

TIMÓTEO MONTEIRO DA SILVA

AVALIAÇÃO TAXONÔMICA E GEOGRÁFICA DE UMA COLEÇÃO  
BIOLÓGICA E A CONTRIBUIÇÃO DE UM BANCO DE DADOS  
INFORMATIZADO PARA O CONHECIMENTO DA BIODIVERSIDADE,  
UM ESTUDO DE CASO EM COLEOPTERA (INSECTA)

BELÉM

2017

TIMÓTEO MONTEIRO DA SILVA

AVALIAÇÃO TAXONÔMICA E GEOGRÁFICA DE UMA COLEÇÃO  
BIOLÓGICA E A CONTRIBUIÇÃO DE UM BANCO DE DADOS  
INFORMATIZADO PARA O CONHECIMENTO DA BIODIVERSIDADE,  
UM ESTUDO DE CASO EM COLEOPTERA (INSECTA)

Trabalho de Conclusão de Curso apresentado ao  
Colegiado do Curso de Licenciatura em Ciências  
Biológicas, Modalidade Biologia da Universidade  
Federal do Pará, como requisito parcial para a  
obtenção do grau de Licenciado em Biologia.  
Orientadores: Dr. Cléverson Ranniéri Meira dos  
Santos - Coordenação de Zoologia – MPEG e  
Dr<sup>a</sup> Emília Zoppas de Albuquerque -  
Coordenação de Ciências da Terra e Ecologia –  
MPEG

BELÉM

2017

TIMÓTEO MONTEIRO DA SILVA

AVALIAÇÃO TAXONÔMICA E GEOGRÁFICA DE UMA COLEÇÃO BIOLÓGICA E A CONTRIBUIÇÃO DE UM BANCO DE DADOS INFORMATIZADO PARA O CONHECIMENTO DA BIODIVERSIDADE, UM ESTUDO DE CASO EM COLEOPTERA (INSECTA)

Trabalho de Conclusão de Curso apresentado ao Colegiado do Curso de Licenciatura em Ciências Biológicas, Modalidade Biologia da Universidade Federal do Pará, como requisito parcial para a obtenção do grau de Licenciado em Biologia.

Orientador: Dr. Cléverson Ranniéri Meira dos Santos  
Coordenação de Zoologia, MPEG

Coorientadora: Dr<sup>a</sup> Roberta de Melo Valente,  
ICB/UFPA

Avaliador: Prof. Dr. Marcos Paulo Alves de Souza.  
Serviço de Tecnologia da Informação, MPEG.

Avaliador: Dr<sup>a</sup>. Jéssica Herzog Viana  
Coordenação de Zoologia, MPEG.

BELÉM

2017

## APRESENTAÇÃO DA REVISTA:

TEMA: BIODIVERSIDADE

REVISTA: **ZooKeys**/Pensoft - é uma revista *peer-reviewed* de livre acesso, de rápida disseminação. A revista visa acelerar pesquisas através da livre troca de informações nas áreas da taxonomia, filogenia, biogeografia e evolução dos animais. ZooKeys, também, visa aplicar as mais recentes tendências e metodologias na publicação e preservação de materiais digitais para satisfazer os mais elevados padrões da época na área da *cyber* taxonomia.

## SUMÁRIO

**RESUMO**

**ABSTRACT**

<b>INTRODUÇÃO</b>	1
<b>MATERIAL E MÉTODOS</b>	2
<b>Informatização</b>	2
<b>Publicação de dados</b>	5
<b>Análise de dados</b>	5
<b>RESULTADOS</b>	10
<b>Informatização e publicação de dados da coleção de Coleoptera do Museu Paraense Emílio Goeldi</b>	10
<b>Abrangência taxonômica</b>	12
<b>Distribuição geográfica</b>	17
<b>DISCUSSÃO</b>	23
<b>CONCLUSÃO</b>	27
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	28

## LISTA DE FIGURAS

- Figura 1: Campos das colunas da planilha de dados correlacionadas com os campos do sistema, automaticamente. ----- 5
- Figura 2: Ferramenta de busca no programa “Specify” com a seleção dos campos e critérios utilizados na realização de uma busca no banco de dados da coleção. ----- 7
- Figura 3: Dados retornados, em planilha, através da busca realizada no banco de dados da coleção. ----- 7
- Figura 4: Diagrama de interações “front end” e “back end” utilizados para o gerenciamento e armazenamento dos dados das coleções biológicas do Museu Paraense Emílio Goeldi (MPEG). ----- 10
- Figura 5: Mapa com informação do local de coleta georreferenciado para os espécimes da Coleção de Coleoptera do Museu Emílio Goeldi. ----- 18
- Figura 6: Mapa de municípios do estado do Pará amostrados para a Coleção de Coleoptera do MPEG. ----- 19
- Figura 7: Mapas de distribuição gêneros analisados presentes na coleção de Coleoptera do Museu Paraense Emílio Goeldi pelos estados e países de ocorrência. **A:** Espécies de *Diabrotica*; **B:** Espécies de *Maecolaspis*, **C:** Espécies de *Syphraea* e **D:** espécies de *Typophorus*. ----- 21
- Figura 8: Mapas de distribuição dos gêneros analisados presentes na coleção de Coleoptera do Museu Paraense Emílio Goeldi por município brasileiro. **A:** Espécies de *Diabrotica*; **B:** Espécies de *Maecolaspis*, **C:** Espécies de *Syphraea* e **D:** espécies de *Typophorus*. ----- 22
- Figura 9. Número de espécimes de besouros coletadas ao longo de 209 anos ----- 23

# Avaliação taxonômica e geográfica de uma coleção biológica e a contribuição de um banco de dados informatizado para o conhecimento da biodiversidade, um estudo de caso em Coleoptera (Insecta)

Timóteo Monteiro da Silva<sup>1</sup>

<sup>1</sup> Museu Paraense Emílio Goeldi, Av. Perimetral, 1901 - Terra Firme, 66077-830, Belém, PA, Brasil

Autor Correspondente: Timóteo Monteiro da Silva (timoteomsilva@gmail.com)

## Resumo

Coleções biológicas hospedam e catalogam registros da biodiversidade mundial e agrupam uma série de informações úteis tanto em análises científicas bem como ao fomentar informação para o auxílio na gestão, monitoramento e prevenção de impactos ambientais. Contudo, muitos acervos possuem informações não atualizadas em termos de taxonomia e tampouco dados georreferenciados. O Museu Paraense Emílio Goeldi, dentre os vários acervos institucionais, possui a Coleção de Coleoptera com mais de 29 mil registros de besouros depositados ao longo de mais de 150 anos. Este acervo teve seus dados submetidos ao processo de informatização em massa, para que fosse possível armazenar e auxiliar no gerenciamento da curadoria e, por conseguinte publicar os dados para o livre acesso. O presente estudo fornece a descrição de uma metodologia de informatização, com o objetivo de gerar de forma mais eficaz informações provindas de um banco de dados, assim como descreve a abrangência taxonômica e geográfica do acervo de Coleoptera do Museu Goeldi e o uso destes dados para uma avaliação do banco através de um protocolo de construção de mapas de ocorrência. O protocolo de informatização do acervo inclui a digitação de dados a partir de fotos dos rótulos informativos, ambos sendo incorporados em um banco de dados através do programa de gerenciamento “Specify” com uso de outras ferramentas auxiliares. Isto permitiu organizar de forma estruturada as informações para que se realizasse a publicação *on-line* do acervo e também pesquisar taxonomicamente e geograficamente através da mineração de dados em softwares “business intelligence”. No total 29.029 registros foram informatizados e disponibilizados em plataformas que hospedam dados de biodiversidade como o Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr) e o The Global Biodiversity Information Facility (GBIF). Apenas 4,8 % dos dados possuíam as coordenadas geográficas de coleta, porém utilizando softwares, tipo “business intelligence”, foi possível avaliar o nível de ocorrências de grandes grupos bem como áreas sem amostragem. Como estudo de caso, utilizando quatro gêneros de Chrysomelidae, foi possível identificar erros de grafia nomenclatural e avaliar a representatividade taxonômica do acervo para cada grupo ao se comparar com o catálogo da fauna do Brasil e com o GBIF. Todos os gêneros analisados

demonstraram ter mais registros no acervo do Museu Goeldi que nas bases consultadas. Agregando os dados taxonômicos e geográficos em plotagem de mapas é possível demonstrar a distribuição dos registros de determinado táxon em estados e nas províncias da América do Sul, isto permite verificar a ocorrência em diferentes biomas. Em outra simulação, para a Amazônia, mapas demonstram a ocorrência de determinados táxons nos municípios e permitem identificar lacunas de amostragens biológicas, o que pode contribuir em futuros planejamentos científicos de coleta. Assim, o processo de informatização de um acervo biológico e o uso aqui empregado de softwares para mineração de dados mostrou-se uma ferramenta eficaz para caracterização do status de uma coleção científica biológica, e tais podem ser aplicados em outros bancos de dados, sejam de projetos de pesquisa ou mesmo aqueles que agreguem diversas fontes sobre a biodiversidade como o SiBBr.

**Palavras-chave:** mineração de dados, Coleoptera, banco de dados, coleções biológicas, informatização.



## **Abstract**

Biological collections host and catalog records of global biodiversity and group gather useful information for scientific analysis and to foment management, monitoring and prevention of environmental impacts. The Museu Paraense Emílio Goeldi, among the several institutional collections, has the Coleoptera Collection with more than 29 thousand records of beetles deposited over more than 150 years. This collection had its data submitted to the process of mass computerization, and through it, was possible to store and assist the management of the curatorship and, therefore, to publish the data for free access. The present study provides a description of computerization methodology, with the aim of generating information from a database more efficiently, as well as describing the taxonomic and geographical coverage of the Coleoptera collection of the Museu Goeldi and the use of this data for a avaluation of a database through a protocol and construction of maps of occurrence. The computerization protocol used Includes data typing and photograph of the specimen original labels. Those records are imported to a database through tools of the Specify Software. This allowed to organize the information in a structured way so that the online publication of the collection was carried out and also to search taxonomically and geographically through data mining in business intelligence software. In total, 29,029 records were computerized and made available on platforms that host biodiversity data such as the Brazilian Biodiversity Information System (SiBBr) and The Global Biodiversity Information Facility (GBIF). Only 4.8% of the data had the geographical coordinates of collection, but using business intelligence software it was possible to evaluate the level of occurrences of large groups as well as areas without sampling. As a case study, using four genera of Chrysomelidae, it was possible to identify nomenclatural errors and to evaluate the taxonomic representativeness of the collection for each group when compared with the “Catálogo da Fauna do Brasil” and with the GBIF. All genera analyzed showed more records in the collection of the Museu Goeldi than in the databases consulted. By adding the taxonomic and geographical data in mapping it is possible to demonstrate the distribution of the records of a given taxon in states and in the provinces of South America, this allows verify the occurrence in different biomes. For the Amazon, maps show the occurrence of certain taxa in the municipalities and allow identify gaps in biological sampling, which may contribute to future scientific planning of collection. Thus, the process of computerizing a biological collection and the use of data mining software used here has proven to be an effective tool for characterizing the status of a biological scientific collection, and can be applied to other databases, whether Research projects or even those that aggregate diverse sources on biodiversity such as SiBBr.

**Keywords:** datamining, coleoptera, database, biological collections, computerizing.

## INTRODUÇÃO

As coleções dos Museus de História Natural tem a missão de reunir informações sobre a diversidade biológica dos organismos (fósseis e atuais), sua distribuição, composição e servir como fonte para a educação e pesquisa, além de poder subsidiar a tomada de decisões por parte do poder público (Zaher & Young 2003; Winker 2004). Hoeksema (2011) explica que nestes museus estão hospedadas coleções históricas de espécimes botânicos e zoológicos que, juntos, são uma importante ferramenta na pesquisa em biodiversidade. De acordo com Shanmughavel (2007) a biodiversidade está distribuída por todo o mundo, mas a informação científica sobre ela está concentrada principalmente nos grandes centros, como nos museus de história natural, universidades e institutos de pesquisa. Essas informações já coletadas sobre biodiversidade mundial estão espalhadas em muitos bancos de dados, artigos ou em outros materiais que não permitem a pesquisa interativa e tão pouco foi armazenado em meios digitais (Shanmughavel, 2007).

O valor do conhecimento contido nessas coleções não é estritamente acadêmico ou de documentação, pois como esclarece Magalhães *et al.* (2001) coleções científicas biológicas possuem imenso potencial em termos econômicos e também culturais, com diversas possibilidades a oferecer, sejam estas em entretenimento, em divulgação de valores culturais de uma região ou relacionados à elementos da fauna e da flora. Magalhães *et al.* (2001) afirmam também que de posse do conhecimento existente, desde que organizados adequadamente, integrados e relacionados a informações de estudos que são aplicados, é possível oferecer embasamento de valor para planejar, estabelecer e acompanhar políticas públicas, programas, alterações ambientais, políticas conservacionistas e de manejo de recursos naturais, com destaque especial para a identificação do que ele chama de componentes da diversidade biológica capazes de conduzir à descoberta de novos recursos e possibilidades. Os dados presentes nos Museus de História Natural também podem contribuir de forma significativa, para a conservação de espécies de interesse, previsão da propagação de espécies invasoras, consequências da mudança climática para a biodiversidade (Graham, 2004). Um dos meios de centralizar informações sobre biodiversidade e de fomentar o uso desse recurso como fonte útil é a informatização. Apesar da necessidade de informatizar coleções científicas De Azevedo (1992) acrescenta que, essas iniciativas pessoais e/ou institucionais não podem dispensar uso de metodologia padronizada e adequada ao tratamento dos dados e que propicie uma recuperação de informações (através de buscas) que atendam plenamente as necessidades informacionais dos pesquisadores.

No Brasil, uma das principais ações relativas à informatização é o Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr) que integra as bases de dados das principais instituições brasileiras detentoras de dados biológicos à rede do Sistema Integrado de Biodiversidade Mundial (The Global Biodiversity Information Facility - GBIF). O SiBBr é uma plataforma *on-line* que facilita e fomenta a integração de dados, acesso à informação e publicação dessas informações. Associado a iniciativa deste sistema está o Museu Paraense Emílio Goeldi (MPEG), que desde 2014, é uma instituição parceira dessa plataforma. Nos últimos dois anos o MPEG publicou mais de 550.000 registros, distribuídos em 15 coleções pertencentes às áreas de Paleontologia, Zoologia e Botânica e que estão disponíveis à comunidade científica e ao público em geral para consulta, como fruto de uma política de informatização e de parcerias.

Um dos grandes desafios quanto à informatização de acervos biológicos está nas coleções entomológicas, uma vez que ultrapassam demasiadamente as demais em número de registros o que demanda uma enorme quantidade de trabalho para transmitir os dados de rótulos para bases computacionais. Assim, protocolos de informatização em massa devem ser construídos a fim de que otimizem o esforço agregado à qualidade da informação informatizada. Neste contexto, este trabalho visa analisar o processo de informatização em massa do acervo de Coleoptera da Coleção Entomológica do Museu Goeldi utilizando o *software* de gerenciamento de dados “Specify” como um estudo de caso a fim apresentar os principais aspectos da informatização e da publicação de dados (disponibilização “on line”). Além disso, apresenta formas de utilização deste banco de dados gerado para a exploração de informações sobre a abrangência taxonômica e geográfica do banco informatizado, através do uso de softwares de “business intelligence” para mineração de dados, com a finalidade de avaliar sua aplicabilidade em futuras ações de pesquisa sobre a biodiversidade.

## **MATERIAL E MÉTODOS**

### **Informatização**

Para a informatização foi implementado um sistema “open source” de gerenciamento, chamado “Specify” (Specify Software Project), voltado para coleções que utiliza banco de dados relacional, escalável e faz uso do padrão de dados biológico Darwin core (DWC). Dados estruturados em DWC com compartimentalizações da informação, unidades de medida semelhantes e conteúdos obrigatórios permitem a comparação de dados,

sem a necessidade de conversões, ou de aplicar retratamento aos dados para que estes sejam equivalentes.

O “Specify” foi desenvolvido no Departamento de Informática do Museu de História Natural da Universidade de Kansas em 2012. Foi concebido para ser multiplataforma (podendo ser executando em Windows, Mac OS X e Linux), sob a licença pública geral e é utilizado por mais de 247 museus de História Natural em todo o mundo (Miller, 2012). Foi desenvolvido para se tornar um banco de dados de gerenciamento de coleção altamente utilizável para coleções biológicas, com capacidade, não só para gerenciar coleções, mas também para facilitar as investigações sobre o material depositado nas coleções (Miller, 2012).

Cada coleção, a ser informatizada com o “Specify”, possui dados específicos para serem registrados. Essas peculiaridades surgem de acordo com o domínio da coleção e áreas de estudo associadas a elas. Em geral, coleções científicas biológicas possuem uma maneira única de curadoria, que estabelece as categorias que serão usadas para agrupar e armazenar as informações da coleção. Assim, antes da informatização propriamente dita é necessário estabelecer uma modelagem de dados que crie e categorize as informações presentes nas coleções. Essa modelagem predefinida é utilizada no sistema de informatização (*software*) e deve seguir modelos de dados internacionais, para permitir que estes dados possam ser utilizados para futuras comparações, e que possam ser relacionados com outros conjuntos de dados semelhantes. Para este processo foi preciso tanto o conhecimento da arquitetura do *software* “Specify” e seu esquema utilizado e da tendência atual de categorização de informação como o conhecimento sobre quais dados são armazenadas e quais dados seriam úteis se fossem adicionados aos já utilizados pela coleção fornecidos pelos curadores, técnicos de coleção e usuários.

O processo de informatização em massa de dados com o programa “Specify”, usado nesse estudo de caso, pode ser feito de duas formas: via importação de planilhas com dados tratados de acordo com o modelo proposto para a coleção ou pela inserção direta de dados via formulário. Um formulário é uma maneira de entrar e visualizar os dados. É uma interface do usuário para dados em uma tabela. Um formulário é essencialmente um modelo digital semelhante a um formulário de papel que pode ser preenchido para documentar um objeto de coleção (Miller, 2012). Os digitadores de dados fazem uso do formulário para registrar e salvar dados recuperados das etiquetas e livros tombo inserindo no sistema através da interface de entrada de dados do sistema escolhido. Para isso, esse ator do processo de informatização necessita de treinamento ou conhecimento prévio sobre os campos utilizados para preencher corretamente cada campo, com o tipo de dado que ele deve receber no domínio

de dados que corresponde à tabela de armazenamento correta. A desvantagem desse processo pode ser a geração de erros de digitação. É inevitável considerar o erro humano associado aos sistemas de *software*, pois os mecanismos do erro humano são inerentes ao processo cognitivo e por isso é esperado que este tipo de falha aconteça (Sandhof, 2006).

Para otimizar o processo de digitação, foi estabelecido um protocolo, onde os bolsistas digitadores, fotografavam as etiquetas de registro dos espécimes armazenados em uma gaveta, e partir destes registros fotográficos os dados foram conferidos e digitados no formulário de entrada de dados. Essa metodologia, reduz o manuseio dos espécimes e possibilita que a fonte de dados, as etiquetas originais, possam ser adicionadas como anexo aos registros e conferidas diretamente no sistema em caso de dúvidas.

A importação de planilhas de dados eletrônicas de outras fontes podem ser realizadas através de um protocolo de mapeamento pré-determinado para que os dados sejam os mesmos correspondentes ao banco previamente criado. Dessa forma, cada coluna de dados possui um cabeçalho com o nome da categoria do campo da tabela do banco de dados que irá receber a informação para que ocorra um auto mapeamento garantindo que os dados tratados serão armazenados no lugar específico. Este tipo de entrada de dados, é útil especialmente para coleções que possuem dados digitalizados em *softwares* como Microsoft Access ou Excel. Neste caso, o tratamento de dados é feito nas próprias planilhas, dividindo os dados por categorias, corrigindo dados e se possível realizando o preenchimento das lacunas de dados faltantes. Esse tratamento, divisão e correção dos dados garante que estarão de acordo com os padrões de publicação de dados biológicos institucionais para serem importados em lotes de no máximo 7.000 linhas, via a ferramenta de importação “*workbench*” do “Specify”. As colunas da planilha de dados são correlacionadas com campos de uma tabela do sistema (Figura 1), que pertence a um tipo de dado. Essa correlação pode ser feita automaticamente pelo sistema caso os cabeçalhos das colunas estejam com o mesmo nome dos campos a serem usadas ou essa relação pode ser feita manualmente.

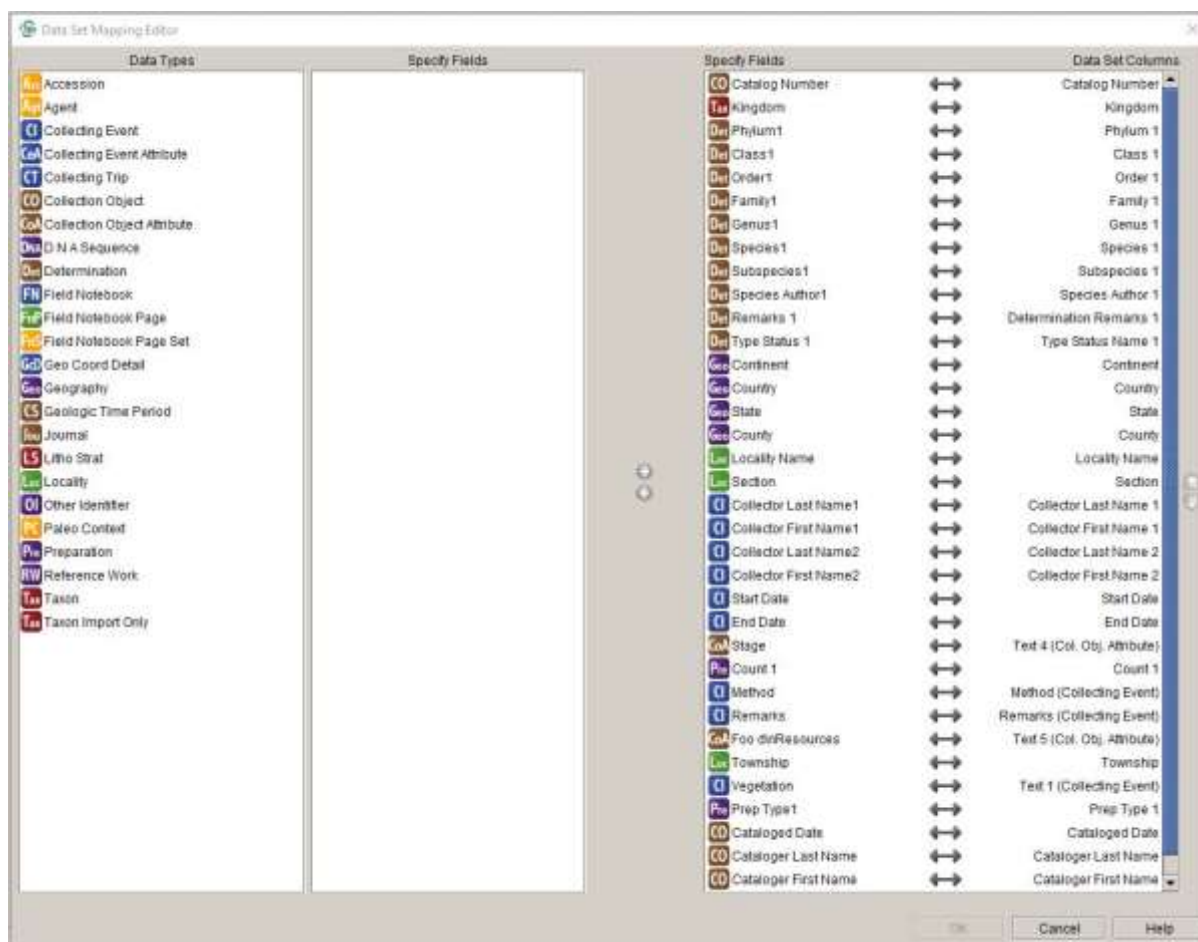


Figura 1. Campos das colunas da planilha de dados correlacionadas com os campos do sistema, automaticamente através da ferramenta apropriada inclusa no programa “Specify”. A esquerda aparecem os tipos de dados os quais podem ser associados e do lado direito, as colunas da planilha de importação, chamado “dataset” já relacionadas com os respectivos campos do programa.

Uma das vantagens de realizar a importação de dados através do “upload” de planilhas estruturadas é redução da possibilidade do erro, pelo fato de não depender da digitação da informação correta, uma a uma, nos campos específicos de cada registro.

## Publicação de dados

Foram exportados dados do “Specify”, através do ferramenta (“Data exporter”), que permite que todos os campos selecionados da base sejam exportados em formato de planilha, texto ou pacote de dados para serem publicados No SiBBR e GBIF. Essa publicação é feita através do “Toolkit Integrated Publishing” -IPT (Wieczorek *et al.* 2012). Essa ferramenta do GBIF é um pacote de *software* baseado em Java que provê a comunidade que lida com questões

de biodiversidade com um meio simples para executar muitas funções necessárias para publicar conjuntos de dados de biodiversidade na *web*, visando a usabilidade (Robertson *et al.*, 2014).

Os dados extraídos para publicação foram: Número de catálogo, Hierarquia Geográfica, Hierarquia Taxonômica, Informações de determinação, Informações do Evento de coleta e Preparação da amostra. Entretanto, além dos dados de ocorrência, é necessário que sejam adicionados metadados. De acordo com Feagraus *et al.* (2005) os metadados abrangem informações como: i) localização da instituição e curadoria, ii) descrição da abrangência geográfica, iii) direitos de uso, iv) imagem e v) lista das categorias geográficas e taxonômicas que aquela coleção hospeda. Este último metadado, a categoria taxonômica, facilita a busca de espécies, porque ferramentas de busca na “web” são capazes de retornar como resultado da busca de uma espécie, o metadado com o nome daquele nível taxonômico que está relacionado ao “dataset” publicado.

Para o cadastro dessa lista de todas a taxonomia da coleção no IPT pode ser feita através de um campo que importa todos os itens de um arquivo de texto em formato de lista, entretanto o nível taxinômico ao qual aquela informação pertence não pode ser adicionada em lote da mesma forma, o que significa que todos os elementos importados da lista precisarão ter essa informação preenchida. Como por exemplo, a lista de gêneros presente na coleção pode ser importada em lote, mas seria necessário clicar em cada elemento dessa lista depois de importado para através de um “picklist” marcá-los como “genus”. Por esse motivo, as instituições que utilizam o IPT, não preenchem essa informação do metadado sobre quais taxa estão presentes no “dataset”, só é possível saber essa informação baixando/consultado o conjunto de dados. Foi produzido, então, um código em “javascript” que faz o preenchimento automático e para isso necessita apenas ser executado no “browser”.

### **Análise de dados**

Após a inserção dos dados na base, eles estão disponíveis para a consulta através do programa “Specify”. Esse programa possui ferramentas que auxiliam nas buscas e geração de estatísticas, que permitem a seleção dos campos e os critérios de busca a serem utilizados em cada campo (Figura 2).

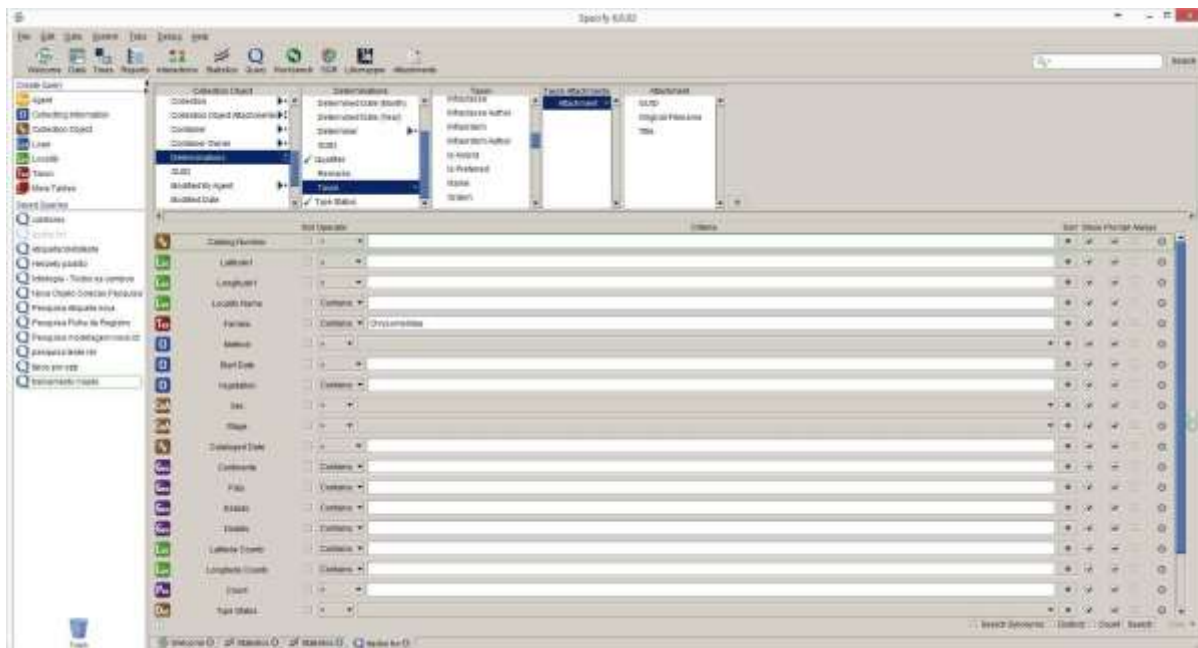


Figura 2. Ferramenta de busca no programa “Specify” com a seleção dos campos e critérios utilizados na realização de uma busca no banco de dados da coleção.

Os resultados das buscas realizadas na ferramenta “Query” do “Specify” são exibidos em uma tabela que pode ser exportada como planilha ou salva no próprio sistema como um conjunto de dados. Esses dados salvos podem ser acessados e utilizados como fonte para geração de documentos de empréstimo de espécimes, chamados “invoices” e etiquetas. As buscas são construídas com a seleção de vários campos da base de dados, e campos gerados pelo “Specify” a partir da interpretação dos campos originais. Essas buscas podem ser salvas, se estas forem de utilização recorrente para evitar que seja preciso construir uma nova pesquisa que atenda aos mesmos requisitos. Um exemplo de retorno de dados, após uma busca pode ser observado na Figura 3, onde cada coluna corresponde a um campo, exibindo os valores nelas armazenadas que atendiam aos critérios selecionados pelo usuário:

Catalog Number	Latitude	Longitude	Locality Name	Family	Method	Start Date	Vegetation	Sex	Stage	Cataloged Date	Collector	Country	State	Locality Code	Language Code	Type Status
MPIC HCO 1500118	-8.48888889	-50.25555556	Serra Nogueira	Chrysomelidae	Manual	28.05.1984				28.05.2011	América do Sul	Brasil	PA			Parasítico
MPIC HCO 1500119	-8.48888889	-50.25555556	Serra Nogueira	Chrysomelidae	Manual	28.05.1984				28.05.2011	América do Sul	Brasil	PA			Parasítico
MPIC HCO 1500117	-7.41628889	-50.53944444	Serra Nogueira	Chrysomelidae	Manual	10.05.1984				28.05.2011	América do Sul	Brasil	PA			Parasítico
MPIC HCO 1500119	-8.48888889	-50.25555556	Serra Nogueira	Chrysomelidae	Manual	28.05.1984				28.05.2011	América do Sul	Brasil	PA			Parasítico
MPIC HCO 1500118	-7.41628889	-50.53944444	Serra Nogueira	Chrysomelidae	Manual	27.11.1984				28.05.2011	América do Sul	Brasil	PA			Parasítico
MPIC HCO 1500115	-7.88055556	-51.88000000	Serra Nogueira	Chrysomelidae	Manual	17.05.1984				28.05.2011	América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501009			Marabá	Chrysomelidae		20.12.1982				28.05.2011	América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501001			Uruçá	Chrysomelidae		29.04.1981					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501009			Marabá	Chrysomelidae		27.08.1981					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501003			Marabá	Chrysomelidae		27.08.1981					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501004			Marabá	Chrysomelidae		27.08.1981					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501006			Marabá	Chrysomelidae		27.08.1981					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501005			Marabá	Chrysomelidae		27.11.1981					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501007			Marabá	Chrysomelidae		17.06.1982					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501008			Marabá	Chrysomelidae		19.06.1982					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501009			Opapitá dos Orombezes	Chrysomelidae		22.01.1981					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501008			Viana	Chrysomelidae		29.02.1981					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501001			Marabá	Chrysomelidae		10.06.1982					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501013			Marabá	Chrysomelidae		10.06.1982					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501015			Marabá	Chrysomelidae		10.06.1982					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501014			Marabá	Chrysomelidae		10.06.1982					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501015			Marabá	Chrysomelidae		10.06.1982					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501017			Marabá	Chrysomelidae		10.06.1982					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501018			Marabá	Chrysomelidae		10.06.1982					América do Sul	Brasil	PA			Parasítico
MPIC HCO 1501019			Marabá	Chrysomelidae		10.06.1982					América do Sul	Brasil	PA			Parasítico

Figura 3. Dados retornados, em planilha, através da busca realizada no banco de dados da coleção.



Outra manipulação dos registros, a mineração de dados, vai além das informações fornecidas pelo “software” de gerenciamento utilizado, ela busca informações que surgem das interações dos dados armazenados. Uma dessas tarefas de mineração possível é a análise exploratória de dados (EDA) ideal para filtrar dados e encontrar as relações inesperadas, apesar da limitação com “datasets” muito grandes. Essa técnica utiliza a análise de conjuntos de dados observacionais para encontrar relações insuspeitas e para resumi-los em novas formas que são compreensíveis e úteis para o proprietário dos dados (Hand & Manila, 2001).

Por esse motivo foi utilizado análises através de agrupamentos e comparações com a utilização de macros e fórmulas matemáticas em softwares de planilhas e no sistema de mineração de dados com software de “business intelligence”, que são “softwares” de T.I voltados para a análise e representação de dados, ideais para a geração de gráficos, e outras inferências sobre os dados informatizados. Associando dois ou mais campos em tabelas dinâmicas ou no sistema de mineração foi possível chegar a resultados que caracterizam a coleção e que permitem fazer reflexões e até mesmo, a tomada, de decisão sobre o status da coleção. Os “softwares” de “business intelligence” utilizados no processo de mineração, associação de categoria e filtragem de dados assim como a geração de mapas foram o Tableau software 9.3, CartoDB e o “software” de planilhas Excel.

O Tableau, “software” proprietário para “desktop”, é capaz de correlacionar campos de diversas categorias, como: informações geográficas, cadeia de caracteres, número inteiros, datas entre outras para a geração de novas formas de visualizar essas informações, inclusive gerando, para alguns casos, coordenadas a partir de campos de texto ou agregações. Vinte três tipos de representações de dados estão disponíveis no “software”, desde que as categorias de dados que são requisito para cada tipo de análise sejam fornecidas. Para este trabalho a foram gerados de mapas a partir de campos de texto com o nome de regiões conhecidas e georreferenciáveis. Foi utilizada uma ferramenta para representar a distribuição biogeográfica de alguns grupos importantes ou representativos da coleção. Além da representação da relação da ocorrência de todas as espécies do banco de dados por município, para os quatro gêneros mais abundantes, foram confeccionados mapas demonstrando a distribuição geográfica dessas espécies, pelos países (quando possível), estados ou províncias e municípios, dessa forma a ocorrência de uma espécie em uma destas categorias geográficas, permite identificá-la como área de ocorrência por gênero. Para complementar, a grande quantidade de registros (linhas) dos grupos taxonômicos alvo, foram representados de forma tabular, mostrando para espécie do grupo presente no banco em quais municípios qual era a

ocorrência, seguindo uma hierarquia, país, estado e município que ela ocorria, usando o banco de dados do MPEG como referência.

Com o relatório gerado através do sistema “Specify” e de usos de “softwares” de tratamento de dados, foi possível validar informações taxonômicas de famílias, identificando se estas eram nomenclaturalmente válidas (atuais e aceitas), e atribuir, caso estivesse ausente nos dados originais a autoria da descrição do nível taxonômico em questão. Para a aplicação deste estudo de caso, foram escolhidos os cinco gêneros de Chrysomelidae que estão associados ao maior número de registros, a fim de estabelecer a quantidade de determinações ausentes dentro dos referidos grupos. A validação taxonômica das famílias de Coleoptera presentes foi baseada no artigo de Bouchard *et al.* (2011), que apresenta uma revisão sobre as famílias de Coleoptera, enquanto a validação das espécies pertencentes aos gêneros mais abundantes, foi realizada com a consulta ao Catálogo “on line” da Fauna do Brasil (CTFB) e o GBIF “backbone taxonomy”. As informações taxonômicas foram contrastadas com as informações disponíveis “on line”, para determinar se haviam correspondências. Assim, as listas das espécies dos gêneros mais abundantes, foram caracterizadas como ausentes ou presentes na base do GBIF e CTFB. Quando presente, a informação de autoria também era verificada e comparada com o registro presente no banco de dados da coleção.

Outra inferência relevante foi a análise da quantidade de registros adicionados a coleção por ano, sendo possível determinar os anos de maior fluxo de coleta ou adesão de espécimes e correlacionar essa quantidade de dados com os coletores referenciados no banco.

Com o *software* de planilhas foi possível quantificar com precisão uma categoria de dados do banco, como os tipos mostrando as quantidades exatas, o que contribui para a caracterização da coleção, através da geração de gráficos ou tabelas.

Uma ferramenta de geração de mapas, que também foi utilizada para esse trabalho, foi a ferramenta “free” e “on line”, CartoDB. Esta é uma possível solução para a geração de mapas com pouco processamento, pois ao invés do Tableau que é executado localmente com os recursos de memória e processamento disponíveis, essa ferramenta “web”, pode receber dados em planilhas e processar em “background” nos servidores da aplicação e retornar uma informação útil. A limitação dessa ferramenta é a rede dados, que precisa comportar o envio de toda base de dados ou partes dela, para as plotagens.

## RESULTADOS

### Informatização e publicação de dados da coleção de Coleoptera do Museu Paraense Emílio Goeldi

Os registros foram armazenados numa arquitetura produzida para comportar o volume de dados e as necessidades de acesso de forma descentralizada ao bando de dados das coleções para isso foi implementada uma arquitetura cliente-servidor. Este sistema possui uma aplicação com a arquitetura cliente-servidor, que é definida por Lewis (1996) como “computação distribuída onde a aplicação é dividida em pelo menos duas partes: uma é executada por um ou mais computadores servidores e a outra por um ou mais computadores clientes”. Nesta arquitetura o sistema de “*front end*” roda na estação de trabalho (*desktops*) das coleções realizando requisições ao serviço que funciona no servidor em *back end* (servidor), que processa, executa e responde às solicitações. Estas duas camadas estão conectadas por uma rede, que pode ou não ser local. Na camada de “*back end*” opera o sistema de gerenciamento de banco de dados, o MySQL 5.7.1 onde as informações são centralizadas. O diagrama a seguir representa a as interações *front end e back end* (Figura 4).

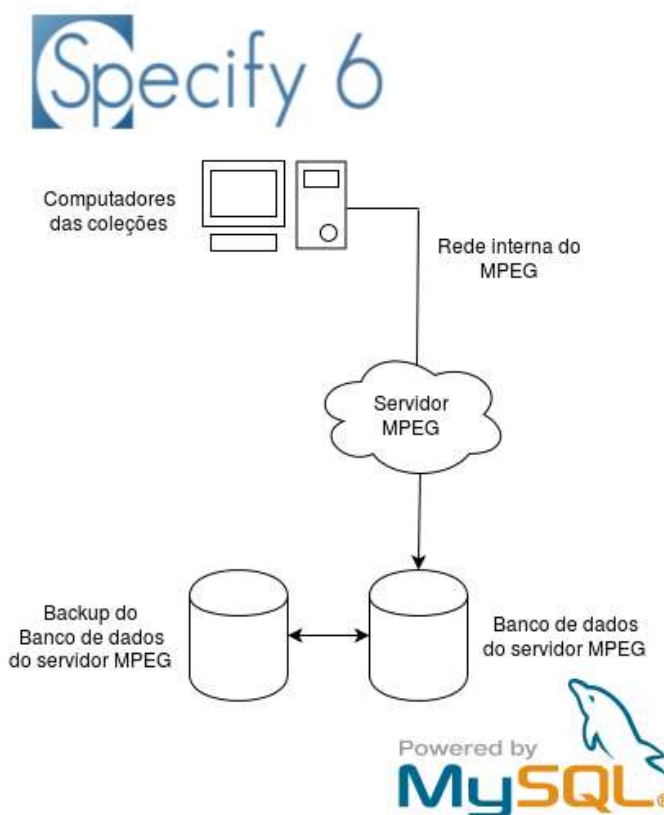


Figura 41: Diagrama de interações *front end e back end* utilizados para o gerenciamento e armazenamento dos dados das coleções biológicas do Museu Paraense Emílio Goeldi (MPEG).

Com essa arquitetura criada, o “Specify” foi configurado e conectado ao MySQL e a coleção de Coleoptera foi introduzida no banco, ficando esta disponível no banco para consultas, análises e publicação. Do volume de dados total, 21% dos registros da coleção foram importados por planilhas de dados, os outros 79% foram digitados por bolsistas associados ao projeto de informatização, seguindo o protocolo citado na metodologia. Mais de 22.980 imagens das etiquetas de tombamento dos espécimes foram geradas e estão anexadas aos registros no “Specify”. Após os esforços para a informatização em massa da coleção realizada ao longo de 12 meses, o número total de registros informatizados passou para 29.029. Os registros que ainda não estão publicados serão disponibilizados em breve. Os dados importados por planilha da coleção de Coleoptera do MPEG, 6.217 registros, estão publicados no SiBBr e no GBIF que também está disponível na página do IPT da instituição (<http://ipt.museu-goeldi.br/ipt/resource?r=museuparaenseemiliogoeldi-collection-entomologiacoleoptera>).

Para a publicação destes, um servidor hospedado no MPEG recebeu a instalação da ferramenta web IPT do GBIF. O museu paraense Emílio Goeldi foi cadastrado como Publicador do SiBBr e por consequência do GBIF, a partir desse momento foi possível criar eventos de publicação como o de Coleoptera. Dados, metadados e nomes dos campos foram preenchidos nessa plataforma para serem associados ao lote de dados.

Para facilitar o processo criação do metadado de abrangência taxonômica, foi desenvolvido e implementado, um código “Javascript” para automatizar o preenchimento de “picklists” com o valor desejado, para que fosse possível atribuir um valor de texto, contendo o nível taxonômico ao qual um valor listado no metadado pertence. Facilitando e possibilitando a visualização de quais grupos estão presentes no “dataset”, sem a necessidade de “download” dos dados.

Eis o código produzido a ser executado no “Browser”:

```
var count = (“começo”);
var ultimo = (“fim”);
var part1 = 'eml.taxonomicCoverages[0].taxonKeywords[';
var part2 = '].rank';
for(count; count <= ultimo; count++) { document.getElementById(part1+ count +
part2).innerHTML = "<option value='species'> species </option>"; }
"<option value='species'> species </option>"
```

Onde o “option value” é preenchido com o valor em texto do nível taxonômico a ser atribuído (Ex: “Genus”, “species”, etc.) para todos os itens que precisam dessa informação no metadado.

### **Abrangência Taxonômica**

Em termos taxonômicos a coleção de Coleoptera possui mais de 1.390 espécies, pertencentes a 551 gêneros distribuídos em 25 famílias. Dentre os espécimes registrados no banco de dados, 8.983 não estão identificados em nível específico, sendo possível, por isso, que o número de espécies presentes na coleção aumente à medida que esforços de identificação ocorram.

A coleção de tipos possui 468 espécies e hospeda 2.443 espécimes em diferentes modalidades: 225 são holótipos, 1.903 são parátipos, 169 são alótipos e 146 pertencentes a outras modalidades. Na base de dados da coleção foram encontrados 143 holótipos, 555 parátipos, 131 alótipos e outros 144 tipos associados a Jan Karel Bechyné e Bohumila Springlova Bechyné, pesquisadores que desenvolveram suas atividades de pesquisa na década de 1960 no MPEG. Com esses dados foi possível determinar que 63% dos holótipos, 29% parátipos, 77% dos alótipos e 78% dos outros tipos da coleção de Coleoptera foram descritos com o trabalho desses pesquisadores.

Das 211 famílias nomenclaturalmente válidas para Coleoptera (Bouchard *et al.*, 2011), a coleção possui representantes de 25 famílias, listadas a seguir juntamente com o número de registros associados: Anobiidae (1), Anthribidae (8), Attelabidae (22), Biphylidae (10), Brentidae (15), Cantharidae (23), Carabidae (62), Cerambycidae (382), Chrysomelidae (23.783), Cleridae (2), Coccinellidae (171), Curculionidae (3843), Elateridae (25), Endomychidae (3), Erotylidae (53), Histeridae (21), Hydrophilidae (16), Megalopodidae (4), Oedemeridae (1), Passalidae (254), Ptilodactylidae (1), Scarabaeidae (258), Scirtidae (7), Staphylinidae (1), Tenebrionidae (61); sendo Chrysomelidae e Curculionidae as mais abundantes em termos de registros. Do total de famílias da coleção, seis delas (Anthribidae, Biphylidae, Cleridae, Endomychidae, Megalopodidae e Scirtidae) possuem menos de 10 espécimes representantes e quatro outras (Anobiidae, Oedemeridae, Ptilodactylidae e Staphylinidae) têm apenas um representante. Esses dois grupos abrangem 44% das famílias da coleção e foram coletados até a década de 1990. Podemos inferir também que as famílias que possuem entre um e 61 espécimes presentes na coleção abrangem 77,8% das famílias do acervo de Coleoptera.

Do total de registros presentes na base de dados, 2 não estão identificados ao nível de família, 4.946 registros não estão identificados ao nível de gênero e 6.674 não possuem a identificação da espécie. Os gêneros mais abundantes na coleção são: *Typophorus* (1.209), *Syphraea* (1.228), *Maecolaspis* (1.052), *Diabrotica* (668) e *Parasyphraea* (638), todos pertencentes à Chrysomelidae. Para *Typophorus*, apenas 28 registros não estão identificados em nível específico, para *Syphraea*, *Maecolaspis* e *Parasyphraea* todos os registros estão identificados em espécie, enquanto para *Diabrotica* 141 espécimes ainda carecem de determinação específica.

Das dez espécies de *Typophorus* apenas uma estava presente nas duas bases *on line* (Catálogo da Fauna do Brasil e *Global Biodiversity Facility*) verificadas para a validação da existência taxonômica. A espécie, *Typophorus nigrinus*, foi a única entre todas as espécies analisadas que possuía correspondência nestes dois sistemas, ou seja, os nomes das espécies eram exatamente iguais. As informações sobre as autorias (autor e ano de descrição) de seis espécies desse gênero puderam ser atribuídas através do CTFB, enquanto o GBIF contribuiu com a informação de autoria de apenas uma espécie (Tabela 1). A comparação entre os nomes das espécies presentes no banco de dados e as informações correspondentes a elas encontradas, ou não, nas duas bases consultadas estão representadas na Tabela 1.

Tabela 1. Comparação entre as espécies de *Typophorus* presentes na base de dados da coleção de Coleoptera do MPEG e as informações presentes nas bases de dados do CTFB e GBIF, para validação taxonômica.

Coleção do MPEG	Espécies no CTFB	Espécies no GBIF
<i>Typophorus aculus</i>	-	-
<i>Typophorus basalis</i>	<i>Typophorus basalis</i> Baly, 1859	-
<i>Typophorus fasciatus</i>	<i>Typophorus fasciatus</i> (Germar, 1824)	-
<i>Typophorus florigradus</i>	<i>Typophorus florigradus</i> Bechyně & Špringlová, 1961	-
<i>Typophorus inflatus</i>	<i>Typophorus inflatus</i> Bechyně, 1958	-
<i>Typophorus nigrinus</i>	<i>Typophorus nigrinus</i>	<i>Typophorus nigrinus</i> (Fabricius, 1801)
<i>Typophorus notunadipennis</i>	-	-
<i>Typophorus rolundipennis</i>	-	-
<i>Typophorus signatus</i>	<i>Typophorus signatus</i> (Lefèvre, 1891)	-
<i>Typophorus simplex</i>	<i>Typophorus simplex</i> Lefèvre, 1884	-

Das 28 espécies representadas para o gênero *Syphraea*, metade pôde ser relacionada com registros do CTFB, acrescentando para todas as espécies a autoria, que no banco de dados do MPEG estava ausente. Duas possíveis entradas de dados errada foram identificadas no banco de dados do MPEG. A espécie *Syphraea striatipennis* não foi validada

em nenhuma das bases, mas outra, com nome similar *Syphraea strigatipennis*, que possui um caractere a mais, foi encontrada podendo representar correto. O GBIF não possuía nenhuma informação sobre essas espécies (Tabela 2). Outro caso encontrado foi da espécie *Syphraea cardiaca* que apesar de estar com a grafia errada em nosso banco de dados, o CTFB retornou a informação com a grafia correta, autor e ano de descrição sendo assim possível corrigi-la.

Tabela 2. Comparação entre as espécies de *Syphraea* presentes na base de dados da coleção de Coleoptera do MPEG e as informações presentes nas bases de dados do CTFB e GBIF, para validação taxonômica.

<b>Coleção do MPEG</b>	<b>Espécies no CTFB</b>	<b>Espécies no GBIF</b>
<i>Syphraea acervata</i>	-	-
<i>Syphraea alegrensis</i>	<i>Syphraea alegrensis</i> Bechyně, 1956	-
<i>Syphraea angularis</i>	<i>Syphraea angularis</i> Bechyně & Bechyně, 1961	-
<i>Syphraea bordoni</i>	-	-
<i>Syphraea calcorifera</i>	-	-
<i>Syphraea cardiaca</i>	<i>Syphraea cardiaca</i> Bechyně & Bechyně, 1961	-
<i>Syphraea crepidoderoides</i>	-	-
<i>Syphraea dira</i>	-	-
<i>Syphraea dirce</i>	<i>Syphraea dirce</i> Bechyně & Bechyně, 1961	-
<i>Syphraea erepidoderoides</i>	-	-
<i>Syphraea iricolor</i>	<i>Syphraea iricolor</i> Bechyně, 1956	-
<i>Syphraea jacareatinga</i>	-	-
<i>Syphraea melanomela</i>	-	-
<i>Syphraea melas</i>	<i>Syphraea melas</i> Bechyně, 1955	-
<i>Syphraea modesta</i>	<i>Syphraea modesta</i> (Jacoby, 1888)	-
<i>Syphraea multiimpressa</i>	<i>Syphraea multiimpressa</i> Bechyně, 1956	-
<i>Syphraea nelcallona</i>	-	-
<i>Syphraea nigrita</i>	<i>Syphraea nigrita</i> (Jacoby, 1899)	-
<i>Syphraea nygis</i>	-	-
<i>Syphraea opalina</i>	-	-
<i>Syphraea sparsata</i>	<i>Syphraea sparsata</i> Bechyně & Bechyně, 1961	-
<i>Syphraea speciosa</i>	<i>Syphraea speciosa</i> (Olivier, 1808)	-
<i>Syphraea sponea</i>	-	-
<i>Syphraea striatipennis</i>	-	-
<i>Syphraea strigatipennis</i>	<i>Syphraea strigatipennis</i> Bechyně & Bechyně, 1964	-
<i>Syphraea uberabensis</i>	<i>Syphraea uberabensis</i> Bechyně, 1955	-
<i>Syphraea utinga</i>	<i>Syphraea utinga</i> Bechyně & Bechyně, 1961	-
<i>Syphraea vulpia</i>	-	-

Das 33 espécies de *Maecolaspis* presentes no banco de dados do MPEG nenhuma informação sobre estas pôde ser encontrada no CTFB e no GBIF. Dessa forma não foi possível atribuir informações sobre o autor e o ano da descrição das espécies, nem verificar a grafia correta dos nomes.

As 32 espécies do gênero *Diabrotica* presentes no MPEG foram comparadas com as espécies presentes nas bases de referência, o que resultou em 18 validações nomenclaturais por correspondência com o CTFB e nenhuma correspondência destes dados com o GBIF. Restaram 14 espécies, que não puderam ser validadas segundo o mesmo critério em nenhuma base de dados, entretanto, deste total ainda foi possível identificar no CTFB, por similaridade ou correspondência parcial (compatibilidade de quase todos os caracteres do nome da espécie), os possíveis nomes válidos de quatro espécies grafadas de maneira incorreta no banco de dados do MPEG: *Diabrotica caricollis*, *Diabrotica emorsitaus*, *Diabrotica glaucilla*, *Diabroticas inderrupda*. Os nomes válidos destas espécies e as informações sobre as autorias obtidas, de acordo com o CTFB, são: *Diabrotica cavicollis* Baly, 1890; *Diabrotica emorsitans* Baly, 1890; *Diabrotica glaucina* Baly, 1889; *Diabrotica interrupta* Baly, 1889 (Tabela 3). Outra espécie foi encontrada escrita de duas de forma diferentes no banco de dados: *Diabrotica manaeusis* e *Diabrotica manaensis*, destas duas formas a primeira foi considerada um erro de inserção no banco e a segunda foi validada com o CTFB e por isso foi considerada correta. No total, 22 espécies desse grupo puderam ser validadas e 21 tiveram as informações de autoria adicionadas. (Tabela 3).

Tabela 3. Comparação entre as espécies de *Diabrotica* presentes na base de dados da coleção de Coleoptera do MPEG e as informações presentes nas bases de dados do CTFB e GBIF, para validação taxonômica.

Coleção do MPEG	Espécies no CTFB	Espécies no GBIF
<i>Diabrotica alegrens</i>	-	-
<i>Diabrotica aracatuba</i>	<i>Diabrotica aracatuba</i>	-
<i>Diabrotica atromaculata</i>	<i>Diabrotica atromaculata</i> Baly, 1889	-
<i>Diabrotica caricollis</i>	<i>Diabrotica cavicollis</i> Baly, 1890	-
<i>Diabrotica clarki</i>	<i>Diabrotica clarki</i> Weise, 1916	-
<i>Diabrotica distincta</i>	<i>Diabrotica distincta</i> Jacoby, 1882	-
<i>Diabrotica egleri</i>	<i>Diabrotica egleri</i> Bechyně & Bechyně, 1961	-
<i>Diabrotica elata</i>	<i>Diabrotica elata</i> (Fabricius, 1801)	-
<i>Diabrotica emorsitaus</i>	<i>Diabrotica emorsitans</i> Baly, 1890	-
<i>Diabrotica glaucilla</i>	<i>Diabrotica glaucina</i> Baly, 1889	-
<i>Diabrotica gracilentata</i>	-	-
<i>Diabrotica ilearpi</i>	-	-
<i>Diabrotica illigeri</i>	<i>Diabrotica illigeri</i> Baly, 1889	-
<i>Diabrotica inderrupda</i>	<i>Diabrotica interrupta</i> Baly, 1889	-
<i>Diabrotica limitata</i>	<i>Diabrotica limitata</i> (Sahlberg, 1823)	-
<i>Diabrotica liturata</i>	-	-
<i>Diabrotica macna</i>	-	-
<i>Diabrotica manaensis</i>	<i>Diabrotica manaensis</i> Weise, 1921	-
<i>Diabrotica manaeusis</i>	-	-
<i>Diabrotica melanocephala</i>	-	-



<i>Diabrotica nafolimbata</i>	-	-
<i>Diabrotica nitidicollis</i>	<i>Diabrotica nitidicollis</i> Baly, 1889	-
<i>Diabrotica pentamaculata</i>	-	-
<i>Diabrotica piceicornis</i>	-	-
<i>Diabrotica recki</i>	<i>Diabrotica recki</i> Marques, 1941	-
<i>Diabrotica samouella</i>	<i>Diabrotica samouella</i> Bechyně, 1956	-
<i>Diabrotica speciosa</i>	<i>Diabrotica speciosa</i> (Germar, 1824)	-
<i>Diabrotica tarcisia</i>	<i>Diabrotica tarcisia</i> Bechyně, 1971	-
<i>Diabrotica tijuquensis</i>	<i>Diabrotica tijuquensis</i> Marques, 1941	-
<i>Diabrotica travassosi</i>	<i>Diabrotica travassosi</i> Marques, 1941	-
<i>Diabrotica trifovedata</i>	<i>Diabrotica trifoveolata</i> Baly, 1890	-
<i>Diabrotica viridimaculata</i>	<i>Diabrotica viridimaculata</i> Jacoby, 1878	-

A validação taxonômica de *Parasyphraea* resultou em uma correspondência de 20 espécies presentes no CTFB, todas com informação sobre autoria. Novamente, o GBIF não retornou nenhuma informação (Tabela 4).

Tabela 4. Comparação entre as espécies de *Parasyphraea* presentes na base de dados da coleção de Coleoptera do MPEG e as informações presentes nas bases de dados do CTFB e GBIF, para validação taxonômica.

Coleção do MPEG	Espécies no CTFB	Espécies no GBIF
<i>Parasyphraea alaida</i>	<i>Parasyphraea alaida</i> Bechyně & Bechyně, 1961	-
<i>Parasyphraea attenuaticornis</i>	<i>Parasyphraea attenuaticornis</i> Bechyně & Bechyně, 1961	-
<i>Parasyphraea bicostulata</i>	<i>Parasyphraea bicostulata</i> Bechyně & Bechyně, 1961	-
<i>Parasyphraea biolena</i>	<i>Parasyphraea biolena</i> Bechyně, 1967	-
<i>Parasyphraea bordoni</i>	-	-
<i>Parasyphraea citrirubra</i>	<i>Parasyphraea citrirubra</i> Bechyně & Bechyně, 1964	-
<i>Parasyphraea desvia</i>	<i>Parasyphraea desvia</i> Bechyně & Bechyně, 1964	-
<i>Parasyphraea egléri</i>	<i>Parasyphraea egléri</i> Bechyně & Bechyně, 1965	-
<i>Parasyphraea granulosa</i>	<i>Parasyphraea granulosa</i> Bechyně & Bechyně, 1961	-
<i>Parasyphraea melanoxantha</i>	-	-
<i>Parasyphraea mesomera</i>	<i>Parasyphraea mesomera</i> Bechyně & Bechyně, 1964	-
<i>Parasyphraea nigriceps</i>	<i>Parasyphraea nigriceps</i> (Boheman, 1859)	-
<i>Parasyphraea nitrosimilis</i>	-	-
<i>Parasyphraea pantana</i>	<i>Parasyphraea pantana</i> Bechyně, 1967	-
<i>Parasyphraea phrymocera</i>	-	-
<i>Parasyphraea phrynocera</i>	-	-
<i>Parasyphraea pleurica</i>	<i>Parasyphraea pleurica</i> Bechyně & Bechyně, 1961	-
<i>Parasyphraea retroversa</i>	<i>Parasyphraea retroversa</i> Bechyně & Bechyně, 1961	-
<i>Parasyphraea rubricata</i>	<i>Parasyphraea rubricata</i> (Illiger, 1807)	-
<i>Parasyphraea sampaia</i>	<i>Parasyphraea sampaia</i> Bechyně & Bechyně, 1961	-
<i>Parasyphraea selecta</i>	<i>Parasyphraea selecta</i> Bechyně & Bechyně, 1961	-
<i>Parasyphraea sinuatella</i>	<i>Parasyphraea sinuatella</i> Bechyně & Bechyně, 1961	-
<i>Parasyphraea sulcifera</i>	<i>Parasyphraea sulcifera</i> Bechyně & Bechyně, 1961	-
<i>Parasyphraea ultrasimilis</i>	<i>Parasyphraea ultrasimilis</i> Bechyně, 1954	-
<i>Parasyphraea yeda</i>	<i>Parasyphraea yeda</i> Bechyně & Bechyně, 1961	-

## Distribuição geográfica

A coleção possui exemplares coletados em um intervalo de 209 anos, entre os anos de 1706 e 2015. Os registros do ano de 2016 ainda não estavam disponíveis no banco de dados e por isso não foram analisados.

Em termos de abrangência geográfica, a coleção apresenta representantes das Américas do Sul e Central, com 10 países representados, por quantidades distintas de registros: Argentina (23), Bolívia (136), Brasil (28.358), Chile (9), Colômbia (4), Guiana Francesa (2), Panamá (39), Peru (44), Uruguai (12) e Venezuela (62).

Os espécimes tipos também puderam ser quantificados e relacionados ao país de origem, sendo a maior parte destes representantes da biodiversidade brasileira. A proporção de tipos por país pode ser descrita da seguinte forma: Argentina (1), Bolívia (2), Brasil (2.382), Colômbia (2), Chile (1), Guiana Francesa (2), Peru (1), Uruguai (4) e Venezuela (48).

Para o Brasil, a análise de distribuição geográfica das espécies por estado mostrou que há espécimes na coleção de Coleoptera do MPEG, coletados em 23 estados brasileiros, além do Distrito Federal, não possuindo amostras para os estados de Alagoas, Pernambuco e Sergipe. Dentre os estados amostrados, o Pará é o mais representado com 14.016 besouros coletados, o que representa 48% do total. Cinco estados brasileiros onde houveram coletas possuem pelo menos um dado georreferenciado, entretanto, a maior parte deste tipo de dado também é para estado do Pará (Tabela 6).

Tabela 5. Número de registros com informações sobre as coordenadas geográficas por estado brasileiro.

<b>Estado</b>	<b>Número de registros com coordenadas geográficas</b>
Amazonas	12
Goiás	01
Mato Grosso	03
Pará	1.369
Santa Catarina	03
Não determinado	26

Analisando as informações sobre georreferenciamento foi identificado que apenas depois do ano 2.000, ocorreram coletas que associavam coordenadas aos espécimes, representando um total de 1.412 registros. Deste total, 679 foram coletados no município de Parauapebas que possui a maior amostragem com este tipo de dado para o estado do Pará. Dos registros, 87% não possuem informações sobre coordenadas geográficas e foram coletados até o final da década de 1990 (Figura 9).



Figura 5. Mapa com informação do local de coleta georreferenciado para os espécimes da Coleção de Coleoptera do Museu Emílio Goeldi.

As análises de distribuição geográfica revelaram que apenas 178 espécimes, 71 com coordenadas, foram coletados na Floresta Nacional de Caxiuanã onde o MPEG possui uma base de pesquisa científica desde 1993. Esses espécimes foram fruto de coletas que ocorreram em dois períodos, de maio de 1991 até 2000 e de setembro de 2010 até outubro de 2015 com 41 e 137 espécimes respectivamente. Desse último grupo todos são parátipos de *Andranthobius setirostris* Valente & Silva, 2014. Esta é a primeira descrição de uma espécie para esse gênero na Amazônia, estando o holótipo, alótipo e 320 parátipos depositados na coleção Entomológica do Museu Paraense Emílio Goeldi.

O uso da ferramenta de “business intelligence” permitiu a plotagem de um mapa de distribuição das coletas no estado do Pará utilizando o município de coleta como referência (Figura 6).

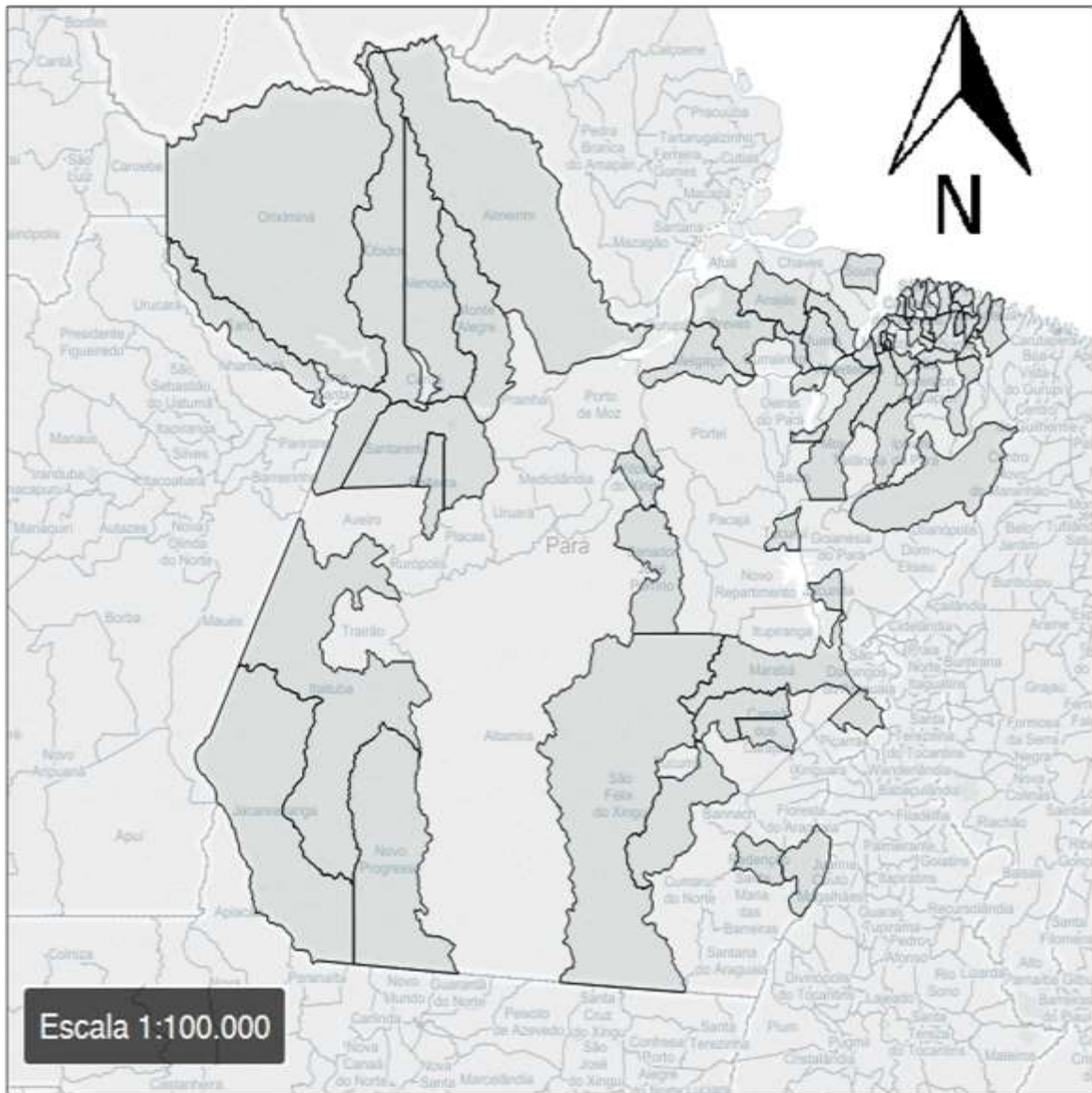


Figura 6. Mapa de municípios do estado do Pará amostrados para a Coleção de Coleoptera do MPEG.

Municípios, demarcados: Abaetetuba, Acará, Alenquer, Almeirim, Anajás, Ananindeua, Barcarena, Belém, Belterra, Benevides, Bragança, Breves, Bujaru, Canaã dos Carajás, Capanema, Capitão Poço, Castanhal, Conceição do Araguaia, Curuçá, Faro, Igarapé-Açu, Itaituba, Jacareacanga, Jacundá, Juruti, Limoeiro do Ajuru, Marabá, Maracanã, Marapanim, Marituba, Melgaço, Mocajuba, Moju, Monte Alegre, Muaná, Novo Progresso, Óbidos, Oriximiná, Ourém, Ourilândia do Norte, Ponta de Pedras, Paragominas, Parauapebas, Peixe Boi, Primavera, Redenção, Salinópolis, Santa Bárbara do Pará, Santa Izabel do Pará, Santarém, Santarém Novo, Santo Antônio do Tauá, São Caetano de Odivelas, São Domingos do Capim, São Felix do Xingu, São Francisco do Pará,

São Geraldo do Araguaia, São João de Pirabas, São Miguel do Guamá, São Sebastião da Boa Vista, Senador José Porfírio, Soure, Tomé-Açú, Tracuateua, Tucuruí, Vigia, Vitória do Xingu e Viséu.

Outras análises específicas foram realizadas para determinar distribuição geográfica para os quatro gêneros mais abundantes da coleção. Estes resultados foram representados em mapas (Figura 7).

O gênero, mais abundante, *Typophorus*, possui uma ampla distribuição no Brasil, podendo ser considerado um gênero de distribuição cosmopolita. Este ocorre na maior parte dos estados brasileiros, exceto pelos estados do Acre, Mato Grosso do Sul, Rio Grande do Sul, Rio de Janeiro, Tocantins, Piauí, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas e Sergipe. Não há registros deste gênero para outros países nas Américas, exceto pela Argentina. O segundo gênero mais abundante na coleção, *Syphraea*, possui ocorrência, para Bolívia, Venezuela e Brasil, em oito estados ou províncias, e sete municípios brasileiros. *Maecolaspis*, também ocorre na Bolívia, nas províncias de La paz e Aragua, assim como no Peru e Brasil, onde não há registros de ocorrência apenas para dez estados: Acre, Mato Grosso do Sul, Tocantins, Piauí, Rio Grande do Norte, Pernambuco, Sergipe, Alagoas, Maranhão e Bahia (Figura 7).

O último gênero analisado, *Diabrotica*, possui ocorrências para três países além do Brasil, sendo estes: Panamá, Bolívia e Argentina, entretanto as informações dos estados e municípios estavam presentes somente para o Brasil, referentes a 13 estados e para o Distrito Federal (Figura 7).

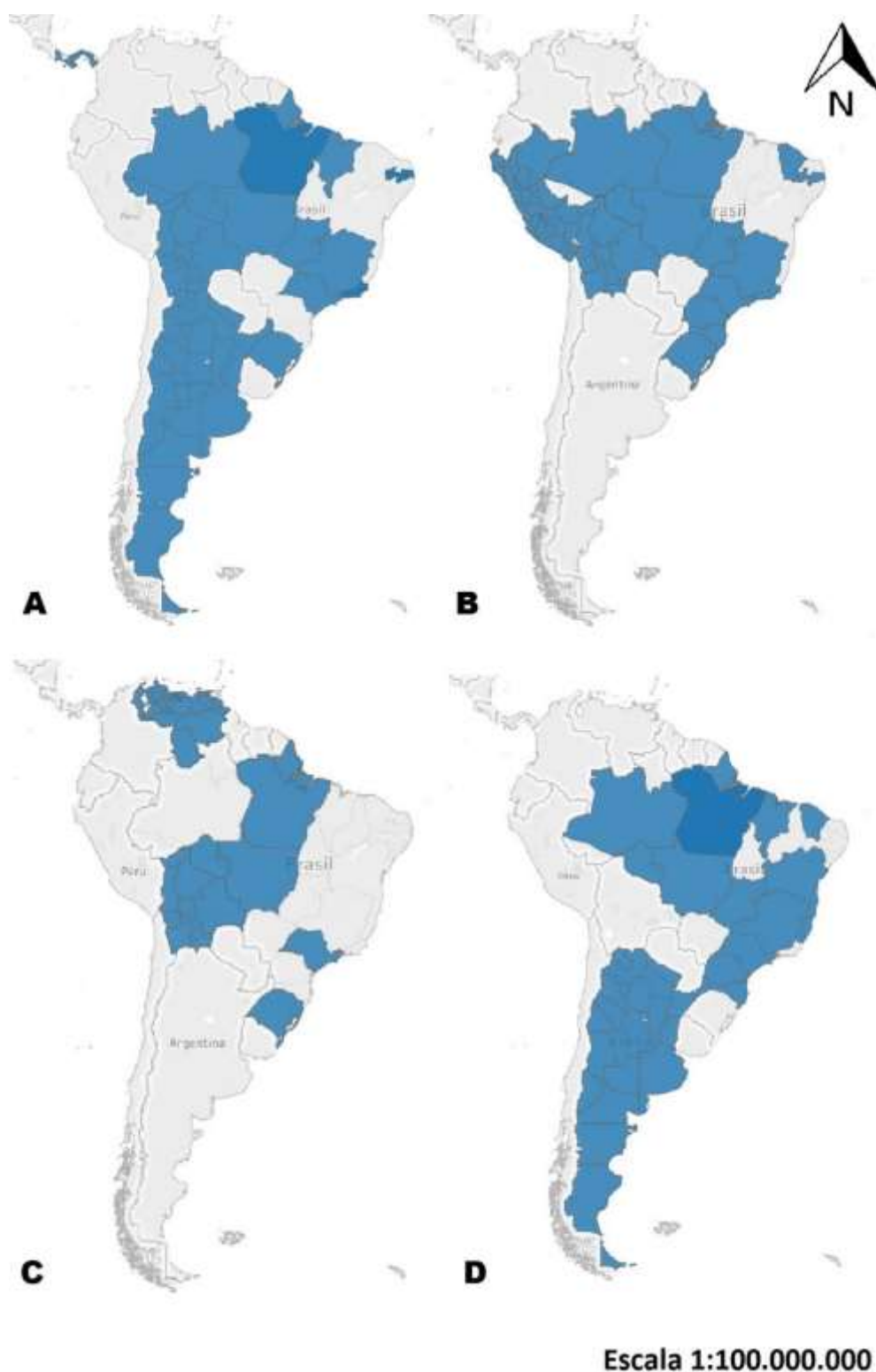


Figura 7. Mapas de distribuição gêneros analisados presentes na coleção de Coleoptera do Museu Paraense Emílio Goeldi pelos estados e países de ocorrência. **A:** Espécies de *Diabrotica*; **B:** Espécies de *Maecolaspis*, **C:** Espécies de *Sypbraea* e **D:** espécies de *Typophorus*.

A distribuição dos gêneros por municípios amostrados deixou evidente as regiões não amostradas para esses grupos.

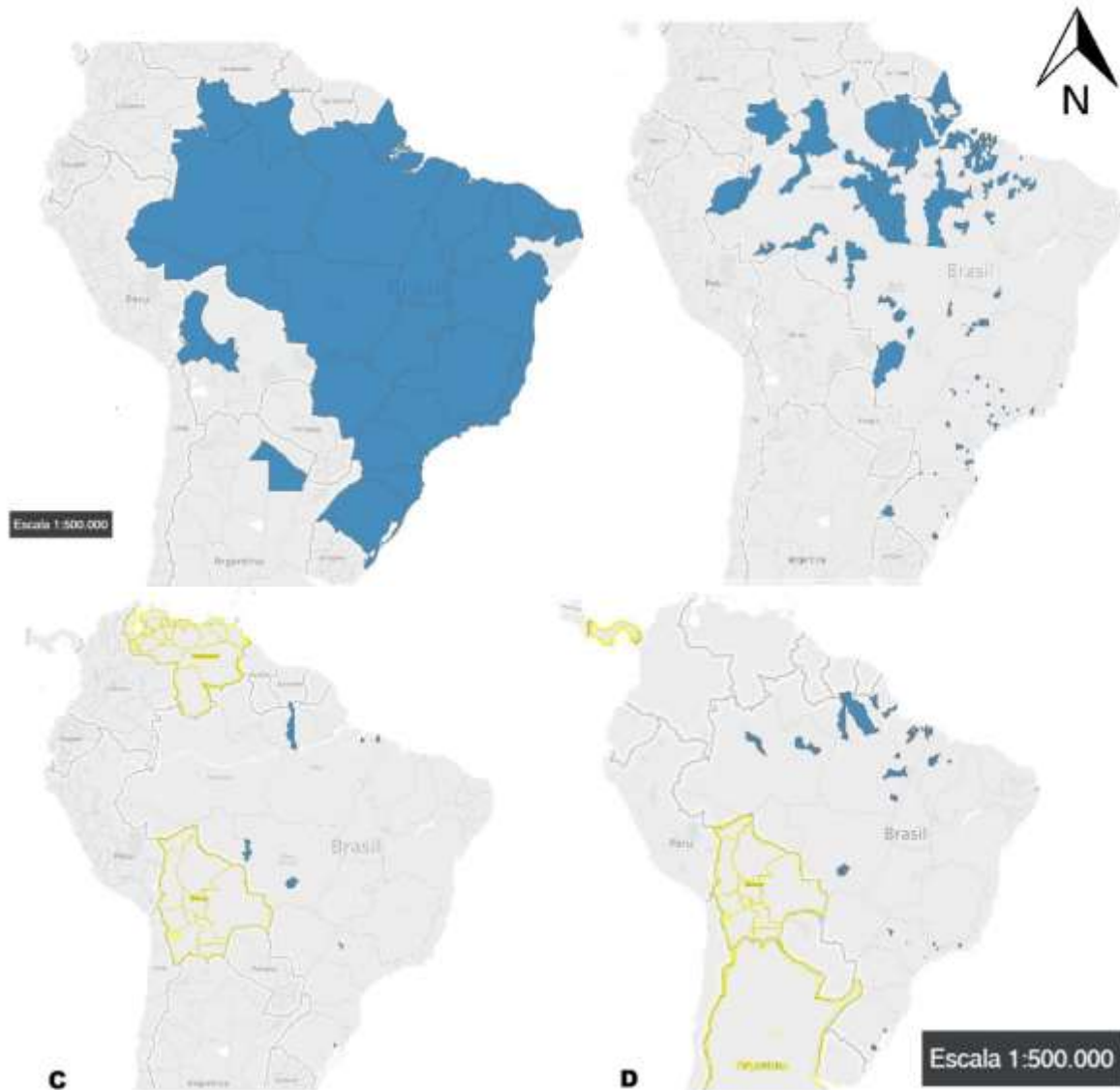


Figura 8. Mapas de distribuição dos gêneros analisados presentes na coleção de Coleoptera do Museu Paraense Emílio Goeldi por município brasileiro. **A:** Espécies de *Diabrotica*; **B:** Espécies de *Maecolaspis*, **C:** Espécies de *Syphraea* e **D:** espécies de *Typophorus*.

Cinco países, além do Brasil, possuem ocorrências de besouros de *Chrysomelidae* registradas na base de dados da coleção de Coleoptera: Chile, Peru, Bolívia, Argentina, Venezuela e Panamá. Três estados brasileiros permanecem sem nenhum registro, Pernambuco, Sergipe e Alagoas. Cerca de 143 municípios brasileiros foram identificados e marcados como detentores de ocorrência (Figura 8).

Este estudo permitiu identificar a intensidade de coleta por ano, e determinou o maior pico de coleta de coleópteros para a coleção do MPEG, durante a década de 1960, quando foram coletados 17.617 espécimes (60,66%) da coleção (Figura 9). Houve ainda outros dois picos de eventos de coleta de menor impacto nos intervalos entre os anos 1976 a 1984 e 1984 a 1988, com 3.237 e 2.938 mil espécimes respectivamente (figura 14). Com essas informações podemos dizer que mais de 80% da coleção foi constituída até o ano de 1988.



Figura 9. Número de espécimes de besouros coletadas e depositadas na Coleção de Coleoptera do Museu Paraense Emílio Goeldi ao longo de 209 anos.

## DISCUSSÃO

A coleção de Coleoptera do MPEG possui cerca de 110.000 espécimes (Costa *et al.*, 2000; Marinoni *et al.*, 2005). Deste volume foram informatizados em um ano 29.029 espécimes, que compõem atualmente a coleção de Coleoptera no sistema de gerenciamento “Specify”. Esse volume, quando publicado, contribuirá com novos taxa à rede do GBIF e SiBBr, fornecendo a esses sistemas novas ocorrências e registros, o que permitirá a identificação de possíveis erros na base por taxonomistas de diversas partes e representará, aproximadamente, 44% dos dados de Coleoptera publicados no SiBBr.

Mesmo não sendo o maior acervo de Coleoptera do Brasil e da Amazônia, a coleção MPEG justifica o seu valor pela sua abrangência histórica, taxonômica e destaca-se pelo processo de informatização em massa descrito neste trabalho. Outras importantes coleções hospedam um número maior de espécimes, entretanto, ainda não publicam esses registros em plataformas como o SiBBr e GBIF e todos os valores são estimados em número de registros. Como exemplo, podemos citar o Instituto Nacional de Pesquisas da Amazônia (INPA) e o Museu Nacional do Rio de Janeiro (MNRJ), que possuem mais de 2.000.000 coleópteros cada (Marinoni, 2005).

O site do SiBBr registra, no total, 91 instituições publicadoras de dados. Podemos destacar dentre estas instituições colaboradoras, as maiores fornecedoras de dados de ocorrência: 1) Instituto de Pesquisas Jardim Botânico do Rio de Janeiro com 969.843 registros,



2) Instituto Nacional de Pesquisas da Amazônia com 286.593 registros, 3) Universidade Estadual de Feira de Santana com 213.490 registros, 4) Instituto de Botânica da USP com 197.375 registros, 5) Museu Nacional com 138.242 registros, 6) Universidade Estadual de Campinas - Instituto de Biologia com 136.359 registros, 7) Universidade Federal de Minas Gerais com 123.023 registros, 8) Escola Superior de Agricultura Luiz de Queiroz com 122.786 registros, 9) Comissão Executiva do Plano da Lavoura Cacaueira – Centro de Pesquisas do Cacau com 121.420 registros, e 10) Museu Paraense Emílio Goeldi 106.497 registros. Todos os metadados dos conjuntos de dados das instituições citadas acima, nenhum, com a exceção do MPEG, possuem os campos da cobertura taxonômica preenchida até o epíteto específico, o que fornece uma descrição detalhada sobre os grupos presentes no “dataset” publicado. Isso facilita o uso dos dados já que é possível checar se determinada família, gênero ou espécie está presente na coleção, sem a necessidade de “download” dos dados, e ainda permite que ferramentas de busca na *web*, encontrem e retornem o recurso de publicação da coleção, mesmo com a busca de quem sabe, um táxon específico. Este baixo índice de preenchimento dessa categoria do metadado relaciona-se com a dificuldade de inserir essas informações nos softwares desenvolvidos pelo GBIF (IPT) e pelo SiBBR (AttaPublica), para publicação de dados de biodiversidade, pois ambos só permitem atribuir manualmente, através de “picklists”, o nível taxonômico correspondente a cada categoria taxonômica distinta presente em cada conjunto de dados. Portanto, o esforço para a produção desta seção do metadado de publicação seria muito grande para as coleções que hospedam grandes volumes de dados, por esse motivo estas não relacionam os gêneros e epítetos específicos e até mesmo as famílias, listando apenas, nos metadados, os níveis taxonômicos superiores (Reino, Filo e Classe).

O Museu Paraense Emílio Goeldi superou a dificuldade de cadastrar toda a cobertura taxonômica da coleção de Coleoptera com a criação de um código em “javascript” a ser executado no próprio browser. Esta pode ser uma implementação-solução para facilitar e popularizar o preenchimento destes campos de categoria taxonômica no metadado. Soluções como esta apresentada facilitam a disponibilização de dados e surgem com a implantação da informatização e publicação em massa de dados de biodiversidade. Quando nos deparamos com grandes massas de dados, surgem novos problemas que quando analisados e estudados podem trazer avanço para o conhecimento científico (Hand *et al.*, 2001).

O processo de validação taxonômica descrito utilizou a comparação dos dados de espécies presentes no banco, com uma base nacional e uma internacional, o Catálogo da Fauna do Brasil e o “Global Biodiversity Facility”, respectivamente. A primeira base mostrou-se eficiente para verificar a existência e validade taxonômica das espécies dos gêneros analisados,

para a adição de autoria e ano de descrição das espécies e para eliminar erros de digitação do banco. A segunda base, apesar de ser a maior base de dados de biodiversidade disponível na internet, provendo acesso a mais de 600 milhões de registros de espécies e ocorrência de espécies (Horn, 2016), pouco pôde contribuir para o processo de validação e incremento dos dados. A dificuldade de usar bases internacionais que hospedam milhões de registros, para validar dados taxonômicos de espécies que ocorrem em regiões ou biomas específicos pode ser justificada pela ausência de publicação de dados referentes ao grupo estudado para estes ambientes, nessas plataformas. Há bilhões de espécimes nos museus ao redor do mundo e apenas 10% dos registros de espécimes estão disponíveis em formato digital e uma muito menor está disponível através da internet em padrões tais como o Darwin *core* (Heidorn, 2011). A medida que informações das coleções, como a coleção do MPEG, forem adicionadas ao sistema, outros poderão usufruir destes registros para estudos e validação de dados através dessa plataforma.

Com a análise dos dados da coleção em “softwares” de mineração de dados/“business intelligence” também foi possível obter informações relevantes para identificar áreas sub-amostradas, para região amazônica e para o Brasil: 1) Outros países pertencentes a Amazônia, estão pouco ou não amostrados, como: Guiana, Suriname e Equador que permanecem sem registros e a Venezuela representada por pouco mais de 65 espécimes; 2) Apenas três estados brasileiros não possuem amostras na coleção, Pernambuco, Alagoas e Sergipe; 3) No Pará, há pouca amostragem (menos de 200 registros) de besouros na região de Caxiuanã, onde há a Estação Científica Ferreira Pena, que é uma base de pesquisas do MPEG e 4) Não há nenhum registro de Chrysomelidae na coleção oriundo do município de Altamira.

Estas informações também são relevantes à gestão, pois uma vez que o curador inicia os trabalhos à frente de um acervo ao qual ele é especialista, deve imediatamente ficar ciente do grau de representatividade geográfica e taxonômica da coleção, para poder estabelecer uma política que vise preencher as lacunas e melhorar a sua representatividade (De Vivo, 2014).

O conhecimento da distribuição das espécies mesmo quando as informações de coordenadas estão ausentes é importante, pois algumas espécies além de informações de ocorrência ou ecológica traz consigo uma importância financeira. A espécie *Typophorus nigrinus*, presente na coleção, está relacionada ao ataque e danos causados às plantações de batata doce *Ipomoea batatas* (Guimarães *et al.*, 2014). *Diabrotica speciosa* (larva-alfinete) também ocorre na coleção e é considerada uma praga da soja, assim como o cascudinho-verde (*Maecolaspis calcarifera*.) (Moreira & Aragão, 2009). Outras espécies, entretanto, possuem

potencial para controle biológico, como a *Syphraea uberabensis* (Wikler & Souza, 2005). Por isso, a plotagem de mapas utilizando os municípios de ocorrência, pode suprir estas informações de distribuição produzindo referências a partir de informações presentes no banco de dados.

Os estudos sobre a abrangência geográfica de Coleoptera são uma contribuição significativa para o conhecimento e distribuição de espécies deste grupo, já que em diferentes áreas essa distribuição não é totalmente conhecida, e o conhecimento da diversidade e densidade de coleópteros para essa região pode ser usada como bioindicador das mudanças ambientais (Da Silva & Da Silva, 2011). A coleta de informações, análises e usos podem ajudar cientistas, cidadãos e políticos a tomar decisões (Heidorn, 2011), e o conhecimento prévio da ocorrência dessas espécies pragas numa determinada área pode antecipar a necessidade da realização de manejo.

Com o uso de ferramentas para informatização e gerenciamento de coleções também é possível produzir informações que caracterizam e simplificam os grandes volumes de dados para um levantamento histórico da coleção, quantificando categorias de dados para comparação com as estimativas conhecidas e informações disponíveis sobre a coleção na literatura. Segundo Magalhães (2001) para que se extraia conhecimento a partir dos dados pode ser necessária uma integração feita de forma correta entre ferramentas analíticas, de simulação, geografia, e visualização de dados científicos, e nesse caso, o usuário analisador de dados conta com a mineração e descoberta de dados para realizar essa tarefa. Antes da informatização, na coleção de Coleoptera as inferências eram realizadas através da consulta direta aos espécimes ou aos livros tomo.

Foi possível, com a informatização, atualizar a informação do catálogo das espécies tipo do MPEG onde estão identificadas 208 espécies. Parte deste trabalho foi realizado por Jan Bechyné e Bohumila Bechyné, que trabalharam no MPEG de novembro de 1960 a dezembro de 1963, coletando crisomelídeos e outros insetos, organizando as coleções entomológicas e publicando os resultados científicos em 24 artigos (Nascimento & Overal, 1979). Atualmente, 468 espécies tipo estão registradas na base de dados, um crescimento de 128% em 19 anos, uma informação quantitativa exata extraída do banco de dados.

Dentre as informações sobre o registro histórico da coleção, podemos citar os registros de espécimes coletados por naturalistas famosos para a história da Entomologia como, Jan Bechyné, Bohumila Bechyné, Adolpho Ducke, Hermann von Lüderwaldt e Richard von Diringshofen. Com este estudo ficou determinado a importância histórica de pesquisadores

visitantes. Até hoje os pesquisadores são os principais coletores e determinadores mais influentes dessa coleção e de muitas outras no Brasil (Marinoni & Marinoni, 2012).

Os avanços mais significativos em incremento do número de espécies coletados presente na coleção de Coleoptera não são recentes, pois os quatro grandes picos de incremento da coleção, são a décadas de 1910, 1960, o final de 1970 e começo de 1980. A ausência de crescimento contínuo no volume de coleta pode ser relacionada à ausência, por longo período, de um especialista em Coleoptera na instituição. Desde a presença do casal Bechyné na década de 1960, somente a partir de 2013, através do Programa de Capacitação Institucional, houve um especialista em besouros no MPEG: Dra. Flávia Rodrigues Fernandes que desenvolveu trabalhos com Chrysomelidae e curadoria de coleções até o ano de 2015, e em 2016, através do mesmo programa a Dra. Jéssica Herzorg Viana que também desenvolveu trabalhos relacionados à Chrysomelidae no MPEG.

## CONCLUSÃO

Historicamente, o crescimento das coleções brasileiras de invertebrados tem sido casual ou condicionado por prioridades específicas. Raramente existe um planejamento para ampliação dos acervos, o que levaria em consideração as lacunas na representatividade taxonômica e biogeográfica em cada coleção (Magalhães, 2010). Essa perspectiva pode justificar as lacunas de amostragem e o fato dos incrementos mais acentuados terem ocorrido em períodos específicos e associadas a pesquisadores específicos.

Os grandes responsáveis por manter e analisar as informações presentes nas coleções são os institutos de pesquisa ou universidades, fiéis depositários desse legado. Com esse objetivo. Estudos de caso como este, podem auxiliar outros núcleos de pesquisa de informática para biodiversidade que desejem seguir o mesmo exemplo descrito no trabalho. Para a coleção de Coleoptera, o status e abrangência apresentados serão informações úteis para guiar os novos objetivos para o acervo.

De forma ideal, todos os dados gerados durante o processo de informatização, devem ser submetidos a análises através de “softwares” de mineração de dados. Isso representa um desafio, pois volume de informações crescente a ser processado demanda recursos computacionais capazes de realizar essa tarefa e armazenar essas informações, requerendo investimentos na modernização dos equipamentos de TI. Contudo esse esforço é necessário para que seja possível continuar interpretando as informações e para que o conhecimento produzido auxilie a pesquisa e a sociedade.

Avaliei que os processos de informatização podem facilitar, também, o processo de agrupamento e categorização de dados para que seja possível submetê-los a revisão e validação, seja através de avaliação de taxonomistas, comparação com a literatura ou o uso de ferramentas de validação de dados associadas às plataformas de publicação de informações sobre biodiversidade. Também ficou evidente que é possível trabalhar com “softwares” de representação de dados para fins de distribuição de espécies.

Finalmente a publicação de dados de ocorrência presentes na coleção, representa um acréscimo ao conhecimento público sobre a Amazônia, que fica disponível a todos que desejaram utilizar desse conhecimento em prol dos objetivos de pesquisa como a conservação e ecologia. Compartilhando essas informações o Museu Paraense Emílio Goeldi cumpre seu papel institucional de fornecer informações e bases científicas a conservação e desenvolvimento sustentável da Amazônia.

### Referências bibliográficas

**Citar catalogo on line da fauna do brasil.** Disponível em:

<http://fauna.jbrj.gov.br/fauna/listaBrasil/ConsultaPublicaUC/ConsultaPublicaUC.do>

Acessado em: 15 de Novembro de 2016.1

BOUCHARD, P.; BOUSQUET, Y.; DAVIES, A.; ALONSO-ZARAZAGA, M.; LAWRENCE, J.; LYAL, C.; NEWTON, A.; REID, C.; SCHMITT, M.; SLIPINSKI, A. & SMITH, A. Family-Group Names In Coleoptera (Insecta). **ZooKeys**, v. 88, p. 1-972. doi: 10.3897/zookeys.88.807. 2011.

**CartoDB.** Contribuidores do © OpenStreetMap com licença Open Data Commons Open Database License e CC-BY-SA <http://www.openstreetmap.org/copyright> Disponível em <<https://cartodb.com/>> Acessado em 15 de março de 2016.

COSTA, C.; IDE, S.; ROSADO-NETO, G. H.; GALILEO, M. H. M.; DA FONSECA, C. R.; V. VALENTE, R. M. & MONNÉ, M. A. Diagnóstico del conocimiento de las principales colecciones brasileñas de Coleoptera. In: Proyecto Iberoamericano de Biogeografía y Entomología Sistemática: PRIBES 2000: trabajos del 1er taller iberoamericano de entomología sistemática. **Sociedad Entomológica Aragonesa**, SEA. p. 115-136. 2000.

**CRIA - Centro de Referência em Informação Ambiental.** Datasets. Disponível em <<http://ipt1.cria.org.br/ipt/>> Acessado em 15 de março de 2016.

DA SILVA, P. G. & DA SILVA, F. C. G. BESOUROS (INSECTA: COLEOPTERA) UTILIZADOS COMO BIOINDICADORES. **Revista Congrega URCAMP (CD-Rom)**, v. 5, n.1, p. 1-16. 2011.

DE AZEVÊDO, A. D. R. P. Tratamento da informação em coleções científicas de Zoologia. **Ciência da informação**, v. 21, p. 1-3. 1992.

DE VIVO, M.; SILVEIRA, L. F. & DO NASCIMENTO, F. O. Reflexões sobre coleções zoológicas, sua curadoria e a inserção dos Museus na estrutura universitária brasileira. **Arquivos de Zoologia**, v. 45, p. 105-113. 2014.

FEGRAUS, E. H.; ANDELMAN, S.; JONES, M. B. & SCHILDHAUER, M. Maximizing the value of ecological data with structured metadata: an introduction to Ecological Metadata Language (EML) and principles for metadata creation. **Bulletin of the Ecological Society of America**, v. 86, n. 3, p. 158-168. 2005.

FEARNSIDE, Philip M. Desmatamento na Amazônia brasileira: história, índices e conseqüências. **Megadiversidade**, v. 1, n. 1, p. 113-123, 2005.

GRAHAM, C. H., FERRIER, S., HUETTMAN, F., MORITZ, C., & PETERSON, A. T. New developments in museum-based informatics and applications in biodiversity analysis. **Trends in ecology & evolution**, v. 19, n. 9, p. 497-503, 2004.

**GBIF - Global Biodiversity Information Facility.** Disponível em <<http://www.gbif.org>> Acessado em 7 de março de 2016.

**GBIF Secretariat: GBIF Backbone Taxonomy.** doi:10.15468/39omei. Acessado via <<http://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c>> em 17-05-2016.

GUIMARAES, J. A.; MADEIRA, N. R.; MOURA, A. P. de; FERNANDES, F. R. & AMARO, G. B. Ocorrência e danos do negro da batata-doce *Typophorus nigrinus* no Distrito

**Federal.** Brasília, DF: Embrapa, 2014. Disponível em <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/106016/1/12-48-COT-97.pdf> Acessado em 15 de setembro de 2016.

HAND, D. J.; MANNILA, H. & SMYTH, P. Principles of data mining. **Massachusetts**: MIT press. p.6. 2001.

HEIDORN, P. B. Biodiversity informatics. **Bulletin of the American Society for Information Science and Technology**, v. 37, n. 6, p. 38-44. 2011.

HOEKSEMA, B. W.; VAN DER LAND, J.; VAN DER MEIJ, S. E.; VAN OFWEGEN, L. P.; REIJNEN, B. T.; VAN SOEST, R. W. & DE VOOGD, N. J. Unforeseen importance of historical collections as baselines to determine biotic change of coral reefs: the Saba Bank case. **Marine Ecology**, v. 32, n. 2, p. 135-141. 2011.

HORN, T. Integrating Biodiversity Data into Botanic Collections. **Biodiversity Data Journal**, v. 4, p. e7971. 2016.

INPA - Instituto Nacional de Pesquisas da Amazônia. Datasets. Disponível em <[ipt.inpa.gov.br/ipt](http://ipt.inpa.gov.br/ipt)> Acessado em 15 de março de 2016.

**JBRJ - Instituto de Pesquisas Jardim Botânico do Rio de Janeiro**: RB – Herbário do Jardim Botânico do Rio de Janeiro. doi:10.15468/7ep9i2. Datasets. Disponível em: <<http://ipt.jbrj.gov.br/jbrj/>> Acessado em 15 de março de 2016.

LEWIS, T. Deploying distributed business software. **SIGS Publications**, Inc., 1996.

MAGALHÃES, C; DOS SANTOS, J. L. C & SALEM, J. I. Automação de coleções biológicas e informações sobre a biodiversidade da Amazônia. **Parcerias Estratégicas**, v. 6, n. 12, p.294-312. 2001.

MAGALHÃES, C.; KURY, A. B.; BONALDO, A. B.; HAJDU, E. & DESIMONE, L. **Coleções de invertebrados do Brasil**. São Paulo, v. 8, n. 5, p. 1-19. 2010. Disponível em

<<http://www.cria.org.br/cgee/documentos/ColecoesdeInvertebradosMagalhaesBonaldoKuryHadju.pdf>> Acessado em 7 de março de 2016.

MARINONI, L.; COURI, M. S.; ALMEIDA, L. M. D.; GRAZIA, J.; & MELO, G. **Coleções entomológicas brasileiras**: estado da arte e perspectivas para dez anos. Documento de trabalho. Projeto Diretrizes e Estratégias para a Modernização de Coleções Biológicas Brasileiras e a Consolidação de Sistemas Integrados de Informações sobre Biodiversidade, p. 1-28. 2005.

MARINONI, R.C. & MARINONI, L. Breve histórico da entomologia brasileira. In: RAFAEL, J.A., MELO, G.A.R., CARVALHO, C.J.B. DE, CASARI, S.A., CONSTANTINO, R. (Orgs.). *Insetos do Brasil*. Ribeirão Preto: Holos, 2012, p. 2-20.

MILLER, T. M. **Specify for Archaeology: A Proposed Data Model for Archaeological Collection Database Management**. 2012. P.138. Tese (Mestrado em Antropologia) - University of Kansas, Kansas. 2012.

**MPEG - Museu Paraense Emílio Goeldi** – Coleção de Coleoptera. doi: 10.15468/i7hmns. Datasets. Disponível em <<http://ipt.museu-goeldi.br/ipt/>> Acessado em 15 de março de 2016.

MOREIRA, HJ da C. & ARAGÃO, F. D. Manual de pragas da soja. **Campinas: FMC**, 2009.

**MySQL community server**. MySQL versão 5.7.1. Oracle©, 2016. Disponível em <<http://dev.mysql.com/downloads/mysql/>> Acessado em 7 de março de 2016.

NASCIMENTO, P.; OVERAL, W. L. Catálogo de tipos entomológicos da coleção do Museu Goeldi : Coleoptera : Chrysomelidae (Insecta). **Boletim do Museu Paraense Emílio Goeldi: Zoologia**. Nova série. 97.. v. 29, n. 3, p. 1; 23. 1979.

ROBERTSON, T.; DÖRING, M.; GURALNICK, R.; BLOOM, D.; WIECZOREK, J. & BRAAK, K. The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. **PLoS ONE**, v. 9, n. 8: e102623, p. 1-7. doi: 10.1371/journal.pone.0102623. 2014.



SANDHOF, K. & FILGUEIRAS, L. Defeitos de software como erros humanos. In: **II Workshop Um olhar sociotécnico sobre a Engenharia de Software**, 2, Espírito Santo. Anais do Workshop Um olhar sociotécnico sobre a Engenharia de Software. Espírito Santo. PESC/COPPE/UFRJ. 2006. p. 75. 2006.

SHANMUGHAVEL, P. An overview on biodiversity information in databases. **Bioinformatics**, v. 1, n. 9, p. 367. 2007.

GADELHA, L., GUIMARÃES, P., & MOURA, A. M. SiBBr: Uma Infraestrutura para Coleta, Integração e Análise de Dados sobre a Biodiversidade Brasileira. In **VIII Brazilian e-Science Workshop (BRESCI 2014)**. 2014.

**SPECIFY Software Project**, versão 6.6.4. U.S. National Science Foundation Grants, 2016. Disponível em <<http://specifyx.specifysoftware.org/>> Acessado em 23 de março de 2016.

TABARELLI, M. & A. VICENTE. Conhecimento sobre plantas lenhosas da Caatinga: lacunas geográficas e ecológicas. In: J.M.C. Silva, M. Tabarelli, M.T. Fonseca & L.V. Lins (Orgs.). Biodiversidade da Caatinga: áreas e ações prioritárias para a conservação. Brasília: Ministério do Meio Ambiente e Universidade Federal de Pernambuco, p. 101-111. 2004.

**TABLEAU DESKTOP**, versão 9.3. Tableau©, 2003. Disponível em <[www.tableau.com](http://www.tableau.com)> Acessado em 7 de março de 2016.

VALENTE, R. M. & DA SILVA, P. A. L. The first Amazonian species of *Andranthobius* Kuschel (Coleoptera: Curculionidae), with records of new host palms for the genus. **Zootaxa**, 3786, n. 4, p. 458–468, apr, 2014.

WIECZOREK, J.; BLOOM, D.; GURALNICK, R.; BLUM, S.; DÖRING, M.; GIOVANNI, R.; ROBERTSON, T. & VIEGLAIS, D. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. **PLoS ONE**, v. 7, n. 1, p. 4:e29715. doi:10.1371/journal.pone.0029715. 2012.

WIECZOREK, J.; DÖRING, M.; DE GIOVANNI, R.; ROBERTSON, T. & VIEGLAIS, D. **Darwin Core Terms**: A quick reference guide. Disponível em <http://rs.tdwg.org/dwc/terms/guides/text/index.html>. 2009

WINKER, K. Natural history museums in a postbiodiversity era. **BioScience**, v. 54, n. 5, p.455-459. 2004.

WIKLER, C. & SOUZA, P. G. Estudos bioecológicos de *Syphrea uberabensis* (Coleoptera: Chrysomelidae) Bechyné 1956. **Ambiência**, v. 1, p. 103-112, 2005.

ZAHER, H. & YOUNG, P.S. As coleções zoológicas brasileiras: panoramas e desafios. **Ciência e Cultura**, v. 55, p. 24-26. 2003.