **gestor de dados de biodiversidade** poderá ler o guia para conhecer o contexto tecnológico que envolve a interoperabilidade de dados de biodiversidade, considerando os **conceitos, padrões e tecnologias** relevantes na área;
2. Um **analista ou desenvolvedor de sistemas**, poderá ler o documento e estender seus estudos às referências e *links* apresentados durante o texto e no final de cada capítulo, obtendo o conhecimento necessário para a implantação de um ou mais componentes que compõem a arquitetura apresentada; e
3. Um **pesquisador** ou **estudante** poderá compreender os principais conceitos, padrões e tecnologias empregadas na área de informática para biodiversidade na atualidade, além de obter informações sobre instituições, projetos e iniciativas existentes. Estas informações definem uma visão geral da área, que poderá apoiar o pesquisador na execução de projetos de pesquisa relacionados ao tema de integração e gestão de dados de biodiversidade.

¹ Portal do GBIF <http://www.gbif.org>

² Portal do DataONE <https://dataone.org>

³ Portal do TDWG <http://www.tdwg.org>

⁴ Portal do Atlas of Living Australia <http://www.ala.org.au>

Este documento está estruturado nos seguintes grandes temas:

Conceitos e iniciativas internacionais de **infraestrutura para gestão de dados de biodiversidade**;

Apresentação de uma **arquitetura padrão para interoperabilidade de dados de biodiversidade** e os **componentes** que a compõem;

Conceitos de **metadados** e definição dos **tipos de dados** geridos pela arquitetura apresentada;

Descrição dos **padrões** utilizados para a interoperabilidade de **dados** e **metadados** de biodiversidade;

Descrição dos **protocolos** e **ferramentas** que apoiam o processo de **compartilhamento e publicação dos dados** de biodiversidade.

1.2 GT de Integração de Dados do Ministério do Meio Ambiente

Esta publicação é uma iniciativa do Grupo Técnico de Integração de Dados de Biodiversidade do Ministério do Meio Ambiente (GT-MMA) e faz parte de um projeto mais amplo que busca a definição de uma arquitetura computacional brasileira para a integração dos diversos provedores e consumidores de dados de biodiversidade do País, sejam eles de instituições ligadas ao Ministério do Meio Ambiente, instituições não-governamentais, universidades e/ou instituições privadas.

No próximo capítulo são apresentadas algumas das principais iniciativas mundiais na gestão de dados de biodiversidade.

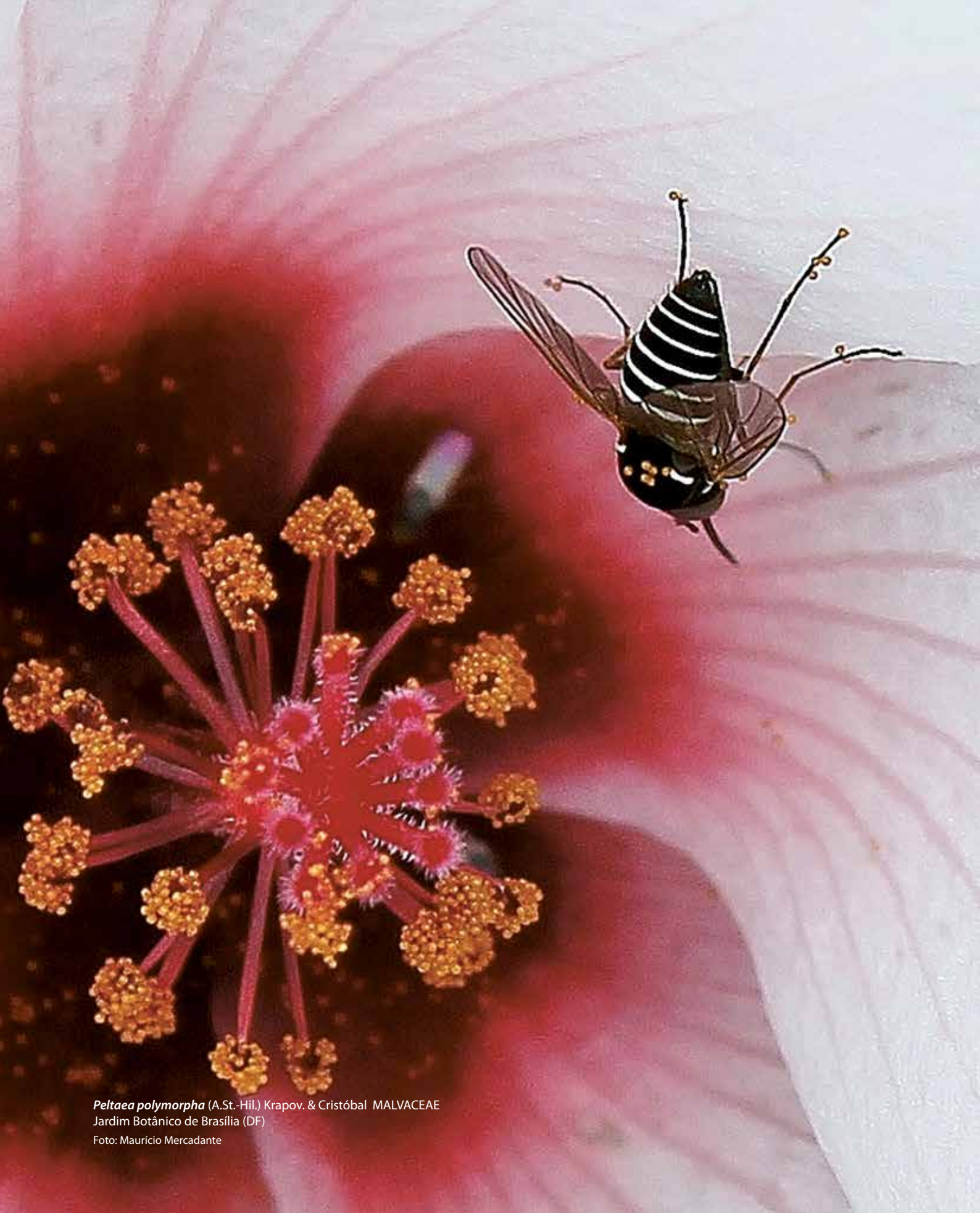
1.3 Referências

ALA (*Atlas of Living Australia*). Disponível em: <<http://www.ala.org.au/>>. Acesso em: 30 jun. 2013.

DATAONE (*Data Observation Network for Earth*). Disponível em: <<https://www.dataone.org/>>. Acesso em: 30 jun. 2013.

GBIF (*Global Biodiversity Information Facility*). Disponível em: <<http://www.gbif.org/>>. Acesso em: 30 jun. 2013.

TDWG (*Biodiversity Information Standards*). Disponível em: <<http://www.tdwg.org/>>. Acesso em: 30 jun. 2013.



Peltaea polymorpha (A.St.-Hil.) Krapov. & Cristóbal MALVACEAE
Jardim Botânico de Brasília (DF)
Foto: Maurício Mercadante

2. Infraestruturas para gestão de dados de biodiversidade

Uma infraestrutura para gestão de dados de biodiversidade, também chamada de rede (*network*), é um conjunto de instituições, sistemas e serviços, que são interligados de forma estruturada para possibilitar a interoperabilidade e o acesso unificado a diversos recursos computacionais distribuídos como nós em uma rede. Estas redes também se apoiam em normas, padrões e protocolos que possibilitam a interoperabilidade dos diversos ambientes, sistemas e seus dados heterogêneos.

2.1 Iniciativas para gestão de dados de biodiversidade

Neste livro são consideradas algumas das principais iniciativas para gestão de dados de biodiversidade:

- GBIF - *Global Biodiversity Information Facility*;
- NBII - *National Biological Information Infrastructure*⁵;
- DATAONE - *Data Observation Network for Earth*;
- IABIN PTN – *Inter American Biodiversity Information Network – Pollinators Thematic Network*⁶.

2.1.1 GBIF

O **GBIF** foi estabelecido como uma iniciativa científica global com o objetivo de promover a utilização eficiente do conhecimento sobre a diversidade biológica do planeta Terra, um dos grandes desafios do século XXI. O GBIF apresenta como sua visão “um mundo onde a informação sobre biodiversidade é gratuita e universalmente disponível para a ciência, para a sociedade e para um futuro sustentável”. A missão do GBIF é se tornar a mais importante fonte global de informações sobre biodiversidade e ser um gerador de soluções inteligentes para o meio ambiente e o bem estar. Para alcançar estes objetivos o GBIF encoraja uma grande variedade de fontes/publicadores de dados de todo o mundo a descobrir, organizar, digitalizar e publicar seus dados por meio de sua rede de biodiversidade.

⁵ Portal do NBII Archive Team <http://www.archive-team.org/index.php?title=NBII>

⁶ Portal de dados da IABIN PTN <http://www.biocomp.org.br/dataportal/>

2.1.2 NBII

O **NBII** foi um programa coordenado pelo setor de *Biological Informatics* da **United States Geological Survey (USGS)**⁷. Seu objetivo principal foi facilitar o acesso aos dados e informações dos provedores de dados biológicos dos Estados Unidos, englobando agências do governo, instituições acadêmicas, organizações não-governamentais e organizações privadas.

O NBII coordenou uma pesquisa na comunidade científica americana sobre sistemas de informação que armazenavam dados biológicos relevantes e que pudessem ser integrados e disponibilizados. Em seguida foram desenvolvidas uma organização metodológica e uma infraestrutura computacional, baseadas em padrões abertos, para a integração e gestão destes dados.

Para o usuário final (pesquisadores, instituições e o público em geral), toda a infraestrutura de integração era transparente, pois o acesso aos dados dos diversos provedores era feito de forma integrada, por meio de um Portal WEB.

O NBII foi finalizado no dia 15 de janeiro de 2012, mas deixou um legado de boas práticas e estratégias para organização, agregação e distribuição de dados biológicos.

2.1.3 DATAONE

O **DATAONE** é um programa patrocinado pela **National Science Foundation (NSF)**⁸ dentro do programa **DataNet**⁹. Seu objetivo principal é o armazenamento de dados ecológicos e ambientais produzidos por cientistas do mundo inteiro. A meta do DATAONE é preservar e dar acesso a dados multiescalas, multidisciplinares e multinacionais.

O DATAONE interliga uma infraestrutura computacional existente para disponibilizar um modelo distribuído, que utiliza tecnologias robustas para gestão dos dados, com foco na preservação de dados de biodiversidade a longo prazo.

2.1.4 IABIN PTN

O **IABIN PTN** é uma rede temática de polinizadores que compõe uma das seis redes temáticas estabelecidas no âmbito da **Inter-American Biodiversity Information Network (IABIN)**¹⁰. Seu objetivo principal é desenvolver um banco de dados interligado e integrado entre as principais fontes de

⁷ Portal do USGS <http://www.usgs.gov>

⁸ Portal do National Science Foundation <http://www.nsf.gov>

⁹ Wikipédia – Datanet <http://en.wikipedia.org/wiki/Datanet>

¹⁰ Portal da IABIN <http://www.iabin.net/>

dados de polinizadores e membros IABIN que compartilham conteúdo crítico relacionado a polinizadores por meio de um conjunto comum de padrões e protocolos de compartilhamento de dados.

O IABIN PTN disponibiliza um catálogo *online* dinâmico e interligado de polinizadores do hemisfério ocidental, que inclui dados sobre (1) nomes (*checklists*) de abelhas, beija-flores, morcegos e outras espécies polinizadores de grande importância; (2) exemplares nas principais coleções; (3) contato dos polinizadores e especialistas; (4) associações polinizadores-plantas; (5) literatura sobre polinizadores e (6) demais informações relacionadas (como, por exemplo, dados geográficos, códigos de barras genéticos etc.).

2.2 Sistema de Informação de Biodiversidade do MMA

O Sistema de Informação para Integração de Dados de Biodiversidade do Ministério do Meio Ambiente é responsável por integrar, compartilhar, publicar e sintetizar dados de biodiversidade gerados e manipulados no âmbito do MMA e em suas instituições vinculadas (**Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (Ibama)**¹¹, **Instituto Chico Mendes de Conservação de Biodiversidade (ICMBio)**¹², **Jardim Botânico do Rio de Janeiro (JBRJ)**¹³, **Agência Nacional de Águas (ANA)**¹⁴ e o **Serviço Florestal Brasileiro (SFB)**¹⁵, de acordo com a **Figura 1**.

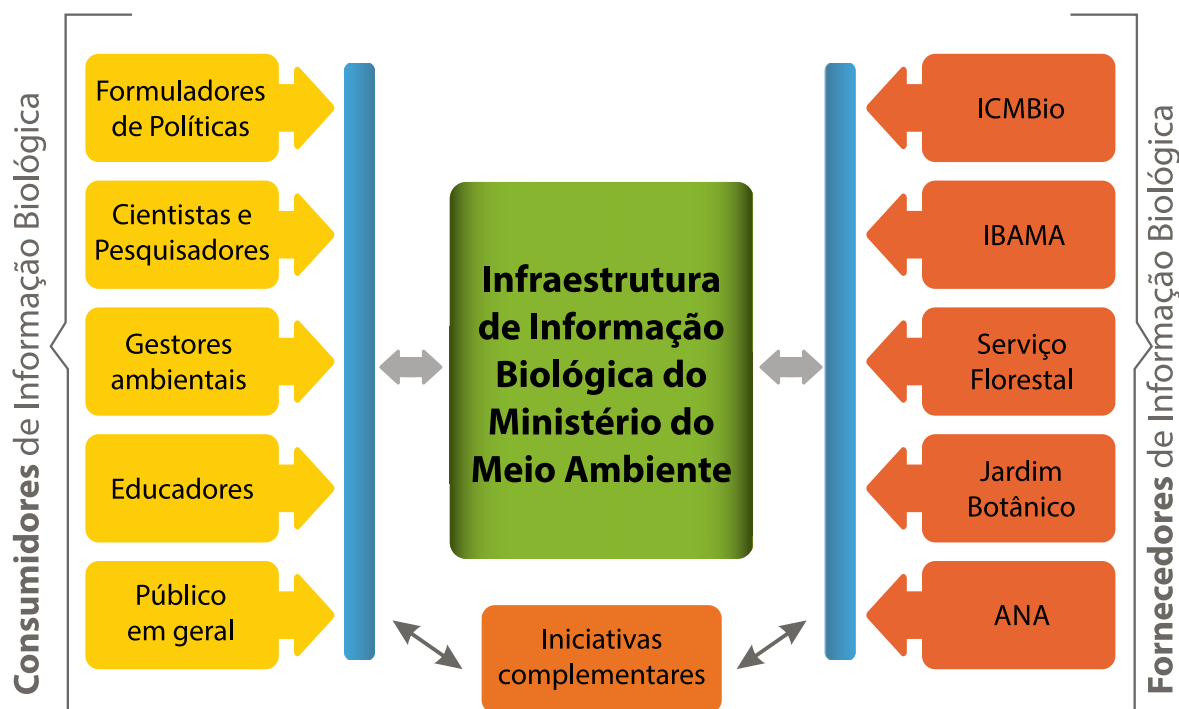


Figura 1 – Visão Geral do Sistema de Informação para Integração de Dados do MMA (Corrêa, 2013).

¹¹ Portal do Ibama <http://www.ibama.gov.br>

¹² Portal do ICMBio <http://www.icmbio.gov.br>

¹³ Portal do JBRJ <http://www.jbrj.gov.br>

¹⁴ Portal da ANA <http://www.ana.gov.br>

¹⁵ Portal do SFB <http://www.florestal.gov.br>

O Sistema de Informação de Biodiversidade do MMA disponibiliza uma infraestrutura de informação integrando fornecedores e consumidores de informação biológica, articulada com sistemas de informação relacionados a iniciativas complementares como o **Global Biodiversity Information Facility (GBIF)**, o **Sistema Brasileiro de Biodiversidade (SIBBr)**¹⁶, a **Infraestrutura Nacional de Dados Espaciais (INDE)**¹⁷, o **Specieslink**¹⁸ e a **SciELO**¹⁹.

A **Figura 2** apresenta a visão da arquitetura de referência e de suas camadas funcionais proposta para o Sistema de Informação de Biodiversidade do MMA.

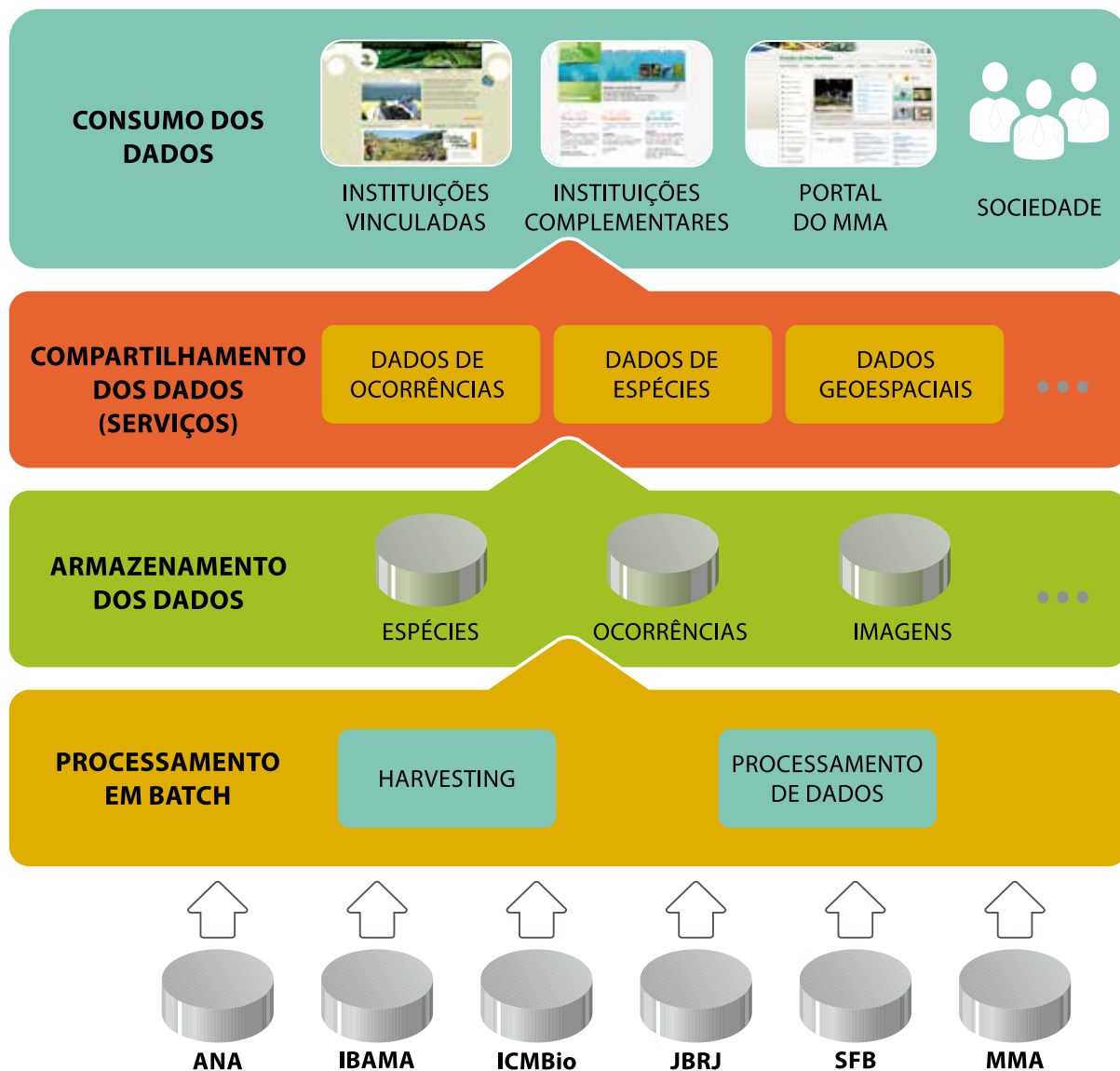


Figura 2 - Camadas funcionais da arquitetura de referência proposta para o Portal de Biodiversidade do MMA (Silva e Corrêa, 2013).

¹⁶ Portal do SIBBr <http://www.sibbr.gov.br>

¹⁷ Portal da INDE <http://www.inde.gov.br>

¹⁸ Portal de dados do SpeciesLink <http://splink.cria.org.br>

¹⁹ Portal do SciELO <http://www.scielo.org>

Os **Consumidores de Informação Biológica** são aqueles que gerenciam, estudam e utilizam dados biológicos, recursos e ferramentas computacionais. Os consumidores podem ser dos setores público e privado: cientistas, formuladores de políticas públicas, analistas ambientais do MMA e outros órgãos federais e estaduais, governos locais, da indústria, professores, estudantes e cidadãos.

Os requisitos dos consumidores de informação biológica são variados e os principais são descritos abaixo:

- **Cientistas e Pesquisadores** precisam ter acesso a dados de alta qualidade e em quantidade para planejar e direcionar suas pesquisas;
- Os gestores de **Órgãos Públicos e Instituições Vinculadas** ao **MMA** precisam ter acesso fácil aos dados biológicos disponíveis, tanto para o planejamento e priorização de políticas públicas voltadas à conservação da biodiversidade, como também para o monitoramento das ações realizadas;
- Para o gerenciamento de áreas protegidas, tais como Unidades de Conservação;
- O **Setor Privado** precisa de informações para que a produção de bens e serviços possa ter impactos reduzidos na biodiversidade. Além disso, a informação sobre a biodiversidade é essencial no processo de licenciamento ambiental;
- **Educadores**, em todos os níveis, precisam de informações atualizadas, além de recursos disponíveis para preparação de material didático adequado para o processo de ensino-aprendizagem de seus alunos;
- O **Público em Geral** precisa de informação **qualificada** e **contextualizada** sobre tendências regionais, locais e globais, abordando os principais impactos gerados pela sociedade e suas consequências para a biodiversidade e para o próprio ser humano;
- A **Sociedade Civil Organizada** (ONGs, Fundações etc.) que utiliza a informação para acompanhar, sugerir e fiscalizar a tomada de decisão e formulação de políticas públicas.

A **Camada de Consumo dos Dados** é representada pelas ferramentas computacionais que consomem os dados disponibilizados pelos serviços da Camada de Compartilhamento. Os portais de dados e as ferramentas de análise, são alguns das aplicações existentes nesta camada.

A **Camada Funcional de Compartilhamento (dos Dados)** é responsável por disponibilizar serviços computacionais para acesso aos dados de biodiversidade. Essa camada é organizada em módulos, conforme o domínio dos dados: dados de espécimes, dados de espécies, dados geoespaciais, referências bibliográficas, etc.

A **Camada de Armazenamento** é responsável pelo armazenamento e gestão dos dados consolidados, que foram previamente integrados e processados por meio da camada de Processamento em Batch.

A **Camada de Processamento em Batch** é responsável pelo *Harvesting* (Coleta) e Indexação dos dados replicados das bases de dados dos sistemas de informação de biodiversidade das instituições vinculadas e de órgãos do MMA (Nós de Informação do MMA).

2.3 Arquitetura Básica de Interoperabilidade

As iniciativas para gestão de dados apresentadas anteriormente possuem arquiteturas computacionais distintas, mas que consideram componentes em comum, existentes nas principais arquiteturas para interoperabilidade de dados de biodiversidade.

Na **Figura 3** é apresentada uma arquitetura padrão para a interoperabilidade de dados de biodiversidade. Em seguida, serão descritos os componentes desta arquitetura e detalhadas suas funções. É importante considerar que estes componentes podem ser implementados de diferentes formas dentro de uma infraestrutura computacional, podendo ser agrupados em um mesmo servidor/aplicação ou distribuídos em diversos servidores dentro de uma arquitetura distribuída.

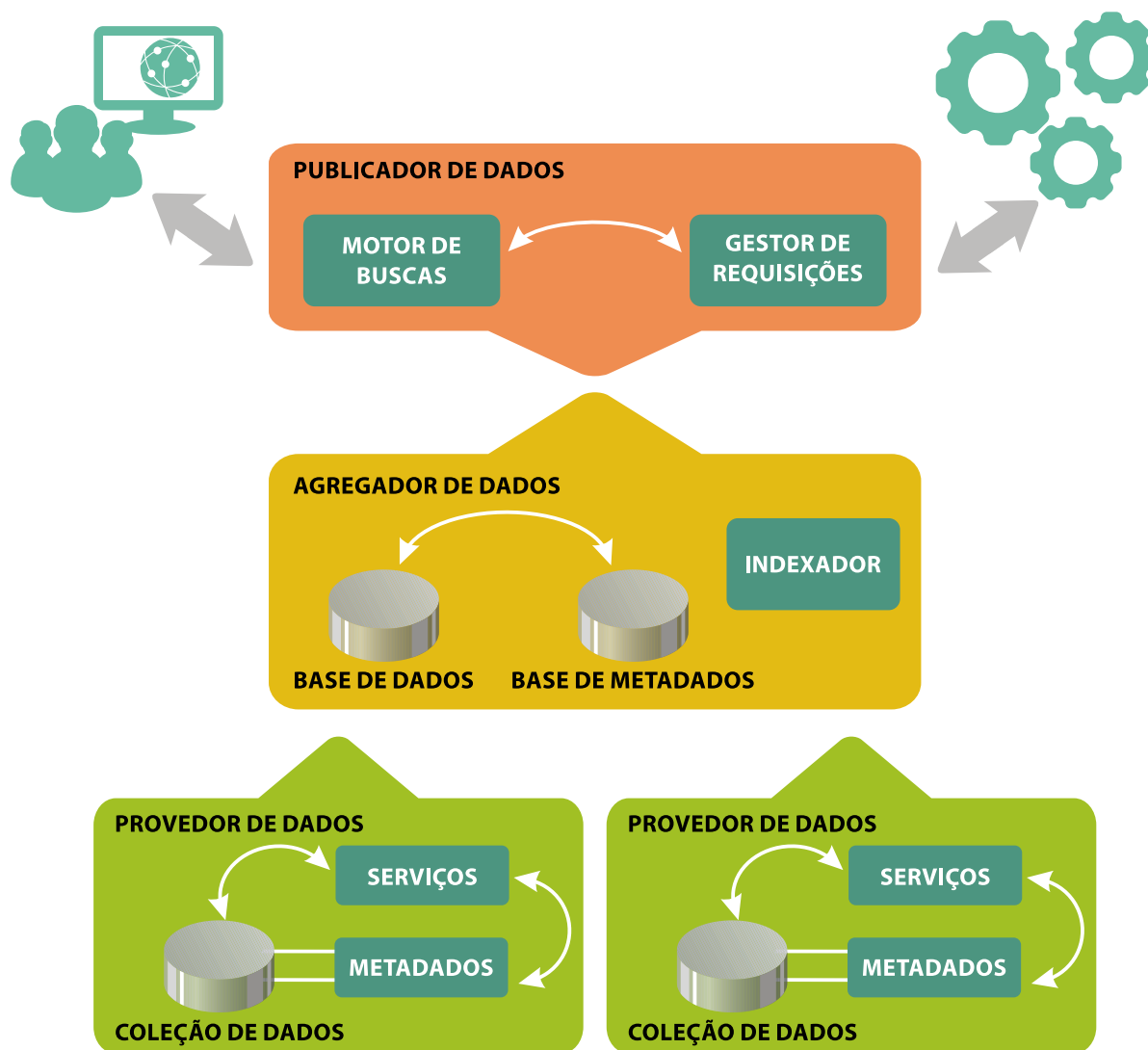


Figura 3 – Arquitetura para interoperabilidade de dados de biodiversidade (Silva, 2013).

2.3.1 Provedor de Dados

Os provedores de dados são os responsáveis por **coletar**, **armazenar** e **disponibilizar**, por meio de interfaces padronizadas, **dados de biodiversidade** e os **metadados** que descrevem estes dados, garantindo a pesquisa e o acesso aos mesmos.

Estes provedores utilizam **formatos comuns** para a troca de dados, com **estruturas padronizadas** que claramente definem como a informação será compartilhada, facilitando assim, o trabalho realizado pelos **agregadores de dados** no agrupamento e indexação das informações de diversos provedores.

A **padronização no compartilhamento dos dados** permite que diversas instituições disponibilizem seus dados de forma organizada e eficiente, independente da tecnologia e da estrutura dos dados existentes nos sistemas e nas suas bases de dados locais.

Como alternativa para a implementação dessas interfaces de compartilhamento, existem **ferramentas computacionais** criadas exclusivamente para essa finalidade. Essas ferramentas mapeiam os dados de um provedor, com base em padrões de metadados, e disponibilizam os dados para outros sistemas por meio de serviços Web ou arquivos com estruturas padronizadas. São exemplos destas ferramentas o **TAPIRLink**²⁰, baseado no protocolo TAPIR, e o *Integrated Publishing Toolkit* (IPT), que serão detalhados no **Capítulo 4**.

2.3.2 Agregador de Dados

Os agregadores de dados são os responsáveis por **gerenciar** o processo de **harvesting (coleta)** dos **dados** e **metadados** provenientes dos **provedores de dados**. Após a coleta dos dados dos diversos provedores, o agregador realiza a organização e indexação destes dados em uma base de dados comum. O mesmo ocorre com os metadados, que serão utilizados principalmente para o processo de descoberta e descrição dos dados.

Os **metadados** não são obrigatoriamente associados a dados digitalizados e/ou disponíveis para consulta *on-line*. Os metadados podem simplesmente identificar a existência de um objeto em um provedor de dados e descrever seus detalhes, possibilitando que o pesquisador interessado realize uma consulta “in loco” na instituição.

²⁰ TDWG Wiki – TapirLink <http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirLink>

2.3.3 Publicador de Dados

Os publicadores de dados são os responsáveis por **publicar** os dados de uma forma **integrada e padronizada**. A publicação é realizada em geral seguindo duas estratégias. A primeira é a utilização de um **Portal Web**²¹, onde os usuários realizam pesquisas nos dados e metadados, com o objetivo de localizar as informações de interesse. Esses portais também disponibilizam ferramentas para a apresentação, análise e *download* dos dados pesquisados.

A segunda estratégia é a disponibilização de **serviços Web**²² para a consulta e recuperação automatizada de informações. Em geral, os mesmos dados disponibilizados pelo Portal Web podem ser acessados via serviços Web, possibilitando que outros sistemas computacionais utilizem esses serviços para o consumo dos dados.

Ferramentas de análise também utilizam os serviços para a aquisição de dados para experimentos. Um exemplo de ferramenta é o **Open Modeller**²³, utilizado na modelagem computacional de nichos ecológicos. A versão 1.1.0 da ferramenta possui *plug-ins* para acesso aos serviços Web disponibilizados pelo Portal de dados do **GBIF** e do **SpeciesLink**²⁴.

O **SpeciesLink** é um sistema distribuído de informação que integra dados primários de coleções científicas. O sistema foi desenvolvido a partir da parceria de diversas instituições nacionais.

2.3.4 Portal Web

Os **Portais Web de Biodiversidade** são sistemas baseados na Internet que possuem como principal objetivo permitir o acesso a uma ou mais bases de dados de biodiversidade. Sua principal aplicação é a disponibilização de dados de ocorrências a partir de coleções de espécimes e observações. Existem também Portais Web de biodiversidade que disponibilizam informações taxonômicas, como lista de nomes (*checklists*), e aplicações de mapas, que utilizam dados geográficos de ocorrências juntamente com informações ambientais para a análise dessas informações.

Alguns exemplos de Portais Web de biodiversidade, e seus respectivos endereços para acesso, são citados a seguir:

- GBIF Data Portal: <http://www.gbif.org>
- IABIN PTN: <http://www.biocomp.org.br/dataportal>
- Integrated Taxonomic Information Systems (ITIS)²⁵: <http://www.itis.gov>

²¹ Wikipédia – Portal Web [http://pt.wikipedia.org/wiki/Portal_\(internet\)](http://pt.wikipedia.org/wiki/Portal_(internet))

²² Wikipédia – Serviços WEB http://pt.wikipedia.org/wiki/Web_service

²³ Portal do Open Modeller <http://openmodeller.sourceforge.net>

²⁴ Portal do SpeciesLink <http://splink.cria.org.br>

- Atlas of Living Australia (ALA): <http://www.ala.org.au>
- Sistema de Información sobre Biodiversidad de Colombia (SiB Colombia)²⁶: <http://www.sibcolombia.net/web/sib>
- Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO)²⁷: <http://www.conabio.gob.mx/informacion/gis>
- SpeciesLink Portal: <http://www.splink.org.br/index?lang=pt>
- Lista da Flora Brasileira: <http://floradobrasil.jbrj.gov.br>
- Portal de Serviços de Geoprocessamento do ICMBio: <http://www.icmbio.gov.br/portal/servicos/geoprocessamento.html>
- Sistema de Informação sobre a Biodiversidade Brasileira (SiBBR): <http://www.sibbr.gov.br/>

2.3.5 Serviços WEB

Os **serviços WEB** são oferecidos para possibilitar o consumo dos dados de biodiversidade de forma automatizada por outros sistemas. Conforme os objetivos do publicador de dados, os serviços podem oferecer recursos tanto para a validação de dados taxonômicos, quanto para a consulta de ocorrências e dados geográficos. Alguns sistemas também disponibilizam serviços com foco em qualidade e análise de dados.

Para possibilitar uma maior abrangência e facilidade na utilização destes serviços, é recomendada a utilização de padrões e boas práticas na sua implementação. Também é recomendada a publicação na WEB de um manual para descrever e apoiar o processo de utilização dos serviços disponibilizados.

Alguns exemplos de sistemas de biodiversidade que disponibilizam serviços para a consulta de dados são citados a seguir, juntamente com os endereços de acesso aos seus manuais:

- GBIF: <http://data.gbif.org/ws/rest/network>
- Integrated Taxonomic Information Systems (ITIS): http://www.itis.gov/ws_description.html
- Atlas of Living Australia (ALA): <http://www.ala.org.au/about-the-atlas/downloadable-tools/web-services>

²⁵ Portal do ITIS <http://www.itis.gov>

²⁶ Portal do SiB Colombia <http://www.sibcolombia.net/web/sib>

²⁷ Portal do CONABIO <http://www.conabio.gob.mx>

2.4 Definição de Conceitos e Nomenclaturas

Considerando o escopo de uma arquitetura para interoperabilidade de dados de biodiversidade, é importante a definição dos seguintes conceitos e nomenclaturas (**Figura 4**):



Figura 4 - Representação das coleções e seus metadados em um provedor de dados de biodiversidade.

- Os dados de biodiversidade publicados por meio de uma rede de interoperabilidade são organizados em **Coleções** (*Datasets*) provenientes de **Provedores de Dados** (*Data Providers* ou *Data Sources*).
- Uma Coleção é um **conjunto de Registros de Dados** sobre um determinado assunto.
- As Coleções são descritas a partir de **Metadados**, que disponibilizam informações sobre a origem, as características e o propósito destes dados.
- Um **registro de dados** é um **conjunto de termos**, também chamados de elementos, propriedades ou atributos. Um registro de dados pode descrever, por exemplo, uma observação de um espécime. Um dos termos mais comuns em um registro de dados é o **"nome científico"** (*scientific Name*).
- Um **termo** contém o valor da informação. Um exemplo de valor do termo "nome científico" poderia ser "*Panthera onca*".
- **Dados primários** (do inglês *primary data* ou *primary biodiversity data*) são considerados os registros de dados textuais ou multimídia que detalham fatos sobre uma instância (ocorrência) de um organismo (o que, onde, quando, como e por quem).

- **Dados secundários** são dados sintetizados ou interpretativos, gerados normalmente com base em grande quantidade de dados primários. Estes dados são apresentados normalmente em formas de mapas, registros e publicações.

2.5 Metadados

Metadados são “dados que descrevem dados” e são considerados essenciais em um sistema de gerenciamento e compartilhamento de informações, descrevendo vários aspectos das coleções e seus dados:

- **Quem** criou estes dados?
- **Qual** é o conteúdo dos dados?
- **Quando** os dados foram gerados?
- **Onde** os dados foram gerados?
- **Como** os dados foram gerados?
- **Porque** os dados foram gerados?

O registro dos metadados preservam a utilidade dos dados ao longo do tempo, detalhando os métodos de coleta dos dados e a criação de bancos de dados. Os metadados minimizam a duplicação de esforços na coleta dos dados e facilita o compartilhamento de dados digitais e o rastreamento da informação.

No contexto dos sistemas de interoperabilidade de dados de biodiversidade, várias questões podem ser respondidas durante o compartilhamento de dados por meio dos seus metadados:

- Quais processos foram utilizados na criação dos dados?
- Quando foi a última atualização dos dados?
- Qual a escala foi utilizada para a geração de um Mapa?
- Quem pode ser contatado para a obtenção de mais informações sobre o dado publicado?
- Quais são as limitações dos dados?

A digitalização dos metadados é fundamental no processo de preparação dos dados para a sua publicação em uma rede de biodiversidade, conforme demonstrado na **Figura 5**.



Figura 5 - Publicação de dados, com seus respectivos metadados, em uma rede de interoperabilidade de dados de biodiversidade.

2.5.1 Importância dos Metadados

Os metadados possuem múltiplas finalidades e são importantes para os **geradores dos dados**, para os **consumidores dos dados** e também para as **organizações envolvidas**.

Para os **geradores dos dados** os registros de metadados ajudam a evitar a duplicação de esforços, pois os pesquisadores podem determinar se dados semelhantes já existem antes da realização de um projeto de coleta de dados. Os pesquisadores podem compartilhar informações confiáveis, identificando a origem, a qualidade dos dados e seu método de coleta. Além disso, os metadados garantem a divulgação dos geradores das informações, agregando valor ao trabalho realizado e reconhecimento aos profissionais responsáveis pelos dados.

Para os **consumidores dos dados** estes metadados permitem a pesquisa e recuperação das informações provenientes dos provedores. As buscas realizadas nos metadados possibilitam resultados mais ricos e detalhados, por meio da utilização de restrições temporais, geográficas e organizacionais durante as buscas. Permitem também que os consumidores possam decidir pela utilização ou não destes dados em uma aplicação específica por meio de critérios como qualidade, metodologia de coleta e responsáveis.

Para as **organizações envolvidas**, a manutenção destes metadados contribui para a garantia dos seus investimentos na gestão dos dados, por meio da retenção de informações relevantes sobre como os dados foram coletados, processados e qualificados, criando um registro permanente da história destes dados. Com isso, as informações deixam de depender de pessoas e passam a gerar conhecimento na organização, garantindo sua reutilização em projetos futuros, por novos pesquisadores e parceiros de projetos.

Os metadados também podem descrever dados em diferentes níveis de um recurso digital. Embora geralmente as descrições sejam realizadas de forma individual (uma foto, um arquivo de vídeo, um registro de observação de espécie), também podem ser realizadas a partir da agregação de recursos (*datasets* de registros de ocorrência, listas de verificação de espécies).

Além da utilização em diferentes níveis, os metadados podem descrever recursos em diferentes camadas (fases). Por exemplo, na **Tabela 1**, a pintura original da Mona Lisa é descrita com metadados, em seguida são descritas a reprodução da obra de arte em um slide e a sua representação digital (arquivo digital).

O desafio na criação dos metadados é decidir quais níveis e camadas precisam ser descritos, a quantidade de detalhes em cada uma delas e como os metadados serão organizados.

Tabela 1 - Diferentes camadas de informações identificadas pelos Metadados.

			010010100101 010010100101 001010101010 100001001010 101010101001 010100101010 011001010101
	Imagem Original	Imagem em Slides	Imagem Digital
	Leonardo da Vinci	Jane Smith [Fotógrafo]	John Brown [Técnico de Digitalização]
	Pintura	Transparência fotográfica	Imagem .JPEG
	Museu de Louvre	Coleção de Slide ABC	A:\images\0023.jpg
...	...	...	...

Fonte: JISC Digital Media²⁸.

2.5.2 O que é um padrão de metadados?

Um **padrão de metadados** prevê **elementos comuns** que são utilizados de forma universal, permitindo que o compartilhamento destas informações se torne mais prático e seguro. As **normas estabelecidas** também garantem a definição de **regras consistentes** e **vocabulários padronizados**, que possibilitam uma comunicação eficiente entre os geradores e consumidores de dados. A estrutura padronizada permite também a pesquisa unificada nestes dados.

2.5.3 Estrutura de um padrão de metadados

Um **esquema de metadados** ou **estrutura de metadados** é formado por um conjunto de elementos projetados para uma finalidade específica, como a descrição de um tipo particular de informação. A definição ou significado dos elementos é chamada de **semântica do esquema**. Já os valores inseridos em cada elemento são chamados de **conteúdo**.

²⁸ JISC Digital Media <http://www.jiscdigitalmedia.ac.uk/guide/an-introduction-to-metadata>

Os esquemas de metadados normalmente definem os nomes dos elementos e sua semântica. Opcionalmente são definidas regras de como o conteúdo deve ser identificado, descrito e os respectivos valores permitidos para o conteúdo.

Também podem ser definidas no esquema as regras de sintaxe de como os elementos e seu conteúdo podem ser codificados. Um esquema de metadados sem regras de sintaxe é chamado de independente. Os metadados podem ser codificados em qualquer sintaxe definida, como **SGML**²⁹(*Standard Generalized Markup Language*), **XML**³⁰(*Extensible Markup Language*), entre outras.

A **Figura 6** apresenta uma representação dos componentes de um esquema do padrão **Darwin Core**, descrito com mais detalhes no **Capítulo 3.1**. O esquema de metadados **Darwin Core** possui vários grupos que organizam os elementos dentro do esquema. O grupo “Record-level Terms” possui o elemento “dwc:basisOfRecord”. Por sua vez, este elemento é associado a um vocabulário controlado que define quais informações são permitidas no seu conteúdo.

Um **esquema de metadados** é projetado para o **compartilhamento de informações de biodiversidade** e não para a gestão destes dados. Por isso, os **bancos de dados** dos sistemas computacionais de biodiversidade e as **planilhas** para registro destes dados devem possuir, primeiramente, uma estrutura baseada nas necessidades dos projetos e da instituição, e não simplesmente uma cópia de um esquema de metadados. Com o conhecimento dos padrões de metadados existentes, é obtida uma **visão mais ampla e padronizada** de importantes elementos a serem considerados durante a definição de um modelo de dados de biodiversidade e como os mesmos podem ser compartilhados mais tarde.



Figura 6 - Representação estrutural de um elemento do Darwin Core e seu vocabulário controlado.

²⁹ Wikipédia – SGML <http://pt.wikipedia.org/wiki/SGML>

³⁰ Wikipédia – XML <http://pt.wikipedia.org/wiki/XML>

2.5.4 Vocabulários Controlados

Os vocabulários controlados são uma maneira de organizar o conhecimento para posterior recuperação. Eles são usados em esquemas de indexação de assuntos, cabeçalhos de registros, thesaurus, ontologias e em outros sistemas de organização do conhecimento. Esquemas de vocabulário controlado obrigam a utilização de termos autorizados, que foram pré-selecionados pelo criador do vocabulário. Ao contrário do vocabulário de linguagem natural, onde não há restrição ao vocabulário utilizado. Algumas das vantagens na utilização destes vocabulários são:

- Consistência;
- Acurácia;
- Simplificação na entrada dos dados;
- Qualidade;
- Interoperabilidade;
- Facilidade na procura e descoberta dos dados.

Um vocabulário controlado pode ser incorporado facilmente em processos automatizados. Em um sistema computacional, um vocabulário pode simplificar a entrada e contribuir para a qualidade dos dados, fornecendo aos usuários ou outros sistemas a lista de entradas permitidas para os elementos dos metadados, além de ser usado para a validação de dados importados de outros sistemas.

Na **Tabela 2** é apresentado o vocabulário controlado do elemento **basisOfRecord** (<http://rs.tdwg.org/dwc/terms/basisOfRecord>) do **Darwin Core**. Este elemento é responsável por especificar a natureza do registro de dado armazenado. Informações adicionais sobre este vocabulário podem ser encontradas no documento **“Darwin Core Type Vocabulary”**³¹.

Tabela 2 – Termos recomendados no vocabulário controlado do elemento **BasisOfRecord** do **Darwin Core**.

Termo	Definição
Occurrence	Um recurso que descreve uma instância da Classe Ocorrência.
Event	Uma ocorrência não persistente, com base no tempo. Para Darwin Core, um recurso que descreve uma instância da Classe Evento.
Location	Um recurso que descreve uma instância da Classe Localização.
Taxon	Um recurso que descreve uma instância da Classe Taxonomia.
PreservedSpecimen	Um recurso que descreve um Espécime preservado.
FossilSpecimen	Um recurso que descreve um Espécime fossilizado.
LivingSpecimen	Um recurso que descreve um Espécime vivo.
HumanObservation	Um recurso que descreve uma observação realizada por uma ou mais pessoas.
MachineObservation	Um recurso que descreve uma observação realizada por uma máquina.
NomenclaturalChecklist	Um recurso que descreve uma lista de nomes.

³¹ Darwin Core Type Vocabulary <http://rs.tdwg.org/dwc/terms/type-vocabulary/index.htm>

2.5.5 Padrões sugeridos pelo GT de Integração de Dados da MMA

Os padrões sugeridos pelo GT-MMA e considerados neste livro são:

- Darwin Core;
- Dublin Core;
- EML;
- Audubon Core;
- Plinian Core;
- Padrões OGC (OpenGIS).

Esses padrões estão apresentados e detalhados no **Capítulo 3**.

2.6 Tipos de dados

Os principais tipos de dados considerados em uma infraestrutura computacional para biodiversidade são os **dados taxonômicos das espécies, ocorrências, provedores de dados e bancos de dados, dados geoespaciais, dados bibliográficos e arquivos multimídia**.

2.6.1 Dados Taxonômicos

São dados que se referem a informações de um táxon e não de uma instância específica (ocorrência) de um indivíduo dentro desse táxon. O **táxon** é uma unidade taxonômica, essencialmente associada a um sistema de **classificação científica**³². O táxon pode estar em qualquer nível de um sistema de classificação.

Um exemplo de coleção de dados taxonômicos seria um **checklist de espécies de aves**, onde cada registro de dados disponibiliza a informação sobre uma única espécie. Na **Figura 7** é apresentado um exemplo de táxon (*Oryzoborus angolensis*) disponibilizado pela **Encyclopedia of Life (EOL)**³³:

³² Wikipédia – Classificação Científica http://pt.wikipedia.org/wiki/Classifica%C3%A7%C3%A3o_cient%C3%ADfica

³³ Portal da Encyclopedia of Life <http://eol.org>



Taxon recognized by

[Avibase - IOC World Bird Names \(2011\):](#)

[Animalia](#) ±

[Chordata](#) ±

[Aves](#) ±

[Passeriformes](#) ±

[Thaupidae](#) ±

[Oryzoborus](#) ±

Oryzoborus angolensis

[Oryzoborus atriostris](#)

[Oryzoborus crassirostris](#)

[Oryzoborus funereus](#)

[Oryzoborus maximiliani](#)

[Oryzoborus nuttingi](#)

Figura 7 - Dados do Táxon *Oryzoborus angolensis* (<http://eol.org/pages/1050362/overview>).

Um dos padrões utilizado na definição destas informações é o **Darwin Core (DwC)**³⁴, que está detalhado no **Capítulo 3.1** deste livro.

2.6.2 Dados de Ocorrências

Os dados de ocorrências são relacionados a uma **instância** específica de um **táxon**, na natureza ou em uma coleção de uma instituição. Estes dados determinam a presença ou ausência de espécies no espaço e no tempo. Um exemplo destes dados seria uma **coleção de observações de aves** que descrevem detalhes sobre os avistamentos de diversas espécies de aves. Outro exemplo seria o registro de uma **exsicata** armazenada em um herbário.

Um dos padrões utilizado para definir estas informações é o **Darwin Core (DwC)**.

A **ocorrência de espécies biológicas** em termos **espaciais** e **temporais** são os dados fundamentais para os serviços de análise destes dados.

³⁴ Portal do Darwin Core Group <http://www.tdwg.org/activities/darwincore>

³⁵ Portal do Missouri Botanical Garden <http://www.missouribotanicalgarden.org>

Na **Figura 8** é apresentado um registro de ocorrência disponibilizado pelo *Missouri Botanical Garden*³⁵ e publicado a partir do *Data Portal* do GBIF.

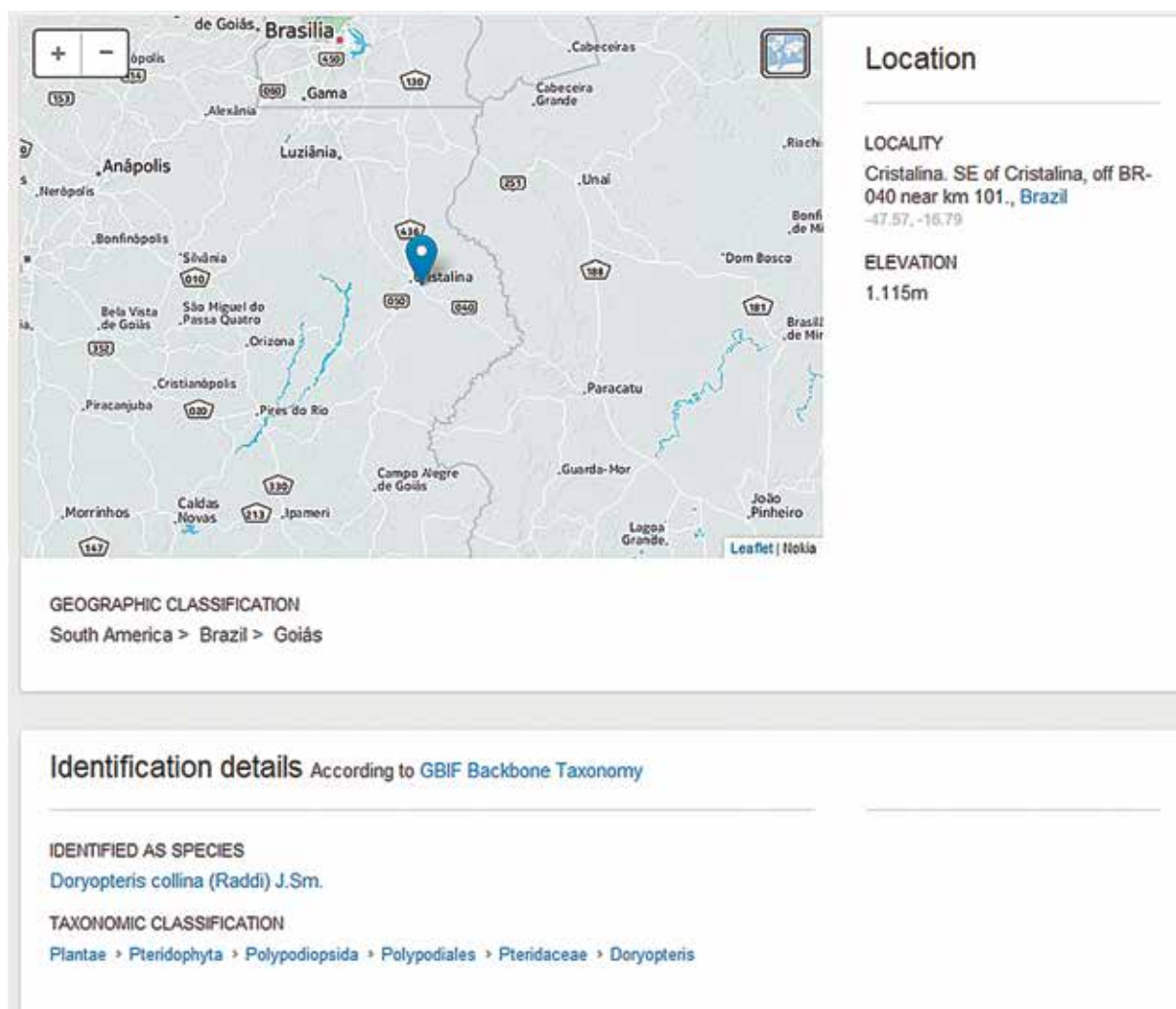


Figura 8 - Registro de dados de uma ocorrência consultada no GBIF Data Portal.

2.6.3 Coleções e Provedores de Dados

As informações das **coleções** e **provedores** de dados de biodiversidade são compartilhadas a partir dos seus **metadados**. Estes metadados descrevem informações sobre as coleções e provedores, e como acessá-las. Também são disponibilizadas informações sobre os dados, como e por quem foram coletados, seus responsáveis, o seu uso indicado, entre outras informações.

Existem diversos padrões de metadados para a descrição de coleções e provedores de dados de biodiversidade. O GBIF utiliza o padrão **GBIF Metadata Profile (GMP)**³⁶, e sua utilização é detalhada no guia “*GBIF Metadata Profile – How-to-Guide*”³⁷. O padrão de metadados sugerido pelo GT-MMA para a disponibilização destes dados é o **Ecological Metadata Language (EML)**³⁸, que será detalhado posteriormente no **Capítulo 3.3**.

2.6.4 Dados Geoespaciais

Os **dados geográficos, geoespaciais** ou **georeferenciados** são dados em que a dimensão espacial refere-se ao seu posicionamento no planeta Terra e no seu espaço próximo, num determinado instante ou período de tempo.

Os **dados espaciais** são quaisquer tipos de dados que descrevem fenômenos que estejam associados a alguma dimensão espacial. A medida observada de um fenômeno ou ocorrência sobre ou sob a superfície terrestre é o que se denomina **dado geográfico**.

Os dados geográficos possuem características espaciais, não espaciais e temporais. As características espaciais informam a posição geográfica do fenômeno e sua geometria. As características não espaciais descrevem o fenômeno, já as características temporais informam o tempo de validade dos dados geográficos e suas variações sobre o tempo.

Os dados geográficos estão cada vez mais presentes na área da biodiversidade, e grande parte das análises realizadas nesta área consideram estas informações.

Os padrões para a descrição, disponibilização e consumo de dados espaciais sugerido pelo GT-MMA são os padrões definidos pela **Open Geospatial Consortium (OGC)**³⁹, que serão apresentados no **Capítulo 3.5**.

2.6.5 Dados Bibliográficos

Em uma arquitetura de dados de biodiversidade os **dados bibliográficos** contemplam as informações sobre **publicações (livros, manuais, artigos científicos e documentos em geral)**, digitais ou não digitais. Estas publicações podem ser relacionados a dados taxonômicos ou ocorrências de espécies, possibilitando a localização destes dados para sua utilização em análises científicas e demais estudos.

³⁶ The GBIF Metadata Profile <http://rs.gbif.org/schema/eml-gbif-profile>

³⁷ GBIF Metadata Profile – http://links.gbif.org/gbif_metadata_profile_how-to_en_v1

³⁸ EML Specification <http://knb.ecoinformatics.org/software/eml>

³⁹ Portal da OGC <http://www.opengeospatial.org/standards/is>

Estes tipos de dados podem ser descritos por meio de um padrão de metadados criado especificamente para a descrição de fontes de informação chamado de **Dublin Core**⁴⁰. Este padrão será detalhado posteriormente no **Capítulo 3.2**.

2.6.6 Multimídia

As imagens e demais recursos multimídia (sons, desenhos, vídeos, etc.) estão frequentemente presentes em sistemas de biodiversidade, seja no registro de observações ou em exsicatas e amostras de coleções biológicas. Para possibilitar a descoberta destes recursos e seu relacionamento com os demais dados de biodiversidade, é necessária a utilização de metadados para descrição destes objetos. Os padrões mais utilizados para esta finalidade são o **Dublin Core**, o **Darwin Core**, a partir de uma de suas extensões, e o **Audubon Core**⁴¹, que será detalhado no **Capítulo 3.4**. Este último é um padrão criado a partir de uma parceria entre o **GBIF** e o **Biodiversity Information Standards (TDWG)**⁴², onde foi definido um conjunto de termos para representar recursos multimídia de biodiversidade.

2.7 Referências

ANA (Agência Nacional de Águas). Disponível em: <<http://www2.ana.gov.br/Paginas/default.aspx>>. Acesso em: 30 jun. 2013.

JISC DIGITAL MEDIA. **An Introduction to Metadata**. Disponível em: <<http://www.jiscdigitalmedia.ac.uk/guide/an-introduction-to-metadata/>>. Acesso em: 30 jun. 2013.

AUDUBON Core Term List. **DRAFT of 1.0 normative - GBIF-Terms**. Disponível em: <[http://terms.gbif.org/wiki/Audubon_Core_Term_List_\(1.0_normative\)](http://terms.gbif.org/wiki/Audubon_Core_Term_List_(1.0_normative))>. Acesso em: 30 jun. 2013.

WIKIPÉDIA (Enciclopédia livre). **Classificação científica**. Disponível em: <http://pt.wikipedia.org/wiki/Classifica%C3%A7%C3%A3o_cient%C3%ADfica>. Acesso em: 30 jun. 2013.

CONABIO (Comissão Nacional da Biodiversidade). Disponível em: <<http://www.conabio.gob.mx/>>. Acesso em: 30 jun. 2013.

DARWIN Core Type Vocabulary. Disponível em: <<http://rs.tdwg.org/dwc/terms/type-vocabulary/index.htm>>. Acesso em: 30 jun. 2013.

⁴⁰ Portal do DCMI <http://dublincore.org>

⁴¹ Audubon Core Term List http://terms.gbif.org/wiki/Audubon_Core_Term_List

⁴² Portal do TDWG <http://www.tdwg.org>

CORRÊA, P. L. P. **Arquitetura para Integração de Sistemas de Informação de Biodiversidade**: Assessoria do Ministério do Meio Ambiente na Integração de Sistemas de Informação em Biodiversidade no âmbito do PROBIO II. Brasília: MMA, maio. 2013.

WIKIPEDIA (Encyclopedia free). **Datanet**. Disponível em: <<http://en.wikipedia.org/wiki/Datanet>>. Acesso em: 30 jun. 2013.

DCMI (Dublin Core Metadata Initiative). Disponível em: <<http://dublincore.org/>>. Acesso em: 30 jun. 2013.

EML (Ecological Metadata Language). Disponível em: <<http://knb.ecoinformatics.org/software/eml/>>. Acesso em: 30 jun. 2013.

EML (Ecological Metadata Language). Disponível em: <<http://knb.ecoinformatics.org/software/eml/>>. Acesso em: 30 jun. 2013.

ENCYCLOPEDIA OF LIFE. **Animals - Plants - Pictures & Information**. Disponível em: <<http://eol.org/>>. Acesso em: 30 jun. 2013.

FRAME, M. **DataONE Science Librarian Assessment**. In: Problems in Information Sciences. University of Tennessee - Information Science, 2013.

GBIF Resources (*Global Biodiversity Information Facility*). Disponível em: <<http://rs.gbif.org/schema/eml-gbif-profile/>>. Acesso em: 30 jun. 2013.

IABIN PTN (Inter American Biodiversity Information Network – *Pollinators Thematic Network Pollinators Thematic Network*). Disponível em: <<http://www.biocomp.org.br/dataportal/>>. Acesso em: 30 jun. 2013.

IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis). Disponível em: <<http://www.ibama.gov.br/>>. Acesso em: 30 jun. 2013.

INDE (**Infraestrutura Nacional de Dados Espaciais**). Disponível em: <<http://www.inde.gov.br/>>. Acesso em: 30 jun. 2013.

ICMBIO (Instituto Chico Mendes de Conservação da Biodiversidade). Disponível em: <<http://www.icmbio.gov.br/portal/>>. Acesso em: 30 jun. 2013.

JBRJ (Instituto de Pesquisas Jardim Botânico do Rio de Janeiro). Disponível em: <<http://www.jbrj.gov.br/>>. Acesso em: 30 jun. 2013.

ITIS (Integrated Taxonomic Information System). Disponível em: <http://www.itis.gov/>. Acesso em: 30 jun. 2013.

IABIN (Inter-American Biodiversity Information Network). Disponível em: <<http://www.oas.org/en/sedi/dsd/iabin/>>. Acesso em: 30 jun. 2013.

MISSOURI Botanical Garden. Disponível em: <<http://www.missouribotanicalgarden.org/>>. Acesso em: 30 jun. 2013.

NBII- ARQUIVETEAM (*National Biological Information Infrastructure*). **Archiveteam**. Disponível em:<<http://www.archiveteam.org/index.php?title=NBI>>. Acesso em: 30 jun. 2013.

NSF (National Science Foundation). Disponível em: <<http://www.nsf.gov/#1>>. Acesso em: 30 jun. 2013.

Ó TUAMA, E.; BRAAK, K.; REMSEN, D. **GBIF Metadata Profile, How-to guide**. Copenhagen: Global Biodiversity Information Facility, 2011.

OGC (**Open Geospatial Consortium**). Disponível em: <<http://www.opengeospatial.org/standards/is>>. Acesso em: 30 jun. 2013.

OPEN Modeller. Disponível em: <http://openmodeller.sourceforge.net/>. Acesso em: 30 jun. 2013.

WICIPÉDIA (Enciclopédia livre). Disponível em:<[http://pt.wikipedia.org/wiki/Portal_\(internet\)](http://pt.wikipedia.org/wiki/Portal_(internet))>. Acesso em: 30 jun. 2013.

SIB COLOMBIA (Sistema de Información Sobre Biodiversidad de Colombia). Disponível em: <<http://www.sibcolombia.net/web/sib/>>. Acesso em: 30 jun. 2013.

REMSSEN, D.; DORING, M.; ROBERTSON, T. **Getting Started, Overview of data publishing in the GBIF Network**. Copenhagen: Global Biodiversity Information Facility, 2011.

SCIELO (Scientific Electronic Library Online). Disponível em: <<http://www.scielo.org/php/index.php>>. Acesso em: 30 jun. 2013.

SFB (Serviço Florestal Brasileiro). Disponível em: <<http://www.florestal.gov.br/>>. Acesso em: 30 jun. 2013.

SIBBR (Sistema de Informação sobre a Biodiversidade Brasileira). Disponível em: <<http://www.sibbr.gov.br/>>. Acesso em: 30 jun. 2013.

WIKIPÉDIA (Enciclopédia livre). SGML. Disponível em: <<http://pt.wikipedia.org/wiki/SGML>>. Acesso em: 30 jun. 2013.

SPECIESLINK. **Sistema de Informação Distribuído para Coleções Biológicas**. Disponível em: <<http://splink.cria.org.br/>>. Acesso em: 30 jun. 2013.

TAPIRLINK. TWiki. Disponível em: <<http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirLink>>. Acesso em: 30 jun. 2013.

TDWG (*Taxonomic Databases Working Group*). DarwinCore. Disponível em: <<http://www.tdwg.org/activities/darwincore/>>. Acesso em: 30 jun. 2013.

TDWG (*Taxonomic Databases Working Group*). Disponível em: <<http://www.tdwg.org/>>. Acesso em: 30 jun. 2013.

WIKIPÉDIA (Enciclopédia livre). **Web service**. Disponível em: <http://pt.wikipedia.org/wiki/Web_service>. Acesso em: 30 jun. 2013.

USGS/US (United States Geological Survey Geological Survey). Disponível em: <http://www.usgs.gov/>. Acesso em: 30 jun. 2013.

WIKIPÉDIA (Enciclopédia livre). **XML**. Disponível em: <<http://pt.wikipedia.org/wiki/XML>>. Acesso em: 30 jun. 2013.



Reserva Biológica de Poço das Antas (RJ)
Foto: Wigold B. Schäffer

3. Padrões de Dados e Metadados

Neste capítulo serão descritos os padrões de metadados considerados na arquitetura computacional definida pelo GT de interoperabilidade do Ministério do Meio Ambiente.

3.1 Darwin Core

3.1.1 Apresentação

O **Darwin Core** é um conjunto de normas definido pelo *Biodiversity Information Standards* ou *Taxonomic Databases Working Group* (TDWG)⁴³, baseado nas definições do *Dublin Core Metadata Initiative* (DCMI)⁴⁴. A especificação do padrão inclui um **glossário de termos**, também chamados de elementos, atributos ou conceitos, destinados a facilitar o compartilhamento de informações sobre a biodiversidade, fornecendo definições de referência, exemplos e comentários sobre os mesmos.

O Darwin Core é baseado principalmente em **dados taxonômicos** e em sua **ocorrência na natureza**, documentada por meio de **observações de espécimes e amostras**. Nas normas do padrão são incluídos documentos que descrevem como estes termos podem ser utilizados, gerenciados e como um conjunto destes termos pode ser estendido para novas necessidades.

3.1.2 Versões do Darwin Core

O Darwin Core foi criado originalmente como um perfil Z39.50 pelo ZBIG (*Z39.50 Biology Implementers Group*), financiado por fundos do *USA National Science Foundation Award*. O nome Darwin Core foi definido por Allen Allison durante a primeira reunião do ZBIG, realizada na Universidade de Kansas em 1998, ao comentar a semelhança conceitual com o perfil Dublin Core. O perfil Darwin Core, mais tarde, foi definido como um documento de esquema XML para uso com o protocolo **Distributed Generic Information Retrieval (DiGIR)**⁴⁵. Posteriormente, o grupo de trabalho TDWG foi criado para revisar esse padrão, e uma ratificação do Darwin Core foi publicado oficialmente em 9 de outubro de 2009 (Darwin Core Terms).

⁴³ Taxonomic Databases Working Group (TDWG) <http://www.tdwg.org/activities/darwincore>

⁴⁴ Dublin Core Metadata Initiative (DCMI) <http://dublincore.org/documents/dcml-terms>

⁴⁵ Distributed Generic Information Retrieval (DiGIR) <http://digir.net>

Apesar de publicado como um padrão do TDWG, o Darwin Core possuiu várias versões anteriores, conforme descrito na **Tabela 3**.

Tabela 3 - Versões publicadas do Padrão Darwin Core.

Nome	Número de termos	Data de Publicação
Darwin Core 1.0	24	1998
Darwin Core 1.2 (Classic)	46	11/Nov/2001
Darwin Core 1.21 (MaNIS/HerpNet /ORNIS/FisNet2)	63	15/Mar/2003
Darwin Core OBIS	27	10/Jul/2005
Darwin Core 1.4 (Draft Standard)	45	10/Jul/2005
Darwin Core Terms (properties)	172	09/Out/2009

Mais informações sobre a história do Darwin Core podem ser consultadas na página do padrão na Wikipédia: http://en.wikipedia.org/wiki/Darwin_Core.

Para participar das discussões e do processo de evolução do padrão ou para se manter informado sobre as atualizações do Darwin Core, acesse o **Portal do projeto Darwin Core**⁴⁶.

3.1.3 Objetivos

O padrão Darwin Core (também abreviado como DwC) é um formato padrão para a codificação de informações relacionadas a espécies. Este formato de dados foi criado para facilitar o compartilhamento de informações sobre a ocorrência espaço-temporal de espécies e a existência de espécimes em coleções físicas e digitais. Atualmente este padrão se tornou bastante amplo e versátil. Extensões para o padrão Darwin Core fornecem um mecanismo para o compartilhamento de informações adicionais, podendo considerar dados de domínios específicos ou apenas complementares às informações já existentes no núcleo do padrão.

O DwC possui um glossário de termos que podem ser usados para descrever a maioria das informações comuns sobre um espécime. Por meio de definições semânticas estáveis, este formato tem como objetivo fornecer um alto grau de reutilização das informações em diferentes contextos.

Os termos descritos no Darwin Core são parte de um conjunto maior de vocabulários e especificações técnicas em desenvolvimento e são mantidas pelo *Biodiversity Information Standards* (TDWG)⁴⁷.

⁴⁶ Portal do projeto Darwin Core <https://code.google.com/p/darwincore>

⁴⁷ Biodiversity Information Standards (TDWG) <http://www.tdwg.org/standards>

O **Darwin Core** é utilizado em geral para o compartilhamento de **dados de espécies (taxonômicos)** e de suas **ocorrências na natureza (observações e amostras)**. Com a evolução de sua estrutura o mesmo passou a ser utilizado também como base para a definição de outros formatos para compartilhamento de dados.

3.1.4 Detalhamento da Estrutura

O Darwin Core possui um conjunto de termos de semântica bem definida, que podem ser compreendidos por pessoas ou interpretados por computadores, tornando possível determinar o uso apropriado dos dados nele codificados. Os termos são organizados em nove categorias, que correspondem aos termos definidos como *Classes* (termos que possuem outros termos para detalhamento). Destas categorias, seis abrangem aspectos mais amplos (evento, localização, contexto geológico, ocorrência, táxon e identificação) no domínio da biodiversidade. As demais categorias abrangem relações com outros recursos, medidas e informações genéricas sobre os registros.

Especialmente na categoria *Record-Level Terms*, o Darwin Core recomenda o uso de diversos termos originários do padrão Dublin Core (*type, modified, language, rights, rightsHolder, accessRights, bibliographicCitation, references*). Os nomes destes termos iniciam com *dcterms*, que identificam que os mesmos são controlados pelo DCMI. Na **Tabela 4** são apresentadas as categorias do Darwin Core e as informações gerais contidas em cada uma delas.

Tabela 4 - Categorias do Darwin Core: O *Simple Darwin Core* considera apenas as sete primeiras categorias.

Record-level Terms	Termos do Dublin Core, instituições, coleções, natureza do registro de dados.	Darwin Core Simple (Simple DwC)
Occurrence	Evidência de espécies na natureza, observações, comportamentos, mídias associadas, referências.	
Event	Protocolos e métodos de amostragem, data, hora, notas de campo.	
Location	Geografia, descrições de localização, dados espaciais.	
Identification	Ligação entre <i>Taxon</i> e <i>Occurrence</i> .	
Taxon	Nomes científicos, nomes populares, nomes utilizados, conceitos do taxon e o relacionamento entre eles.	
GeologicalContext	Tempo geológico, cronoestratigrafia, bioestratigrafia, litoestratigrafia.	Darwin Core Genérico
ResourceRelationship	Relacionamento explícito entre os recursos identificados (exemplo: um organismo para o outro, taxon e localização).	
MeasurementOrFact	Medições, fatos, características, afirmações e referências.	

Embora o Darwin Core seja insuficiente para as necessidades biológicas de todas as disciplinas, ele pode ser adaptado para servir a novos propósitos. O padrão pode ser estendido pela adição de novos termos para o compartilhamento de informações. Para fazer isso, é necessário estar familiarizado com as recomendações e procedimentos definidos na **Política de Nomes do Darwin Core**⁴⁸. Basicamente, antes da proposta de um novo termo, deve-se considerar os termos existentes no Darwin Core e em outros padrões compatíveis, para se determinar se este novo conceito poderá ser acomodado em um termo já existente, por meio de uma simples revisão da descrição e dos comentários, sem a perda do significado original do termo.

Um dos princípios fundamentais deste padrão é manter a definição dos termos **independente da tecnologia utilizada para compartilhamento**, por exemplo, XML, TXT ou RDF.

3.1.5 Definição dos termos

Em sua documentação, os termos do Darwin Core são descritos geralmente pelos seguintes atributos:

- **Nome do termo:** O nome pelo qual a versão mais recente do termo é conhecido;
- **Identificador:** O identificador exclusivo para a versão mais recente do termo;
- **Classe:** A categoria principal de termos (*Class*) que este termo descreve;
- **Definição:** Uma descrição da natureza do conteúdo do termo;
- **Comentário:** Comentários adicionais sobre o uso do termo, com exemplos e referências para mais informações;
- **Detalhes:** Um *link* para as informações completas sobre o termo, incluindo o acesso ao seu histórico.

A **Tabela 5** apresenta a descrição do termo *catalogNumber*:

Tabela 5 - Descrição do Termo *catalogNumber* no “Darwin Core Terms: A quick reference guide”.

Term Name: <i>catalogNumber</i>	
Identifier	http://rs.tdwg.org/dwc/terms/catalogNumber
Class	http://rs.tdwg.org/dwc/terms/Occurrence
Definition	An identifier (preferably unique) for the record within the data set or collection.
Comment	Examples: “2008.1334”, “145732a”, “145732”. For discussion see http://code.google.com/p/darwincore/wiki/Occurrence
Details	<i>catalogNumber</i>

⁴⁸ Política de Nomes do Darwin Core <http://rs.tdwg.org/dwc/terms/namespace>

O conjunto completo dos termos atuais do Darwin Core com suas descrições e demais informações estão disponíveis no Guia “**Darwin Core Terms: A quick reference guide**”⁴⁹, publicado pelo TDWG. Outro importante documento é o “**Darwin Core Terms: A complete historical record**”⁵⁰, que inclui a descrição de todos os termos, o histórico de mudanças entre as versões do Darwin Core e seu relacionamento com outros termos e padrões.

3.1.6 Exemplo de registro Darwin Core

Na **Listagem 1** é apresentado o trecho de um arquivo Darwin Core que descreve o registro de uma observação de ave.

```

1 <?xml version="1.0"?>
2 <dwc:DarwinRecordSet
3   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4   xsi:schemaLocation="http://rs.tdwg.org/dwc/dwcrecord/ http://rs.tdwg.org/dwc/xsd/tdwg_dwc_classes.xsd"
5   xmlns:dcterms="http://purl.org/dc/terms/"
6   xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
7   xmlns:dwcr="http://rs.tdwg.org/dwc/dwcrecord/"
8   <dcterms:Location>
9     <dwc:locationID>http://guid.mvz.org/sites/arg/127</dwc:locationID>
10    <dwc:country>Argentina</dwc:country>
11    <dwc:countryCode>AR</dwc:countryCode>
12    <dwc:stateProvince>Neuquén</dwc:stateProvince>
13    <dwc:locality>Valle Limay, Estancia Rincon Grande, 48 ha area with centroid at this point</dwc:locality>
14    <dwc:decimalLatitude>-40.97467</dwc:decimalLatitude>
15    <dwc:decimalLongitude>-71.0734</dwc:decimalLongitude>
16    <dwc:geodeticDatum>epsg:4326</dwc:geodeticDatum>
17    <dwc:coordinateUncertaintyInMeters>200</dwc:coordinateUncertaintyInMeters>
18  </dcterms:Location>
19  <dwc:Event>
20    <dwc:eventID>http://guid.mvz.org/events/2006/11/26/17</dwc:eventID>
21    <dwc:samplingProtocol>area count</dwc:samplingProtocol>
22    <dwc:eventDate>2006-11-26</dwc:eventDate>
23    <dwc:locationID>http://guid.mvz.org/sites/arg/127</dwc:locationID>
24  </dwc:Event>
25  <dwc:Occurrence>
26    <dwc:occurrenceID>urn:catalog:AUDCLO:EBIRD:OBS64515288</dwc:occurrenceID>
27    <dcterms:type>Event</dcterms:type>
28    <dcterms:modified>2009-02-17T07:33:04Z</dcterms:modified>
29    <dcterms:rightsHolder>Avian Knowledge Network</dcterms:rightsHolder>
30    <dcterms:rights>Creative Commons License</dcterms:rights>
31    <dwc:institutionCode>AUDCLO</dwc:institutionCode>
32    <dwc:collectionCode>EBIRD</dwc:collectionCode>
33    <dwc:basisOfRecord>HumanObservation</dwc:basisOfRecord>
34    <dwc:individualCount>2</dwc:individualCount>
35    <dwc:eventID>http://guid.mvz.org/events/2006/11/26/17</dwc:eventID>

```

Listagem 1 - Trecho de um registro Darwin Core codificado em XML.

⁴⁹ Darwin Core Terms: A quick reference guide <http://rs.tdwg.org/dwc/terms/index.htm>

⁵⁰ Darwin Core Terms: A complete historical record <http://rs.tdwg.org/dwc/terms/history/index.htm>

3.1.7 Simple Darwin Core

O **Simple Darwin Core**⁵¹ ou Darwin Core Simplificado é um subconjunto predefinido dos termos que possuem uso comum em uma ampla variedade de aplicações de biodiversidade. Os termos utilizados neste esquema são encontrados nas classes do Darwin Core de nomes taxonômicos, lugares e eventos, e possibilitam o registro de ocorrências biológicas. Os dois princípios deste esquema são a **simplicidade** e a **flexibilidade**.

Sua simplicidade decorre da necessidade de uma estrutura plana (*flat*), utilizando apenas linhas e colunas. Desta forma os nomes dos termos podem ser considerados como os nomes dos campos e assim um registro do Simple Darwin Core pode ser definido em uma planilha ou em uma única tabela em um banco de dados.

O Simple Darwin Core não possui campos obrigatórios, o que possibilita o compartilhamento de qualquer combinação significativa de campos. Por exemplo, apenas nomes ou lugares, observações individuais realizadas na natureza em um determinado local e tempo. Essa flexibilidade promove a reutilização dos termos e mecanismos de compartilhamento por uma ampla variedade de serviços.

Na **Listagem 2** é apresentado um trecho do registro de ocorrência de um espécime preservado utilizando o padrão Simple Darwin Core.

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <SimpleDarwinRecordSet
3      xmlns="http://rs.tdwg.org/dwc/dwcrecord/"
4      xmlns:dc="http://purl.org/dc/terms/"
5      xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
6      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
7      xsi:schemaLocation="http://rs.tdwg.org/dwc/dwcrecord/ ../../xsd/tdwg_dwc_simple.xsd">
8      <SimpleDarwinRecord>
9          <dc:type>PhysicalObject</dc:type>
10         <dc:modified>2009-02-12T14:22Z</dc:modified>
11         <dc:language>en</dc:language>
12         <dc:rights>Content licensed under Creative Commons Attribution 3.0 United States License</dc:rights>
13         <dc:rightsHolder>Museum of Vertebrate Zoology, UC Berkeley</dc:rightsHolder>
14         <dc:accessRights>open</dc:accessRights>
15         <dc:bibliographicCitation xsi:nil="true"/>
16         <dc:institutionID xsi:nil="true"/>
17         <dc:collectionID>urn:lsid:biocol.org:col:34818</dc:collectionID>
18         <dc:datasetID xsi:nil="true"/>
19         <dc:institutionCode>MVZ</dc:institutionCode>
20         <dc:collectionCode>Mammal</dc:collectionCode>
21         <dc:datasetName xsi:nil="true"/>
22         <dc:ownerInstitutionCode xsi:nil="true"/>
23         <dc:basisOfRecord>PreservedSpecimen</dc:basisOfRecord>

```

Listagem 2 - Trecho de um registro Simple Darwin Core em XML.

⁵¹ Padrão Simple Darwin Core <http://rs.tdwg.org/dwc/terms/simple/index.htm>

As regras e demais informações necessárias para a utilização do Simple Darwin Core podem ser consultadas no documento de referência do padrão.

3.1.8 Darwin Core Archive (DwC-A)

O **Darwin Core Archive (DwC-A)** é um padrão para o compartilhamento de dados de biodiversidade que faz uso dos termos do Darwin Core para produzir um conjunto de registros de **ocor-rências** e/ou de **dados taxonômicos**. Por meio da utilização de extensões outras informações podem ser agregadas ao esquema. Estas extensões são descritas no **GBIF Resources**⁵².

O DwC-A é composto por um conjunto de arquivos texto, com os dados separados por vírgulas ou tabulação, juntamente com um arquivo descritor em XML (meta.xml), para informar como os arquivos texto são organizados, e um arquivo de metadados em EML (eml.xml), para descrever o conjunto de dados publicado. Este conjunto de arquivos são compactados em um único arquivo (*Archive*). Este formato é o preferido pelo GBIF para a publicação de dados na rede GBIF Network.

A ideia central de um *Archive* é que os arquivos com os dados estejam logicamente organizados em um modelo estrela ou *star-schema* (**Figura 9**), onde um arquivo principal (núcleo) é rodeado por diversos arquivos de extensões. Cada arquivo de extensão (ou linha do arquivo de extensão) aponta para um registro do arquivo de núcleo. Desta forma, muitos registros de extensão podem existir para cada registro do núcleo.

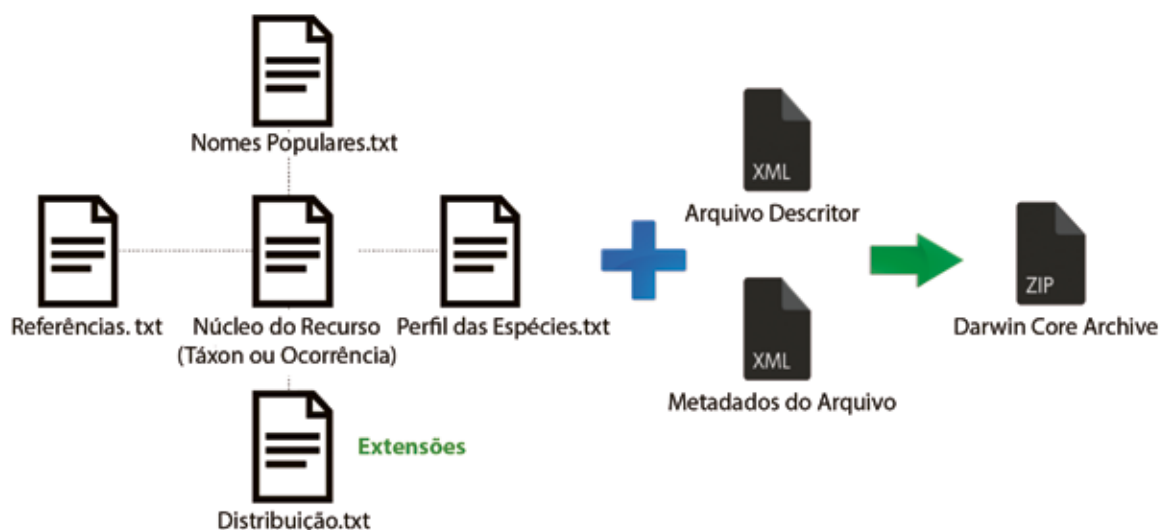


Figura 9 - Componentes de um arquivo Darwin Core Archive (DwC-A).

⁵² Portal do GBIF Resources <http://rs.gbif.org>

Compartilhar um ou mais conjuntos de dados completos em um *archive*, em substituição a utilização de serviços Web que utilizam arquivos XML, permite uma transferência de dados mais eficiente, pois os arquivos são transferidos em arquivos texto compactados, o que diminui o tamanho dos dados transferidos.

Lembrando que tanto as técnicas que utilizam os arquivos XML quanto as que utilizam arquivos texto possuem vantagens e desvantagens, que devem ser consideradas na escolha da estratégia mais adequada para cada aplicação.

Conforme descrito anteriormente, o arquivo do núcleo pode ser um arquivo texto, com os dados separados por tabulação, conforme a **Listagem 3**:

1	id	type	modified	language	collectionID	institutionCode	collectionCode	basisOfRecord	occurrence
2	urn:catalog:JBRJ:RB:852708	Collection	2014-08-19	16:35:54.492851	pt	urn:lsid:biocol.org:col:15528	JBRJ		
3	urn:catalog:JBRJ:RB:561192	Collection	2011-03-29	12:13:56.780423	pt	urn:lsid:biocol.org:col:15528	JBRJ		
4	urn:catalog:JBRJ:RB:307915	Collection	2014-05-16	15:15:13.616052	pt	urn:lsid:biocol.org:col:15528	JBRJ		
5	urn:catalog:JBRJ:RB:650654	Collection	2013-06-18	13:02:18.076937	pt	urn:lsid:biocol.org:col:15528	JBRJ		
6	urn:catalog:JBRJ:RB:368781	Collection	2006-12-18	06:46:48.178672	pt	urn:lsid:biocol.org:col:15528	JBRJ		
7	urn:catalog:JBRJ:RB:439493	Collection	2007-04-18	10:52:14.249131	pt	urn:lsid:biocol.org:col:15528	JBRJ		
8	urn:catalog:JBRJ:RB:789213	Collection	2014-07-08	01:19:32.666501	pt	urn:lsid:biocol.org:col:15528	JBRJ		
9	urn:catalog:JBRJ:RB:871437	Collection	2014-08-19	17:13:07.323105	pt	urn:lsid:biocol.org:col:15528	JBRJ		
10	urn:catalog:JBRJ:RB:346298	Collection	2014-05-23	13:58:05.445339	pt	urn:lsid:biocol.org:col:15528	JBRJ		
11	urn:catalog:JBRJ:RB:871500	Collection	2014-08-19	19:48:12.116057	pt	urn:lsid:biocol.org:col:15528	JBRJ		
12	urn:catalog:JBRJ:RB:378753	Collection	2006-12-18	06:33:28.755867	pt	urn:lsid:biocol.org:col:15528	JBRJ		
13	urn:catalog:JBRJ:RB:868858	Collection	2014-08-12	13:10:14.451274	pt	urn:lsid:biocol.org:col:15528	JBRJ		
14	urn:catalog:JBRJ:RB:368154	Collection	2007-01-03	15:21:17.290821	pt	urn:lsid:biocol.org:col:15528	JBRJ		
15	urn:catalog:JBRJ:RB:373800	Collection	2006-05-27	00:15:45.015896	pt	urn:lsid:biocol.org:col:15528	JBRJ		
16	urn:catalog:JBRJ:RB:83281	Collection	2010-12-11	13:35:31.839089	pt	urn:lsid:biocol.org:col:15528	JBRJ		
17	urn:catalog:JBRJ:RB:578415	Collection	2009-12-10	15:37:44.060377	pt	urn:lsid:biocol.org:col:15528	JBRJ		
18	urn:catalog:JBRJ:RB:364765	Collection	2006-12-18	09:19:35.704047	pt	urn:lsid:biocol.org:col:15528	JBRJ		
19	urn:catalog:JBRJ:RB:366416	Collection	2006-12-18	06:49:39.551289	pt	urn:lsid:biocol.org:col:15528	JBRJ		
20	urn:catalog:JBRJ:RB:871556	Collection	2014-08-20	09:34:56.044064	pt	urn:lsid:biocol.org:col:15528	JBRJ		
21	urn:catalog:JBRJ:RB:657276	Collection	2014-07-07	21:34:51.813512	pt	urn:lsid:biocol.org:col:15528	JBRJ		
22	urn:catalog:JBRJ:RB:779616	Collection	2014-07-08	01:10:46.469369	pt	urn:lsid:biocol.org:col:15528	JBRJ		
23	urn:catalog:JBRJ:RB:849691	Collection	2014-08-19	16:26:26.245209	pt	urn:lsid:biocol.org:col:15528	JBRJ		
24	urn:catalog:JBRJ:RB:425009	Collection	2011-10-06	15:42:19.226487	pt	urn:lsid:biocol.org:col:15528	JBRJ		
25	urn:catalog:JBRJ:RB:849681	Collection	2014-08-19	16:23:28.633348	pt	urn:lsid:biocol.org:col:15528	JBRJ		
26	urn:catalog:JBRJ:RB:369495	Collection	2006-12-18	06:23:41.943094	pt	urn:lsid:biocol.org:col:15528	JBRJ		
27	urn:catalog:JBRJ:RB:348077	Collection	2014-05-23	13:37:19.44319	pt	urn:lsid:biocol.org:col:15528	JBRJ		
28	urn:catalog:JBRJ:RB:604813	Collection	2014-07-07	20:13:12.665095	pt	urn:lsid:biocol.org:col:15528	JBRJ		
29	urn:catalog:JBRJ:RB:372477	Collection	2006-05-26	23:54:20.868557	pt	urn:lsid:biocol.org:col:15528	JBRJ		
30	urn:catalog:JBRJ:RB:346694	Collection	2014-05-21	18:15:57.722223	pt	urn:lsid:biocol.org:col:15528	JBRJ		
31	urn:catalog:JBRJ:RB:365104	Collection	2007-04-18	10:52:17.821103	pt	urn:lsid:biocol.org:col:15528	JBRJ		
32	urn:catalog:JBRJ:RB:606934	Collection	2013-06-11	16:56:43.080107	pt	urn:lsid:biocol.org:col:15528	JBRJ		

Listagem 3 - Exemplo de arquivo do núcleo do DwC-A.

O arquivo do núcleo, apresentado na Listagem 3, e os demais arquivos de extensão considerados devem ser descritos por meio de um arquivo chamado de meta.xml. Este arquivo (**Listagem 4**) descreve a estrutura de cada um dos arquivos que compõem o modelo estrela.

```

1 <archive xmlns="http://rs.tdwg.org/dwc/text/" metadata="eml.xml">
2   <core encoding="UTF-8" fieldsTerminatedBy="\t" linesTerminatedBy="\n" fieldsEnclosedBy="\" ic
3   <files>
4     <location>occurrence.txt</location>
5   </files>
6   <id index="0" />
7   <field index="1" term="http://purl.org/dc/terms/type"/>
8   <field index="2" term="http://purl.org/dc/terms/modified"/>
9   <field index="3" term="http://purl.org/dc/terms/language"/>
10  <field index="4" term="http://rs.tdwg.org/dwc/terms/collectionID"/>
11  <field index="5" term="http://rs.tdwg.org/dwc/terms/institutionCode"/>
12  <field index="6" term="http://rs.tdwg.org/dwc/terms/collectionCode"/>
13  <field index="7" term="http://rs.tdwg.org/dwc/terms/basisOfRecord"/>
14  <field index="8" term="http://rs.tdwg.org/dwc/terms/occurrenceID"/>
15  <field index="9" term="http://rs.tdwg.org/dwc/terms/catalogNumber"/>
16  <field index="10" term="http://rs.tdwg.org/dwc/terms/recordNumber"/>
17  <field index="11" term="http://rs.tdwg.org/dwc/terms/recordedBy"/>
18  <field index="12" term="http://rs.tdwg.org/dwc/terms/year"/>
19  <field index="13" term="http://rs.tdwg.org/dwc/terms/month"/>
20  <field index="14" term="http://rs.tdwg.org/dwc/terms/day"/>
21  <field index="15" term="http://rs.tdwg.org/dwc/terms/fieldNumber"/>
22  <field index="16" term="http://rs.tdwg.org/dwc/terms/fieldNotes"/>
23  <field index="17" term="http://rs.tdwg.org/dwc/terms/country"/>
24  <field index="18" term="http://rs.tdwg.org/dwc/terms/stateProvince"/>
25  <field index="19" term="http://rs.tdwg.org/dwc/terms/municipality"/>
26  <field index="20" term="http://rs.tdwg.org/dwc/terms/locality"/>
27  <field index="21" term="http://rs.tdwg.org/dwc/terms/verbatimLatitude"/>
28  <field index="22" term="http://rs.tdwg.org/dwc/terms/verbatimLongitude"/>
29  <field index="23" term="http://rs.tdwg.org/dwc/terms/identifiedBy"/>
30  <field index="24" term="http://rs.tdwg.org/dwc/terms/dateIdentified"/>
31  <field index="25" term="http://rs.tdwg.org/dwc/terms/typeStatus"/>
32  <field index="26" term="http://rs.tdwg.org/dwc/terms/scientificName"/>

```

Listagem 4 - Exemplo de arquivo (meta.xml) que descreve a estrutura dos demais arquivos de um DwC-A.

Informações adicionais sobre o **Darwin Core Archive** podem ser consultadas nos guias “**Darwin Core Text Guide**”⁵³ e “**Darwin Core Archive – How to Guide**”⁵⁴.

⁵³ Darwin Core Text Guide <http://rs.tdwg.org/dwc/terms/guides/text/index.htm>

⁵⁴ Darwin Core Archive – How to Guide http://imgbif.gbif.org/CMS_ORC/?doc_id=2816&download=1

3.1.9 Referências

DARWIN CORE Namespace Policy. Disponível em: <<http://rs.tdwg.org/dwc/terms/namespace/>>. Acesso em: 30 jun. 2013.

DARWIN CORE Terms: **A complete historical record**. Disponível em: <<http://rs.tdwg.org/dwc/terms/history/index.htm>>. Acesso em: 30 jun. 2013.

DARWIN CORE Terms: **A quick reference guide**. Disponível em: <<http://rs.tdwg.org/dwc/terms/index.Htm>>. Acesso em: 30 jun. 2013.

DARWIN CORE Text Guide. Disponível em: <<http://rs.tdwg.org/dwc/terms/guides/text/index.htm>>. Acesso em: 30 jun. 2013.

DARWIN CORE. **Project site for discussion and development - Google Project Hosting**. Disponível em: <<https://code.google.com/p/darwincore/>>. Acesso em: 30 jun. 2013.

DCMI Metadata Terms. Disponível em: <<http://dublincore.org/documents/dcmi-terms/>>. Acesso em: 30 jun. 2013.

DIGIR - Table of Contents. Disponível em: <<http://digir.net/>>. Acesso em: 30 jun. 2013.

DÖRING, M.; ROBERTSON, T.; REMSEN, D. **Darwin Core Archive Format, Reference Guide to the XML Descriptor File**. Copenhagen: Global Biodiversity Information Facility, 2011.

FRAME, M. Darwin Core. In: **Problems in Information Sciences**. University of Tennessee - Information Science, 2013.

GBIF Resources. Disponível em: <<http://rs.gbif.org/>>. Acesso em: 30 jun. 2013.

GBIF: Darwin Core Archives. Disponível em: <<http://www.gbif.org/informatics/standards-andtools/publishing-data/data-standards/darwin-core-archives/>>. Acesso em: 30 jun. 2013.

REMSSEN, D.; DÖRING, M.; ROBERTSON, T. **GBIF Darwin Core Archive, How-to Guide**. Copenhagen: Global Biodiversity Information Facility, 2010.

REMSEN, D. P.; DÖRING, M.; ROBERTSON, T. **Create your own Darwin Core Archive, Step-by-Step Guide**. Copenhagen: Global Biodiversity Information Facility, 2010.

REMSEN, D.; SOOD, R. **Darwin Core Archive Assistant, User Guide**. Copenhagen: Global Biodiversity Information Facility, 2010.

SIMPLE Darwin Core. Disponível em: <<http://rs.tdwg.org/dwc/terms/simple/index.htm>>. Acesso em: 30 jun. 2013.

TDWG: Darwin Core. Disponível em: <<http://www.tdwg.org/activities/darwincore/>>. Acesso em: 30 jun. 2013.

TDWG: Standards. Disponível em: <<http://www.tdwg.org/standards/>>. Acesso em: 30 jun. 2013.

WIECZOREK, J.; De GIOVANNI, R.; VIEGLAIS, D.; REMSEN, D. P.; DÖRING, M.; ROBERTSON, T. **Darwin Core Quick Reference Guide**. Copenhagen: Global Biodiversity Information Facility, 2010.

WIECZOREK, J.; BLOOM, D.; GURALNICK, R.; DÖRING, M.; et al. Darwin Core. An Evolving Community-Developed Biodiversity Data Standard. **PLoS ONE**, v. 7, n. 1, p. 29715, 6 jan. 2012.

3.2 Dublin Core

3.2.1 Apresentação

O **Dublin Core Metadata Initiative (DCMI)**⁵⁵ é uma organização dedicada a promover a adoção de padrões de interoperabilidade de metadados e desenvolver vocabulários especializados para descrever fontes que tornem mais inteligentes os sistemas para descoberta de informações.

O padrão **Dublin Core**⁵⁶ é um vocabulário de quinze elementos para a descrição de recursos. O nome “Dublin” é devido a sua origem ter ocorrido em um workshop na cidade de Dublin (República da Irlanda) em 1995. O “Core” é devido os seus elementos serem amplos e genéricos, podendo ser utilizados na descrição de uma ampla gama de recursos.

Estes elementos foram aprovados pelos seguintes padrões de referência:

- ISO Standard 15836:2009⁵⁷;

⁵⁵ Dublin Core Metadata Initiative (DCMI) <http://dublincore.org>

⁵⁶ Wikipédia - Dublin Core http://pt.wikipedia.org/wiki/Dublin_Core

⁵⁷ ISO 15836:2009 <http://www.iso.org/iso/search.htm?qt=15836&searchSubmit=Search&sort=rel&type=simple&published=on>

- ANSI/NISO Standard Z39.85-2007⁵⁸;
- IETF RFC 5013⁵⁹.

Os quinze elementos do Dublin Core fazem parte de um conjunto maior de vocabulários de metadados e especificações técnicas mantidas pela DCMI. O conjunto completo de vocabulários também inclui um conjunto de recursos de classes, esquemas para codificação de vocabulários e sintaxes. Mais informações podem ser consultadas no Guia “**DCMI Metadata Terms**”⁶⁰.

3.2.2 Objetivos

O **Dublin Core** possui como objetivo a descrição de objetos, tais como documentos (livros, artigos, padrões, protocolos, relatórios, etc.), vídeos, sons, imagens, textos e sites da Web. Aplicações do Dublin Core utilizam o **XML** e o **RDF (Resource Description Framework)**.

O Dublin Core é um conjunto de metadados para descoberta de recursos. Ele não é adequado para a descrição e compreensão de um recurso, pois não possui elementos de descrição de atributos, metodologia, qualidade, entre outras informações necessárias no detalhamento de recursos.

Em **sistemas de biodiversidade** a utilização mais comum do **Dublin Core** é na descrição de referências bibliográficas de documentos (livros, artigos científicos, relatórios técnicos). Outra utilização do padrão é a descrição de recursos em sistemas de gerenciamento de conteúdo com o objetivo de facilitar a pesquisa, descoberta e interpretação dos mesmos.

3.2.3 Detalhamento da Estrutura

O Dublin Core possui dois níveis: **Simple** e **Qualificado**. O nível simples inclui quinze elementos, já o qualificado possui três elementos adicionais (audiência, proveniência e detentor de direitos), juntamente com um grupo para refinamento dos elementos existentes (também chamados de qualificadores), que refinam a semântica dos elementos de maneira que sejam úteis no processo de descoberta dos recursos.

Todos os seus elementos são opcionais e podem ser repetidos, além disso, não existem restrições quanto a ordem dos elementos. Na **Figura 10** são descritos os diversos elementos do **Dublin Core Qualificado**.

⁵⁸ ANSI/NISO Standard Z39.85-2007 <http://www.niso.org/standards/z39-85-2007>

⁵⁹ IETF RFC 5013 <http://www.ietf.org/rfc/rfc5013.txt>

⁶⁰ DCMI Metadata Terms <http://dublincore.org/documents/dcmi-terms>

Elementos	Refinamentos		Codificações	Tipos
1. Identifier	Abstract	Is referenced by	Box	Collection
2. Title	Access rights	Is replaced by	DCMIType	Dataset
3. Creator	Alternative	Is required by	DDC	Event
4. Contributor	Audience	Issued	IMT	Image
5. Publisher	Available	Is version of	ISO3166	Interactive
6. Subject	Bibliographic citation	License	ISO639-2	Resource
7. Description	Conforms to	Mediator	LCC	Moving Image
8. Coverage	Created	Medium	LCSH	Physical Object
9. Format	Date accepted	Modified	MESH	Service
10. Type	Date copyrighted	Provenance	Period	Software
11. Date	Date submitted	References	Point	Sound
12. Relation	Education level	Replaces	RFC1766	Still Image
13. Source	Extent	Requires	RFC3066	Text
14. Rights	Has format	Rights holder	TGN	
15. Language	Has part	Spatial	UDC	
	Has version	Table of contents	URI	
	Is format of	Temporal	W3CTDF	
	Is part of	Valid		

Figura 10 – Elementos, refinamentos, codificações e tipos que compõem o padrão Dublin Core.

3.2.4 Definição dos termos

Em sua documentação, os termos do Dublin Core são descritos geralmente pelos seguintes atributos:

- **Nome:** O nome inserido no URI do namespace do DCMI para criar o URI do termo;
- **URI (Uniform Resource Identifier):** Identificador e denominador do termo na Internet;
- **Rótulo:** O nome pelo qual o termo é conhecido;
- **Definição:** Uma descrição da natureza do conteúdo do termo;
- **Tipo do Termo:** O tipo do termo conforme descrito no **DCMI Abstract Model (DCAM)**⁶¹.

Quando aplicáveis, outros atributos são disponibilizados com informações adicionais sobre o termo, como:

- **Comentário:** Comentários adicionais sobre o uso do termo;
- **Referências:** Referência do dicionário utilizado no termo;
- **Versão:** A descrição histórica de um termo.

A **Tabela 6** apresenta a descrição do termo *coverage*:

⁶¹ DCMI Abstract Model (DCAM) <http://dublincore.org/documents/abstract-model/>

Tabela 6 - Descrição do Termo *coverage* no “DCMI Metadata Terms”.

Term Name: coverage	
URI	http://purl.org/dc/elements/1.1/coverage
Label	Coverage
Definition	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Comment	Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates. Temporal topic may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. Recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names [TGN]. Where appropriate, named places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.
References	[TGN] http://www.getty.edu/research/tools/vocabulary/tgn/index.html

A descrição e demais informações dos termos que fazem parte do nível simplificado do Dublin Core estão disponíveis no Guia “**Dublin Core Metadata Element Set, Version 1.1**”⁶². Já o Dublin Core qualificado está disponível no Guia “**DCMI Metadata Terms**”⁶³.

3.2.5 Exemplo de registro Dublin Core

Na **Listagem 5** é apresentado o trecho de um arquivo Dublin Core que descreve o Guia “Darwin Core Quick Reference” do GBIF (**Figura 11**).

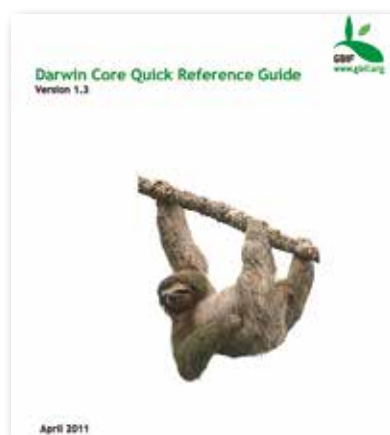


Figura 11 – Guia “Darwin Core Quick Reference” publicado pelo GBIF.

⁶² Dublin Core Metadata Element Set, Version 1.1 <http://dublincore.org/documents/dces/>

⁶³ DCMI Metadata Terms <http://dublincore.org/documents/dcmi-terms/>

```

1 <?xml version="1.0"?>
2 <metadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:dc="http://purl.org/dc/elements/1.1/">
3   <dc:title>Darwin Core Quick Reference Guide</dc:title>
4   <dc:description>Description: A printable reference guide to the ratified Darwin Core terms found at
5     http://rs.tdwg.org/dwc/terms/index.htm. Abstract: The Darwin Core is body of standards that includes
6     a glossary of terms. The Darwin Core is primarily based on taxa, their occurrence in nature as documented
7     by observations, specimens, and samples, and related information. The terms are organized by categories in
8     this printed guide.
9   </dc:description>
10  <dc:publisher>GBIF Secretariat</dc:publisher>
11  <dc:creator>Author(s):Wieczorek , J., De Giovanni , R., Vieglaes , D. Remsen D.P., Döring, M, Robertson, T,
12  Contributor(s):David Remsen (dremsen@gbif.org)</dc:creator>
13  <dc:date>2012-03</dc:date>
14  <dc:subject><a href="http://www.gbif.org/orc/?sword=&sortby=NumDownloads&advanced_search=1&keyword=10">Biodiversity
15  data standards protocols</a>, <a href="http://www.gbif.org/orc/?sword=&sortby=NumDownloads&advanced_search=1&keyword=3">
16  Biodiversity informatics</a></dc:subject>
17  <dc:source></dc:source>
18  <dc:language>English</dc:language>
19  <dc:identifier>http://links.gbif.org/gbif_dwc-a_guide_en_v1.1</dc:identifier>
20  <dc:format>application/msword</dc:format>
21  <dc:relation>http://www.gbif.org/orc/?doc_id=2803</dc:relation>
22  <dc:type>Digital document</dc:type>
23  <dc:coverage></dc:coverage>
24  <dc:contributor>Requests related to this metadata record should be addressed to the Global Biodiversity Information
25  Facility Secretariat, gbif@gbif.org, +4535321470</dc:contributor>
26  <dc:rights>Rights: This document is governed by the standard legal, copyright, licensing provisions and disclaimers
27  issued by the Taxonomic Databases Working Group. Right holders: GBIF Secretariat</dc:rights>
28 </metadata>

```

Listagem 5 - Exemplo de arquivo Dublin Core referente ao arquivo digital do “Darwin Core Quick Reference Guide”.

3.2.6 Referências

ANSI/NISO Z39.85-2007. **The Dublin Core Metadata Element Set - National Information Standards Organization**. Disponível em: <http://www.niso.org/apps/group_public/project/details.php?project_id=57>. Acesso em: 30 jun. 2013.

CDP METADATA WORKING GROUP. **Dublin Core Metadata Best Practices**. Denver: Collaborative Digitization Program at the University of Denver, 2006. Disponível em: <<http://www.mndigital.org/digitizing/standards/metadata.pdf>>. Acesso em: 30 jun. 2013.

DCMI (Dublin Core Metadata Initiative). **Abstract Model**. Disponível em: <<http://dublincore.org/documents/abstract-model/>>. Acesso em: 30 jun. 2013.

DCMI (Dublin Core Metadata Initiative). Disponível em: <<http://dublincore.org/>>. Acesso em: 30 jun. 2013.

DCMI (Dublin Core Metadata Initiative). **Metadata Terms**. Disponível em: <<http://dublincore.org/documents/dcmi-terms/>>. Acesso em: 30 jun. 2013.

WIKIPÉDIA (Enciclopédia Livre). **Dublin Core** – Disponível em: <http://pt.wikipedia.org/wiki/Dublin_Core>. Acesso em: 30 jun. 2013.

DUBLIN CORE Metadata Element Set, Version 1.1. Disponível em: <<http://dublincore.org/documents/dces/L>>. Acesso em: 30 jun. 2013.

FRAME, M. Dublin Core. In: **Problems in Information Sciences**. University of Tennessee - Information Science, 2013.

ISO 15836:2009. **Information and documentation - The Dublin Core metadata element set**. Disponível em: <http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=52142>. Acesso em: 30 jun. 2013.

KUNZE, J. A.; BAKER, T. **RFC 5013 - The Dublin Core Metadata Element Set The IETF Trust**, 2007. Disponível em: <<http://www.ietf.org/rfc/rfc5013.txt>>. Acesso em: 30 jun. 2013.

QUAM, E. **Minnesota metadata guidelines for dublin core metadata**: training manual, 2002. Disponível em: <<http://cdigital.uv.mx/handle/123456789/10372>>. Acesso em: 30 jun. 2013.

3.3 EML

3.3.1 Apresentação

O **Ecological Metadata Language (EML)**⁶⁴ é uma especificação de metadados aberto para facilitar o compartilhamento, preservação e interpretação de dados científicos entre instituições e grupos de pesquisa. Ele é baseado nos trabalhos realizados pelo **Ecological Society of America**⁶⁵, juntamente com alguns parceiros de pesquisa, como o **Knowledge Network for Biocomplexity**⁶⁶. O EML foi estruturado de forma modular e extensível para descrever diversos tipos de dados provenientes das ciências ecológicas. Estes dados abrangem uma série de domínios distintos da ciência, incluindo estudos de biodiversidade, funções do ecossistema, meteorologia, ciências do solo, hidrologia, sensoriamento remoto e muitas outras.

O EML permite ao cientista fornecer descrições padronizadas de dados desses diversos domínios em um formato de texto acessível. As descrições do EML podem ser extremamente resumidas

⁶⁴ Wikipédia - Ecological Metadata Language (EML) http://en.wikipedia.org/wiki/Ecological_Metadata_Language

⁶⁵ Portal da Ecological Society of America <http://www.esa.org/esa>

⁶⁶ Portal do Knowledge Network for Biocomplexity <http://knb.ecoinformatics.org/index.jsp>

(**Listagem 6**) ou incrivelmente ricas (incluindo descrições completas da sintaxe dos dados, esquema e semântica). Assim, o EML pode ser aplicado em uma grande variedade de aplicações de pesquisa, da descoberta de dados simples até o processamento automatizado de dados detalhados.

O EML foi criado com base nas seguintes premissas:

- **Aberto:** para possibilitar a leitura humana e facilitar o arquivamento de dados em longo prazo;
- **Modular:** para promover o reuso das estruturas/sessões dos metadados;
- **Extensível:** para possibilitar a inclusão de metadados que não estão definidos originalmente no EML;
- **Estruturado:** para possibilitar o processamento computacional realizado por aplicações de análise e outras aplicações de software;
- **Fácil de implementar:** para minimizar os requisitos e esforços necessários na utilização do padrão.

3.3.2 Objetivos

A arquitetura EML foi projetada para atender as necessidades da comunidade científica e possibilitar o benefício gerado pelo reaproveitamento de trabalhos realizados em outros padrões de metadados. O EML possui características de diversos padrões existentes, além de facilitar o processamento e a integração automática de conjuntos de dados por meio de seus metadados.

O EML tem grande valor para o armazenamento de dados em longo prazo, por ser um arquivo de texto que pode ser lido por cientistas e gestores dos dados, mas que também possui uma ampla utilização em aplicações de software devido seu formato estruturado, que possibilita o processamento computacional de seus dados.

Na **Listagem 6**, um arquivo simplificado em EML é apresentado. Este arquivo descreve um conjunto de dados a partir do seu nome (título) e das informações do seu criador. Este nível de detalhamento é útil apenas para a descoberta de dados pela comunidade científica. Para outras aplicações, devem ser criados documentos com maior riqueza de informações.

```

1  <?xml version="1.0"?>
2  <eml:eml
3    packageId="eml.1.1" system="http://knb.ecoinformatics.org"
4    xmlns:eml="eml://ecoinformatics.org/eml-2.0.0">
5
6    <dataset>
7      <title>Biodiversity surveys for Lesser Tree Frogs at Barro Colorado Island (BCI) from 1994
8      to 1999 </title>
9      <creator id="23445" scope="document">
10        <individualName>
11          <givenName>Jane</givenName>
12          <surName>Smith</surName>
13        </individualName>
14        <electronicMailAddress>jane@data.org</electronicMailAddress>
15      </creator>
16      <contact>
17        <references>23445</references>
18      </contact>
19    </dataset>
20  </eml:eml>

```

Listagem 6 - Trecho de arquivo EML simplificado para descoberta de conjunto de dados.

Informações mais detalhadas são desejáveis em grande parte das aplicações que utilizam estes metadados. No entanto, devido a geração e documentação de metadados consumirem uma quantidade considerável de recursos, estas atividades devem ser planejadas desde a concepção dos projetos. Além disso, também é aconselhável uma definição prévia do nível de detalhamento dos metadados considerados, com base nas diretrizes e necessidades da instituição ou projeto de pesquisa.

O principal repositório internacional que considera a utilização do EML é o Knowledge Network for Biocomplexity (KNB). Alguns repositórios nacionais que também consideram a infraestrutura da rede KNB são gerenciados pelo Instituto Nacional de Pesquisas da Amazônia (INPA), como os repositórios do PPBio e do PELD.

O **EML** é um conjunto de documentos XML, que podem ser utilizados de forma **modular** e **extensível** para **documentar dados ecológicos**. Cada módulo EML é projetado para descrever uma parte dos metadados que são inclusos nos conjuntos de dados ecológicos. O seu principal objetivo é **disponibilizar** para a **comunidade científica** um **padrão de metadados extensível e flexível** para utilização na análise e armazenamento dos dados.

3.3.3 Detalhamento da Estrutura

Um documento EML fornece metadados para um único recurso, que pode ser um conjunto de dados, uma citação de literatura, um protocolo científico ou um software. Para um determinado recurso descrito, uma vasta variedade de outros tipos de metadados, incluindo outros recursos, podem ser considerados. Por exemplo, ao descrever um recurso de conjunto de dados, é comum a referência a uma ou mais citações da literatura, bem como as descrições dos modelos de dados e seus valores.

Na **Tabela 7** são descritos os módulos do EML (versão 2.0.0):

Tabela 7 - Descrição dos módulos do EML (versão 2.0.0).

General modules	
eml	Módulo global contendo metadados EML e metadados adicionados por meio de extensões.
eml-resource	Descrição geral de um recurso, contendo campos comuns a todos os recursos, como quem os criou, título e palavras-chave.
eml-dataset	Campos para descrever recursos de dados, incluindo tabelas, imagens espaciais e outros tipos de entidades.
eml-literature	Campos para descrição de recursos literários relacionados às suas citações bibliográficas.
eml-protocol	Campos para descrição de protocolos científicos para a geração ou processamento de dados.
eml-software	Campos para descrição de recursos de software.
Entities and Attributes	
eml-entity	Campos gerais para descrição de todos os tipos de entidades no conjunto de dados, como tabelas e imagens espaciais.
eml-attribute	Campos para descrição de todos os atributos (variáveis) em uma entidade de dados.
eml-constraint	Campos para definição das restrições de integridade (principalmente relacional) entre entidades.
eml-physical	Campos para descrição das características físicas de um objeto de dados (arquivos e streams), bem como as informações para a sua distribuição.
eml-dataTable	Campos especializados para descrição de dados tabulares.
eml-spatialRaster	Campos especializados para descrição de dados espaciais do tipo Raster.
eml-spatialReference	Campos especializados para descrição de dados espaciais.
eml-spatialVector	Campos especializados para descrição de dados vetoriais espaciais.
eml-storedProcedure	Campos especializados para descrição de procedimentos armazenados encontrados em um sistema de gerenciamento de dados.
eml-view	Campos especializados para descrição das visões encontradas em um sistema de gerenciamento de dados.
Research Context	
eml-party	Campos para identificação e fornecimento de informações de contato das pessoas e organizações envolvidas.
eml-coverage	Campos para descrição taxonômica, espacial e temporal do conjunto de dados.
eml-methods	Campos para descrição de métodos científicos (coleta dos dados, qualificação e experimentos)
eml-project	Campos para descrição de projetos de pesquisa que estão associados a um recurso de dados ou outros tipos de recurso.

Tabela 7 - Descrição dos módulos do EML (versão 2.0.0), (continuação).

Miscellaneous	
eml-access	Campos para especificação de restrições de acesso a dados e metadados.
eml-text	Campos para fornecimento de descrições de textos estruturados e não estruturados a serem utilizadas nos diversos módulos do EML.
stmml	A linguagem para definir as unidades de medida personalizadas e os fatores de conversão das unidades relacionadas.

O EML foi criado com o objetivo de permitir uma fácil implementação, por isso, a maioria dos módulos da **Tabela 7** são opcionais. Sendo que o fornecimento apenas dos metadados mínimos não é considerado uma boa prática em uma perspectiva de gerenciamento de dados.

Em um mundo ideal, os dados relevantes devem ser descritos por meio do conjunto completo do EML, mas devido a limitação de recursos, muitos cientistas podem fornecer apenas descrições gerais de seus conjuntos de dados. Sendo que mesmo os metadados mínimos, quando fornecidos em EML, possibilitam a geração de uma base para a construção de serviços de compartilhamento e processamento dos dados, serviços analíticos, serviços de visualização baseados na Web, entre outras aplicações.

3.3.4 Definição dos elementos

Em sua documentação, o detalhamento do padrão EML é organizado conforme a sua estrutura. São descritos os módulos, os elementos, os atributos dos módulos e os seus tipos complexos.

Os módulos do EML são detalhados por meio dos seguintes atributos (**Tabela 8**):

- **Utilização recomendada:** Demonstra quais as situações recomendadas para a utilização do módulo;
- **Importado:** Identifica os elementos/padrões importados para utilização neste módulo;
- **Importado por:** Identifica os módulos que importam o módulo atual;
- **Stand-alone:** Informa se o módulo pode ser utilizado de forma isolada;
- **Visualização de Imagem do esquema:** Apresenta um diagrama que representa a estrutura do módulo (**Figura 12**).

Tabela 8 - Descrição do módulo eml-access.

Module details: eml-access	
Recommended Usage:	all data where controlling user access to the dataset is an issue
Stand-alone:	Yes
Imports:	eml-documentation, eml-resource
Imported By:	
View an image of the Schema:	eml-access image

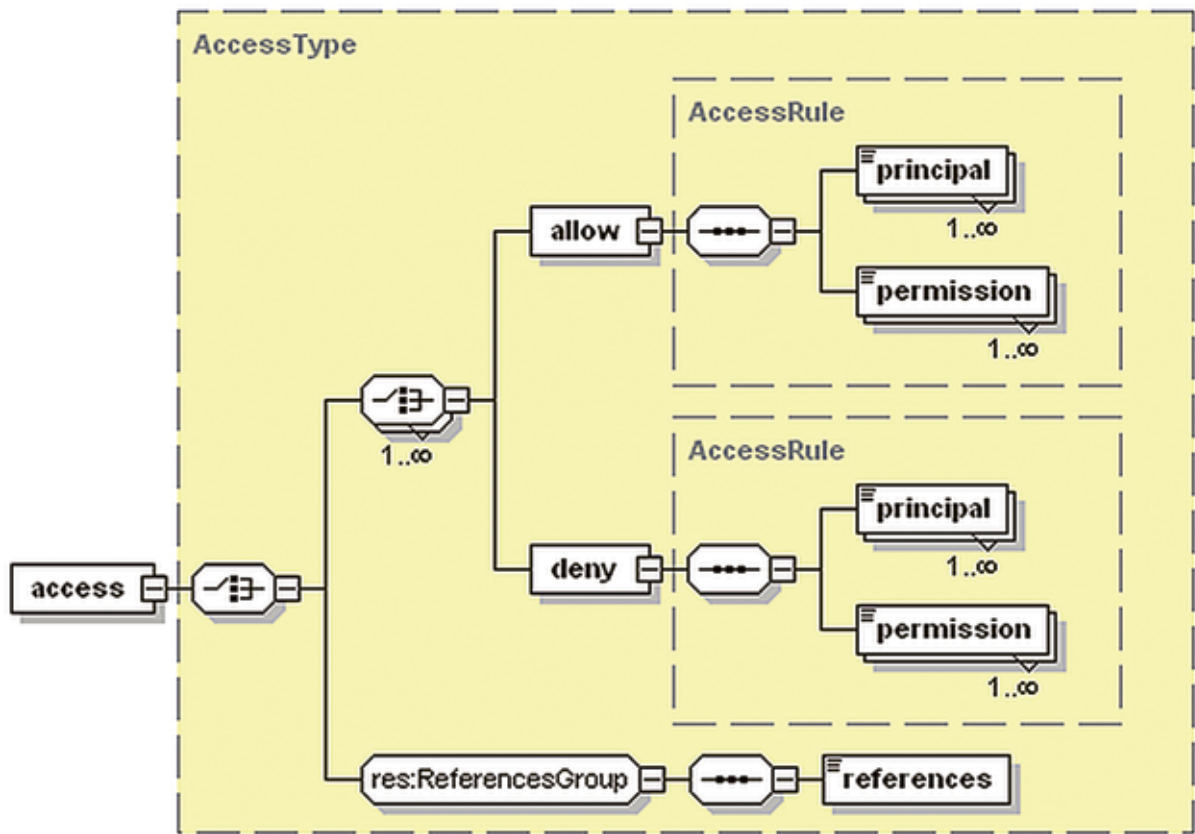


Figura 12 - Diagrama do esquema do módulo eml-access.

Já os elementos, são documentados por meio dos seguintes atributos:

- **Nome do elemento:** O nome pelo qual o elemento é conhecido;
- **Valor padrão:** Valor padrão definido para o campo;
- **Descrição do elemento:** Descreve o objetivo do campo e quais informações podem ser utilizadas, incluindo também alguns exemplos de preenchimento;
- **Tipo:** O tipo do dado utilizado no preenchimento do campo. Este tipo pode ser simples ou complexo. Os tipos complexos também são detalhados na documentação do EML.

A **Tabela 9** apresenta a descrição do elemento *Access*:

Tabela 9 - Descrição do elemento *Access*.

Element Definitions	
Name	Access
Default Value	This element has no default value.
Type	AccessType
Description of this field	The access element contains a list of rules defining permissions for this resource. For descriptions of the individual elements, see the AccessType. The permission rules defined here can be overridden by rules added to an access tree in the PhysicalDistributionType at the entity level.

A especificação detalhada do padrão EML, versão 2.1.1, pode ser consultada no Guia *“Ecological Metadata Language (EML) Specification”*⁶⁷.

⁶⁷ Ecological Metadata Language (EML) Specification <http://knb.ecoinformatics.org/software/eml/eml-2.1.1/index.html>

3.3.5 Exemplo de registro EML

```
1 <?xml version="1.0"?>
2 <eml:eml
3   packageId="eml.1.1" system="knb"
4   xml:lang="pt_BR"
5   xmlns:eml="eml://ecoinformatics.org/eml-2.1.1"
6   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
7   xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.1 eml.xsd">
8
9   <dataset id="ds.1">
10
11     <!-- English title with Portuguese translation -->
12     <title xml:lang="en_US">
13       Sample Dataset Description
14       <value xml:lang="pt_BR">Exemplo Descrição Dataset</value>
15     </title>
16     ...
17     <!-- Portuguese abstract with English translation -->
18     <abstract>
19       <para>
20         Neste exemplo, a tradução em Inglês é secundário
21         <value xml:lang="en_US">In this example, the English translation is secondary</value>
22       </para>
23     </abstract>
24     ...
25     <!-- two keywords, each with an equivalent translation -->
26     <keywordSet>
27       <keyword keywordType="theme">
28         árvore
29         <value xml:lang="en_US">tree</value>
30       </keyword>
31       <keyword keywordType="theme">
32         água
33         <value xml:lang="en_US">water</value>
34       </keyword>
35     </keywordSet>
36     ...
37   </dataset>
38 </eml:eml>
```

Listagem 7 - Trecho de um arquivo EML que descreve os campos título, resumo e palavras-chave de um dataset por meio de duas línguas (inglês e português).

3.3.6 Referências

EML (Ecological Metadata Language). Wikipedia, the free encyclopedia. Disponível em: <http://en.wikipedia.org/wiki/Ecological_Metadata_Language>. Acesso em: 30 jun. 2013.

EML (Ecological Metadata Language). **Specification**. Disponível em: <<http://knb.ecoinformatics.org/software/eml/eml-2.1.1/index.html>>. Acesso em: 30 jun. 2013.

ECOLOGICAL SOCIETY OF AMERICA. Disponível em: <<http://www.esa.org/esa/>>. Acesso em: 30 jun. 2013.

EML (Ecological Metadata Language). Best Practices for LTER Sites. [s.l.] LTER Information Management, 2011. Disponível em: <http://im.lternet.edu/sites/im.lternet.edu/files/emlbestpractices-2.0-FINAL-20110801_0.pdf>. Acesso em: 30 jun. 2013.

FEIGRAUS, 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, n. 3, p. 158-168, 2005.

KNB (The Knowledge Network for Biocomplexity). Disponível em: <<http://knb.ecoinformatics.org/index.jsp>>. Acesso em: 30 jun. 2013.

LEITE, M. de Q. **Arquitetura de sistemas orientada a serviços para análise de experimentos ambientais em Weblabs de abelhas**. Dissertação (Mestrado). Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3141/tde-08122011-160506/pt-br.php>>. Acesso em: 30 jun. 2013.

MICHENER, W. K.; BRUNT, J. W. Nongeospatial Metadata for the Ecological Sciences. **Ecological Applications**, v. 7, n. 1, p. 330-342, 1 fev. 1997.

MICHENER, W. K.; PORTER, J.; SERVILLA, M.; VANDERBILT, K. Long term ecological research and information management. **Ecological Informatics**, v. 6, n. 1, p. 13-24, jan. 2011.

BLANKMAN, D.; MCGANN, J. **Ecological Metadata Language: Practical Application for Scientists**. In: ALBUQUERQUE, N. M. **LTER Network Office**, 2003. Disponível em: <<http://im.lternet.edu/sites/im.lternet.edu/files/emlHandbook.pdf>>. Acesso em: 30 jun. 2013.

3.4 Audubon Core

3.4.1 Apresentação

O **Audubon Core (AC)**⁶⁸ é um conjunto de metadados projetados para representar recursos e coleções multimídia de biodiversidade. Recursos multimídia são objetos físicos ou digitais que normalmente abrangem mais do que texto puro. Eles incluem fotos, ilustrações, desenhos, som, vídeo, animações, materiais de apresentação e mídia on-line interativa. A coleção multimídia é um conjunto destes objetos, que passaram ou não por um processo de curadoria, eletronicamente acessíveis ou não. A especificação deste padrão de metadados considera como recurso multimídia tanto as coleções com diversos recursos multimídia quanto um recurso único isolado. A versão 1.0 do **AC** está em processo de revisão e avaliação pelo **TDWG**.

O Audubon Core também se tornou uma extensão do Darwin Core, chamada de **Audubon Core v1.0.1**, e pode ser consultada no **GBIF Vocabularies – Audubon Core**⁶⁹.

3.4.2 Objetivos

Os vocabulários do **AC** descrevem informações que identificam se os recursos multimídia estão aptos para serem utilizados em aplicações científicas específicas, antes mesmo da obtenção destes recursos. Entre diversas informações, os vocabulários do padrão descrevem questões relacionadas ao gerenciamento do arquivo ou da coleção, descrições do seu conteúdo, sua taxonomia, geografia, cobertura temporal e a forma indicada para recuperação, reprodução e utilização dos mesmos.

O **AC** compartilha as terminologias e as preocupações consideradas em outros padrões conhecidos durante a descrição de recursos como o *Dublin Core*, o *Darwin Core*, o *Adobe Extensible Metadata Platform (XMP)*, o *International Press and Telecommunications Council (IPTC)*, o *Metadata Working Group Schema (MWG)* e o *Natural Collections Schema (NCD)*.

Muitas coleções multimídia de biodiversidade já são descritas em *Darwin Core* ou *Dublin Core*. Mas o Audubon Core surge como uma alternativa e pretende tornar mais fácil a reutilização das descrições já existentes, possibilitando o detalhamento com informações adicionais quando necessário.

O **Audubon Core** foi projetado para facilitar a **publicação e descoberta de recursos multimídia** relacionados a biodiversidade, por meio da descrição consistente de um recurso ou de um conjunto de recursos.

⁶⁸ Audubon Core (AC) http://terms.gbif.org/wiki/Audubon_Core_Term_List

⁶⁹ GBIF Vocabularies – Audubon Core <http://vocabularies.gbif.org/node/126782>

3.4.3 Detalhamento da Estrutura

O **Audubon Core** possui na sua estrutura duas camadas de termos, também chamadas de *Layers*, que caracterizam principalmente o nível de detalhes na descrição de um recurso:

- **Camada 1 (Layer 1):** Fornece o vocabulário mais básico para descrever o recursos multimídia com os termos do AC.
- **Camada 2 (Layer 2):** Fornece vocabulários adicionais e mais ricos para descrever os recursos.

Os termos da Camada 1 são considerados os mais comuns e descrevem a grande maioria dos recursos multimídia, mesmo possuindo poucos termos obrigatórios. Durante a utilização do **AC** deve-se considerar a utilização do maior número possível de termos da Camada 1. Já os termos da Camada 2 são necessários em tipos particulares de recursos ou em aplicações que necessitam de descrições altamente detalhadas dos seus recursos.

O AC possui 80 termos no seu esquema, sendo apenas seis obrigatórios, conforme a **Tabela 10**.

Tabela 10 - Termos obrigatórios do Audubon Core.

Termo	Definição
Identifier (dcterms:identifier)	Um código exclusivo para o recurso, sendo uma coleção ou um item multimídia. É obrigatório um identificador para as coleções e é recomendado um identificador para os itens de mídia.
Type (dcterms:type)	São aceitos tipos definidos pelo DCMI (http://dublincore.org/documents/dcmi-type-vocabulary/#H7).
Title (dcterms:title)	Título, nome ou rótulo da instituição, coleção ou recurso individual.
Metadata Language (ac:MetadataLanguage)	Linguagem da descrição e outros metadados, representado na ISO639-1 ou ISO639-3.
Copyright Owner (xmpRights:Owner)	O nome do proprietário dos direitos. "Unknown" é um valor aceito.
Copyright Statement (dcterms:rights)	Informação sobre os direitos existentes relacionados ao recurso. O texto completo da declaração de direitos autorais, conforme exigido pela legislação nacional do detentor do copyright. Em coleções isso se aplica a todos os objetos contidos, a menos que um objeto unitário possua uma própria declaração diferente.

Mais informações sobre a estrutura do Audubon Core podem ser consultadas no Guia **"Audubon Core (DRAFT of 1.0 Normative)"**⁷⁰.

⁷⁰ Audubon Core (DRAFT of 1.0 Normative) <http://terms.gbif.org/w/index.php?oldid=9000>

3.4.4 Definição dos termos

Em sua documentação, os termos do AC são descritos por meio dos seguintes atributos:

- **URI normativo:** URL que identifica o termo;
- **Rótulo:** Identificação do termo no AC;
- **Atributos:** Identifica o nível do termo, sua obrigatoriedade de preenchimento e se o mesmo pode ser repetido;
- **Definição:** Uma descrição sobre a função do termo e seus possíveis valores;
- **Definido por:** Termo ou vocabulário no qual o atributo foi baseado. Em geral é baseado em elementos de outros padrões de metadados como o Dublin Core;
- **Notas:** Descreve informações adicionais sobre a utilização do termo e seu mapeamento com os dados do recurso descrito.

A **Tabela 11** apresenta a descrição do termo *Identifier*:

Tabela 11 - Descrição do termo *Identifier*.

Termo Identifier (dcterms:identifier)	
Normative URI:	http://purl.org/dc/terms/identifier
Label	Identifier
	Layer: 1 — Required: Yes for media collections, No for media resources (but preferred if available). — Repeatable: Yes
Definition:	An arbitrary code that is unique for the resource, with the resource being either a provider, collection, or media item.
Defined by:	dcterms:identifier
Notes	Using multiple identifiers implies that they have a same-as relationship, i.e. they all identify the same object (e. g. an object may have all of an http-URL, an Isid-URI, and a UUID).

A descrição e demais informações dos termos que fazem parte do Audubon Core estão disponíveis no Guia **“Audubon Core Term List”**⁷¹.

⁷¹ Audubon Core Term List http://terms.gbif.org/wiki/Audubon_Core_Term_List_%281.0_normative%29

3.4.5 Exemplo de registro Audubon Core

```

1  <MEDIA_METADATA_CONTAINER>
2    <dcterms:identifier>http://example.com/pictures/thePicture.jpg</dcterms:identifier>
3    <dcterms:title>A red beech leaf</dcterms:title>
4    <dcterms:format>jpg</dcterms:format>
5    <ac:accessURI>http://example.com/fullres/thePicture.jpg</ac:accessURI>
6    ...
7  <MEDIA_METADATA_CONTAINER>
8  <MEDIA_METADATA_CONTAINER>
9    <dcterms:identifier>http://example.com/pictures/thePicture.jpg</dcterms:identifier>
10   <dcterms:format>png</dcterms:format>
11   <ac:accessURI>http://example.com/fullres/thePicture-hires.png</ac:accessURI>
12   ...
13 </MEDIA_METADATA_CONTAINER>

```

Listagem 8 - Trecho de um arquivo Audubon Core que descreve uma imagem na Internet.

3.4.6 Referências

AUDUBON Core Term List (DRAFT of 1.0 normative) - GBIF-Terms. Disponível em: <[http://terms.gbif.org/wiki/Audubon_Core_Term_List_\(1.0_normative\)](http://terms.gbif.org/wiki/Audubon_Core_Term_List_(1.0_normative))>. Acesso em: 30 jun. 2013.

AUDUBON Core (DRAFT of 1.0 normative) - GBIF-Terms. Disponível em: <<http://terms.gbif.org/w/index.php?oldid=9000>>. Acesso em: 30 jun. 2013.

AUDUBON Core v 1.0.1 | GBIF Vocabularies. Disponível em: <<http://vocabularies.gbif.org/node/126782>>. Acesso em: 30 jun. 2013.

3.5 Plinian Core

3.5.1 Apresentação

O **Plinian Core**⁷² é um padrão de dados desenvolvido para organizar e compartilhar informações de espécies. A principal diferença deste esquema em relação aos outros esquemas que funcionam de maneira similar (como o *Taxon Concept Schema* – TCS ou o Darwin Core), é o fato de que o Plinian Core considera não apenas informações taxonômicas, mas inclui também informações referentes ao comportamento, natureza, influência e genética de uma espécie.

⁷² Portal do Plinian Core <http://www.pliniancore.org>

3.5.2 Objetivos

O desenvolvimento do Plinian Core foi iniciado em 2005, nas filiais espanhola e costa-riquenha do GBIF. O objetivo inicial foi realizar um mapeamento das informações associadas as espécies que são publicadas na web e uma avaliação dos padrões existentes que consideravam estes dados. O resultado dessa pesquisa demonstrou a necessidade da criação de um padrão de dados universal, capaz de integrar informações de biodiversidade importantes e disponibilizar uma variedade de informações advindas de diferentes bancos de dados de forma padronizada.

O **Plinian Core** apresenta uma estrutura de dados simples, de fácil entendimento, e com um nível de detalhes suficiente para atender os campos de interesse e necessidades de todo o público-alvo, que inclui profissionais ligados à biodiversidade, pesquisadores, professores, alunos, formadores de opinião e profissionais de outras áreas que podem se beneficiar desse conhecimento.

3.5.3 Detalhamento da Estrutura

O esquema Plinian Core faz uso de um arquivo .XSD para estruturar os elementos de informação sobre espécies, definindo os dados a serem publicados pelos provedores de dados. Nesse esquema, as informações são subdivididas na seguinte estrutura:

- Metadados
- Informações taxonômicas
- Elementos básicos
- Nomenclatura e Classificação
- Descrição
- História Natural
- Habitat e Distribuição
- Democracia e Conservação
- Uso e Gestão
- Documentação

Cada subdivisão da estrutura apresenta uma série de elementos (obrigatórios e não obrigatórios) para a padronização no esquema Plinian Core. O objetivo do padrão de dados é permitir o mapeamento destas informações em diferentes bancos de dados e possibilitar que estes dados sejam distribuídos de maneira padronizada.

3.5.4 Definição dos Termos

Em sua documentação, os termos do Plinian Core são descritos por meio dos seguintes atributos:

- **Nome:** Identificação do elemento no Plinian Core;
- **Descrição:** Descrição de características do termo (estrutura, obrigatoriedade);
- **Referência:** Informações sobre o preenchimento do termo, seu objetivo, origem e público-alvo. Os dicionários controlados (quando existirem) também são detalhados nesse atributo;
- **Classe:** Detalhamento do tipo da informação do termo.

A **Tabela 12** apresenta a descrição do termo *TaxonRecordID*:

Tabela 12 – Descrição do termo *TaxonRecordID*.

TaxonRecordID
Description
Obligatory element
Quick Reference:
A unique identifier inside the database, for the species record.
Class:
http://code.google.com/p/pliniancore/wiki/NormalString

A descrição e demais informações dos termos que fazem parte do Plinian Core estão disponíveis no **Repositório do padrão Plinian Core**⁷³. Os arquivos .XSD com o esquema da estrutura de dados do Plinian Core também podem ser encontrados no Portal do Plinian Core. Seguindo a estrutura definida nestes arquivos, um provedor de informações de biodiversidade pode disponibilizar seus dados no padrão Plinian Core.

⁷³ Repositório do padrão Plinian Core <https://code.google.com/p/pliniancore>

3.5.5 Exemplo de registro Plinian Core

```

1  <?xml version="1.0" encoding="utf-8" ?>
2  <dataset
3      xmlns="http://www.eol.org/transfer/content/0.3"
4      xmlns:xsd="http://www.w3.org/2001/XMLSchema"
5      xmlns:dc="http://purl.org/dc/elements/1.1/"
6      xmlns:dcterms="http://purl.org/dc/terms/"
7      xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
8      xmlns:dwc="http://rs.tdwg.org/dwc/dwcore/"
9      xmlns:adw="http://animaldiversity.ummz.umich.edu/morphology/"
10     xmlns:eol="http://www.eol.org/transfer/content/0.3"
11     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
12     xsi:schemaLocation="PlinianCore_AbstractModel_v3.1.xsd">
13     <TaxonRecord>
14         <BaseElements>
15             <TaxonRecodId>Phocoena sinus Norris & McFarland, 1958</TaxonRecodId>
16             <dwc:taxonConceptID>20676727</dwc:taxonConceptID>
17             <GlobalUniqueIdentifier></GlobalUniqueIdentifier>
18             <abstract></abstract>
19         </BaseElements>
20         <dc:language>es</dc:language>
21         <Version>
22         <Revision>
23         <TaxonRecordName>
24         <Synonyms>
25         <CommonNames>
26             <CommonNamesAtomized>
27                 <Name xml:lang="es">vaquita</Name>
28                 <language>es</language>
29             </CommonNamesAtomized>
30             <CommonNamesAtomized>
31                 <Name >vaquita marina</Name>
32                 <language>es</language>
33             </CommonNamesAtomized>
34             <CommonNamesAtomized>
35             <CommonNamesAtomized>
36             <CommonNamesAtomized>
37         </CommonNames>
38         <Hierarchy>
39             <dwc:Kingdom>Animalia</dwc:Kingdom>
40             <dwc:Phylum>Chordata</dwc:Phylum>
41             <dwc:Class>Mammalia</dwc:Class>
42             <dwc:Order>Cetacea</dwc:Order>
43             <dwc:Family>Phocoenidae</dwc:Family>
44             <dwc:Genus>Phocoena</dwc:Genus>
45         </Hierarchy>
46         <BriefDescription> summaryDestription-Taxon
47         </BriefDescription>
48         <FullDescriptionUnstructured>fullDescription-Taxon
49         </FullDescriptionUnstructured>
50     </TaxonRecord>
51 </dataset>

```

Listagem 9 - Trecho de um arquivo Plinian Core que descreve a espécie *Phocoena sinus*.

3.5.6 Referências

DATA PROVIDER. SSTN – IABIN (Species Specimen Thematic Network – American Biodiversity Information Network). Disponível em: <<http://larus.inbio.ac.cr/dokuwiki/sstn/doku.php?id=dataproducer>>. Acesso em: 30 jun. 2013.

PLINIAN CORE. Disponível em: <<http://www.pliniancore.org/>>. Acesso em: 30 jun. 2013.

PLINIAN CORE. **PlC 3.0 - Google Project Hosting**. Disponível em: <<https://code.google.com/p/pliniancore/>>. Acesso em: 30 jun. 2013.

SUFROSTICO/DwC-A-PHP-Library GitHub. Disponível em: <<https://github.com/Sufrostico/DwC-A-PHP-Library>>. Acesso em: 30 jun. 2013.

3.6 Padrões OGC

3.6.1 Apresentação

As funcionalidades para manipulação de mapas e dados georeferenciados são requisitos essenciais em grande parte dos sistemas de biodiversidade. Estes recursos são considerados em diversos processos, como na inserção e visualização de dados primários, além da utilização na geração de dados secundários, a partir das atividades como consolidação e análise de dados. Para a adoção destes recursos é recomendada a utilização de padrões de dados geográficos.

As duas principais organizações mundiais de padronização de dados geográficos são: a **International Organization for Standardization (ISO)**⁷⁴ e a **Open Geospatial Consortium (OGC)**⁷⁵.

A **ISO** foi fundada em 1947 e é composta por 156 países. A organização possui um comitê especial para a normatização de dados geográficos digitais (**ISO/TC 211**)⁷⁶ que está envolvido em projetos de modelagem de dados geográficos, metadados, referências e outros.

Já a **OGC** é uma organização internacional que lidera o desenvolvimento e a padronização

⁷⁴ International Organization for Standardization (ISO) <http://www.iso.org/iso/home.html>

⁷⁵ Open Geospatial Consortium (OGC) <http://www.opengeospatial.org>

⁷⁶ (ISO/TC 211) <http://www.isotc211.org>

de conteúdos nas áreas geoespacial e de serviços baseados em localização (LBS). Esta organização também é composta por diversos membros, sendo eles empresas e órgãos governamentais, companhias privadas e universidades, que participam da definição de especificações públicas.

As especificações **OpenGIS**, uma marca registrada da **OGC**, servem de apoio a soluções interoperáveis que habilitam os serviços geográficos para diversos ambientes tecnológicos. Por meio da definição de padrões e procedimentos, o OGC auxilia os produtores e usuários de geotecnologias no acesso, manipulação e exploração de dados geográficos, independente do sistema ou tecnologia utilizada.

Dentre as principais especificações existentes estão:

- *Geography Markup Language (GML)*⁷⁷;
- *Web Feature Service (WFS)*⁷⁸;
- *Web Map Service (WMS)*⁷⁹;
- *Web Coverage Service (WCS)*⁸⁰;
- *OpenGIS Catalog Service*⁸¹;
- *Keyhole Markup Language (KML)*⁸²;
- *Simple Features Interface Standard (SFS)*⁸³;

Mais informações sobre os padrões OGC podem ser consultadas no guia **“OGC Reference Model”**⁸⁴ ou nas documentações específicas de cada padrão.

Além da adoção dos padrões definidos pela OGC, para fazer parte da **Infraestrutura Nacional de Dados Geoespaciais (INDE)**⁸⁵, os provedores de dados geográficos nacionais devem seguir as normas definidas pela **Comissão Nacional de Cartografia (CONCAR)** e pelo **e-PING (Programa de Interoperabilidade do Governo Eletrônico)**⁸⁶. Estas normas podem ser consultadas nos documentos disponibilizados na área de **Normas e Padrões do Portal da INDE**⁸⁷.

⁷⁷ Geography Markup Language (GML) http://portal.opengeospatial.org/files/?artifact_id=20509

⁷⁸ Web Feature Service (WFS) http://portal.opengeospatial.org/files/?artifact_id=8339

⁷⁹ Web Map Service (WMS) http://portal.opengeospatial.org/files/?artifact_id=14416

⁸⁰ Web Coverage Service (WCS) http://portal.opengeospatial.org/files/?artifact_id=41437

⁸¹ OpenGIS Catalog Service http://portal.opengeospatial.org/files/?artifact_id=20555

⁸² Keyhole Markup Language (KML) http://portal.opengeospatial.org/files/?artifact_id=27810

⁸³ Simple Features Interface Standard (SFS) http://portal.opengeospatial.org/files/?artifact_id=25355

⁸⁴ OGC Reference Model <http://www.opengis.net/doc/orm/2.1>

⁸⁵ Infraestrutura Nacional de Dados Geoespaciais (INDE) <http://www.inde.gov.br>

⁸⁶ e-PING (Programa de Interoperabilidade do Governo Eletrônico) <http://eping.governoeletronico.gov.br>

⁸⁷ Normas e Padrões do Portal da INDE http://www.inde.gov.br/?page_id=70

3.6.2 Objetivos

O objetivo principal dos padrões OGC é possibilitar o acesso a dados espaciais, localmente ou remotamente, sem a preocupação com os modelos de dados proprietários, formatos de arquivos ou Sistemas de Informação Geográfica (SIGs) que gerenciam estes dados em seus provedores.

Considerando estes padrões nos sistemas de biodiversidade, suas aplicações são variadas e algumas das principais utilizações são:

- Os dados podem ser obtidos e referenciados como pontos individuais no espaço geográfico, no caso de ocorrências de espécies;
- Os dados podem ser transformados em informações categorizadas, camadas (layers), tais como as que são usadas para produzir mapas de recursos naturais, como os tipos de solo, categorias de vegetação e clima;
- Os dados podem ser utilizados como parâmetros espaciais na pesquisa dos dados;
- Os dados podem ser disponibilizados e consumidos por meio de serviços e padrões de codificação para a utilização em sistemas GIS, sistemas de análise e modelagem de dados de biodiversidade.

3.6.3 Detalhamento da Estrutura

As especificações dos padrões OGC são definidas a partir de diversos documentos. Os principais são:

- **Especificação Abstrata (*Abstract Specification*)**⁸⁸: Fornece a visão conceitual para as atividades de desenvolvimento dos padrões OGC. Os padrões OGC são construídos e referenciados por meio destas especificações abstratas, permitindo assim a interoperabilidade entre diferentes empresas e sistemas de processamento espacial distintos.
- **Implementação do Padrão (*Implementation Standard*)**⁸⁹: Documentos técnicos que detalham as interfaces e codificações definidas pelo OGC. Desenvolvedores de software utilizam estes documentos para a construção de interfaces abertas e codificações nos seus produtos e serviços. Estas especificações são detalhadas em nível de implementação, possibilitando que implementações distintas, realizadas por diferentes desenvolvedores, sejam compatíveis.
- **Melhores práticas (*Best Practices*)**⁹⁰: Documentos contendo as melhores práticas relacionadas ao uso e/ou implementação dos padrões OGC.

⁸⁸ OGC Abstract Specification <http://www.opengeospatial.org/standards/as>

⁸⁹ OGC Implementation Standard <http://www.opengeospatial.org/standards/is>

⁹⁰ OGC Best Practices <http://www.opengeospatial.org/standards/bp>

3.6.4 Referências

CARTILHA Técnica para Publicação de Dados Abertos no Brasil v1.0. Portal Brasileiro de Dados Abertos. Disponível em: <<http://dados.gov.br/cartilha-publicacao-dados-abertos/>>. Acesso em: 30 jun. 2013.

CEMG-CONCAR. (Comitê de Estruturação de Metadados Geoespaciais). **Perfil de Metadados Geoespaciais do Brasil (Perfil MGB)**. Comissão Nacional de Cartografia, nov. 2009. Disponível em: <http://www.concar.ibge.gov.br/arquivo/Perfil_MGB_Final_v1_homologado.pdf>. Acesso em: 30 jun. 2013.

ISO (International Organization of Standardization). **ISO 19115 - Geographic information - Metadata**. Suíça: International Organization of Standardization, 1 maio. 2003. Disponível em: <<https://www.ghrsst.org/files/download.php?m=documents&f=ISO%2019115%20.pdf>>. Acesso em: 30 jun. 2013.

ISO (International Organization for Standardization). Disponível em: <<http://www.iso.org/iso/home.html>>. Acesso em: 30 jun. 2013.

OGC (Open Geospatial Consortium). **Reference Model**. Disponível em: <<http://www.opengeospatial.org/>>. Acesso em: 30 jun. 2013.

WIKIPEDIA (Encyclopedia free). **List of geographic information systems software**. Disponível em: <http://en.wikipedia.org/wiki/List_of_geographic_information_systems_software>. Acesso em: 30 jun. 2013.

OGC (Open Geospatial Consortium). **Reference Model**. Open Geospatial Consortium, 19 dez. 2011. Disponível em: <https://portal.opengeospatial.org/files/?artifact_id=47245>. Acesso em: 30 jun. 2013.

PING (Padrões de Interoperabilidade de Governo Eletrônico). Disponível em: <<http://eping.governoeletronico.gov.br/>>. Acesso em: 30 jun. 2013.

SANTANA, F. S. **Uma infraestrutura orientada a serviços para a modelagem de nicho ecológico**. Tese (Doutorado). Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3141/tde-13072009-165044/pt-br.php>>. Acesso em: 30 jun. 2013.



Reserva Extrativista de Cassurubá (BA)
Foto: Wigold B. Schäffer

4. Ferramentas e Protocolos para Manipulação de Dados de Biodiversidade

A arquitetura de interoperabilidade de sistemas de biodiversidade apresentada no **Capítulo 2.2** sugere a utilização de protocolos e ferramentas livres e de código aberto que garantem a comunicação entre os diversos sistemas envolvidos.

Os **provedores de dados** podem utilizar estas ferramentas para a partilha dos seus dados científicos com outras instituições e com a comunidade científica.

Os **agregadores de dados** utilizam estas ferramentas para a consulta automatizada dos dados partilhados pelos diversos provedores existentes e, em diversas situações, realiza a coleta (*harvesting*) destes dados nas suas bases centralizadas.

Por fim, os **publicadores de dados** utilizam estas ferramentas geralmente para a publicação dos dados em Portais na Web, que disponibilizam ferramentas para a localização, manipulação e reutilização dos dados. Os agregadores de dados muitas vezes também fazem o papel de publicadores de dados e disponibilizam as informações (provenientes de diversos provedores) em um portal na Web.

4.1 TAPIR

4.1.1 Apresentação

O **TAPIR (TDWG Access Protocol for Information Retrieval)**⁹¹ é um protocolo de web services para a realização de consultas em banco de dados distribuídos de estrutura física e lógica variada. O TAPIR destina-se a comunicação entre aplicações, utilizando o HTTP como o mecanismo de transporte. Suas funcionalidades estão disponíveis por meio de cinco operações de solicitação/resposta:

- Recuperar metadados de serviço;
- Recuperar as configurações de serviço;
- Inspeccionar conteúdo disponível;
- Realizar consultas;
- Monitorar a disponibilidade de serviço.

⁹¹ TAPIR (TDWG Access Protocol for Information Retrieval) <http://wiki.tdwg.org/TAPIR>

O **protocolo TAPIR** é baseado na **troca de mensagens** de requisição/resposta entre um **provedor** (que possui os dados e os disponibiliza por meio dos serviços Web) e um **cliente (consumidor)**, que conhece o protocolo e requisita as informações destes **provedores**.

O TAPIR especifica um padrão sem estado (stateless), transmitido por HTTP, baseado em um protocolo de pedido e resposta em XML, para acessar dados estruturados que podem estar armazenados em um grande número de bases de dados distribuídas com variadas estruturas físicas e lógicas. O TAPIR é considerado uma evolução dos protocolos **BioCaSe**⁹² e **DiGIR**⁹³ como alternativa para a comunicação entre aplicações clientes e provedores de dados, utilizando a internet.

As principais premissas deste protocolo são:

- Sem vínculo a um específico sistema operacional;
- Sem vínculo com um específico padrão de metadados;
- Sem vínculo com um específico modelo de saída dos dados;
- Flexível na utilização de extensões dos padrões;
- Gratuito e de uso livre.

O TAPIR foi desenvolvido para ser utilizado com coleções de dados de biodiversidade, mas pode ser aplicado em diversos domínios, devido à flexibilidade no uso de padrões de metadados.

4.1.2 Funcionamento

O protocolo TAPIR e os padrões de metadados (Exemplo: Darwin Core ou Plinian Core) não possuem dependências diretas entre si, pois ambos são isolados de forma a terem suas funções definidas e independentes, conforme descrito a seguir:

- **Padrões de metadados:** São utilizados para determinar quais campos e formatos de campo serão mapeados e transportados durante a comunicação entre os sistemas. Possibilita o uso de extensões dos metadados que permitem o mapeamento de uma maior quantidade de campos, com o objetivo de detalhar as informações disponibilizadas.
- **Protocolo:** Fornece meios, através do uso de rede de computadores, do recebimento de solicitações de um requisitante e da criação de pacotes de dados padronizados (baseado no padrão de metadados) para transmiti-los do provedor para o cliente.

⁹² Portal do BioCaSe <http://www.biocase.org/products/protocols>

⁹³ Portal do DiGIR <http://digir.net>

Baseado nestes conceitos foi definida uma arquitetura abrangente que permite a integração entre o protocolo e diferentes versões de padrões (**Figura 13**).

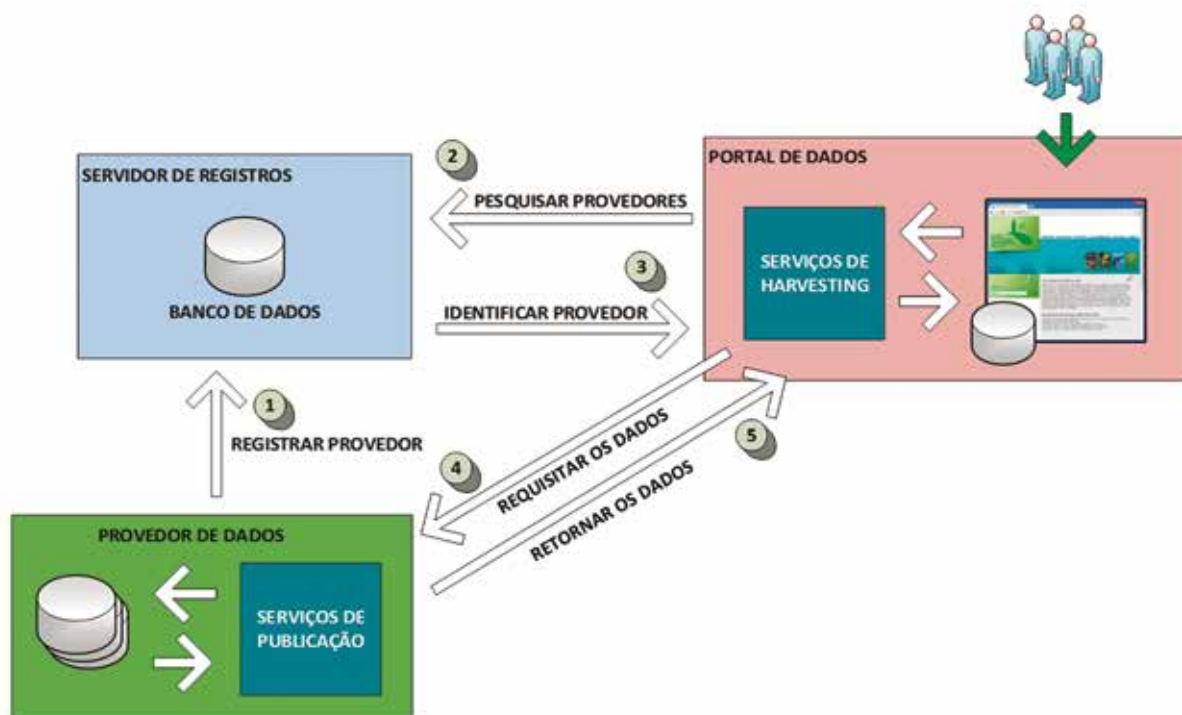


Figura 13 - Arquitetura de alto nível do TAPIR/DiGIR (DIGIR, 2013).

A comunicação nesta arquitetura é toda baseada em XML, para facilitar tanto as solicitações como os retornos de dados. O protocolo é aplicado tanto no **Provedor** quanto no **Cliente**, representado pelo **Portal** na **Figura 13**. O serviço de “**Registro dos Provedores**” é opcional nesta arquitetura, pois pode ser realizado pelo próprio **Cliente**.

Nos provedores TAPIR os objetivos são mais específicos:

- Efetuar a comunicação com a base de dados local;
- Realizar o mapeamento dos dados locais baseados no padrão de metadados utilizado;
- Fornecer serviços de requisição, para que outros sistemas possam solicitar os dados disponíveis.

Já nos clientes TAPIR os objetivos são:

- Fornecer as regras necessárias para a formulação das requisições dos dados, conforma as regras do protocolo;

- Retornar os dados em formato padronizado, baseado no padrão utilizado, de forma que o sistema possa interpretar e armazenar os dados recebidos em sua base de dados local.

Em relação ao mapeamento e disponibilização de dados utilizando o protocolo TAPIR, as tecnologias e os modelos de dados dos provedores TAPIR são irrelevantes para os clientes TAPIR durante as suas requisições, pois todas as consultas são direcionadas aos elementos (conceitos) da **camada de abstração dos dados** (esquemas conceituais), e não diretamente nos dados armazenados no banco de dados do provedor (**Figura 14**).

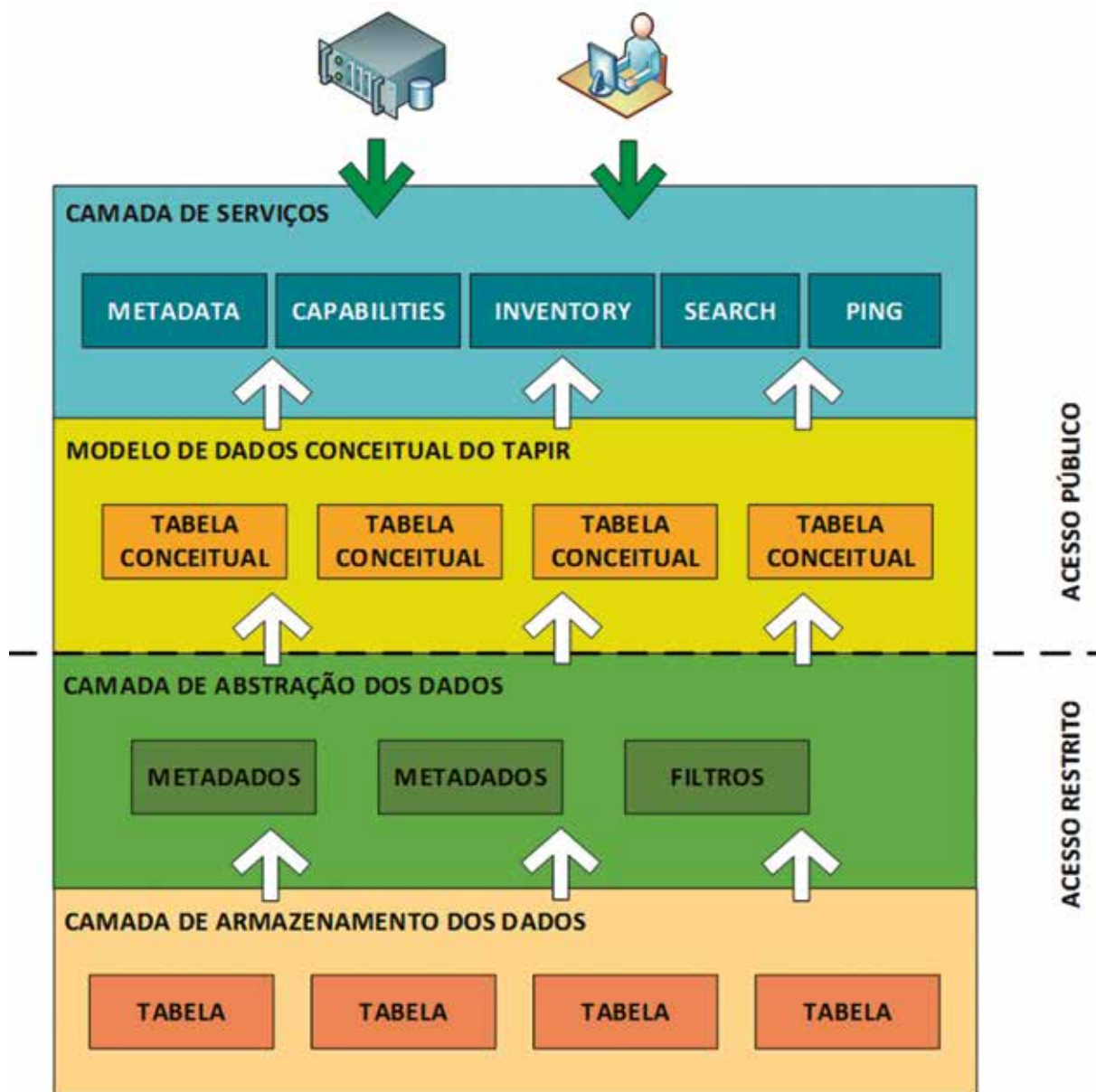


Figura 14 – Camada de Abstração de dados na configuração do Provedor TAPIR (SILVA; CORREA, 2012).

Conforme apresentado na **Figura 14**, por meio de uma **camada de abstração**, os dados armazenados no **modelo de dados do provedor** são mapeados por um esquema conceitual, baseado em um padrão de metadados, como o Darwin Core. Com isso, os dados e a estrutura do modelo de dados do provedor são “escondidos” e os serviços TAPIR passam a visualizar apenas os elementos do **modelo de dados conceitual**.

A partir daí, toda a troca de informações entre os sistemas, realizada pelo protocolo TAPIR, é realizada a partir deste modelo de dados conceitual.

O TAPIR também foi projetado para ser independente de qualquer estrutura particular para as respostas das consultas (conhecidos como modelos de saída). É necessária a definição de pelo menos um modelo de saída, ou referenciar um modelo já existente, para ser utilizado nas respostas das operações de busca.

A especificação completa do protocolo TAPIR pode ser consultada no **documento de especificação do protocolo TAPIR (versão 1.0)**⁹⁴. Outras informações relacionadas ao TAPIR, sua utilização e evolução podem ser consultadas na **Wiki do Protocolo TAPIR**⁹⁵.

4.1.3 TAPIR Wrappers

Os *Wrappers* do TAPIR são classes codificadas em algumas linguagens de programação que implementam as principais funções do protocolo TAPIR. Com a utilização destes *wrappers* é possível incorporar as funcionalidades do protocolo TAPIR em sistemas computacionais existentes ou em novos sistemas, sem a necessidade da implementação completa do protocolo, sendo necessária apenas a realização de algumas customizações. Existem iniciativas (ativas e inativas) de *wrappers* para TAPIR nas seguintes linguagens de programação:

- Java;
- PHP;
- .NET;
- Python.

Os detalhes sobre os *wrappers* do TAPIR podem ser consultados na Wiki do **TapirSoftware**⁹⁶.

⁹⁴ Protocolo TAPIR http://www.tdwg.org/dav/subgroups/tapir/1.0/docs/tdwg_tapir_specification_2010-05-05.htm

⁹⁵ Wiki do Protocolo TAPIR <http://wiki.tdwg.org/TAPIR>

⁹⁶ TapirSoftware <http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirSoftware>

Além dos *wrappers* existem softwares livres que também implementam os serviços do protocolo TAPIR e podem ser utilizados e estendidos de forma simples. O mais utilizado é o **TapirLink**, que será comentado a seguir.

4.1.4 TapirLink

O **TAPIRLink**⁹⁷ é uma ferramenta genérica para provedores TAPIR baseado no **PHP DiGIR provider**. Ele utiliza a biblioteca **PHP ADOdb** para acessar diferentes tipos de bancos de dados relacionais. Uma única instância do TapirLink possibilita o acesso a múltiplas fontes TAPIR, cada uma com seus serviços próprios. O TapirLink também possibilita que cada fonte de dados seja mapeada com uma ou mais camadas de abstração de dados e que as respostas das suas solicitações sejam retornadas em diferentes formatos.

O processo de instalação e configuração do TapirLink pode ser verificado a partir do **Manual do TapirLink**⁹⁸.

4.1.5 Referências

DIGIR (Distributed Generic Information Retrieval). **Table of Contents**. Disponível em: <<http://digir.net/>>. Acesso em: 30 jun. 2013.

BIOCASE PROTOCOLS. Disponível em: <<http://www.biocase.org/products/protocols/>>. Acesso em: 30 jun. 2013.

SALVANHA, P. **Modelo para análise de desempenho do processo de replicação de dados em portais de biodiversidade**. Dissertação (Mestrado). Disponível em: <http://www.teses.usp.br/teses/disponiveis/3/3141/tde-11082010-161042/pt-br.php>. Acesso em: 30 jun. 2013.

SILVA, D. L.; CORREA, P. L. P. **Arquitetura de Interoperabilidade do Herbário Virtual de Plantas Repatriadas do JBRJ**: Projeto REFLORA. Universidade de São Paulo, 13 ago. 2012.

TAPIR (TDWG Access Protocol for Information Retrieval). Disponível em: <http://www.tdwg.org/dav/subgroups/tapir/1.0/docs/tdwg_tapir_specification_2010-05-05.htm>. Acesso em: 30 jun. 2013.

TAPIRLINK. **Tapir TWiki**. Disponível em: <<http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirLink>>. Acesso em: 30 jun. 2013.

⁹⁷ TAPIRLink <http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirLink>

⁹⁸ Manual do TapirLink <http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirLinkManual>

TAPIRLINK. **Tapir Link Manual**. Disponível em: <<http://wiki.tdwg.org/twiki/bin/view/TAPIR/Tapir-LinkManual>>. Acesso em: 30 jun. 2013.

TAPIRLINK. **Tapir Software**. TAPIR TWiki. Disponível em: <<http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirSoftware>>. Acesso em: 30 jun. 2013.

TAPIR TWiki. Disponível em: <<http://wiki.tdwg.org/TAPIR/>>. Acesso em: 30 jun. 2013.

4.2 IPT

4.2.1 Apresentação

O **Integrated Publishing Toolkit (IPT)**⁹⁹ é uma aplicação Web livre e de código aberto para a publicação de dados de biodiversidade. Desenvolvida pelo GBIF, com o apoio de diversas instituições de biodiversidade, é a ferramenta recomendada pelo GBIF para a publicação de dados no *GBIF Network*, embora também sejam aceitos pelo GBIF dados publicados por outras aplicações e protocolos, como TAPIR, DiGIR ou BioCASE.

O IPT é desenvolvido na linguagem de programação Java e possibilita a publicação de conteúdo de bancos de dados e arquivos texto utilizando os padrões abertos Darwin Core (DwC) e Ecological Metadata Language (EML).

4.2.2 Funcionamento

O IPT realiza o gerenciamento e a publicação de três tipos de dados:

- **Dados Primários de Biodiversidade (dados de ocorrências):** Consistem em textos ou dados multimídia que detalham a ocorrência de uma espécie, tais como descrição, local, data, coletor e condições da coleta.
- **Nomes de espécies (dados taxonômicos):** Trata-se de um “*checklist* taxonômico”, ou seja, um conjunto de nomes válidos (científicos e comuns) de espécies conhecidas.
- **Metadados de recursos:** Descrição das coleções de dados. Incluem a localização, proveniência, proprietário e qualidade dos dados das coleções, podendo também abranger aspectos técnicos.

⁹⁹ Integrated Publishing Toolkit (IPT) <http://www.gbif.org/ipt>

As principais funcionalidades do IPT são:

- Criação de coleções de dados a partir da importação de registros provenientes de bancos de dados diversos, arquivos texto estruturados (CSV), arquivos no formato EXCEL (XLS e XLSX) ou arquivos Darwin Core Archive (DwC-A);
- Gerenciamento dos metadados das coleções na própria interface Web da ferramenta. Os metadados são estruturados de acordo com o **GBIF Metadata Profile**¹⁰⁰, um subconjunto de elementos do padrão EML;
- Publica a coleção de dados no padrão Darwin Core Archive, no padrão *Ecological Metadata Language* (EML V2.1.1) e RSS (apenas para informações sobre as atualizações nas coleções);
- Registra o conjunto de dados/instituição no **GBIF Registry**¹⁰¹, possibilitando que seus dados sejam publicados no *GBIF Network*.

Para a publicação destes dados na rede GBIF, um provedor de dados deve ser criado e descrito por meio de metadados, a origem dos dados deve ser identificada pelo carregamento de arquivos ou pela configuração de uma conexão com um banco de dados, e a realização do mapeamento que possibilitará a correta transformação dos dados nos termos estabelecidos pelo padrão Darwin Core e pelas suas extensões, caso sejam utilizadas.

A documentação completa da ferramenta pode ser consultada no **manual de usuário do IPT**¹⁰². Outras informações relacionadas ao IPT, sua utilização e evolução podem ser consultadas na **Wiki do IPT**¹⁰³.

4.2.3 Referências

Gbif.org: Publishing. Disponível em: <<http://www.gbif.org/ipt>>. Acesso em: 30 jun. 2013.

gbif-providertoolkit - A Java web application for publishing biodiversity data on the web and the GBIF network - Google Project Hosting. Disponível em: <<http://code.google.com/p/gbif-provider-toolkit/>>. Acesso em: 30 jun. 2013.

GBRDS-Registry. Disponível em: <<http://gbrds.gbif.org/index#>>. Acesso em: 30 jun. 2013.

IPT2 Manual Notes - gbif-providertoolkit - A Java web application for publishing biodiversity data on the web and the GBIF network - Google Project Hosting. Disponível em: <<http://code.google.com/p/gbif-providertoolkit/wiki/IPT2ManualNotes?tm=6>>. Acesso em: 30 jun. 2013.

O TUAMA, E.; BRAAK, K. **GBIF Metadata Profile, Reference Guide**. GBIF, 2011.

¹⁰⁰ GBIF Metadata Profile http://www.gbif.org/orc/?doc_id=2820

¹⁰¹ GBIF Registry <http://gbrds.gbif.org>

¹⁰² Manual de usuário do IPT <http://code.google.com/p/gbif-providertoolkit/wiki/IPT2ManualNotes?tm=6>

¹⁰³ Wiki do IPT <http://code.google.com/p/gbif-providertoolkit>

4.3 HIT

4.3.1 Apresentação

O **Harvesting and Indexing Toolkit (HIT)**¹⁰⁴ é uma ferramenta criada pelo GBIF com o propósito de facilitar o trabalho de agregar registros e dados relacionados à biodiversidade. Trata-se de um software de gerenciamento de dados baseado em Java que realiza a construção de um conjunto de dados indexado, com base nos registros coletados de variados provedores de dados.

O conceito do HIT se mostra interessante devido ao cenário atual das coletas de dados de biodiversidade, onde as instituições muitas vezes utilizam diferentes protocolos (e versões) para o compartilhamento dos seus dados, gerando uma heterogeneidade de redes de publicação de difícil gestão.

Uma vez que as fontes de publicações que adotam um protocolo de transferência padrão não costumam adotar novos protocolos na mesma frequência com que estes são desenvolvidos, existe a necessidade de que os agregadores continuem trabalhando com todos os protocolos que ainda possuem suporte ao mesmo tempo em que se adaptam aos novos. Esse cenário de maior complexidade acaba por dificultar o trabalho dos “usuários finais”, responsáveis por coletar e agregar estes dados.

É nesse cenário que surge o HIT. A ferramenta conta com uma interface desenvolvida especialmente para abstrair a complexidade existente na transferência de dados e no conjunto de operações gerenciadas pelo usuário. Ao esconder as diferenças entre os protocolos e agregar todos os registros em uma base de dados única e padronizada, o HIT oferece um mecanismo de fácil acesso para o agregador ou publicador de dados, que tem controle sobre todos os registros coletados em um único ambiente.

4.3.2 Funcionamento

A ferramenta HIT suporta a coleta de registros nos três protocolos de maior utilização em publicações de biodiversidade:

- **DiGIR;**
- **BioCAsE;**

¹⁰⁴ Harvesting and Indexing Toolkit (HIT) <https://code.google.com/p/gbif-indexingtoolkit/>

- **TAPIR.**

Cada um desses protocolos permite ao usuário definir como suas requisições XML serão codificadas. O HIT dá suporte nativo às codificações (versões) listadas abaixo:

- **DiGIR**

- DwC 1.0
- DwC 1.4

- **BioCAsE**

- ABCD 1.2
- ABCD 2.06

- **TAPIR**

- DwC 1.4
- ABCD 1.2
- ABCD 2.06

O HIT também permite a importação de arquivos no padrão **Darwin Core Archive (DwC-A)**.

Esta ferramenta foi desenvolvida de forma a possibilitar uma extensa customização do usuário, de modo que é possível desenvolver de forma simples o suporte a novos protocolos ou novas versões dos protocolos e padrões já existentes.

No entanto, o suporte ao HIT foi descontinuado pelo GBIF, e a sua última versão oficial lançada foi a 1.43-RC1, de março de 2012.

Também foi identificado nesta última versão da ferramenta uma falha de segurança, causada por uma vulnerabilidade no Struts2, o framework computacional utilizado no desenvolvimento do HIT. Por isso, é recomendável a utilização desta ferramenta apenas em redes internas, sem sua disponibilização aberta na internet.

Como alternativas ao HIT, estão disponíveis na internet bibliotecas para processamento de arquivos Darwin Core Archive e para a comunicação utilizando os protocolos TAPIR e DiGIR. Estas bibliotecas podem ser utilizadas no desenvolvimento de ferramentas ou scripts para a agregação de dados provenientes de diversos provedores. Outra alternativa é a utilização de ferramentas já

¹⁰⁵ Manual de usuário do HIT <https://code.google.com/p/gbif-indexingtoolkit/wiki/UserManual>

¹⁰⁶ Wiki do HIT <https://code.google.com/p/gbif-indexingtoolkit/>

desenvolvidas pela comunidade científica, como o Biocache-Store (*Atlas of Living Australia*) e o Canadensys-Harvester (*Canadensys*).

A documentação completa da ferramenta pode ser consultada no **manual de usuário do HIT**¹⁰⁵. Outras informações relacionadas do HIT, sua utilização e evolução podem ser consultadas na **Wiki do HIT**¹⁰⁶.

4.3.3 Referências

GBIF (Global Biodiversity Information Facility). **Gbif.org: Indexing**. Disponível em: <<https://code.google.com/p/gbif-indexingtoolkit/>>. Acesso em: 30 jun. 2013.

GBIF (Global Biodiversity Information Facility). **gbif-indexingtoolkit - The GBIF Harvesting and Indexing Toolkit (HIT) - Google Project Hosting**. Disponível em: <<https://code.google.com/p/gbif-indexingtoolkit/>>. Acesso em: 30 jun. 2013.

GBIF (Global Biodiversity Information Facility). **User Manual - gbif-indexingtoolkit - User Manual aimed at end-users - The GBIF Harvesting and Indexing Toolkit (HIT) - Google Project Hosting**. Disponível em: <<https://code.google.com/p/gbifindexingtoolkit/wiki/UserManual>>. Acesso em: 30 jun. 2013.



Ceiba speciosa (*Chorisia speciosa*) Malvaceae (*Bombacaceae*)
<http://www.gbif.org/page/201>
Foto: Maurício Mercadante

5. Ferramentas de apoio

Na **Tabela 13** são apresentadas algumas das principais ferramentas e serviços de apoio utilizadas nos diversos processos de registro, mapeamento, geração de metadados, validação e apresentação dos dados. A maior parte destas ferramentas são livres e de código aberto (algumas com o projeto de desenvolvido encerrado), mas existem também ferramentas com versões pagas.

Com a utilização destas ferramentas alguns dos problemas mais comuns na manipulação de dados de biodiversidade podem ser simplificados e até solucionados.

Tabela 13 – Ferramentas de apoio no processo de gestão de dados de biodiversidade.

Ferramentas	Descrição	URL
API Google Maps	Biblioteca para a apresentação de dados em mapas nos navegadores de internet.	https://developers.google.com/maps/?hl=pt-br
Arc GIS Explorer	Visualizador de mapas gratuito.	http://www.esri.com/software/arcgis/explorer/download.html
Atlas of Living Australia	Conjunto de aplicações e ferramentas que formam a infraestrutura do Portal de Dados do Atlas of Living Australia.	https://github.com/AtlasOfLivingAustralia
Bibmaster	Ferramenta de base de dados para gestão de informações sobre biodiversidade, com ênfase na literatura e informações nomenclaturais.	http://www.gbif.es/bibmaster/bibmaster.php
BioLink	Ferramenta utilizada para a gestão, manutenção, análise e disseminação de dados taxonômicos, de biodiversidade e informações do meio ambiente.	https://code.google.com/p/biolink/wiki/BioLink
Bionomenclature Glossary	Compilação de termos utilizados na Bionomenclatura.	http://bionomenclature-glossary.gbif.org/
Canadensys	Conjunto de aplicações e ferramentas que formam a infraestrutura do Portal de Dados Canadensys.	https://github.com/Canadensys
CollectiveAccess	Ferramenta baseada na Web para a gestão completa de coleções de museus e demais coleções digitais.	http://collectiveaccess.org/

Ferramentas	Descrição	URL
Conversor	Ferramenta que converte diferentes tipos de representação de coordenadas geográficas e datum's.	http://splink.cria.org.br/conversor?criaLANG=pt
Coordinate conversion	Ferramenta para conversão de coordenadas geográficas de GMMSS para graus decimais.	http://data.canadensys.net/tools/coordinates
C-Squares Mapper	Ferramenta que converte latitudes e longitudes em c-squares (quadrados), que são georeferenciados, representado uma área da superfície terrestre.	http://www.marine.csiro.au/csquares/
Darwin Test	Ferramenta para validação de registros nos formatos Darwin Core, versões 1.2 e 1.4.	http://www.gbif.es/darwin_test/Darwin_Test_in.php
Data Profiler	Ferramenta que possibilita a análise das características de dados vindos de fontes de dados em diversos formatos (XML, JDBC ou CSV).	http://sourceforge.net/projects/dataprofiler/
DataONE Software Tools Catalog	Listagem de vários softwares utilizados nos diversos processos de armazenamento, compartilhamento, publicação, análise e gestão de dados biológicos.	http://www.dataone.org/all-software-tools
Date parsing	Ferramenta que identifica os padrões de datas e os padroniza.	http://data.canadensys.net/tools/dates
DC Meta Toolkit	Ferramenta que transforma dados coletados em diferentes formatos em metadados Dublin Core.	http://sourceforge.net/projects/dcmetatoolkit/
DC-Assist	Ferramenta simples de ajuda, para aplicações que utilizam o Dublin Core.	http://www.ukoln.ac.uk/metadata/dcassist/
DC-dot	Serviço que recupera uma página Web e gera automaticamente um registro de metadados Dublin Core.	http://www.ukoln.ac.uk/metadata/dcdot/
DELTA	Implementação em Java do DELTA (http://delta-intkey.com/). É um método flexível para a codificação de descrições taxonômicas para o processamento computacional.	https://code.google.com/p/open-delta/
DIVA-GIS	Sistema de informação geográfico gratuito.	http://www.diva-gis.org/
Dublin Core Generator	Ferramenta Web que possibilita a geração de um registro no padrão Dublin Core por meio do preenchimento de um formulário.	http://dublincoregenerator.com/generator_nq.html

Ferramentas	Descrição	URL
DwC Reader	Biblioteca Java que realiza a leitura de arquivos DwC-A.	https://code.google.com/p/darwincore/wiki/DarwinCoreArchiveReader
DwC-A Assistant	Assistente para o processo de criação de um arquivo DwC-A.	http://tools.gbif.org/dwca-assistant/
DwC-A Spreadsheet Processor	Assistente para a criação de um arquivo DwC-A a partir de dados armazenados em uma planilha eletrônica.	http://tools.gbif.org/spreadsheet-processor/
DwC-A Validator	Validador de arquivos Darwin Core Archive (DwC-A).	http://tools.gbif.org/dwca-validator/
dwca2sql	Ferramenta para a importação de arquivos Darwin Core Archive para um banco de dados relacional.	https://code.google.com/p/canadensys/wiki/dwca2sql
Earth Point	Disponibiliza grades geográficas (Lat./Lon., UTM, UPS, MGRS e USNG) no Google Earth.	http://www.earthpoint.us/Grids.aspx
Editor-Converter Dublin Core metadata	Ferramenta para edição de registros Dublin Core publicados na internet e sua conversão no formato UNIMARC.	http://www.library.kr.ua/dc/dcreditunie.html
Field Data Software	Pacote de software para Web com foco na comunidade científica para registrar e gerenciar dados de biodiversidade.	http://www.ala.org.au/get-involved/citizen-science/fielddata-software/
FindIt2Darwin Core	Ferramenta que extrai nomes científicos armazenados em documentos PDF, localizados em um endereço Web.	http://sourceforge.net/projects/findit2darwinco/
GBIF Data Portal	Aplicação Web do antigo Portal de Dados do GBIF. Apesar de descontinuado, ainda é bastante utilizado na geração de novos Portais de Dados de Biodiversidade.	URL: https://code.google.com/p/gbif-dataportal/
GBIF Resource Browser	Navegador para pesquisa e consulta de informações referentes aos termos do Darwin Core e suas extensões.	http://tools.gbif.org/resource-browser/
GBIF Vocabularies	Portal que disponibiliza informações sobre a utilização de diversos vocabulários e extensões para padrão Darwin Core.	http://vocabularies.gbif.org/

Ferramentas	Descrição	URL
Geo Locate	Ferramenta para georeferenciamento a partir da análise de descrições textuais de localização.	http://www.museum.tulane.edu/geolocate/
Geo Network	Projeto livre e de código aberto para catalogação de recursos espacialmente referenciados.	http://geonetwork-opensource.org/
geo:truc	Ferramenta para a obtenção de coordenadas a partir do clique em mapas. É possível a geração das coordenadas em diversos padrões.	http://www.geotruc.net/?hl=es_ES
geoLoc	Ferramenta que auxilia o trabalho de georeferenciamento dos acervos das coleções biológicas.	http://splink.cria.org.br/geoloc?criaLANG=pt
GeoServer	Servidor de código aberto que possibilita o compartilhamento, processamento e edição de dados geoespaciais.	http://geoserver.org/display/GEOS/Welcome
GeoTrans	Ferramenta que possibilita a conversão geográfica de coordenadas entre uma ampla variedade de sistemas de coordenadas.	http://earth-info.nima.mil/GandG/geotrans/
Google Earth	Aplicação Desktop e Web para visualização de mapas e dados geoespaciais.	http://www.google.com.br/intl/pt-BR/earth/index.html
Google Earth	Ferramenta para visualização de modelos tridimensionais do globo terrestre.	http://www.google.com.br/intl/pt-BR/earth/index.html
Google Maps	Serviço de Mapas na Web que também disponibiliza fotos de satélite, rotas e diversas informações geográficas.	http://googlemaps.com
GPS Visualizer	Ferramenta para visualizar pontos e trajetos em um mapa, permitindo a correção de coordenadas.	http://www.gpsvisualizer.com/
Herbar	Ferramenta para gestão de herbários.	http://www.gbif.es/herbar/herbar.php
HZL	Ferramenta para informatização e gestão de coleções botânicas e zoológicas.	http://www.gbif.es/hzl/hzl.php

Ferramentas	Descrição	URL
I3GEO	Software baseado no MapServer que possibilita o compartilhamento, processamento e edição de dados geoespaciais.	http://www.softwarepublico.gov.br/ver-comunidade?community_id=1444332
infoXY	Ferramenta que auxilia na validação de dados geográficos.	http://splink.cria.org.br/infoxy?criaLANG=pt
iPlant Collaborative	Ferramenta para validação de nomes científicos de plantas.	http://tnrs.iplantcollaborative.org/
Itislib	Biblioteca Java que disponibiliza uma API para acesso a base de dados taxonômica do ITIS*ca.	http://knb.ecoinformatics.org/software/download.jsp#itislib
Jalama	Ferramenta para a construção flexível e customizada de formulários de entrada de dados.	http://jalama.ecoinformatics.org/
Kepler	Sistema de workflow científico que permite a criação de workflows e sua execução eficiente utilizando diversas abordagens.	https://kepler-project.org/
Lista de Espécies da Flora do Brasil	Sistema de gestão da Lista de Espécies da Flora do Brasil que realiza a validação taxonômica da flora do Brasil.	http://floradobrasil.jbrj.gov.br/
MapServer	Servidor de código aberto que possibilita o compartilhamento, processamento e edição de dados geoespaciais.	http://www.mapserver.org/
Metacat	É uma arquitetura de banco de dados de metadados flexível, que utiliza XML como sintaxe comum para representar padrões de conteúdo pelos ecologistas.	http://knb.ecoinformatics.org/software/metacat/
Monarch	Ferramenta para a análise exploratória de dados. Seu objetivo é apoiar a análise e síntese de arquivos de dados por meio de metadados.	http://knb.ecoinformatics.org/software/monarch
Morpho	Ferramenta que permite a criação e gestão de metadados em EML, além de seu compartilhamento.	http://knb.ecoinformatics.org/morphoportal.jsp
My Meta Maker	Ferramenta <i>online</i> para criação e edição de metadados de publicações <i>online</i> (acesso aberto).	http://www.isn-oldenburg.de/services/mmm/

Ferramentas	Descrição	URL
Name Parser	Serviço Web do ECAT que analisa nomes científicos e realiza a divisão do nome nos seus componentes.	http://tools.gbif.org/nameparser/
Name Parser Access	Versão em MS Access da ferramenta Name Parser, que analisa nomes científicos e os divide em componentes (gênero, espécie, autor etc.)	http://www.gbif.es/name_parser/Name_Parser.php
Omeka	Ferramenta livre, flexível e aberta para a publicação na internet de dados de bibliotecas, museus e coleções acadêmicas.	http://omeka.org/
Open Layers	Biblioteca Javascript código aberto para a visualização de dados em mapas nos navegadores de internet.	http://www.openlayers.org/
Open Modeller	Ferramenta para modelagem de distribuição espacial, utilizada para prever a distribuição de espécies.	http://openmodeller.sourceforge.net/
Open Refine	Antigo Google Refine, é uma ferramenta para manipulação de dados confusos e bagunçados, possibilitando sua transformação e limpeza.	http://openrefine.org/
Photo RDF-Gen	Ferramenta que gera arquivos RDF para descrição de imagens, especialmente fotográficas.	http://www.webposible.com/utilidades/photo_rdf_generator_en.html
PostGIS	Extensão espacial para o banco de dados PostgreSQL, permitindo a manipulação de dados geoespaciais.	http://postgis.net/
Quantum GIS	Desktop GIS livre e código aberto, que possibilita a visualização, edição e análise dos dados.	http://www.qgis.org/
Rapid Color Guides	Repositório de guias de campos.	http://fm2.fieldmuseum.org/plantguides/
SandBox	Ferramenta para validação de arquivos CSV utilizando o padrão Darwin Core.	http://sandbox.ala.org.au/datacheck/
Simple DwC Terms (Access)	Tabela do Simple Darwin Core no MS Access.	http://darwincore.googlecode.com/files/SimpleDwCMSAccess.mdb

Ferramentas	Descrição	URL
Simple DwC Terms (Excel)	Esquema do Simple Darwin Core no MS Excel.	http://darwincore.googlecode.com/files/SimpleDwCMSExcel.xls
speciesBase	Base de dados genérica para gestão de registos taxonômicos.	http://splink.cria.org.br/speciesbase?criaLANG=pt
speciesMapper	Ferramenta que possibilita a visualização de coordenadas em um mapa.	http://splink.cria.org.br/mapper
spOutlier	Ferramenta de detecção de <i>outliers</i> para os parâmetros latitude, longitude e altitude.	http://splink.cria.org.br/outlier?criaLANG=pt
Storage Resource Broker (SRB)	É uma arquitetura cliente-servidor para interligar uma rede distribuída de recursos de dados heterogêneos.	http://www.npaci.edu/DICE/SRB/index.html
Talent Open Profiler	Ferramenta de análise de dados que define o conteúdo, estrutura e qualidade de dados complexos.	http://www.talend.com/products/data-quality
Taxon Tagger	Aplicação Web que localiza, identifica e extrai nomes científicos de páginas Web e arquivos PDF.	http://tools.gbif.org/taxontagger/
TDWG Subgroup on Biological Collection Data	Listagem de vários softwares utilizados na gestão de coleções biológicas (softwares livres e produtos comerciais).	http://www.bgbm.org/TDWG/acc/Software.htm
TerraLib	Biblioteca GIS de código aberto que possibilita que bancos de dados relacionais manipulem dados geoespaciais.	http://www.terralib.org/
Text Extraction Service Help	Serviço para extração de textos de arquivos binários em formatos como PDF, MS Word, MS Excel, MS PowerPoint entre outros.	http://ecat-dev.gbif.org/ws-tika/
Tools API	API para utilização das ferramentas de conversão de coordenadas e análise de datas.	http://data.canadensys.net/tools/api
Viewer-Generator Dublin Core metadata	Ferramenta online para testar se uma URL possui um registro de metadados Dublin Core. Também gera automaticamente um registro DC para páginas que possuem apenas metadados HTML.	http://www.library.kr.ua/dc/lookatdce.html
Zoorbar	Aplicação para gestão de coleções de História Natural.	http://www.gbif.es/zoobar/zoobar.php



Mico-Leão-Dourado
Foto: Wigold B. Schäffer

Por ordem do

 Ministério Federal
do Meio Ambiente, Proteção da Natureza,
Construção e Segurança Nuclear

da República Federal da Alemanha

giz Experteise & Qualität
für internationale
Zusammenarbeit (GIZ) GmbH



Ministério do
Meio Ambiente

GOVERNO FEDERAL
BRASIL
PÁTRIA EDUCADORA