

Sistema de Informação Distribuído para Coleções Biológicas: a Integração do *Species Analyst* e *SinBiota*



dados e ferramentas

Coordenador: Vanderlei Perez Canhos

Processo FAPESP: 2001/02175-5

Relatório Técnico – Outubro 2005



SUMÁRIO EXECUTIVO

O projeto "**Sistema de Informação Distribuído para Coleções Biológicas: a Integração do *Species Analyst* e *SinBiota***" (Processo Fapesp 2001/02175-5) teve como meta contribuir para a consolidação da infra-estrutura compartilhada de dados e informações sobre a biodiversidade do Estado de São Paulo. Teve por objetivo desenvolver um sistema de informação integrando dados dos acervos de coleções biológicas distribuídas pelo Estado de São Paulo e dados da rede *Species Analyst*¹ e do sistema *SinBiota*². Os objetivos também incluíram a pesquisa e o desenvolvimento de aplicativos para a análise, síntese e visualização dos dados disponíveis no sistema de informação. Na primeira etapa do projeto as três linhas básicas de atividade foram:

- pesquisa e desenvolvimento de um sistema distribuído de informação;
- apoio às coleções biológicas na informatização de seus acervos; e,
- pesquisa em modelagem de nicho ecológico.

O resultado dos primeiros dois anos de projeto foi o lançamento da versão *beta* da rede descentralizada *speciesLink*³, totalmente operacional, integrando dados dos acervos de 12 coleções biológicas com os sistemas *SinBiota* e SICol⁴, utilizando, para a busca e recuperação de dados, o protocolo DiGIR⁵, fruto de um trabalho de desenvolvimento colaborativo internacional.

Quanto à pesquisa em modelagem de nicho ecológico, nos primeiros dois anos a pesquisa se concentrou no desenvolvimento e teste do GARP (*Genetic Algorithm for Rule-Set Production*)⁶ em colaboração com a Universidade de Kansas.

Constatado o sucesso da iniciativa e a crescente demanda de novas coleções interessadas em se integrarem à rede, em outubro de 2003 foi solicitada à Fapesp uma prorrogação de prazo e recursos adicionais com o objetivo de dar continuidade à digitalização dos dados dos acervos das coleções participantes na primeira fase e a inclusão de novas coleções de forma a ampliar a rede para integrar 24 acervos de coleções biológicas do Estado de São Paulo. Propôs-se também a continuidade dos trabalhos de pesquisa relativos ao protocolo DiGIR e ao desenvolvimento e avaliação de ferramentas de modelagem, envolvendo, nesta fase, a colaboração da Escola Politécnica da USP. A proposta incluiu também o desenvolvimento de ferramental para a correção de dados.

O prazo do projeto foi estendido por mais dois anos, mas os recursos solicitados tiveram um corte de 40% em relação às despesas de serviços de terceiros. As atividades referentes ao desenvolvimento de um ambiente computacional para modelagem foram então excluídas do projeto. A coordenação do projeto resolveu então submeter para a avaliação da Fapesp uma nova proposta para o desenvolvimento de ferramentas de modelagem em parceria com a Escola Politécnica da USP e o Instituto Nacional de Pesquisas Espaciais. Esta proposta foi aprovada em abril de 2005 (Processo no. 04/11012-0).

A arquitetura desenvolvida para a rede mostrou ser inclusiva, no sentido de permitir a participação plena tanto de coleções com alto grau de informatização e boa conectividade como também de coleções com baixa capacitação em informática e conectividade precária. Ambos os tipos de coleções estão acessíveis e ativas em relação à atualização de seus dados.

¹ <http://speciesanalyst.net>

² <http://sinbiota.cria.org.br>

³ <http://splink.cria.org.br>

⁴ <http://sicol.cria.org.br>

⁵ Distributed Generic Information Retrieval (<http://digir.sourceforge.net/>)

⁶ <http://www.lifemapper.org/desktopgarp/>

A arquitetura da rede possibilitou o total controle das coleções sobre os seus dados, determinando o que poderia ou não ser disponibilizado, promovendo alterações, incluindo e retirando registros, sem alterar a sua rotina de trabalho ou duplicar o seu sistema.

Foram desenvolvidos aplicativos para visualização e avaliação da qualidade dos dados, além de várias ferramentas de monitoramento da evolução de cada coleção e de apresentação de seu perfil, tendo como base a informação on-line. As ferramentas de monitoramento e de identificação de possíveis erros nos registros contribuíram para a consolidação da parceria entre o gestor do sistema (o CRIA) e o provedor de dados (as coleções biológicas). Graças a essa parceria foi possível aprimorar aplicativos existentes e criar novas ferramentas de interesse das coleções.

Acreditamos que por esses desenvolvimentos técnicos concluímos o projeto tendo estabelecido parcerias com 41 coleções pertencentes às 3 Universidades estaduais, a sete institutos de pesquisa do Estado de São Paulo e ao Jardim Botânico do Rio de Janeiro, quando o projeto previa o estabelecimento de parcerias com 24 coleções do estado de São Paulo.

Estamos concluindo o projeto com cerca de 720 mil registros on-line, disponíveis para qualquer pessoa com acesso à Internet de forma livre e aberta. 42% desses registros estão geo-referenciados e o número total de registros on-line representa cerca de 30% do acervo das coleções participantes da rede. Os resultados por grupo taxonômico estão apresentados a seguir.

Coleções Biológicas	Acervo Total	Registros On-line	% on-line	Registros geo-ref	% geo-ref
Herbários e Xilotecas	1.429.161	285.292	20%	69.251	24%
Peixes	128.023	96.803	76%	63.247	65%
Aracnídeos	35.459	27.425	77%	4.904	18%
Répteis e Anfíbios	111.179	93.519	84%	7.124	8%
Mamíferos	34.063	19.753	58%	815	4%
Insetos	508.171	106.155	21%	82.117	77%
Aves	112.224	23.381	21%	12.488	53%
Microrganismos	2.688	2.064	77%	0	0%
Total dos Acervos	2.360.968	654.392	28%	239.946	37%
Dados de Observação	63.586	63.586	100%	63.584	100%
Total Geral	2.424.554	717.978	30%	303.530	42%

O relatório apresenta o trabalho de pesquisa e desenvolvimento realizado nos últimos 4 anos dividido nos seguintes itens:

- pesquisa em sistemas distribuídos de informação;
- apoio às coleções;
- ferramentas de correção e visualização de dados;
- aplicativo para a produção dinâmica de mapas na Internet;
- sistema de monitoramento dos indicadores do projeto;
- modelagem; e,
- manutenção do sistema.

É também apresentada a evolução dos indicadores de acompanhamento estabelecidos para o projeto (volume, uso, qualidade, formação de recursos humanos e divulgação).

Por fim, nos comentários finais o relatório faz uma breve análise da importância de sistemas de informação de acesso livre e aberto à ciência e da necessidade de estudar um mecanismo que garanta a sustentabilidade de uma infra-estrutura de dados de acesso público e permanente.

Ao longo do relatório são apresentadas algumas recomendações à Fapesp e à coordenação do programa Biota/Fapesp com base na manifestação dos curadores das coleções.

Recomenda-se à Fapesp

- manter uma linha de apoio para a digitalização dos acervos das coleções biológicas do Estado de São Paulo. Uma recomendação específica é que sejam criadas bolsas associadas ao Programa Biota/Fapesp para digitalização de acervos visando dar continuidade aos trabalhos, evitando que cada curador tenha que apresentar um projeto para esse fim. Ainda em relação às bolsas, a sugestão é que a bolsa seja concedida ao orientador e não ao candidato.
- solicitar na apresentação de toda proposta visando o apoio à informatização de coleções, uma estratégia para a digitalização de acervos.
- apoiar a produção de catálogos de nomes e *checklists on-line* para servir de referência para a atualização da nomenclatura dos acervos das coleções.
- estudar uma forma de dar suporte à manutenção da rede.

Recomenda-se à coordenação do programa Biota/Fapesp:

- a programação de cursos sobre a utilização dos principais software de gerenciamento adotados pelas coleções como Brahms e Biota, e até aqueles incluídos no Office, como Access e Excel.
- incluir sessões demonstrativas ou mini cursos sobre o uso das ferramentas desenvolvidas para a rede speciesLink (p.ex. a interface de busca, *data cleaning*, geoLoc e conversor) nos simpósios do programa Biota



Índice

Sumário Executivo	i
I. Trabalho Realizado	1
1. Introdução.....	1
2. Pesquisa em sistemas distribuídos de informação.....	2
2.1. Protocolo de comunicação.....	2
2.1.1 Metadados.....	3
2.1.2 Busca	4
2.1.3 Inventário.....	6
2.2. Componentes da rede.....	7
2.2.1 Provedor de dados (provider)	8
2.2.2 Portal.....	10
2.2.3 Servidor regional	10
2.2.4 spLinker.....	13
2.3. Desenvolvimento da interface de busca	15
2.4. O futuro de redes de busca em dados distribuídos sobre biodiversidade.....	22
3. Apoio às coleções.....	22
3.1. Software de gerenciamento	23
3.2. Visitas técnicas.....	25
3.3. Suporte remoto.....	26
3.4. Instalação do equipamento	27
3.5. Treinamento	28
3.6. Bolsas.....	28
3.7. Relatórios de Desempenho.....	30
4. Ferramentas de visualização e correção de dados	30
4.1. speciesMapper	30
4.2. conversor.....	30
4.3. geoLoc.....	30
4.4. spOutlier	31
4.5. infoXY	31
4.6. Banco de dados de coletores.....	31
4.7. Data cleaning	32
4.7.1 Coleção	33
4.7.2 Dados Taxonômicos.....	33
4.7.3 Dados Geográficos.....	36
4.7.4 Monitoramento	42
4.7.5 Perfil do acervo	44
5. Aplicativo para a produção dinâmica de mapas na internet.....	49
5.1. Mapcria Web Service	50
5.2. Mapcria Viewer 3.0	54
6. Sistema de monitoramento dos indicadores do projeto	58
6.1. Lista.....	58
6.2. Monitor	59
6.3. Mapa	60
6.4. Estatísticas	62
6.5. Diagrama.....	63
6.6. Provedores	64
6.7. Notas	64
7. Interoperabilidade e integração de dados com outros sistemas	65
7.1. SinBiota	65
7.2. Coleções biológicas do exterior	65
7.3. Outros sistemas	66
8. Modelagem	70

8.1.	Histórico	70
8.2.	Estudos de caso.....	71
8.2.1	Utilização de diferentes escalas de dados climáticos.....	71
8.2.2	Modelagem de espécies lenhosas de cerrado em todo o estado de São Paulo.....	72
8.2.3	Modelagem de espécies arbóreas do Médio Paranapanema: aplicações para recuperação de áreas degradadas	74
8.2.4	Modelagem de espécies com distribuição restrita ou com poucos pontos de amostragem	77
8.2.5	Modelos de distribuição geográfica de espécies vegetais dispersas pela megafauna: os efeitos da perda de dispersores.....	78
8.2.6	Modelagem de distribuição potencial das espécies da seção típica do gênero <i>Rauvolfia</i> (Apocynaceae) – aplicação como ferramenta auxiliar na taxonomia e estudos biogeográficos.....	79
8.2.7	Controle biológico de <i>Tetranychus evansi</i> (Acari: Tetranychidae) na África.....	83
8.2.8	Distribuição de <i>Parides ascanius</i> no Brasil: Urbanização, áreas protegidas e conservação de uma espécie ameaçada de borboleta	84
8.2.9	Prevenção a invasão da espécie invasora <i>Crotalaria pallida</i> nas unidades de conservação do Brasil.....	86
8.2.10	Invasão de <i>Callithrix jacchus</i> and <i>C. penicillata</i> : contaminação genética do gênero <i>Callithrix</i> e declínio de espécies endêmicas da Mata Atlântica no Brasil	86
9.	Manutenção do sistema	86
9.1.	Infra-estrutura computacional	86
9.2.	Infra-estrutura de software	87
9.3.	Rotina de backup e proteção dos dados do projeto	87
9.4.	Instalação e manutenção dos servidores regionais.....	88
II.	Indicadores	89
1.	Volume	89
2.	Uso	94
3.	Qualidade	96
3.1.	Desenvolvimento de Parcerias	96
3.1.1	Provedores de Dados	97
3.1.2	Parceiros técnicos.....	98
3.2.	Qualidade dos dados	99
4.	Formação de recursos humanos.....	100
5.	Divulgação dos resultados do projeto.....	104
5.1.	Participação em reuniões científicas.....	104
5.2.	Publicações e relatórios técnicos.....	105
5.3.	O projeto speciesLink na mídia.....	105
III.	Resultados	106
1.	Processo de digitalização dos acervos	106
2.	Desenvolvimento da rede.....	112
3.	Conteúdo	113
4.	Parceria com outras iniciativas	115
5.	Definição de um modelo.....	116
IV.	Comentários Finais	117
V.	Equipe	119
	Anexo I Questionário para avaliação da infra-estrutura da coleção para sua participação no projeto speciesLink	122
	Anexo II Participação em eventos científicos.....	124



Eventos no país	124
Eventos no Exterior	128
Anexo II Publicações e Relatórios Técnicos	131
Anexo III. O projeto na mídia	134

I. TRABALHO REALIZADO

1. INTRODUÇÃO

O trabalho procurou atender a demanda de dois atores específicos: o provedor de dados (**coleções biológicas**) e os **usuários**.

As demandas das **coleções biológicas** incluem o desenvolvimento de mecanismos para:

- disseminar os dados de seus acervos na Internet sem alterar a sua rotina de trabalho;
- visualizar seus dados (inventários, distribuição espacial, etc.)
- auxiliar no processo de geo-referenciamento dos dados; e,
- validar os dados (espacial e taxonômico).

Os **usuários**, de maneira bastante genérica demandam:

- um sistema simples, capaz de realizar buscas em sistemas distribuídos na rede, via interface web;
- um sistema integrado, de forma bastante transparente, com outros sistemas relevantes;
- diferentes formas de busca;
- diferentes formatos de saída ou de apresentação dos resultados (HTML, Excel e mapas); e,
- ferramentas para análise preditiva da distribuição geográfica das espécies.

Para atender essas demandas, visando cumprir com os objetivos propostos pelo projeto, foi necessário:

- pesquisar e desenvolver sistemas distribuídos de informação: protocolos de comunicação, componentes da rede e interface de busca;
- desenvolver um sistema interoperável e integrado com outros sistemas de interesse;
- pesquisar e desenvolver ferramentas de visualização e correção dos dados;
- desenvolver serviços web para a produção dinâmica de mapas na Internet;
- desenvolver ferramentas para o monitoramento dos indicadores do projeto;
- desenvolver ferramentas para a modelagem preditiva da distribuição de espécies;
- desenvolver os sistemas de segurança e backup; e,
- manter o sistema em operação.

É importante ter em mente que o projeto não é um trabalho puramente acadêmico que tenha como meta divulgar o conhecimento adquirido. O produto central é um sistema distribuído de dados dos acervos de coleções biológicas do Estado de São Paulo de acesso livre e aberto em plena operação.

Para testar ferramentas e provar o conceito da importância do compartilhamento aberto e livre a dados, foram também desenvolvidos estudos de caso utilizando ferramentas de modelagem preditiva de distribuição de espécies.

Todas as etapas de trabalho estão descritas a seguir.

2. PESQUISA EM SISTEMAS DISTRIBUÍDOS DE INFORMAÇÃO

No início do desenvolvimento dos trabalhos foi lançado no ambiente *sourceforge* um projeto cooperativo *on-line* denominado DiGIR, *Distributed Generic Information Retrieval*⁷. O CRIA resolveu participar do desenvolvimento desse projeto ao invés de desenvolver um protocolo próprio por perceber que essa iniciativa, além de vir de encontro às necessidades do projeto, tinha grande potencial de se tornar um padrão internacional que, uma vez aceito pela comunidade mundial, permitiria o compartilhamento de dados e informações biológicas em escala global. Além disso, a possibilidade de se beneficiar do conhecimento de desenvolvedores de outros países, era extremamente atrativa. A experiência tem sido muito bem sucedida.

O DiGIR é um protocolo cliente/servidor projetado para recuperar informação de fontes distribuídas. Usa o protocolo *Hypertext Transport Protocol* (HTTP) como mecanismo de transferência de dados e o *eXtensible Markup Language* (XML) para codificar as mensagens trocadas entre clientes e servidores. Foi projetado para suportar a recuperação de informação conforme um modelo de dados genérico e arbitrário. O protocolo mantém a independência entre o mecanismo de transmissão de mensagens e o modelo de dados em que a informação é recuperada. Dessa forma é possível utilizar o protocolo para recuperar dados de outros domínios e não apenas de coleções biológicas.

A maneira mais comum de utilizar o protocolo é através de três componentes principais: uma camada de apresentação (responsável pela interação com o usuário final), um portal (responsável pela distribuição das consultas), e o provedor de dado (responsável por interagir com cada banco de dados ligado à rede). A interface entre os três componentes é muito bem definida, permitindo seu desacoplamento e também possibilitando implementações em diferentes linguagens de programação e plataformas computacionais.

As principais características do protocolo são:

- **Transparência de localização:** a complexidade de identificação da localização dos vários provedores de dados disponíveis na rede é função do portal;
- **Transparência de plataforma e modelo local de dados:** cada provedor de dados é responsável pela tradução necessária entre o seu próprio modelo de dados (esquema do banco de dados) e o modelo genérico adotado pela comunidade a que este pertence;
- **Descoberta automática de fontes de dados:** através do *Universal Description, Discovery and Integration* (UDDI), o portal DiGIR pode detectar automaticamente quais são os provedores servindo dados segundo um determinado modelo. Os provedores, por sua vez, ao serem configurados, podem registrar-se em um dos diretórios UDDI públicos disponíveis para que os portais existentes possam descobri-los e associá-los às suas redes.

2.1. PROTOCOLO DE COMUNICAÇÃO

O protocolo DiGIR é especificado através de um esquema XML (*XML Schema*) que define a estrutura das mensagens enviadas e retornadas por provedores DiGIR⁸.

Para desacoplar o protocolo do modelo de dados usado por uma rede DiGIR, são definidos elementos de busca genéricos e fictícios que são substituídos em um outro esquema XML que define o modelo de dados a ser compartilhado por cada comunidade em particular⁹. Para isso são usados dois esquemas XML separados e hierárquicos, um para o protocolo, definindo a estrutura das mensagens trocadas entre os componentes da rede e outro para a

⁷ <http://sourceforge.net/projects/digir/>

⁸ <http://digir.sourceforge.net/prot/beta3/digir.xsd>

⁹ <http://digir.sourceforge.net/fed/beta3/darwin2.xsd>

definição dos conceitos específicos para cada domínio de informação. Nesta seção é descrita a estrutura do primeiro esquema, o esquema XML do protocolo DiGIR.

O esquema XML do protocolo define duas estruturas principais: o elemento `<request>` que contém uma consulta a um provedor DiGIR e o elemento `<response>` que define a estrutura da resposta do provedor à consulta.

O elemento `<request>` contém um elemento `<header>` onde são definidos: a versão do protocolo, o horário em que a consulta está sendo feita, o endereço IP de quem originou a consulta (no caso o IP registrado é o do usuário final para efeito de registro de atividade do sistema), o destinatário da mensagem (pode ser um único provedor ou provedores múltiplos) e o tipo de consulta que está sendo realizada.

Atualmente o protocolo DiGIR suporta três tipos de operação: metadados, busca e inventário.

2.1.1 METADADOS

O pedido de metadados é especificado quando o elemento `<type>` contém o valor "metadata", não sendo necessária a existência do corpo da mensagem.

Abaixo é apresentado um exemplo de pedido de metadados no DiGIR:

```
<request xmlns="http://www.namespaceTBD.org/digir">
  <header>
    <version>&#36;Revision: 1.7 &#36;</version>
    <sendTime>2003-06-04 18:07:34-0300</sendTime>
    <source>200.144.120.37</source>
    <destination>http://splink.cria.org.br/provider/DiGIR.php
    </destination>
    <type>metadata</type>
  </header>
</request>
```

Este pedido obtém como resposta um documento XML descrevendo cada provedor especificado como destinatário da mensagem e cada coleção (*resource*) conectada ao provedor:

```
<response xmlns="http://digir.net/schema/protocol/2003/1.0">
  <header>
    <version>&#36;Revision: 1.7 &#36;</version>
    <sendTime>2003-06-04 18:07:35-0300</sendTime>
    <source>http://splink.cria.org.br:80/provider/DiGIR.php</source>
    <destination>200.144.120.37</destination>
  </header>
  <content>
    <metadata>
      <provider>
        <name>speciesLink_CRIA</name>
        <accessPoint>http://splink.cria.org.br:80/provider/DiGIR.php
        </accessPoint>
        <implementation>&#36;Revision: 1.7 &#36;</implementation>
      </provider>
      <host>
        <name>Centro de Referência em Informação Ambiental</name>
        <code>CRIA</code>
        <relatedInformation>http://www.cria.org.br/
        </relatedInformation>
        <contact type="technical">
          <name>Sidnei de Souza</name>
          <title>Analista de Sistemas</title>
          <emailAddress>sidnei@cria.org.br</emailAddress>
          <phone>+55 19 32880466</phone>
        </contact>
        <abstract>O Centro de Referência em Informação Ambiental, é uma
        sociedade civil, sem fins lucrativos, que pretende disseminar o conhecimento
        científico e tecnológico e promover a educação, visando a conservação e a
```

```

utilização sustentável dos recursos naturais e a formação da
cidadania.</abstract>
</host>
<resource>
  <name>Coleção Bras. de Microrganismos de Ambiente e Indústria</name>
  <code>CBMAI</code>
  <relatedInformation>http://www.cpqba.unicamp.br</relatedInformation>
  <contact type="administrative ">
    <name>Gilson Paulo Manfio</name>
    <title>Curador</title>
    <emailAddress>gmanfio@cpqba.unicamp.br</emailAddress>
    <phone>+55 19 38847500</phone>
  </contact>
  <abstract>A coleção possui cerca de 700 linhagens de arqueas,
bactérias, fungos filamentosos, plasmídeos e organismos geneticamente
modificados (OGM), de interesse industrial e ambiental. As principais atividades
em que a coleção é utilizada são: identificação, treinamento, pesquisa e
accessoria à indústria.
</abstract>
  <keywords>bactérias, fungos</keywords>
  <citation />
  <useRestrictions />
  <conceptualSchema
schemaLocation="http://digir.sourceforge.net/fed/beta3/darwin2.xsd">http://www.n
amespaceTBD.org/darwin2</conceptualSchema>
  <recordIdentifier>CBMAI</recordIdentifier>
  <recordBasis>strain</recordBasis>
  <numberOfRecords>110</numberOfRecords>
  <dateLastUpdated>10/04/2003</dateLastUpdated>
  <minQueryTermLength>0</minQueryTermLength>
  <maxSearchResponseRecords>1000</maxSearchResponseRecords>
  <maxInventoryResponseRecords>1000</maxInventoryResponseRecords>
<defaultRecordFormat>http://digir.sourceforge.net/prov/darwin/darwin2brief.xsd</
defaultRecordFormat>
  <defaultInventoryConcept>darwin:ScientificName</defaultInventoryConcept>
</resource>
</provider>
</metadata>
</content>
<diagnostics>
  <diagnostic code="STATUS_INTERVAL" severity="info">3600</diagnostic>
  <diagnostic code="STATUS_DATA" severity="info">1,3,0</diagnostic>
</diagnostics>
</response>

```

2.1.2 BUSCA

Uma requisição de busca é especificada pelo elemento `<search>` dentro de `<request>`. O elemento `<search>` contém dois outros elementos: um elemento `<filter>` e outro `<records>`. O primeiro define a consulta a ser realizada e o segundo define quais campos devem ser retornados na resposta. Numa comparação com uma consulta SQL, o elemento `<filter>` seria equivalente à cláusula *WHERE*, e o elemento `<records>` ao conjunto de campos que seguem a cláusula *SELECT*.

O filtro (estrutura definida no elemento `<filter>`) define uma consulta a ser realizada no provedor através da combinação de operadores lógicos (LOPs) e comparativos (COPs) numa estrutura em árvore expressa em XML. Os operadores de comparação são usados para comparar o valor de um campo com uma constante enquanto os operadores lógicos possibilitam a comparação lógica entre dois operadores de comparação ou dois operadores lógicos. Os operadores lógicos utilizados são: E lógico, OU lógico e a negação. Os operadores comparativos utilizados são: menor que (<), menor ou igual a (≤), igual a (=), maior que (>), maior ou igual a (≥), diferente de (≠), contém (in), e assemelha-se a (like).

O filtro tem a vantagem de ser uma estrutura independente de linguagem que pode ser facilmente transformada em uma consulta SQL e pode ser analisada sintaticamente através de um dos muitos *parsers* XML disponíveis.

O elemento `<records>` é utilizado para definir a estrutura e os tipos de dados a serem retornados na consulta. Este elemento contém uma lista de elementos presentes no esquema XML conceitual, sendo que cada um define um campo a ser retornado na resposta à consulta. O elemento `<records>` pode opcionalmente especificar uma *URL* que contém o documento que descreve o conjunto de registros a serem retornados na consulta.

A seguir é apresentado um exemplo de consulta do tipo *search*:

```
<request xmlns:digir="http://www.namespaceTBD.org/digir"
  xmlns:darwin="http://www.namespaceTBD.org/darwin2"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns="http://www.namespaceTBD.org/digir">
  <header>
    <version>>$Revision: 1.7 $</version>
    <sendTime>2003-06-05 11:40:50-0300</sendTime>
    <source>129.237.201.120</source>
    <destination
resource="IBSBF">http://splink.cria.org.br/provider/DiGIR.php</destination>
    <type>search</type>
  </header>
  <search>
    <filter>
      <like>
        <darwin:Genus>agrobacterium</darwin:Genus>
      </like>
    </filter>
    <records start="0" limit="10" count="0">
      <structure>
        <xsd:element name="record">
          <xsd:complexType>
            <xsd:sequence>
              <xsd:element ref="darwin:InstitutionCode"/>
              <xsd:element ref="darwin:CollectionCode"/>
              <xsd:element ref="darwin:CatalogNumber"/>
              <xsd:element ref="darwin:Genus"/>
              <xsd:element ref="darwin:Species"/>
            </xsd:sequence>
          </xsd:complexType>
        </xsd:element>
      </structure>
    </records>
  </search>
</request>
```

Outras tecnologias foram consideradas para substituir o filtro como linguagem para expressar as consultas. Entre elas foram avaliadas o XPath, o XQuery e o próprio SQL.

O XQuery não era uma tecnologia madura o suficiente para ser utilizada na época da definição do protocolo. Além disso, tanto o XPath quanto o XQuery exigem transformações não triviais para conversão em SQL, linguagem de consulta nativa utilizada pela maioria dos bancos de dados existentes nos provedores. Essas tecnologias seriam mais apropriadas caso os provedores utilizassem bancos de dados XML nativos.

A linguagem SQL foi descartada pois existem muitos dialetos diferentes no mercado, o que exigiria o desenvolvimento de um analisador sintático para extrair os componentes da consulta e reconstruir a expressão de busca no dialeto padrão ANSI.

Como exemplo de resposta a uma consulta de registros, temos:

```
<response xmlns="http://digir.net/schema/protocol/2003/1.0">
  <header>
    <version>$Revision: 1.7 $</version>
    <sendTime>2003-06-05 11:40:51-0300</sendTime>
    <source
resource="IBSBF">http://splink.cria.org.br:80/provider/DiGIR.php</sou
rce>
    <destination>129.237.201.120</destination>
  </header>
  <content xmlns:darwin="http://www.namespaceTBD.org/darwin2"
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <record>
      <darwin:InstitutionCode>IB</darwin:InstitutionCode>
      <darwin:CollectionCode>IBSBF</darwin:CollectionCode>
      <darwin:CatalogNumber>343</darwin:CatalogNumber>
      <darwin:Genus>Agrobacterium</darwin:Genus>
      <darwin:Species>radiobacter</darwin:Species>
    </record>
    <record>
      <darwin:InstitutionCode>IB</darwin:InstitutionCode>
      <darwin:CollectionCode>IBSBF</darwin:CollectionCode>
      <darwin:CatalogNumber>307</darwin:CatalogNumber>
      <darwin:Genus>Agrobacterium</darwin:Genus>
      <darwin:Species>rhizogenes</darwin:Species>
    </record>
  </content>
  <diagnostics>
    <diagnostic code="STATUS_INTERVAL"
severity="info">3600</diagnostic>
    <diagnostic code="STATUS_DATA"
severity="info">3,13,0</diagnostic>
    <diagnostic code="MATCH_COUNT" severity="info">2</diagnostic>
    <diagnostic code="RECORD_COUNT" severity="info">2</diagnostic>
    <diagnostic code="END_OF_RECORDS"
severity="info">>false</diagnostic>
  </diagnostics>
</response>
```

2.1.3 INVENTÁRIO

Finalmente, o terceiro tipo de consulta é o inventário, e é definido através do elemento *<inventory>*. Através dessa consulta é possível obter uma lista dos valores únicos presentes em um campo num provedor, e o número de registros de cada valor retornado é equivalente a um comando SQL: *SELECT DISTINCT campo, COUNT(campo) AS total*.

O campo sobre o qual o inventário será feito é definido dentro do elemento *<inventory>*. Este tem estrutura semelhante ao mesmo elemento no caso da consulta do tipo *search*, mas no caso do inventário, deve conter apenas um registro. Opcionalmente a requisição de inventário pode conter um filtro, o que permite limitar o alcance do comando a um subconjunto dos registros do provedor.

A seguir é mostrado um exemplo de consulta do tipo *inventory*:

```
<request xmlns="http://www.namespaceTBD.org/digir"
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:darwin="http://www.namespaceTBD.org/darwin2">
  <header>
    <version>>$Revision: 1.7 $ </version>
    <sendTime>2003-06-05 11:54:00-03:00</sendTime>
```

```

    <source>216.91.87.102</source>
    <destination
resource="IBSBF">http://splink.cria.org.br/provider/DiGIR.php</destin
ation>
    <type>inventory</type>
  </header>
  <inventory>
    <darwin:Genus />
    <count>>true</count>
  </inventory>
</request>

```

Uma possível resposta a esta consulta seria:

```

<response xmlns="http://digir.net/schema/protocol/2003/1.0">
  <header>
    <version>$Revision: 1.7 $</version>
    <sendTime>05-06-2003 11:55:56-0300</sendTime>
    <source
resource="IBSBF">http://splink.cria.org.br:80/provider/DiGIR.php</sou
rce>
    <destination>216.91.87.102</destination>
  </header>
  <content xmlns:darwin="http://www.namespaceTBD.org/darwin2"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <record>
      <darwin:Genus count="33">Agrobacterium</darwin:Genus>
    </record>
    <record>
      <darwin:Genus count="25">Bacillus</darwin:Genus>
    </record>
    <record>
      <darwin:Genus count="34">Clavibacter</darwin:Genus>
    </record>
    <record>
      <darwin:Genus count="364">Xanthomonas</darwin:Genus>
    </record>
  </content>
  <diagnostics>
    <diagnostic code="STATUS_INTERVAL"
      severity="info">3600</diagnostic>
    <diagnostic code="STATUS_DATA"
      severity="info">3,14,1</diagnostic>
    <diagnostic code="MATCH_COUNT" severity="info">4</diagnostic>
    <diagnostic code="RECORD_COUNT" severity="info">4</diagnostic>
    <diagnostic code="END_OF_RECORDS"
      severity="info">true</diagnostic>
  </diagnostics>
</response>

```

2.2. COMPONENTES DA REDE

A extensão do prazo de execução do projeto permitiu aprofundar a pesquisa e desenvolvimento do protocolo DiGIR, visando aprimorar as ferramentas disponíveis para a construção da infra-estrutura compartilhada de dados. A figura 1 é um esquema da rede com os componentes principais do protocolo DiGIR, o portal e o provedor.

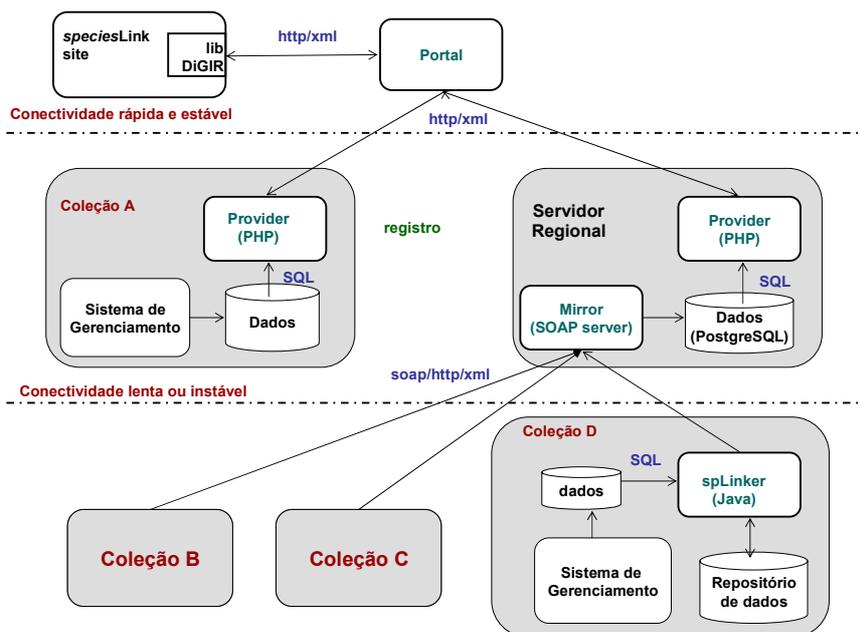


Figura 1. Esquema dos protocolos e software utilizados pela rede speciesLink

2.2.1 PROVEDOR DE DADOS (PROVIDER)

Trata-se da camada responsável por abstrair a heterogeneidade das fontes de dados conectadas à rede. Cada provedor de dados pode estar ligado a uma ou mais bases de dados (denominadas “resources” pelo protocolo). A comunicação é feita através do protocolo DiGIR. Ao receber uma requisição, o provedor faz a tradução da mesma para o padrão SQL, interage com a base de dados especificada na requisição, e produz uma resposta de acordo com o protocolo DiGIR. Normalmente são portais que se comunicam com provedores de dados.

O software utilizado pelo speciesLink para cumprir o papel de provedor de dados faz parte da implementação padrão do protocolo DiGIR, resultado de um esforço internacional já mencionado e do qual nossos desenvolvedores participam ativamente. Seu desenvolvimento foi feito na linguagem PHP¹⁰, por ser multi-plataforma, bem documentada, e particularmente voltada para operar sob HTTP. Além disso, é uma ferramenta de fácil acesso (não proprietária), *open source*, e possui bibliotecas e recursos necessários para trabalhar com XML e abstrair a comunicação com diversos tipos de bases de dados.

Os pré-requisitos para instalação de um provedor de dados são:

- Servidor Web que suporte PHP (já foram testados: Apache e IIS)
- PHP versão maior ou igual a 4.2.3
- Existência de um domínio fixo para o servidor (sem um endereço fixo torna-se impraticável utilizar os serviços de um provedor de dados)

A configuração do provedor é feita através de arquivos no formato XML. O primeiro deles, normalmente chamado *providerMeta.xml* contém os metadados do provedor. O segundo, chamado *resources.xml*, lista todas as coleções conectadas a ele, indicando para cada uma delas o nome do arquivo de configuração. Este último, além de conter os metadados da coleção, possui também instruções para conexão com a base de dados e um mapeamento dos campos em relação ao esquema de dados sendo usado pelo protocolo.

¹⁰ <http://www.php.net/>

O **provedor DiGIR** passou por várias alterações durante o projeto. Três destas alterações merecem destaque: a implementação de filtros locais, a modularização do esquema conceitual e a validação com base no XML Schema do protocolo.

a. FILTROS LOCAIS

Foi implementada pelo CRIA uma maneira de configurar filtros para dados locais, de tal forma que as buscas sejam apenas realizadas nos registros que passam por este filtro, não sendo possível retornar dados de outros registros. Anteriormente este tipo de filtragem só era possível através da criação de “visualizações” (“*views*”) nos bancos de dados. A utilização de “visualizações” nem sempre é a melhor solução, pois não são todos os bancos de dados que permitem esse tipo de construção, e mesmo quando permitem, muitas vezes há restrições para sua criação caso o banco de dados faça parte de um sistema proprietário de gerenciamento de coleções. Nestes casos, os filtros locais apresentam-se como uma boa alternativa. Coleções conectadas indiretamente (através de servidores regionais) já possuíam a capacidade de filtrar os dados enviados através da utilização de expressões regulares ou da inclusão de condições específicas em cláusulas WHERE de comandos em SQL. No entanto, na medida em que mais coleções forem conectadas diretamente (*software* como Biota¹¹ e Specify¹² estarão distribuindo provedores DiGIR em suas próximas versões), a utilização de filtros locais passa a ser uma boa opção para o controle de acesso a dados sensíveis, dados em vias de serem publicados e a dados ainda não validados. Há também uma outra vantagem de ordem prática, uma vez que os servidores regionais armazenam registros de várias coleções biológicas nas mesmas tabelas (de forma normalizada) sendo entretanto apresentados como recursos separados. A manutenção de um grande número de visualizações é mais complexa do que a simples utilização de filtros locais. Tais filtros seguem o mesmo padrão de codificação de filtros do protocolo DiGIR, porém são armazenados localmente e concatenados com eventuais filtros presentes em requisições.

b. MODULARIZAÇÃO DE ESQUEMAS CONCEITUAIS

O esquema conceitual que atualmente é utilizado pelo *speciesLink* é conhecido por DarwinCore, e abrange dados básicos que são comuns a todos os grupos taxonômicos. Mesmo sendo essencial para promover interoperabilidade entre várias fontes de dados distintas, o DarwinCore não é suficiente em situações em que determinados grupos taxonômicos queiram disponibilizar outros dados de caráter mais específico. Para estes casos, são necessárias extensões ao DarwinCore, que são suportadas pelo provedor DiGIR. Um exemplo concreto é a rede OBIS, *the Ocean Biogeographic Information System*¹³, especializada em dados marinhos. O trabalho desenvolvido pela equipe do CRIA, tornou possível a utilização de esquemas conceituais independentes. A atual infra-estrutura de software do *speciesLink*, portanto, já permite extensões e inclusões de esquemas conceituais específicos, tendo inclusive sido montado no CRIA um protótipo de extensão para dados de coleções microbiológicas.

c. VALIDAÇÃO XML

Foram feitas todas as alterações necessárias para que as mensagens retornadas pelos provedores, como respostas às requisições, pudessem ser completamente validadas com base no XML Schema do protocolo. Os provedores DiGIR instalados, um em cada servidor regional e um no servidor principal do CRIA, foram todos atualizados para incorporar as alterações acima descritas.

¹¹ <http://viceroy.eeb.uconn.edu/Biota>

¹² <http://www.specifysoftware.org/>

¹³ <http://www.iobis.org/>

d. OUTRAS ALTERAÇÕES

A nova versão do software provedor instalada em todos os servidores regionais teve as bibliotecas Php.XPath¹⁴, pear¹⁵ e ADOdb¹⁶ atualizadas para versões mais recentes. Houve também melhorias em relação à performance, em particular tratando-se da operação de metadados. A interface de configuração via web foi alterada para melhorar a compatibilidade com diferentes navegadores. Por fim, ao longo do projeto, também foram corrigidos vários erros.

2.2.2 PORTAL

O portal é o módulo responsável por receber as requisições feitas pelos aplicativos de apresentação dos dados e distribuí-las entre os provedores de dados. É também responsável por juntar as respostas dos vários provedores em uma única resposta e devolver ao aplicativo de apresentação que a requisitou. Pode-se pensar no portal como um centralizador de provedores. Os comandos aceitos pelo portal são aqueles definidos pelo protocolo DiGIR, já descritos anteriormente.

No contexto do DiGIR, o portal foi concebido de tal forma que ao implementar um aplicativo de apresentação não seja necessário toda vez reimplementar mecanismos necessários à consulta em bases de dados distintas. Os aplicativos de apresentação não precisam conhecer diretamente os provedores de dados. Eles precisam somente conhecer o endereço do portal e seu protocolo. O portal consegue tratar a complexidade de gerenciar várias requisições em paralelo e unificar a resposta.

O Portal foi implementado como sendo um Java Servlet através da API padrão, descrita no pacote javax.servlet. Para o gerenciamento do servlet que implementa o portal foi escolhido o “servlet container” Tomcat desenvolvido pelo projeto “*Apache Jakarta Project*¹⁷”. O servidor web utilizado para tratar do protocolo HTTP é o Apache¹⁸. Estas tecnologias foram escolhidas por serem amplamente utilizadas no mercado e por serem livres (Java) ou *open source* (Tomcat e Apache).

O portal DiGIR tem sido atualizado ao longo do projeto. A nova versão exigiu a atualização do servidor de aplicações Jakarta Tomcat para a versão 4.1.27 e a atualização do Java¹⁹ para a versão 1.4.2. Tais atualizações aumentaram significativamente a estabilidade e a performance do portal. Uma das novas funcionalidades do novo portal DiGIR é o suporte a requisições e respostas em UTF-8 (anteriormente restrito ao *charset* latin1).

2.2.3 SERVIDOR REGIONAL

Apesar do protocolo DiGIR oferecer bastante flexibilidade em termos de configuração e conexão de bancos de dados de coleções biológicas em diferentes plataformas computacionais, é necessário também que o *hardware* utilizado pelo provedor de dados satisfaça as seguintes condições: i) tenha desempenho suficiente para suportar a carga imposta pelos usuários da rede; ii) esteja conectado ao portal através de uma linha suficientemente rápida e estável; e, iii) esteja disponível 24 horas por dia para consulta.

Embora várias das coleções biológicas do estado de São Paulo tenham modernizado a sua infra-estrutura graças a investimentos governamentais e da própria FAPESP, muitas têm problemas com conectividade e falta de pessoal de apoio em informática. Visando estender a funcionalidade do protocolo DiGIR para esses casos, foi criado pela equipe do CRIA o

¹⁴ <http://www.carrubbers.org/scripts/php/xpath>

¹⁵ <http://pear.php.net/>

¹⁶ <http://adodb.sourceforge.net>

¹⁷ <http://jakarta.apache.org/tomcat/>

¹⁸ <http://www.apache.org>

¹⁹ <http://java.sun.com>

conceito dos nós intermediários de espelhamento, ou servidores regionais. Estes servidores ficam situados entre o provedor de dados DiGIR e o banco de dados da coleção, e servem como espelho dos dados, contendo cópias dos dados da coleção.

Os servidores regionais funcionam como provedores DiGIR comuns, mas ao invés de estarem ligados ao banco de dados de uma coleção em particular, estão ligados a um banco de dados especial que armazena espelhos dos dados de uma ou mais coleções. Estes espelhos são atualizados através de um sistema cliente/servidor instalado no nó intermediário (servidor) e no microcomputador da coleção (cliente). A interface de comunicação com o servidor é definida pelos seguintes métodos a serem invocados através do protocolo de comunicação SOAP²⁰ via HTTP:

GetCollectionId: Retorna a chave primária de uma determinada coleção.

- *Parâmetros:* código da instituição (caracter), código da coleção (caracter).
- *Valor de retorno:* chave primária da coleção (número inteiro maior que zero) ou código de erro.

Reset: Remove todos os registros de uma determinada coleção.

- *Parâmetros:* chave primária da coleção (número inteiro).
- *Valor de retorno:* "1" para indicar sucesso na operação, ou código de erro.

Suspend: Suspende as buscas em dados de uma determinada coleção (função disponível para os administradores da coleção).

- *Parâmetros:* chave primária da coleção (número inteiro).
- *Valor de retorno:* "1" para indicar sucesso na operação, ou código de erro.

Resume: Volta a disponibilizar os dados de uma determinada coleção que estavam previamente suspensos.

- *Parâmetros:* chave primária da coleção (número inteiro).
- *Valor de retorno:* "1" para indicar sucesso na operação, ou código de erro.

Block: Bloqueia alterações nos dados de uma determinada coleção (funcionalidade disponível para os administradores da rede em caso de manutenção).

- *Parâmetros:* chave primária da coleção (número inteiro).
- *Valor de retorno:* "1" para indicar sucesso na operação, ou código de erro.

Unblock: Desbloqueia dados de uma determinada coleção que estavam previamente bloqueados (funcionalidade disponível para os administradores da rede).

- *Parâmetros:* chave primária da coleção (número inteiro).
- *Valor de retorno:* "1" para indicar sucesso na operação, ou código de erro.

RemoveRecords: Remove registros específicos de uma determinada coleção.

- *Parâmetros:* chave primária da coleção (número inteiro), número total de registros a serem removidos (número inteiro), códigos de identificação dos registros a serem removidos separados por "]" (caracter).
- *Valor de retorno:* "1" para indicar sucesso na operação, ou código de erro.

InsertRecords: Inclui registros de uma determinada coleção.

- *Parâmetros:* chave primária da coleção (número inteiro), número total de registros a serem inseridos (número inteiro), registros a serem inseridos, separados por quebra de linha, com valores separados por "]" (caracter).
- *Valor de retorno:* "1" para indicar sucesso na operação, ou código de erro.

²⁰ <http://www.w3.org/TR/SOAP/>

Observação: Caso algum campo contenha o caracter separador (“|”), este deverá vir precedido de barra invertida (“\”).

GetCollectionData: Retorna os metadados da coleção que estão armazenados no servidor.

- **Parâmetros:** chave primária da coleção (número inteiro).
- **Valor de retorno:** Metadados da coleção separados por “|” (caracter: código da instituição | nome da instituição | código da coleção | nome da coleção | nome da pessoa para contato | e-mail da pessoa para contato | endereço (url) do site da coleção | código do status da coleção). Ou código de erro.

Observação: Caso algum campo contenha o caracter separador (“|”), este deverá vir precedido de barra invertida (“\”).

SetCollectionData: Atualiza os metadados de uma determinada coleção.

- **Parâmetros:** chave primária da coleção (número inteiro), metadados da coleção separados por “|” (código da instituição | nome da instituição | código da coleção | nome da coleção | nome da pessoa para contato | e-mail da pessoa para contato | endereço (url) do site da coleção | código do status da coleção).
- **Valor de retorno:** “1” para indicar sucesso na operação, ou código de erro.

Observações: Caso algum campo contenha o caracter separador (“|”), este deverá vir precedido de barra invertida (“\”). Decidiu-se por descartar os três primeiros parâmetros (só poderão ser alterados pelos administradores da rede).

Códigos de erro:

- 1: falha na comunicação com o banco de dados
- 2: operação não realizada
- 3: parâmetros incorretos

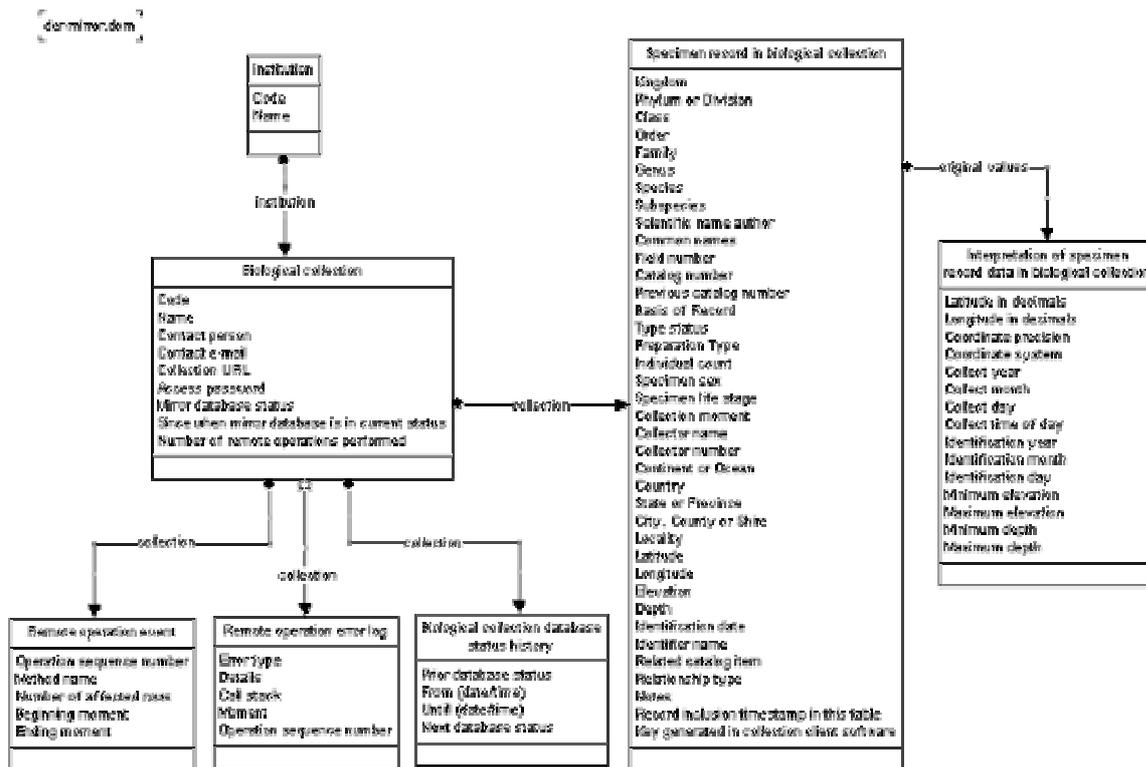
Mais detalhes sobre eventuais erros são armazenados no arquivo de log do servidor.

Nome do domínio para todos os métodos (a ser usado na URI): interface

O servidor foi desenvolvido em linguagem “Perl”, por possuir um dos módulos mais bem conceituados para utilização do protocolo SOAP (chamado SOAP::Lite), por ser uma linguagem fortemente especializada em manipulação de texto, e também por ser *open source* e muito bem documentada. O banco de dados escolhido foi o PostgreSQL²¹ devido a sua robustez, performance, e disponibilidade de recursos como: controle de transações, manutenção da integridade referencial, e disparadores automáticos.

O diagrama entidade-relacionamento do banco de dados instalado no servidor segue o modelo abaixo:

²¹ <http://www.postgresql.org/>



Ao espelhar os dados de uma coleção, todos os valores originais são armazenados intactos. Entretanto, para alguns campos, faz-se necessária uma interpretação dos dados, pois o protocolo DiGIR também estabelece padrões para alguns deles (ex: latitude e longitude devem estar em graus decimais, altitude e profundidade em metros, datas são sempre tratadas separadamente em dia, mês e ano). Desta forma, foi incorporada ao servidor a capacidade de interpretar valores de determinados campos, armazenando o resultado numa tabela para que estes possam ser imediatamente fornecidos caso solicitados em uma requisição DiGIR.

No último ano de projeto, o servidor SOAP²² desenvolvido em Perl²³ e instalado nos servidores regionais com o objetivo de receber as atualizações de dados provenientes das coleções e armazená-los em um banco relacional foi incrementado principalmente no tocante às regras de interpretação de dados. Cada registro enviado para servidores regionais possui seus valores originais armazenados, porém alguns campos que exigem formatação específica de acordo com o DarwinCore (tais como coordenadas e datas) são passíveis de interpretação em seus valores. Ao longo do projeto as regras de interpretação foram melhoradas e expandidas de forma a abranger a maioria das variações de formatação detectadas nos bancos de dados das coleções biológicas.

O componente que atua como cliente é um aplicativo desenvolvido em Java, por nós denominado *spLinker*, descrito a seguir.

2.2.4 SPLINKER

A utilização de um *servidor regional* como provedor equaciona o problema de conectividade, porém gera o problema de envio dos dados da coleção ao *servidor regional*. Para resolver este problema foi criado o aplicativo *spLinker*.

²² <http://www.w3.org/TR/SOAP>

²³ <http://www.perl.org/>

O **spLinker** é um aplicativo desenvolvido pelo CRIA para migrar os dados das coleções que não têm uma conexão Internet capaz de satisfazer os quesitos de velocidade e estabilidade necessários à instalação de um servidor web.

Para a concepção do spLinker foram utilizadas as seguintes premissas:

- Ser o mais independente possível de alguma plataforma. Isto porque não sabemos qual a plataforma utilizada por coleções que queiram entrar no projeto futuramente e é uma premissa do projeto impor o mínimo de mudanças ou exigências técnicas às coleções.
- Exigir a instalação do menor número possível de software auxiliares.
- Ser de fácil entendimento e utilização.

Com base nestas premissas foi escolhido desenvolver o spLinker na linguagem Java. Java é uma das linguagens orientadas a objeto mais independente de plataforma e sua execução exige apenas a instalação da jvm (Java Virtual Machine) que nas versões mais atuais do Windows já é instalada com o sistema.

A figura 2 a seguir mostra a interface do spLinker.

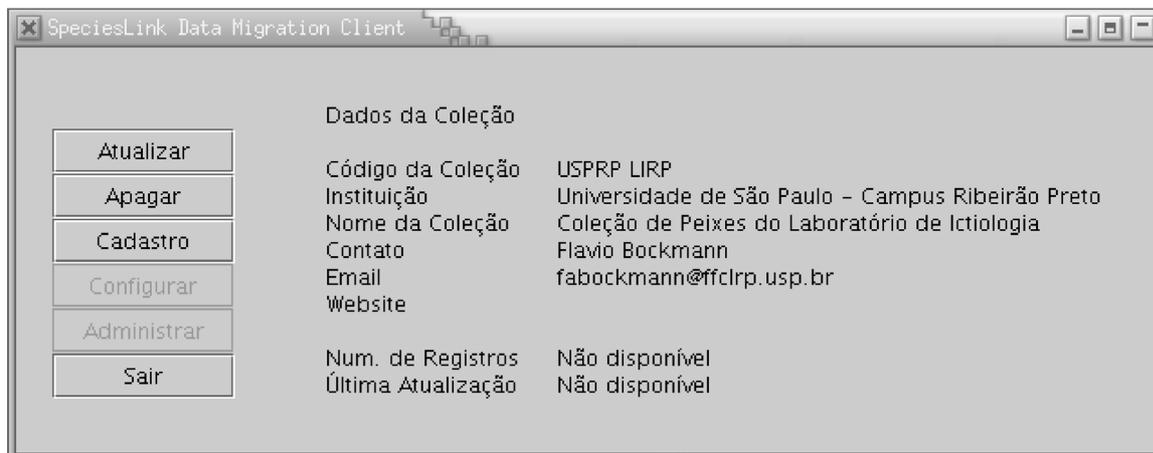


Figura 2. Interface do spLinker

Ao usuário é permitido realizar apenas três operações:

- **Atualizar:** os dados serão lidos na base de dados da coleção, comparados com um repositório local que contém os últimos dados enviados ao servidor regional e serão enviados apenas os registros novos (a inserir) ou os removidos (a remover). Os registros modificados são tratados como uma remoção seguida de uma inserção.
Note que este mecanismo de enviar apenas as alterações da base da coleção, apesar de replicar novamente a base, é fundamental se pensarmos em uma conexão precária da coleção com o servidor regional.
- **Apagar:** este comando envia uma requisição ao servidor regional pedindo que todos os registros pertencentes à coleção em questão sejam apagados. Em seguida, o repositório local é removido.
- **Cadastro:** lê os metadados das coleções cadastradas no servidor regional.

O spLinker foi projetado para ler os dados das coleções através de:

- Conexões aceitas pelo JDBC²⁴: O JDBC é a forma padrão para acesso a bancos de dados SQL quando se programa em Java.
- Arquivo texto: caso não seja possível o acesso aos dados via JDBC, basta conseguir gerá-los em arquivo texto que o *spLinker* será capaz de lê-los e enviá-los ao servidor regional. Esta solução foi implementada pois alguns programas gerenciadores de coleções mantêm os dados em formato proprietário, porém permitem a exportação no formato texto, como é o caso do software Biota.

É importante notar que a introdução de novas formas de leitura dos dados é de fácil implementação, pois é utilizado o paradigma de orientação a objetos.

A configuração do *spLinker* é feita através de arquivos texto. Em sua configuração são preenchidas informações sobre:

- Como se conectar ao servidor regional (URL e URI).
- Os dados sobre a coleção: identificador, nome, instituição, etc.
- O mapeamento entre os nomes dos campos (JDBC) ou colunas (texto) dos dados originais da coleção e os respectivos nomes utilizados no esquema conceitual utilizado pelo DiGIR (a rede *speciesLink* utiliza os campos descritos no DarwinCore).
- O tipo de acesso: JDBC ou Texto
- Informações sobre o acesso via JDBC: URL, driver, usuário e senha da conexão SQL, cláusulas FROM e WHERE para o comando SELECT a ser utilizado para extrair os dados da base da coleção. Note que modificando a cláusula WHERE pode-se filtrar dados que por algum motivo não possam ser disponibilizados.
- Informações sobre o acesso via texto: nome do arquivo, tipo de codificação do arquivo, nome da coluna a ser utilizada como filtro e expressão regular a ser utilizada como filtro.

O *spLinker* sofreu várias alterações ao longo do desenvolvimento dos trabalhos de forma a melhor suportar condições adversas de conexão. Na eventualidade de ocorrer algum erro durante a transmissão de dados, são feitas novas tentativas de comunicação com o servidor regional. É também possível configurar dados de proxy quando este está presente no ambiente de rede da coleção.

2.3. DESENVOLVIMENTO DA INTERFACE DE BUSCA

A interface de busca via Internet utilizada pelo *speciesLink* foi totalmente construída no âmbito deste projeto. Foi desenvolvido um módulo em linguagem Perl²⁵ capaz de interagir com o portal DiGIR, bem como uma série de scripts em Perl que fazem uso deste módulo.

Ao acessar a interface de busca são inicialmente exibidas as condições de uso da rede onde estão enfatizados aspectos como o respeito a eventuais restrições específicas de cada provedor de dados, e a obrigatoriedade de dar créditos à rede e às coleções biológicas caso os dados venham a ser utilizados em trabalho publicado.

A interface de busca exibe todas as coleções participantes, separadas por coleções nacionais e coleções internacionais a serem pesquisadas, oferecendo ao usuário a opção de selecionar a coleção ou coleções onde ele deseja realizar a busca (figura 3).

²⁴ JDBC Java Database Connectivity (java.sun.com/products/jdbc/)

²⁵ <http://www.perl.org/>

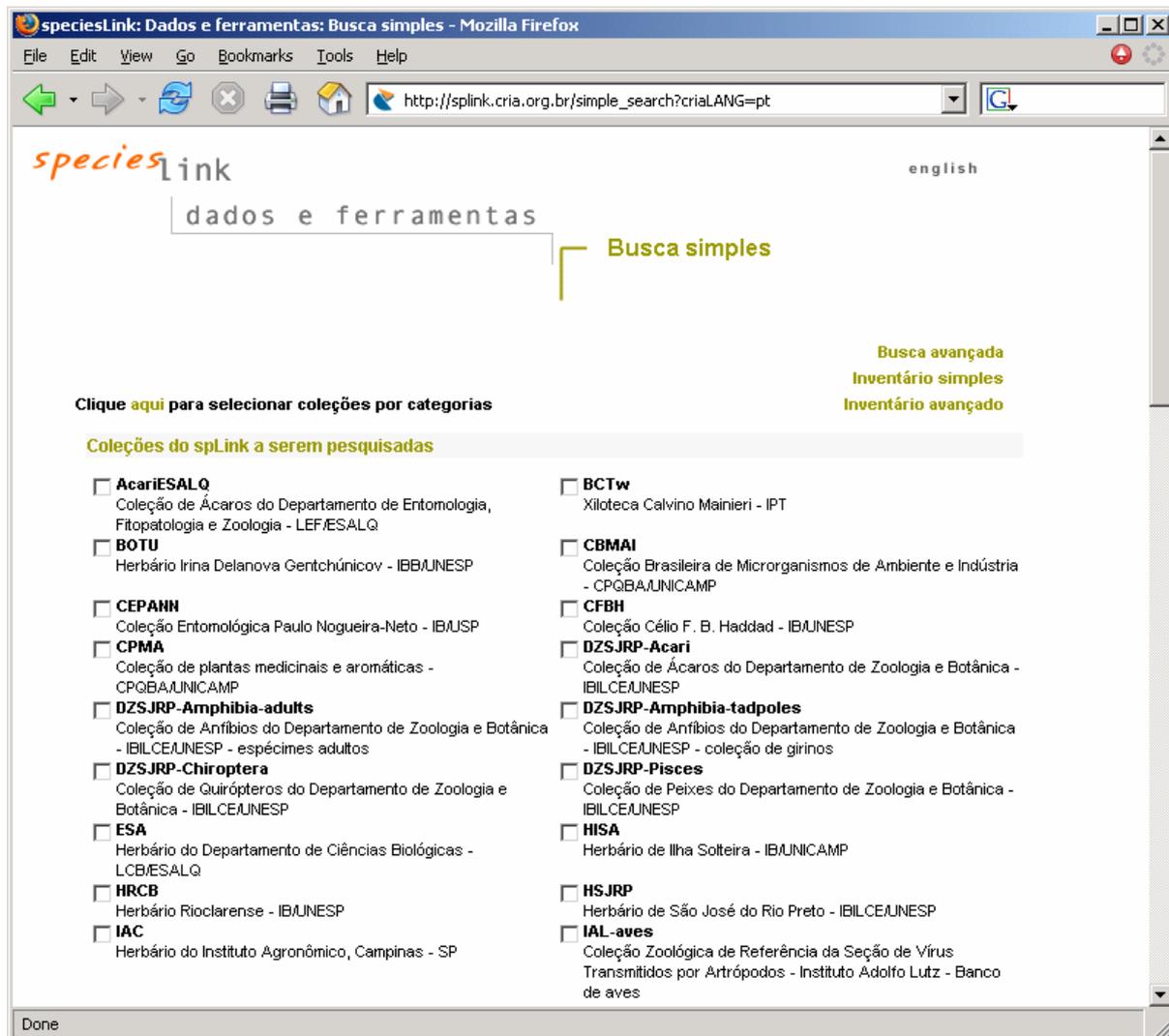
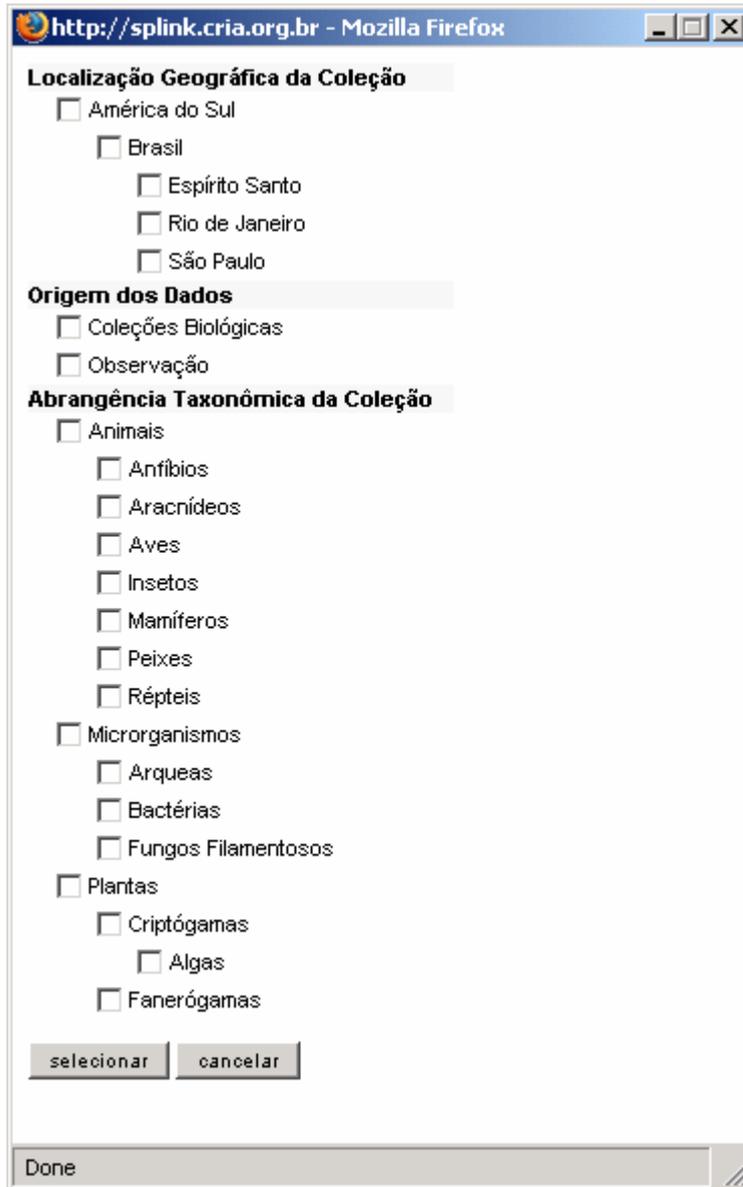


Figura 3. Parte da página inicial de busca simples

As coleções nacionais foram todas categorizadas através da configuração de seus metadados (campo “keywords”), de forma que também é possível selecionar as coleções de de acordo com o grupo taxonômico ou a sua localização geográfica.



http://splink.cria.org.br - Mozilla Firefox

Localização Geográfica da Coleção

- América do Sul
 - Brasil
 - Espírito Santo
 - Rio de Janeiro
 - São Paulo

Origem dos Dados

- Coleções Biológicas
- Observação

Abrangência Taxonômica da Coleção

- Animais
 - Anfíbios
 - Aracnídeos
 - Aves
 - Insetos
 - Mamíferos
 - Peixes
 - Répteis
- Microrganismos
 - Arqueas
 - Bactérias
 - Fungos Filamentosos
- Plantas
 - Criptógamas
 - Algas
 - Fanerógamas

selecionar cancelar

Done

Figura 4. Seleção de coleções por região geográfica ou grupo taxonômico

Foram disponibilizados aos usuários da rede quatro tipos de busca: busca simples, busca avançada, inventário simples e inventário avançado.

Busca simples: A busca simples oferece um conjunto fixo de campos que, se preenchidos, produzem uma requisição de busca com um filtro especificando que todas as condições devem ser satisfeitas. Foram escolhidos os campos utilizados com mais frequência pelos usuários: Grupo taxonômico, Nome científico, Status de tipo, Localidade, Município, Estado, País, e Nome do Coletor. Com exceção do campo Grupo taxonômico, todos os outros campos texto geram condições de busca aproximada (operador “contendo”, ao invés do operador “igual a”).

Filtros

Grupo taxonômico Nome científico Tipo

Localidade Município Estado País

Coletor

Figura 5. Campos que podem ser utilizados na busca simples

Busca Avançada: A busca avançada oferece ao usuário a possibilidade de especificar completamente o critério de busca através de combinações de expressões de comparação (campo do DarwinCore + operador de comparação do DiGIR + valor especificado) com operadores lógicos do DiGIR.

Filtros

Ano da coleta maior ou igual a 2002

e

Ano da coleta menor ou igual a 2005

e

Município igual a Campinas

adicionar condição remover última condição

Figura 6. Exemplo de filtro especificado em uma busca avançada

Ambos os tipos de busca oferecem algumas opções para formatação do resultado:

- Conjunto de dados a retornar: *Coordenadas* (retorna um conjunto de campos suficiente para ser mapeado e utilizado em procedimentos de modelagem), *Resumido* (retorna um conjunto mínimo de campos), ou *Completo* (retorna todos os campos do DarwinCore)
- Número de registros por coleção: 10, 50, 100 ou 5.000; A opção de recuperar todos os registros não está disponível porque pode gerar uma quantidade suficientemente grande para comprometer o bom funcionamento do sistema
- Tipo de documento de resposta: HTML ou planilha Excel
- Opção para visualização dos dados em mapa

As figuras a seguir são exemplos de apresentação do resultado de buscas, no caso, html e mapa.

Sistema de Informação Ambiental - BIOTA/FAPESP

Condições de uso / Como citar

spLink	Código da instituição	Código da coleção	Catálogo	Nome científico	Longitude	Latitude
SP	FAPESP	SinBiota	C10248T86972	leptodactylus cf.podicipinus	-47.0608	-22.9056
SP	FAPESP	SinBiota	C10248T86973	leptodactylus cf.podicipinus	-47.0608	-22.9056
SP	FAPESP	SinBiota	C10325T75904	heliconius ethilla narcaea	-47.05	-22.9
SP	FAPESP	SinBiota	C10325T83804	dynamine atheron maeon	-47.05	-22.9
SP	FAPESP	SinBiota	C10325T87824	enartia lina psamathe	-47.05	-22.9
SP	FAPESP	SinBiota	C10325T87825	melete lycimnia	-47.05	-22.9
SP	FAPESP	SinBiota	C10325T87826	ascia monuste	-47.05	-22.9
SP	FAPESP	SinBiota	C10325T87827	anteos clorinde	-47.05	-22.9
SP	FAPESP	SinBiota	C10325T87828	phoebis sennae	-47.05	-22.9
SP	FAPESP	SinBiota	C10325T87829	phoebis philea	-47.05	-22.9

10 / 93

Figura 7 Exemplo de um resultado de busca em HTML para dados do SinBiota.

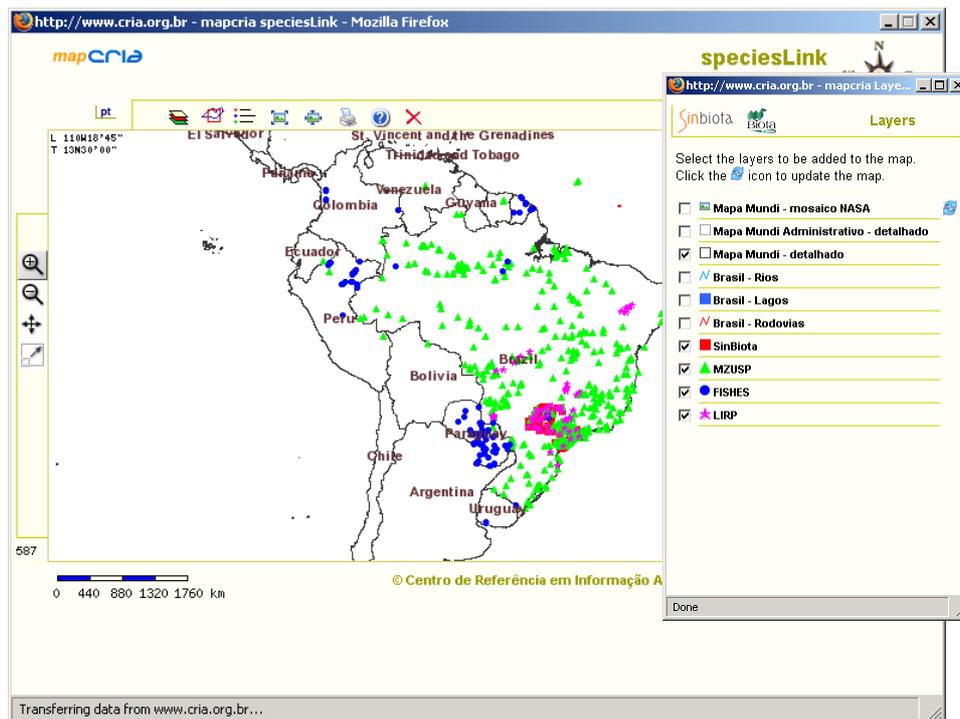


Figura 8. Exemplo da visualização dos dados em mapa (exemplo utilizado: busca pelo gênero *Hoplias*)

Inventário simples: Oferece as mesmas opções de filtro disponíveis na busca simples, porém realiza uma operação de inventário através do protocolo DiGIR. Operações de inventário são feitas em um determinado campo, e o resultado é um conjunto de valores distintos para o campo especificado, seguido da quantidade de registros associada.

Filtros

Grupo taxonômico Nome científico Tipo

Localidade Município Estado País

Coletor

Formatação do resultado

Campo a retornar Documento de resposta

Figura 9. Exemplo de uma busca do tipo inventário para um determinado município (no caso “Campinas”), solicitando a lista de espécies encontradas.

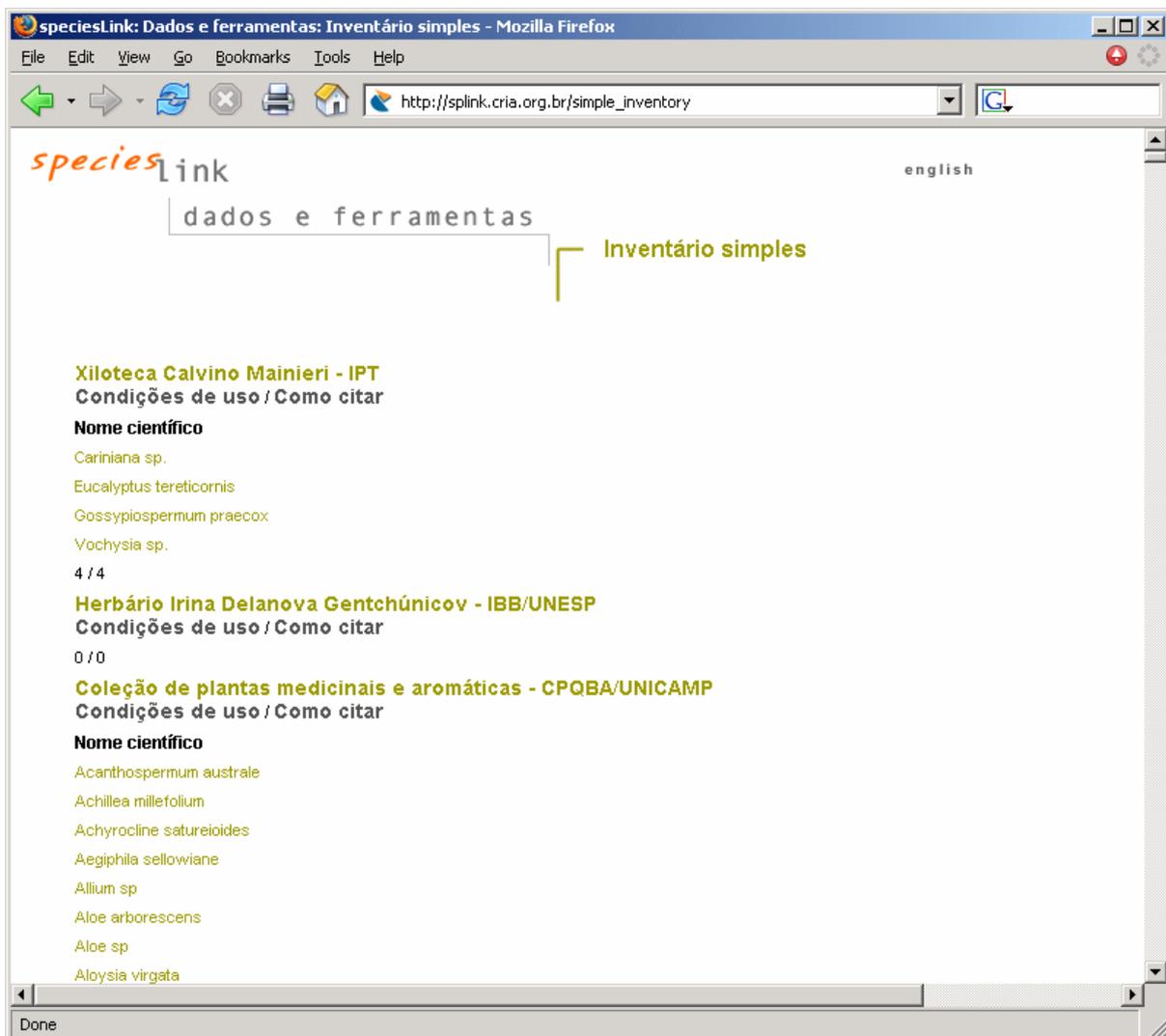


Figura 10. Resultado (parcial) da busca por “Campinas”, solicitando o retorno do campo “nome científico”

Inventário avançado: Oferece as mesmas opções de filtro disponíveis na busca avançada, porém realiza uma operação de inventário através do protocolo DiGIR.

Filtros

Ano da coleta contendo

e

Município contendo

Formatação do resultado

Campo a retornar Documento de resposta

Figura 11. Página de busca no “Inventário Avançado”

É importante observar que cada coleção pode especificar condições de uso e instruções sobre como citar a coleção. Tais informações estão sempre acessíveis através de links que aparecem logo abaixo do nome de cada coleção no resultado de qualquer tipo de busca. Estas informações fazem parte dos metadados de cada coleção. O conjunto principal de metadados de cada coleção pode ser acessado clicando-se sobre o nome da coleção.

IBSP-Herpeto
Coleção Herpetológica Alphonse Richard Hoge - Instituto Butantan

A Coleção Herpetológica "Alphonse Richard Hoge" é uma coleção de serpentes de representatividade mundial, com ênfase nas serpentes Neotropicais e, em especial Brasileiras. É a maior coleção de serpentes Neotropicais do mundo, com quase 80.000 exemplares. Faz parte da coleção principal, uma série de coleções acessórias como: Exemplares tipo, Coleção osteológica, Coleção de hemipênis, Banco de imagens, Banco de tecidos, Coleção do Hospital Vital Brazil do Instituto Butantan (acervo de 8.000 exemplares de serpentes causadoras de acidentes, trazidas juntamente com os acidentados ao Hospital), Cadastro do recebimento das serpentes que entraram no Instituto Butantan desde 1908 até agora. A maior parte do acervo é composta por exemplares conservados em via úmida, ou seja, submersos em álcool etílico 70%, acondicionados em vidros e dispostos em prateleiras metálicas. O acervo está em processo de informatização em Excel, já tendo sido magnetizado dados referentes a cerca de 70% do acervo.

Contato administrativo
 Dr. Francisco Luís Franco - (Curador)
ffranco@butantan.gov.br
 +55 11 3726-7222 r: 2245

Contato técnico
 Sidnei de Souza - (Analista de Sistemas)
sidnei@cria.org.br
 +55 19 32880466

Condições de uso
 Não especificadas

Número de registros disponíveis
 66167

Última atualização em 2005-10-11 14:27:32-03

Figura 12. Exemplo de página com metadados de uma coleção.

2.4. O FUTURO DE REDES DE BUSCA EM DADOS DISTRIBUÍDOS SOBRE BIODIVERSIDADE

O envolvimento com o desenvolvimento do protocolo DiGIR, com pessoas e instituições responsáveis por redes semelhantes, e principalmente com a organização internacional que busca definir padrões e protocolos para informática para biodiversidade (TDWG²⁶) culminou com um papel importante do CRIA na definição de um novo protocolo. O TAPIR (TDWG *Access Protocol for Information Retrieval*) foi criado com o objetivo de unificar os protocolos DiGIR e BioCAsE²⁷, incorporando o que cada um tem de melhor, adicionando novas funcionalidades. No momento estão sendo desenvolvidos dois protótipos de provedores TAPIR (em Python e Java) com o objetivo de validar o conceito do novo protocolo. A etapa seguinte será a de gradativamente estimular as redes de dados de biodiversidade existentes a migrarem para o novo software e adotarem o novo protocolo.

O TDWG está também buscando padronizar outros tipos de dados sobre biodiversidade, como dados descritivos de espécies (SDD, *Structured Descriptive Data*), dados de conceitos taxonômicos (TCS, *Taxonomic Concept Transfer Schema*), dados de observação, e metadados de coleções biológicas. A tendência é que as futuras redes adotem estes padrões com vista a maximizar a interoperabilidade e integração de diferentes tipos de dados.

3. APOIO ÀS COLEÇÕES

Por se tratar de um projeto de pesquisa, onde queríamos testar um modelo de sistema distribuído de dados que fosse o mais inclusivo possível, consideramos importante lidar com coleções nos mais diversos estágios de organização e desenvolvimento. Embora grande parte das principais coleções científicas paulistas tenha se modernizado nos últimos anos devido a incentivos, em especial providos pela Fapesp, a situação das diferentes coleções é bastante heterogênea. Encontramos coleções informatizadas, parcialmente informatizadas e outras que sequer haviam escolhido o *software* a ser utilizado. O único critério comum exigido para que uma coleção pudesse fazer parte da rede era o compromisso de compartilhar os dados através de um sistema de acesso público na Internet.

No início dos trabalhos, várias coleções foram contatadas para verificar se havia interesse em participar dessa iniciativa. Foi um trabalho de convencimento muito grande porque havia, e em alguns casos ainda há, uma resistência ao acesso livre e aberto a dados através da Internet, por qualquer pessoa interessada, sem qualquer tipo de controle.

Demonstrado o interesse em participar e em compartilhar dados não sensíveis, era solicitado da coleção o preenchimento de um questionário (anexo 1) para avaliar o grau de informatização da coleção e a qualidade de sua conexão à Internet. Com base nas respostas tínhamos uma idéia das principais dificuldades que encontraríamos.

A estratégia utilizada para fomentar e auxiliar as coleções na digitalização de seus acervos incluiu:

- Alocar equipamento básico para cada coleção participante;
- Permitir ao curador total autonomia na escolha do software para o gerenciamento do acervo;
- Oferecer apoio e orientação em relação ao software escolhido, inclusive realizando trabalhos de recuperação e importação de dados;
- Auxiliar o curador no desenvolvimento de filtros para conversão de dados (p.ex. para a conversão dos diferentes tipos de representação de coordenadas geográficas) e para o bloqueio de dados sensíveis; e,

²⁶ <http://www.tdwg.org/>

²⁷ <http://www.biocase.org/dev/protocol/index.shtml>

- Abrir às coleções a possibilidade de recrutarem bolsistas para o trabalho de digitação.

No último ano do projeto o CRIA foi procurado por várias coleções de maneira espontânea, solicitando a sua inclusão na rede, mesmo sem o repasse de hardware, mas sempre contando com o apoio técnico do CRIA. Houve, na nossa opinião, uma enorme mudança cultural no sentido dos curadores compreenderem a importância e vantagem de compartilharem os dados de seus acervos.

3.1. SOFTWARE DE GERENCIAMENTO

Para auxiliar as coleções científicas no processo de informatização e para estudar formas de integrar os diferentes acervos, foi realizada uma avaliação preliminar dos *software* disponíveis no mercado para a informatização de coleções biológicas. Foram analisados os seguintes *software*:

BRAHMS - *Botanical Research and Herbarium Management System*: é um programa voltado para pesquisa e para gerenciamento de coleções botânicas. Foi desenvolvido no Department of Plant Sciences, University of Oxford.

BIOTA - *Biodiversity Database Manager*: gerenciador de dados para coleções baseado em informação sobre espécimens e dados biológicos.

Specify - programa gerenciador de dados provenientes de diferentes grupos taxonômicos (plantas, mamíferos, aves, peixes, répteis, anfíbios, vertebrados fósseis, plantas fósseis e insetos).

Mantis - *Manager of Taxonomic Information and Specimens*: software dedicado à entomologia e possui recursos multimídia, como imagens e sons.

BioLink - Este programa tem por objetivo auxiliar no trabalho com informação taxonômica sobre espécimens.

BIOTICA - programa desenvolvido para auxiliar o trabalho dos curadores de coleções em relação a dados nomenclaturais, geográficos, bibliográficos e ecológicos.

ARTEDIAN – aplicação MS-Access2000 para o gerenciamento de coleções ictiológicas que está sendo utilizado pelo Museu de História Natural da Suécia.

A análise completa de todos os *software* está disponível *on-line* no seguinte endereço: http://splink.cria.org.br/collection_management_software.

A arquitetura da rede foi desenhada de forma a permitir às coleções total autonomia quanto à escolha do *software* para o gerenciamento de seus acervos. A tabela 1 traz a relação de todas as coleções integradas à rede, e o *software* de gerenciamento adotado.

Tabela 1. Relação das coleções participantes da rede *speciesLink* e *software* de gerenciamento utilizado.

	Coleção	software
1	Coleção de Ácaros do Departamento de Entomologia, Fitopatologia e Zoologia (AcariESALQ)	Biota
2	Coleção de Ácaros (DZSJRP-Acari)	Biota
3	Coleção de Quirópteros (DZSJRP-Chiroptera)	Biota
4	Coleção de Peixes DZSJRP (DZSJRP-Pisces)	Biota
5	Coleção de Peixes do Laboratório de Ictiologia de Ribeirão Preto (LIRP)	Biota
6	Coleção Científica de Aranhas (Araneae) do Depto. de Zoologia da Unesp, Campus Botucatu (UBTU)	Biota

	Coleção	software
7	Herbário da Escola Superior de Agricultura Luiz de Queiroz (ESA)	Brahms
8	Herbário de Ilha Solteira (HISA)	Brahms
9	Herbário Rioclarense (HRCB)	Brahms
10	Herbário de São José do Rio Preto (HSJRP)	Brahms
11	Herbário Dimitri Sucre Benjamin (JBRJ)	Brahms
12	Coleção de Fanerógamas do Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo" (SP)	Brahms
13	Herbário do Departamento de Botânica, SPF-IB/USP (SPF)	Brahms
14	Coleção de Algas do Departamento de Botânica, SPF-IB/USP (SPF-Algae)	Brahms
15	Xiloteca do Instituto de Biociências da Universidade de São Paulo (SPFw)	Brahms
16	Herbário Dom Bento Pickel (SPSF)	Brahms
17	Herbário da Universidade Estadual de Campinas (UEC)	Brahms
18	Coleção de plantas medicinais e aromáticas (CPMA)	MS-Access
19	Coleção Zoológica de Referência da Seção de Vírus Transmitidos por Artrópodos - Banco de Aves (IAL-aves)	MS-Access
20	Coleção Zoológica de Referência da Seção de Vírus Transmitidos por Artrópodos (IAL-roedores)	MS-Access
21	Coleção Entomológica "Adolph Hempel" do Instituto Biológico (IBSP-IB)	MS-Access
22	Coleção de Anfíbios do Museu de História Natural "Prof. Dr. Adão José Cardoso" (MHN-anfibios)	MS-Access
23	Coleção de Aves do Museu de História Natural "Prof. Dr. Adão José Cardoso" (MHN-aves)	MS-Access
24	Coleção de Mamíferos do Museu de História Natural "Prof. Dr. Adão José Cardoso" (MHN-mamiferos)	MS-Access
25	Coleção de Peixes do Museu de História Natural "Prof. Dr. Adão José Cardoso" (MHN-peixes)	MS-Access
26	Coleção de Répteis do Museu de História Natural "Prof. Dr. Adão José Cardoso" (MHN-repteis)	MS-Access
27	Coleção Camargo (RPSP)	MS-Access
28	Coleção de Algas do Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo" (SP-Algae)	MS-Access
29	Xiloteca Calvino Mainieri (BCTw)	MS-Excel
30	Coleção do Laboratório de Abelhas do IB/USP (CEPANN)	MS-Excel
31	Coleção "Célio F. B. Haddad" (CFBH)	MS-Excel
32	Coleção de Anfíbios (DZSJRP-Amphibia-adults)	MS-Excel
33	Coleção de Anfíbios do Departamento de Zoologia e Botânica - Coleção de Girinos (DZSJRP-Amphibia-tadpoles)	MS-Excel
34	Coleção de Culturas de Fitobactérias do Instituto Biológico (IBSBF)	MS-Excel
35	Coleção Acarológica do Instituto Butantan (IBSP-Acari)	MS-Excel
36	Coleção Herpetológica "Alphonse Richard Hoge" (IBSP-Herpeto)	MS-Excel
37	Coleção do Museu de Entomologia da FEIS/Unesp (MEFEIS)	MS-Excel
38	Herbário "Irina Delanova Gemtchújnicov" (BOTU)	MS-Works
39	Coleção de Peixes do Museu de Zoologia da USP (MZUSP)	Specify
40	Herbário do Instituto Agrônomo de Campinas (IAC)	mySQL
41	Coleção Brasileira de Microrganismos de Ambiente e Indústria (CBMAI)	proprietário

Podemos observar que quase todos os herbários adotaram o *software* **Brahms** para o gerenciamento de seus acervos. Os software da Microsoft, Access e Excel são utilizados por cerca de 50% das coleções e 6 coleções, cerca de 15% do total utilizam o software Biota.

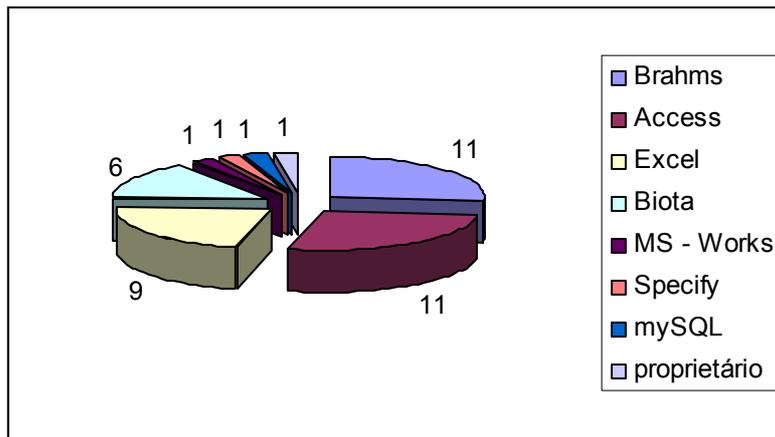


Figura 13. Software de gerenciamento utilizados pelas coleções da rede speciesLink

3.2. VISITAS TÉCNICAS

Para poder prestar apoio às coleções é importante conhecer as características funcionais de cada acervo. O protocolo DiGIR adotado para a integração dos acervos utiliza o DarwinCore²⁸ como modelo de dados. Portanto, para a integração dos acervos à rede é necessário mapear os campos obrigatórios e auxiliares de cada coleção de forma adequada. Isso via-de-regra requer a realização de visitas técnicas a cada coleção, para discutir o processo de informatização e integração à rede speciesLink.

Baseado nos estudos de software o CRIA ofereceu a cada coleção suporte quanto à escolha do software de gerenciamento do acervo, mostrando as principais características e os pontos fortes e fracos de cada sistema. No entanto, sempre coube ao curador escolher o sistema que melhor se adaptasse às suas necessidades.

O projeto ofereceu à maioria das coleções a instalação de uma infra-estrutura mínima (microcomputador, *no-break*, *software*) para viabilizar a sua participação na rede. Na maioria dos casos o equipamento era entregue e instalado na primeira visita. Além da configuração e instalação do software, na maioria dos casos onde a coleção havia optado por um novo software de gerenciamento, o trabalho da equipe incluía a importação dos dados do software anterior.

A tabela 2 a seguir traz a relação de visitas por coleção ao longo do desenvolvimento dos trabalhos.

Tabela 2. Número de visitas às coleções da rede speciesLink

Coleção	Visitas
Coleção de Ácaros do Departamento de Entomologia, Fitopatologia e Zoologia (AcariESALQ)	2
Coleção de Ácaros (DZSJRP-Acari)	2
Coleção de Quirópteros (DZSJRP-Chiroptera)	3
Coleção de Peixes DZSJRP (DZSJRP-Pisces)	2
Coleção de Peixes do Laboratório de Ictiologia de Ribeirão Preto (LIRP)	1
Coleção Científica de Aranhas (Araneae) do Depto. de Zoologia da Unesp, Campus Botucatu (UBTU)	3
Herbário da Escola Superior de Agricultura Luiz de Queiroz (ESA)	6
Herbário de Ilha Solteira (HISA)	2
Herbário Rioclarense (HRCB)	3

²⁸ Darwin Core (<http://darwincore.calacademy.org/>)

Coleção	Visitas
Herbário de São José do Rio Preto (HSJRP)	2
Herbário Dimitri Sucre Benjamin (JBRJ)	0
Coleção de Fanerógamas do Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo" (SP)	7
Herbário do Departamento de Botânica, SPF-IB/USP (SPF)	7
Coleção de Algas do Departamento de Botânica, SPF-IB/USP (SPF-Algae)	2
Xiloteca do Instituto de Biociências da Universidade de São Paulo (SPFw)	4
Herbário Dom Bento Pickel (SPSF)	2
Herbário da Universidade Estadual de Campinas (UEC)	8
Xiloteca Calvino Mainieri (BCTw)	4
Coleção de plantas medicinais e aromáticas (CPMA)	1
Coleção Zoológica de Referência da Seção de Vírus Transmitidos por Artrópodos - Banco de Aves (IAL-aves) e Roedores (IAL-roedores)	4
Coleção Entomológica "Adolph Hempel" do Instituto Biológico (IBSP-IB)	2
Museu de História Natural "Prof. Dr. Adão José Cardoso": Coleções de Anfíbios, Aves, Mamíferos, Peixes e Répteis.	3
Coleção Camargo (RPSP)	2
Coleção de Algas do Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo" (SP-Algae)	5
Coleção do Laboratório de Abelhas do IB/USP (CEPANN)	2
Coleção "Célio F. B. Haddad" (CFBH)	3
Coleção de Anfíbios (DZSJRP-Amphibia-adults) e de Girinos (DZSJRP-Amphibia-tadpoles) do Departamento de Zoologia e Botânica	3
Coleção de Culturas de Fitobactérias do Instituto Biológico (IBSBF)	2
Coleção Acarológica do Instituto Butantan (IBSP-Acari)	3
Coleção Herpetológica "Alphonse Richard Hoge" (IBSP-Herpeto)	3
Coleção do Museu de Entomologia da FEIS/Unesp (MEFEIS)	1
Herbário "Irina Delanova Gemtchújnicov" (BOTU)	2
Coleção de Peixes do Museu de Zoologia da USP (MZUSP)	5
Herbário do Instituto Agrônomo de Campinas (IAC)	1
Coleção Brasileira de Microrganismos de Ambiente e Indústria (CBMAI)	1
Número total de visitas	103

Além de visitas *in loco*, a equipe do CRIA mantém suporte contínuo a todas as coleções via email, telefone e via *NetMeeting* (veja item a seguir).

3.3. SUPORTE REMOTO

Foi utilizado o software *NetMeeting* para que a equipe pudesse oferecer suporte remoto. Através desse software é possível assumir o controle à distância do computador alocado à coleção, permitindo alterar a configuração e atualizar software.

Existem outros software de acesso remoto, porém, optou-se pelo *NetMeeting* por ser um programa do pacote Windows (sistema operacional adotado pelas coleções) e por requerer a autorização de alguém da coleção para permitir o acesso ao computador. Não queríamos que os curadores das coleções perdessem a autonomia e o controle sobre o sistema.

A ferramenta provou ser muito eficiente principalmente em relação às coleções mais distantes de Campinas, resultando em economia de recursos e tempo. O fato de não precisar deslocar a equipe até a coleção faz com que estas tenham mais liberdade para solicitar qualquer apoio específico e o CRIA, por sua vez, consegue ser mais prestativo e eficiente.

3.4. INSTALAÇÃO DO EQUIPAMENTO

O equipamento básico distribuído a cada coleção foi um micro computador, um *no-break* e *software* (MS Windows e Microsoft Office). Durante os quatro anos de projeto foram instalados 35 computadores, 24 nobreaks e 6 servidores. Para acelerar o processo de digitação, foram entregues às coleções 4 computadores do CRIA, já obsoletos para o desenvolvimento de software, mas úteis para a digitação de dados.

A tabela 3 a seguir traz a relação dos equipamentos entregues nas duas fases do projeto.

Tabela 3. Equipamento entregue em comodato às coleções participantes da rede *speciesLink*

Fase	Coleção	Computador	Nobreak	Servidor
Fase 1	Coleção de Algas do Herbário do IBt	1		
	CBMAI - CPQBA – UNICAMP	1		
	Coleção de Fanerógamas do Herbário do IBt	1		
	UEC-UNICAMP	2		
	Col. Peixes Dep. Zoologia e Botânica - DZSJRP – UNESP	1		
	Coleção de Ácari Depto Zoologia e Botânica UNESP de SJRP	1		
	Coleção de Ácari Depto Entomologia, Fitopatologia e Zoologia Agrícola ESALQ-USP	1		
	Herbário Depto. Botânica – USP	1		
	Coleção Peixes MZUSP	1		
	Herbário IAC	1		
	LIRP - Ribeirão Preto *		2	1
	Coleção de Peixes Dep. Zoologia e Botânica - DZSJRP – UNESP			1
	Escola Politécnica de São Paulo			1
	CRIA			1
	Serviço Técnico de Informática do IB / Campus de Botucatu / UNESP			1
Total Fase 1		11	2	5
Fase 2	Herbário ESA / ESALQ	1	1	
	Herbário Depto. Botânica – USP	1	1	
	IBSBF	1	1	
	Museu Prof. Adão José Cardoso / UNICAMP	1	1	
	Coleção Herpetológica Alphonse Richard Hoje / Inst. Butantan	1		1
	Coleção Acarológica do Instituto Butantan	1	1	
	Coleção de Abelhas da USP	1	1	
	Xiloteca do Instituto de Biociências / USP	2	1	
	Xiloteca Calvino Manieri / IPT	1	1	
	Coleção de Anfíbios Célio F. B. Haddad / UNESP Rio Claro	1		1
	Coleção Anfíbios / UNESP São José do Rio Preto	1	1	
	Coleção de Quirópteros / S. José do Rio Preto	1	1	
	Herbario / São Jose do Rio Preto	1	1	
	Herbário Dom Bento Pickel / IF	1	1	
	Herbário Rioclarense / UNESP Rio Claro	1	1	
	Coleção Zoológica de Referencia da Seção de Vírus Transmitidos por Artrópodos – IAL	2		1
	Herbário BOTU ("Irina Delanova Gentchúnicov") / UNESP Botucatu	1		1
	Coleção Científica de Aranhas, Instituto de Biociências, UNESP, Botucatu	1		1
	Coleção Camargo (Abelhas) / USP Rib. Preto	1	1	

Fase	Coleção	Computador	Nobreak	Servidor
	Herbário HISA / UNESP Ilha Solteira	1	1	
	Coleção do Museu de Entomologia da FEIS / UNESP Ilha Solteira	1	1	
	Coleção Entomológica "Adolph Hempel" / Inst. Biológico	1	1	
	NAC-Agronomia / Cache Node Ilha Solteira			1
Total Fase 2		24	22	1
Total Geral		35	24	6

3.5. TREINAMENTO

Um grande problema das coleções é a falta de recursos humanos fixos e a alta rotatividade de alunos e bolsistas. Isso faz com que seja necessário realizar várias visitas para oferecer treinamento no uso do software de gerenciamento adotado pela coleção e das ferramentas de informática desenvolvidas ao longo do projeto. Em algumas coleções, foi necessário oferecer treinamentos do mesmo software em três ou mais épocas diferentes para pessoas diferentes.

Recomendações:

Seria importante programar cursos sobre a utilização dos principais software de gerenciamento adotados pelas coleções como Brahms e Biota, e até aqueles incluídos no Office, como Access e Excel.

Nos simpósios do programa Biota seria importante incluir sessões demonstrativas ou minicursos sobre o uso das ferramentas desenvolvidas para a rede speciesLink (p.ex. a interface de busca, data cleaning, geoLoc e conversor).

3.6. BOLSAS

Além do treinamento do aluno, sempre objeto de qualquer projeto financiado pela Fapesp, as bolsas de treinamento técnico são fundamentais para o processo de informatização dos acervos.

Hoje, 41 Coleções fazem parte do speciesLink. No período de vigência do projeto, 24 coleções solicitaram bolsas e 16 foram contempladas.

Foram solicitadas 47 bolsas, sendo que 26 foram implementadas e concluídas, 15 bolsistas desistiram (ainda no processo de implementação, devido à demora na análise ou pouco tempo depois da bolsa implementada, devido ao aparecimento de novas oportunidades) e 6 pedidos foram negados (devido à inadequação no currículo do candidato ou porque o período de duração da bolsa, durante a análise, acabou sendo inferior a 6 meses).

A tabela a seguir mostra o número de bolsas aprovadas, por coleção.

Tabela 4. Bolsas concedidas aos participantes da rede speciesLink

Coleção	Instituição	Pós-Doc	TT3	TT2	TT1	Total
1. BCTw	Xiloteca Calvin Mainieri - Instituto de Pesquisas Tecnológicas, IPT, Divisão de Produtos Florestais		1		1	2
2. CEPANN	Coleção do Laboratório de Abelhas do IB/USP		1			1

Coleção	Instituição	Pós-Doc	TT3	TT2	TT1	Total
3. DZSJRP-Acari	Coleção de Ácaros, UNESP, Campus São José do Rio Preto, Departamento de Zoologia e Botânica		1			1
4. DZSJRP-Amphibia	Coleção de Anfíbios, Unesp, São José do Rio Preto, IBILCE, Departamento de Zoologia e Botânica		1			1
5. ESA	Herbário da Escola Superior de Agricultura Luiz de Queiroz, ESALQ, USP				1	1
6. HRCB	Herbário Rioclarense, Instituto de Biociências, UNESP, Campus de Rio Claro, SP				1	1
7. IAC	Herbário Instituto Agrônomo de Campinas - IAC		1		2	3
8. IAL	Coleção Zoológica de Referência da Seção de Vírus Transmitidos por Artrópodos (IAL), Instituto Adolfo Lutz		2			2
9. IBSP-Herpeto	Coleção Herpetológica "Alphonse Richard Hoge", Instituto Butantan, Laboratório de Herpetologia		1			1
10. MEFEIS	Coleção do Museu de Entomologia da FEIS/Unesp, Departamento de Fitossanidade		2			2
11. MZUSP	Coleção de Peixes do Museu de Zoologia da USP		2			2
12. RPSP	Coleção Camargo, Universidade de São Paulo, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto		1			1
13. SP	Coleção de Fanerógamas do Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo" (SP) Instituto de Botânica, IBt		1			1
14. SPFW	Xiloteca do Instituto de Biociências da Universidade de São Paulo		1		1	2
15. SPSF	Herbário Dom Bento Pickel, Instituto Florestal de São Paulo		1			1
16. UEC	Herbário da Universidade Estadual de Campinas, Instituto de Biologia, Depto. Botânica	1	2	1		4
Total		1	18	1	6	26

Coleções, cuja solicitação foi negada ou o bolsista desistiu:

BOTU	Herbário "Irina Delanova Gemtchújnicov", Depto. Botânica, Instituto de Biociências, UNESP, Campus de Botucatu, SP
DZSJRP-Chiroptera	Coleção de Quirópteros, UNESP, São José do Rio Preto, SP
HISA	Herbário de Ilha Solteira UNESP, Campus Ilha Solteira
IBSBF	Instituto Biológico de Campinas, Coleção de Culturas de Fitobactérias do Instituto Biológico, Campinas
IBSP-Acari	Laboratório de Parasitologia do Instituto Butantan, São Paulo
SP-Algae	Coleção de Algas - Instituto de Botânica
SPF	IB Botânica, Universidade de São Paulo
UBTU	Coleção Científica de Aranhas (Araneae) do Depto. de Zoologia da Unesp, Campus Botucatu, SP

3.7. RELATÓRIOS DE DESEMPENHO

Visando estimular a parceria, foram produzidos pela equipe do CRIA relatórios individuais de desempenho das coleções baseados apenas nos dados disponíveis *on-line*, utilizando as ferramentas disponíveis, principalmente o *monitor* e *data cleaning*. Esses relatórios serviram para identificar problemas e novas demandas por aplicativos e certamente estimularam a participação das coleções mais ativas.

4. FERRAMENTAS DE VISUALIZAÇÃO E CORREÇÃO DE DADOS

Para a 2ª fase do projeto foi proposto o desenvolvimento de ferramentas para auxiliar as coleções biológicas na tarefa de verificação e correção dos dados. No processo de informatização é freqüente a ocorrência de erros relacionados à grafia, digitação ou à falta de atualização dos dados. Os erros são freqüentemente associados aos nomes dos táxons, às informações de localidade e às coordenadas geográficas associadas e podem comprometer os resultados de trabalhos que utilizam estas informações. O desenvolvimento de ferramentas para auxiliar as coleções na revisão dos dados de seus acervos certamente contribui para a melhoria da qualidade dos dados. Além de detectar possíveis erros, as ferramentas fornecem subsídios para a correção ou complementação dos dados.

4.1. SPECIESMAPPER²⁹

O objetivo dessa ferramenta é permitir aos pesquisadores a visualização espacial de suas coletas de forma simples e rápida, sem a necessidade de utilizar softwares complexos ou de custo elevado.

O speciesMapper permite a visualização de pontos geográficos (coordenadas) em um mapa. Para utilizá-lo, basta inserir um ou mais conjuntos de coordenadas geográficas (longitude e latitude) e em seguida visualizar o(s) ponto(s) em um mapa. A ferramenta disponibiliza mapas de diferentes abrangências: mapa Mundi; América do Sul e Central; Brasil; e, o Estado de São Paulo.

4.2. CONVERSOR³⁰

Um erro bastante freqüente no registro da coordenada geográfica é a falta de especificação do DATUM e o erro na conversão dos dados para graus decimais, unidade definida como padrão do DarwinCore. Este foi o principal motivo que levou o CRIA a desenvolver o *conversor*. A ferramenta tem como objetivo auxiliar os pesquisadores na conversão dos diferentes tipos de representação de coordenadas geográficas (graus, minutos e segundos, graus decimais e UTM) e entre os mais comuns DATUM's (WGS84, SAD69 e Córrego Alegre). A conversão das coordenadas é necessária para a uniformização dos dados, para a adequação de um ponto ao datum mais apropriado ou para possibilitar a visualização dos pontos em mapas (geralmente são utilizadas as coordenadas em graus decimais para esta finalidade). Para a conversão é utilizada a biblioteca PROJ.4 - *Cartographic Projections Library*³¹.

4.3. GEOLOC³²

O *geoLoc* é uma ferramenta desenvolvida com o objetivo de auxiliar as coleções biológicas no geo-referenciamento de seus dados. Permite a busca de coordenadas geográficas a partir do nome, completo ou parcial, de municípios e localidades e, além de exibir as

²⁹ <http://smlink.cria.org.br/mapper>

³⁰ <http://smlink.cria.org.br/conversor>

³¹ <http://www.remotesensing.org/proj>

³² <http://smlink.cria.org.br/geoloc>

coordenadas encontradas no banco de dados, permite a visualização dos pontos em um mapa dinâmico do Brasil.

O banco de dados sobre localidades brasileiras tem como fonte o arquivo “cidades e vilas” do IBGE³³, o GEOnet Names Server³⁴ e a própria rede *speciesLink*/Fapesp. As informações de localidades provenientes da rede do *speciesLink* são atualizadas mensalmente no banco de dados do *geoLoc*. O banco de dados desenvolvido em PostgreSQL, contém atualmente mais de 120.000 registros de localidades brasileiras.

Além da busca de coordenadas a partir dos nomes de localidades, o *geoLoc* possui um algoritmo (baseado no Egaz³⁵) capaz de calcular as coordenadas para uma determinada distância e direção de um ponto. Para isso basta informar a distância no campo "distância (km)", a direção (N, S, E, W, NE, NW, SE ou SW) e clicar sobre a localidade de origem. O mapa resultante exibe o conjunto de coordenadas do ponto selecionado e do ponto calculado, informando o erro inerente ao cálculo da distância.

A ferramenta permite também que a entrada dos dados seja feita através do envio de uma planilha eletrônica.

4.4. SPOUTLIER³⁶

Esta ferramenta usa técnicas modificadas por Chapman (1999)³⁷ para detectar pontos fora do padrão esperado para um determinado conjunto, utilizando dados de latitude, longitude e altitude. Para utilizá-la o usuário simplesmente insere o conjunto de coordenadas geográficas (longitude, latitude) e a ferramenta apontará os pontos “suspeitos”, em formato texto e visualizados em um mapa.

O sistema também identifica pontos suspeitos por ocorrerem “na terra” ou “no mar” (organismos estritamente aquáticos ou terrestres). Selecionando a opção “na terra”, por exemplo, a ferramenta destacará os pontos que ocorrem no mar como “suspeitos”, e vice-versa.

4.5. INFOXY³⁸

O *infoXY* permite a identificação de localidades a partir de coordenadas geográficas possibilitando a detecção de erros ou incoerências, além da complementação de dados de coleções. Através das coordenadas geográficas, a ferramenta retorna informações sobre o ponto, como o nome do país, estado ou região administrativa e, no caso do ponto ocorrer no Brasil, o nome no município.

Esta ferramenta se baseia em um mapa mundi com divisões administrativas e o mapa de municípios do Brasil do IBGE. Esses mapas estão armazenados no banco de dados PostGIS que é utilizado para identificar as coordenadas.

4.6. BANCO DE DADOS DE COLETORES³⁹

O banco de dados de coletores de plantas do Brasil foi desenvolvido para estimular a padronização dos dados. Vem sendo alimentado com informações provenientes de contribuições de pesquisadores e da literatura. A inclusão do *link* do banco de coletores de

³³ <http://www.ibge.gov.br>

³⁴ <http://gcmd.nasa.gov/records/GEONET.html>

³⁵ <http://www.bioblink.csiro.au/egaz.html>

³⁶ <http://splink.cria.org.br/outlier>

³⁷ Chapman, A.D. (1999). Quality control and validation of point-sourced environmental resource data pp.409-418 in Lowell, K. and Jatón, A. (eds). Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources. Chelsea, Michigan: Ann Arbor Press. 455pp

³⁸ <http://splink.cria.org.br/infoxy>

³⁹ http://splink.cria.org.br/collectors_db

plantas do Brasil na página da Sociedade de Botânica do Brasil⁴⁰ favoreceu o aumento das contribuições de pesquisadores. Foram incorporadas as informações dos coletores que participaram da obra *Flora brasiliensis*, de Martius, e da lista dos coletores que participaram do projeto “RadamBrasil”. Os coletores encontrados nos bancos de dados das coleções ligadas ao projeto *speciesLink* foram acrescidos ao banco de dados a partir de listas geradas pela ferramenta de busca “Inventário simples⁴¹”. O banco de dados possui 5.145 registros para abreviaturas de nomes na planilha dos dados, sendo que 2.138 contém os nomes completos dos coletores e outras informações como a Instituição em que o coletor atua, sua especialidade, data de nascimento e período em que mais coletou. Estão disponibilizadas cerca de 50 fotos e a intenção é que este trabalho seja contínuo. Este banco de dados está integrado à ferramenta de *data cleaning* descrita a seguir.

4.7. DATA CLEANING⁴²

O sistema de *data cleaning* reúne os conceitos de várias das ferramentas citadas anteriormente com alguns novos aplicativos para auxiliar as coleções no processo de verificação e correção de erros, na complementação de dados e na análise do processo de informatização. O sistema indica quais os registros suspeitos, cabendo ao provedor do dado corrigir eventuais erros. É importante enfatizar que o sistema não altera os dados da coleção. Cabe ao curador decidir se os dados suspeitos são de fato erros e se devem ser corrigidos. Acessando a página da ferramenta, o usuário visualiza todos os pontos georreferenciados do sistema.



Figura 14. Distribuição geográfica das coletas de todas as coleções nacionais (cor azul) e internacionais (cor laranja)

⁴⁰ <http://www.botanica.org.br/>

⁴¹ http://splink.cria.org.br/simple_inventory

⁴² <http://splink.cria.org.br/dc/>

Selecionando um acrônimo, é apresentada uma página padrão para todas as coleções conforme mostra a figura a seguir.



speciesLink
dados e ferramentas **Data Cleaning**

Essa ferramenta tem a finalidade de auxiliar a detecção de possíveis erros e facilitar a padronização dos dados. Geralmente são erros pequenos e simples, onde uma letra é adicionada, trocada ou omitida. São erros difíceis de serem encontrados em um grande conjunto de dados pela pequena diferença da ortografia. Uma premissa do trabalho é não interferir na informação. Nenhum dado é modificado, o sistema apenas indica os registros "suspeitos", recomendando ao autor que confira e escolha entre corrigir ou não a informação. A ferramenta ainda está em desenvolvimento e todas as sugestões serão bem vindas.

Selecione uma Coleção:

Coleção: UEC	
total de registros indexados	35699
- sem coordenadas geográficas	25819
- georeferenciados	9880
- no mar	743
registros repetidos	
- número do catálogo	9622
- registros inteiros	5920
- nome + número do coletor Novo	6383
última atualização	
- da coleção	17-06-2005
- do data cleaning	22-09-2005

Distribuição geográfica das coletas da coleção



[ver perfil da coleção](#)
[ver estatísticas do data cleaning](#)

dados taxonômicos	
inventário	nome científico - coletor - tipos
família	855 suspeitos
gênero	1063 suspeitos
espécie	889 suspeitos
subespécie	não encontrado
autor	3027 suspeitos
duplicata Novo	73 suspeitos

dados geográficos	
inventário	país - estado - município
nome do país/estado	510 suspeitos
outlier	16 suspeitos
long/lat fora do limite mundial	1 suspeitos
long/lat iguais	2 suspeitos
long ou lat com zero	41 suspeitos
long/lat mar (Brasil)	134 suspeitos
nome do município (Brasil)	6742 suspeitos
Sugestões para campos em branco:	
-long/lat (Brasil)	20148 sugestões 
-nome país/estado	91 sugestões
-nome município (Brasil)	426 sugestões

Figura 15. Página padrão da ferramenta *data cleaning*.

4.7.1 COLEÇÃO

O item "coleção" apresenta as informações atualizadas sobre os registros indexados, o número de registros com e sem coordenadas geográficas, o número de registros com coordenadas geográficas no mar e a data da última atualização realizada pelo provedor. Aponta os registros repetidos considerando o número de catálogo, os registros inteiros e aqueles com o mesmo coletor e número de coleta. Esta ferramenta auxilia o provedor dos dados a localizar problemas no banco de dados ou na coleção.

4.7.2 DADOS TAXONÔMICOS

Fornecer inventários sobre nomes de espécies e de coletores, que podem auxiliar na criação de dicionários úteis para cada coleção e apresenta os nomes de táxons e autores com possíveis erros de grafia. Além disso, aponta registros de duplicatas de material que aparecem com nomes distintos entre as coleções da rede *speciesLink*.

a. ERROS DE GRAFIA

O sistema realiza uma checagem dos campos de família, gênero, espécie e autor, comparando-os e fazendo algumas suposições. Se dois registros têm a mesma família, gênero e espécie, supõe-se que o autor seja o mesmo. Quando isso não ocorre o sistema apresenta os registros como sendo "suspeitos". Para registros com o mesmo nome para famílias e espécies, o sistema supõe que o gênero deve ser o mesmo. Quando não é, são apresentadas as variações fonéticas encontradas e os registros são apresentados como "suspeitos".

Para cada registro suspeito é indicado o número de ocorrências daquele conjunto de dados na coleção e em toda a rede *speciesLink*. O sistema também indica se o nome consta nos dicionários disponíveis no CRIA, como o Catálogo da Vida do Species 2000⁴³, apontando, nesse caso, o nome da família em cor verde. Outras listas de referência poderiam ser utilizadas desde que disponibilizadas eletronicamente.

Exemplos de nomes suspeitos de famílias:

family	genus	ocor_col	ocor_total
[Chroococcaceae]	[Aphanothece]	1	1
[Chroococcaceae]	[Aphanothece]	3	3
[Chroococcaceae]	[Gloeothece]	2	2
[Chroococcaceae]	[Gloeothece]	5	5

Esse exemplo mostra uma variação de grafia para a família "Chroococcaceae". Indica que o nome (família e gênero) *Chroococcaceae Aphanothece* ocorre uma vez no acervo analisado e 1 vez na rede *speciesLink*. Indica também que o nome *Chroococcaceae Aphanothece* ocorre 3 vezes no acervo analisado e em toda a rede *speciesLink* e que consta nos dicionários (palavra em cor verde). O sistema não altera os dados, procura apenas dar elementos para que o responsável pela informação possa decidir se o registro "suspeito" está realmente errado ou não. Se o nome correto for *Chroococcaceae Aphanothece*, o curador pode *clique* na ocorrência do nome *Chroococcaceae Aphanothece* para identificar o número do registro na coleção e alterar o seu banco de dados.

catalognumber	family	genus	species	collector	country	stateprovince	county	locality
238946	Chroococcaceae	Aphanothece	castagnei	L.H.Z. Branco	Brasil	São Paulo	Cananéia	Ilha do Cardoso, em mangue a jusante do Rio Pereque

Nota: algumas colunas foram excluídas do registro por uma questão de formatação

Neste caso, basta ao curador acessar o registro 238946 de seu banco de dados e alterar o nome da família de *Chroococcaceae* para *Chroococcaceae*.

O mesmo conceito é aplicado para gênero.

family	genus	species	subspecies	ocor_col	ocor_total	status_sp2000
[Dictyotaceae]	<i>SP</i> [Lobophora]	[variegata]	[]	70	111	accepted name
[Dictyotaceae]	<i>SP</i> [Lobofora]	[variegata]	[]	8	8	
[Dictyotaceae]	<i>SP</i> [Stypopodium]	[zonale]	[]	22	22	accepted name
[Dictyotaceae]	<i>SP</i> [Stipopodium]	[zonale]	[]	5	5	

⁴³ <http://www.sp2000.org>

O sistema indica que a grafia correta é *Lobophora variegata* e não “Lobofora” uma vez que o primeiro nome consta no dicionário (palavra em cor verde), ocorre 111 vezes na rede speciesLink e é um nome aceito no Catálogo da Vida do Species 2000. A grafia “Lobofora” só ocorre 8 vezes na rede e todas as ocorrências são da coleção que está sendo analisada. Ao “cliquear” no texto da coluna “status_sp2000” o pesquisador tem a descrição completa do nome.

O mesmo conceito utilizado para o gênero se aplica a nomes suspeitos de espécies:

genus	species	subspecies	ocor_col	ocor_total	status_sp2000
sp [Brosimum]	[guianense]	[]	1	25	accepted name
sp [Brosimum]	[guyanense]	[]	1	1	
sp [Calliandra]	[selloi]	[]	7	12	unambiguous synonym
sp [Calliandra]	[selloii]	[]	2	2	

E subespécies:

genus	species	subspecies	ocor_col	ocor_total	status_sp2000
sp [Heliconius]	[erato]	[philis]	1	1	
sp [Heliconius]	[erato]	[phyllis]	28	28	
sp [Eurema]	[nise]	[tenela]	2	2	
sp [Eurema]	[nise]	[tenella]	15	15	
sp [Danaus]	[plexippus]	[erippus]	20	20	
sp [Danaus]	[plexippus]	[eripus]	1	1	
sp [Danaus]	[plexipus]	[erippus]	1	1	

Clicando no símbolo **sp** o sistema passa a buscar outras informações sobre a espécie nos dicionários e sistemas gerenciados pelo CRIA e sites externos como, por exemplo, SciELO, GenBank, ITIS e Google Images, entre outros.

Essa ferramenta mostra a importância da existência de *checklists* de espécies locais com nomes validados por especialistas. É fundamental que as informações de iniciativas como a Flora Fanerogâmica do Estado de São Paulo e outras listas de nomes validados sejam rapidamente disponibilizadas *on-line*.

b. DUPLICATA

Esta ferramenta procura identificar duplicatas de material com nomes distintos entre as coleções para auxiliar no processo de correção de erros ou na atualização dos dados da coleção. Os registros “suspeitos” são apontados quando as informações dos campos “gênero + espécie + subespécie” são distintas para uma mesma combinação dos campos “coletor + número do coletor”. Essa ferramenta é principalmente útil para herbários.

Todos os nomes são então mostrados junto com o nome do identificador. Isso permite que se adote a identificação de um especialista em situações de dúvida ou nos casos onde não havia uma identificação prévia. Esta ferramenta será bastante útil para auxiliar os curadores dos herbários na atualização dos dados de identificação.

collector	collector number	Collection code	genus	species	subspecies	identified by	ocor col
Aguiar, O.T.	176	ESA	<i>SP</i> Securinega	guaraiuva		Baitello, J. B.	1
Aguiar, O.T.	176	SP	<i>SP</i> Savia	dictyocarpa			1
Cardoso-Leite, E.	245	ESA	<i>SP</i> Vernonia	diffusa		Leitão Filho, H.F.	1
Cardoso-Leite, E.	245	HRCB	<i>SP</i> Vernonia	petiolaris			1
Cardoso-Leite, E.	294	ESA	<i>SP</i> Cordyline	terminalis		Cardoso-Leite, E.; Rodrigues, R.R.	3
Cardoso-Leite, E.	294	UEC	<i>SP</i> Cordyline	spectabilis			1
Cardoso-Leite, E.	298	ESA	<i>SP</i> Rollinia	sylvatica		Cardoso-Leite, E. & Rodrigues, R. R.	2
Cardoso-Leite, E.	298	UEC	<i>SP</i> Rollinia	silvatica			1
Cardoso-Leite, E.	334	ESA	<i>SP</i> Dasyphyllum	tomentosum		Cardoso-Leite, E. & Rodrigues, R. R.	1
Cardoso-Leite, E.	334	HRCB	<i>SP</i> Dasyphyllum	brasiliensis			1
Cardoso-Leite, E.	392	ESA	<i>SP</i> Hyeronima	alchorneoides		Cardoso-Leite, E. & Rodrigues, R. R.	1
Cardoso-Leite, E.	392	UEC	<i>SP</i> Hyeronima	alchornioides			1
Chung, F.	79	ESA	<i>SP</i> Ilex	paraguariensis		Grosso Jr., M.	1
Chung, F.	79	UEC	<i>SP</i> Ilex	dumosa			1
Dedecca, D.	9262	ESA	<i>SP</i> Arrabidaea	triplinervea		Lohmann, L.G. (1998)	1
Dedecca, D.	9262	UEC	<i>SP</i> Arrabidaea	triplinervia			1

Nos últimos meses foi feito um trabalho com os “tipos” das coleções. Foi desenvolvido o inventário de tipos por coleção, mostrando o conteúdo do campo tipo e o total de registros na coleção:

Tipo	ocor_col
[Alótipo]	5
[Hólotipo]	21
[Holótipo]	37
[Metátipo]	470
[Paralectótipo]	3
[Parátipo]	2058
[Parátipo*]	183

Aqui também a coleção pode trabalhar na uniformização da nomenclatura.

4.7.3 DADOS GEOGRÁFICOS

a. REGISTROS SUSPEITOS

O sistema fornece inventários para nomes de países, estados e municípios e apresenta os registros suspeitos considerando a inconsistência entre os dados (país vs. estado);

coordenada vs. município), por constituírem “outliers” ou por remeterem a prováveis erros de coordenadas. O sistema também fornece sugestões para a complementação de campos de dados geográficos.

Para a checagem dos dados geo-referenciados, são utilizadas as técnicas das ferramentas *infoXY*, *spOutlier* e *geoLoc*. Além do uso dessas ferramentas, são também verificados os registros com coordenadas geográficas fora do limite mundial, simplesmente testando se a longitude é menor que -180 graus ou maior que 180 graus e a latitude menor que -90 graus ou maior que 90 graus, registros com longitude e latitude iguais e longitude ou latitude igual a zero.

A tabela a seguir apresenta exemplos de registros fora do limite mundial.

country	stateprovince	county	longitude	latitude	ocor_col
Brasil	São Paulo	Ubatuba	-445.8	-23.21	1
	California	Monterey county	36.25	-121.25	1

A ferramenta *infoXY* é utilizada para identificar o nome do país, estado e, no caso de registros no Brasil, município que corresponde às coordenadas registradas na coleção. Esse resultado é comparado com as informações sobre localidade constantes na coleção. Os registros suspeitos, ou seja, aqueles cujo nome do país, estado ou município são diferentes daqueles indicados pela coleção, são apresentados em uma lista e indicados em mapas conforme mostrado na figura abaixo.

country	stateprovince	longitude	latitude	map	country_sug	state_sug	ocor_col
Brasil	Mato Grosso do Sul	-56.559723	-22.587223		Paraguay	Amambay	7

Nesse exemplo, a coleção indica que a coleta foi realizada em Mato Grosso do Sul no Brasil, porém o ponto indicado (longitude e latitude) está localizado no estado de Amambay, no Paraguai.

Clicando no ícone do mapa, é apresentada a localização do ponto (figura abaixo). Clicando na ocorrência na coleção são apresentados todos os registros (no exemplo são sete) que indicam o estado do Mato Grosso do Sul e Brasil como estado e país onde o ponto está localizado. Cabe ao curador alterar ou não o dado.



Figura 16. Localização do ponto “suspeito”, mostrando que ele não ocorre no Brasil

A mesma técnica utilizada para identificar os países suspeitos é utilizada para identificar os municípios suspeitos, sendo que o CRIA só dispõe da base de municípios para o Brasil. Assim, somente registros brasileiros são considerados nesta opção.

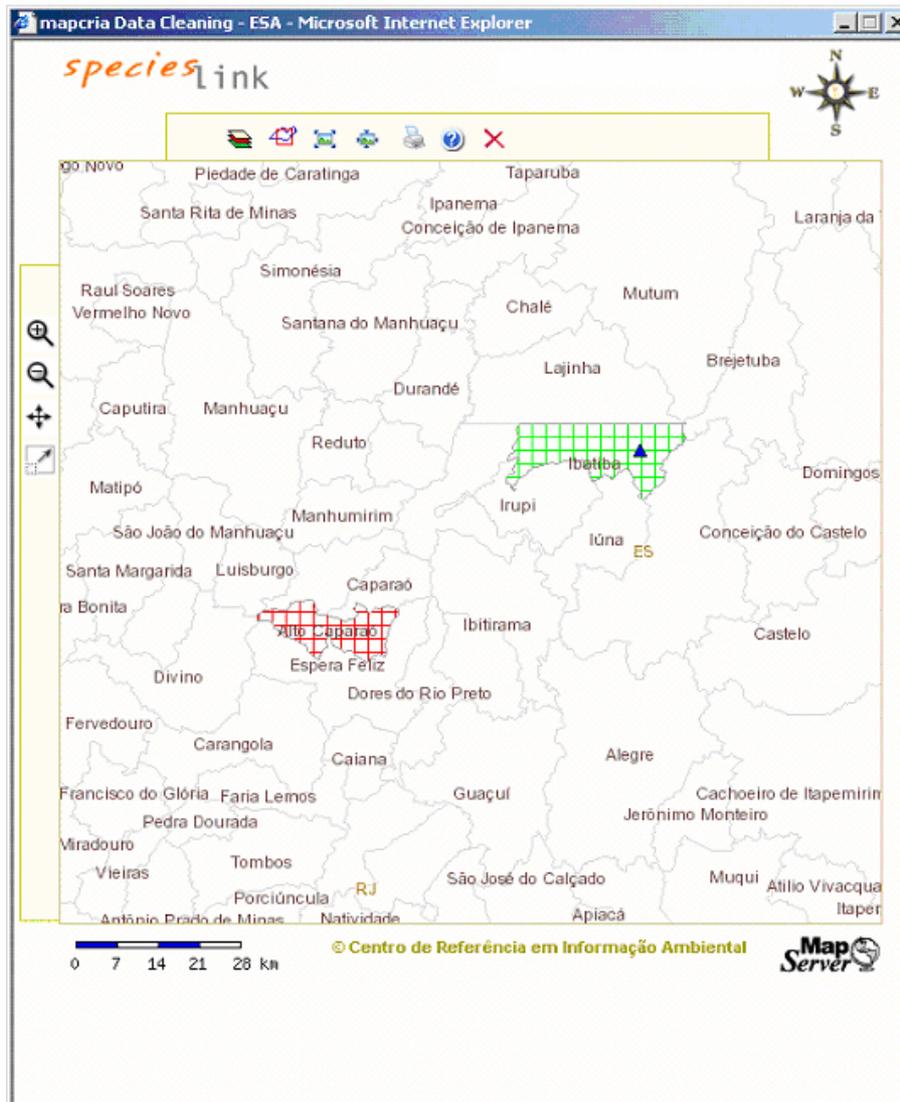


Figura 17. Localização do ponto “suspeito”, mostrando que ele não ocorre no Município de Alto Caparaó, mas em Itatiba

O sistema também verifica se a coordenada do registro está no mar territorial brasileiro (faixa de 200 milhas marítimas) quando o campo país é igual a “Brasil”. A figura a seguir é uma imagem do mapa gerado com os pontos suspeitos, ou seja, com os pontos que estão fora do limite das 200 milhas marítimas, seguida pela tabela com os dados suspeitos.

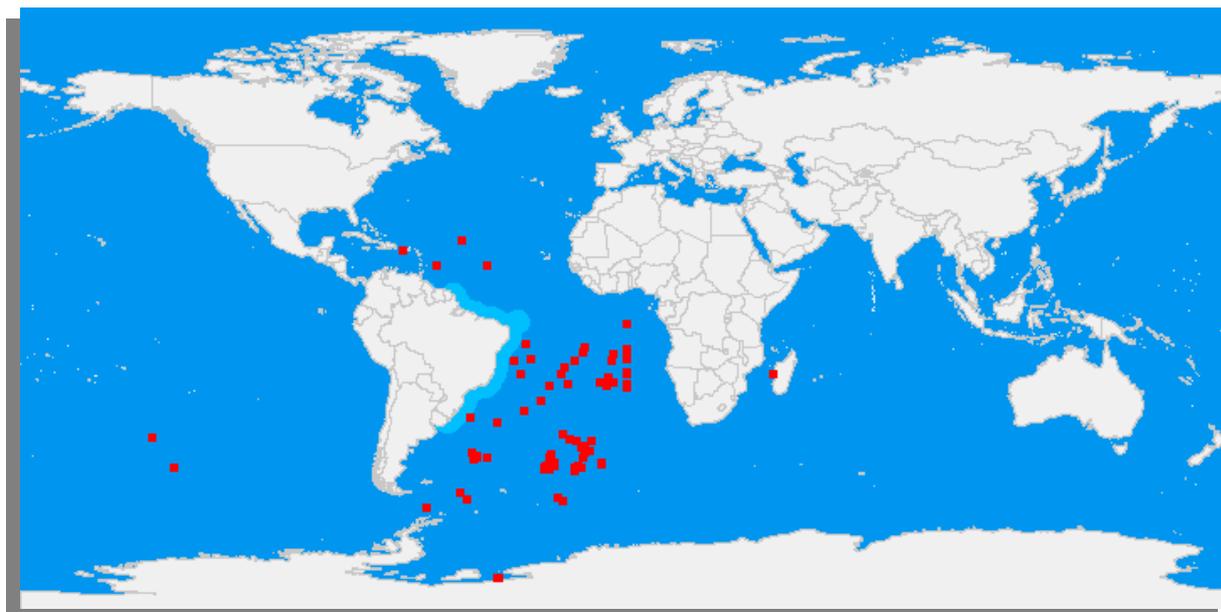


Figura 18. Registros fora do mar territorial brasileiro

id	longitude	latitude	stateprovince	county	ocor_col
3640761	0	-22.3	Rio de Janeiro	Nova Friburgo	1
3639578	0	-19.7	Minas Gerais	Santana do Riacho	2
3643110	0	-19.56	São Paulo	Paulo de Faria	1
3639177	0	-19.2	Minas Gerais	Jaboticatubas	2
3639579	0	-19.19	Minas Gerais	Santana do Riacho	1
3639580	0	-19.1	Minas Gerais	Santana do Riacho	3
3637858	0	-15	Mato Grosso	Barra do Garças	1
3635891	0	-13.48	Bahia	Ituaçu	1
3636194	0	-13.36	Bahia	Rio das Contas	1
....					

Para a detecção dos *outliers* a ferramenta analisa todos os pontos geo-referenciados na coleção e utiliza técnicas estatísticas para identificar aqueles que estão fora do padrão esperado. A figura abaixo mostra o mapa produzido automaticamente para os registros detectados como *outlier*.

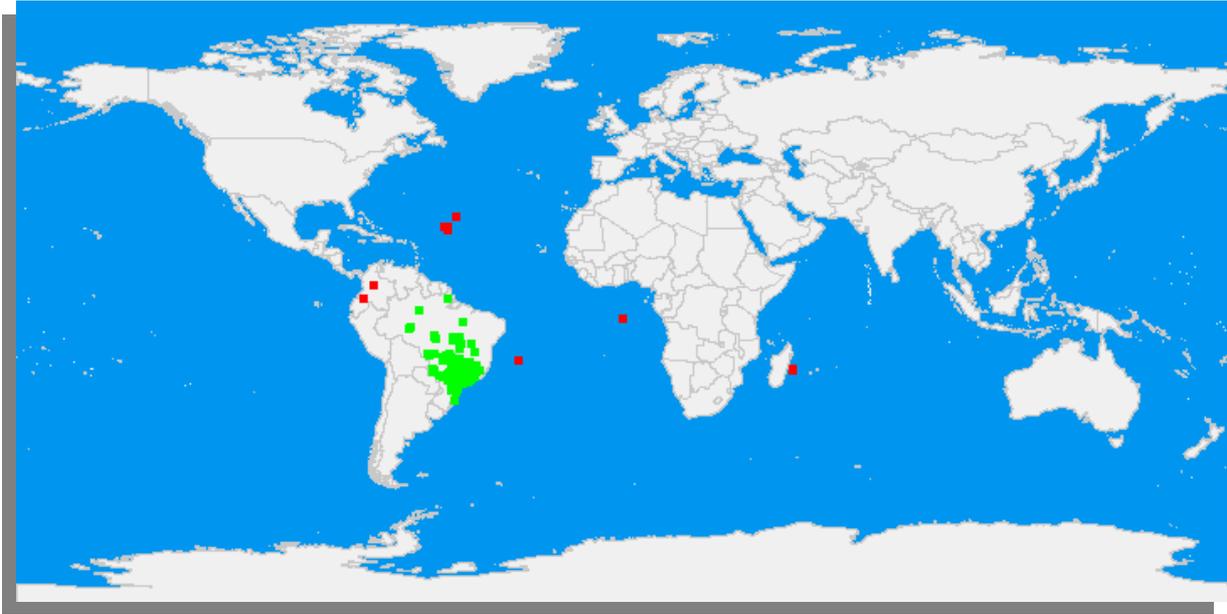


Figura 19. Os pontos em vermelho são os pontos detectados como sendo *outliers*

b. GEO-REFERENCIAMENTO

O sistema sugere valores de longitude e latitude para registros sem coordenadas geográficas baseados no nome do município indicado no registro da coleção. A base de dados de localidades brasileiras do IBGE é a referência utilizada. Como essas coordenadas não são precisas, recomenda-se indicar, no registro de dados, a precisão da coordenada como sendo “município” para que o usuário possa optar pelo uso ou não dessa informação. Para muitos usos essa precisão pode ser adequada. É importante ressaltar que para algumas coleções essa ferramenta chegou a sugerir coordenadas geográficas para mais de 80% dos registros sem coordenadas.

A figura abaixo mostra o mapa produzido automaticamente com os 20.360 registros sem coordenadas, mas que tem o registro do nome dos municípios onde as coletas foram realizadas.

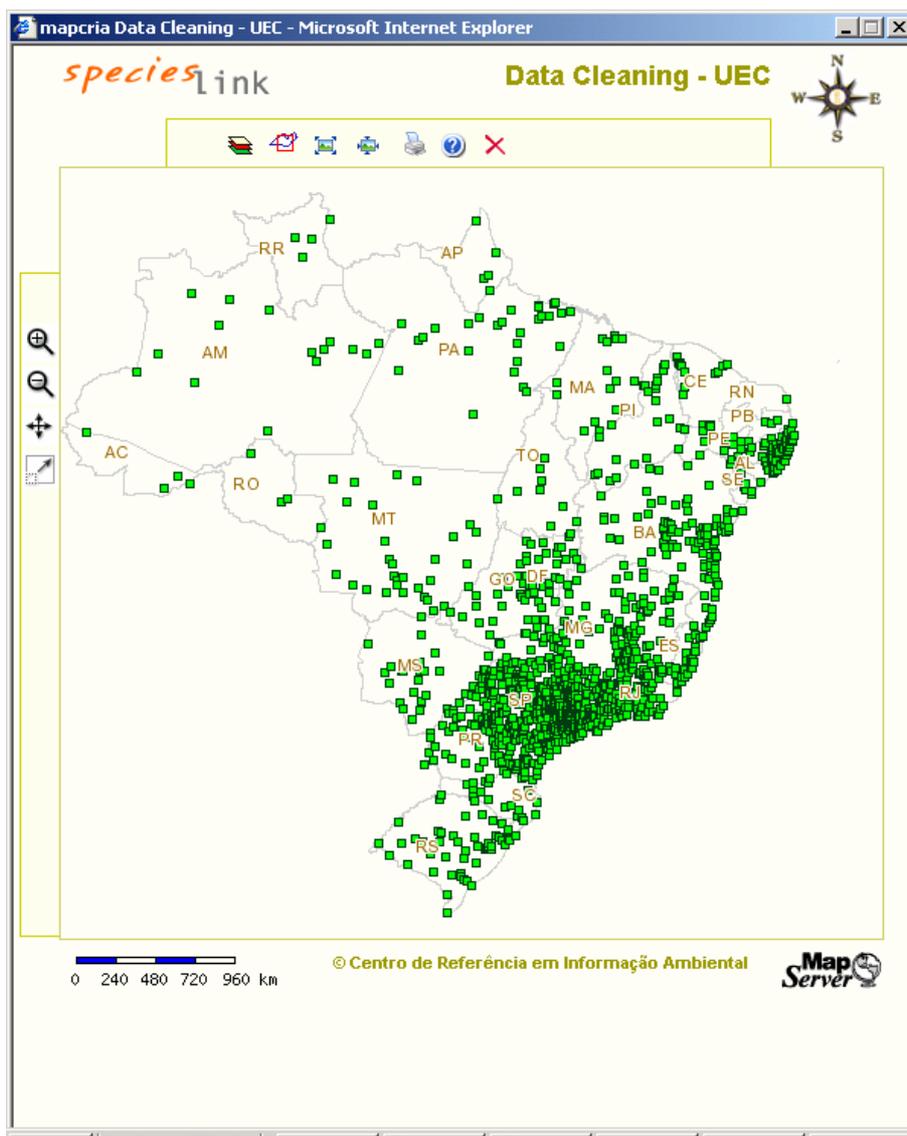


Figura 20. Geo-referenciamento automático de registros sem coordenadas geográficas mas com dados sobre o município da coleta.

Associado ao mapa, o sistema apresenta uma tabela com todos os pontos georeferenciados por município.

Tabela 5. Tabela com os valores de longitude e latitude dos dados que indicam os nomes dos municípios onde as coletas foram realizadas.



country	stateprovince	county	locality	longitude_sug	latitude_sug	datum	map	ocor_col
Brasil		Alpinópolis		-46.3881	-20.8636	SAD69		1
Brasil		Caratinga		-42.1392	-19.7897	SAD69		1
Brasil		Guaraçai	Reserva legal. Próximo a Fazenda Aroeiral.	-51.2	-21.02	SAD69		1
Brasil		Jaboticatubas	Ao longo da rodovia Lagoa Santa-Conceição do Mato Dentro-Diamantina.	-43.745	-19.5136	SAD69		1
Brasil		Januária		-44.3617	-15.4881	SAD69		1
Brasil		Piracicaba		-47.64	-22.72	SAD69		4
Brasil		Porto Firme		-43.0844	-20.6733	SAD69		1
Brasil		Quatro Barras		-49.0769	-25.3656	SAD69		1
Brasil		Rio Formoso		-35.1586	-8.66361	SAD69		1
Brasil		Santana Do Riacho	Pequena Mata seca proximo de afloramento de calcáreo	-43.7144	-19.1689	SAD69		1

Além do georeferenciamento automático, a ferramenta *data cleaning* também oferece nomes para o preenchimento de municípios, estados e países quando esses dados estão em branco. Nesse caso, o sistema utiliza as coordenadas geográficas indicadas, identifica os registros onde os campos da localidade não foram preenchidos (município, estado, país) e sugere os nomes em formato de tabela.

4.7.4 MONITORAMENTO

O sistema monitora o uso da ferramenta *data cleaning* por parte das coleções. A figura a seguir mostra a o uso da ferramenta, utilizando como exemplo o Instituto Agrônomo de Campinas - IAC.



Figura 21. Estatística de Correção de Registros suspeitos pelo IAC

É evidente que esses gráficos precisam ser avaliados caso a caso, uma vez que nem todo registro “suspeito” é de fato um registro com erro. De qualquer maneira, os gráficos indicam que a relação IAC – *speciesLink* está ativa.

4.7.5 PERFIL DO ACERVO

Na reunião com os curadores das coleções foi solicitado ao CRIA o desenvolvimento de uma ferramenta capaz de apresentar o perfil do acervo da coleção. Estamos apenas iniciando este processo e estas informações provavelmente irão fazer parte da página de metadados de cada coleção.

A ferramenta desenvolvida apresenta o mapa mundi com a distribuição geográfica dos registros geo-referenciados, gráficos de representatividade do acervo com respeito às famílias, estados brasileiros, países, coletores (os 10 mais citados), número de registros por ano de coleta e a lista dos tipos da coleção.

Usando como exemplo a Coleção de Peixes do Museu de Zoologia da USP - MZUSP, o sistema produz os seguintes gráficos de forma dinâmica (figuras 22 a 26).

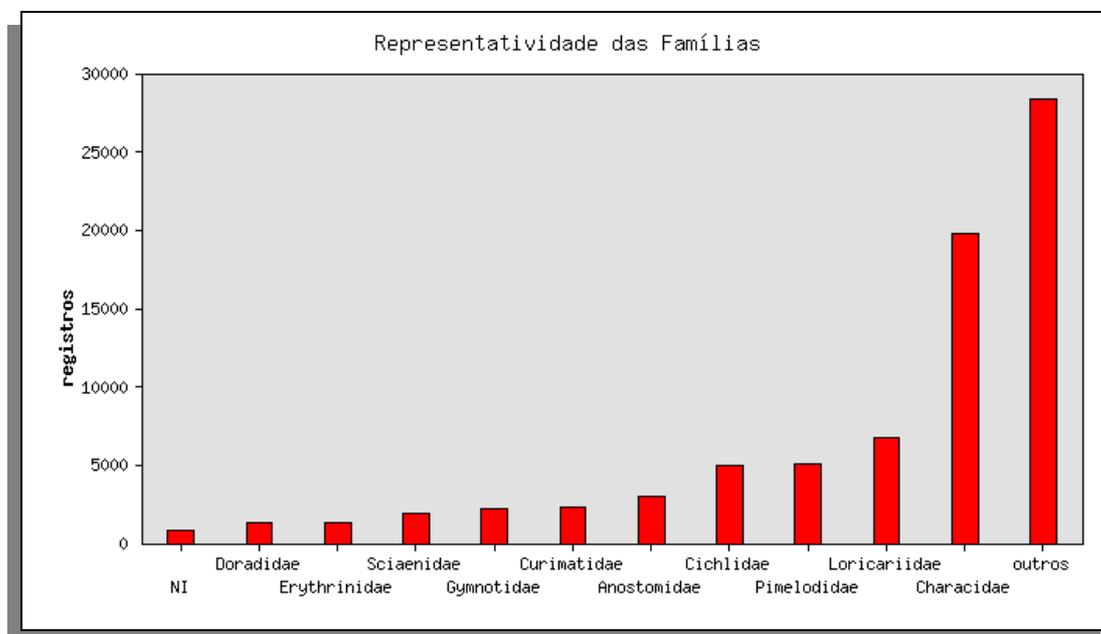


Figura 22. As 10 famílias mais representativas da coleção

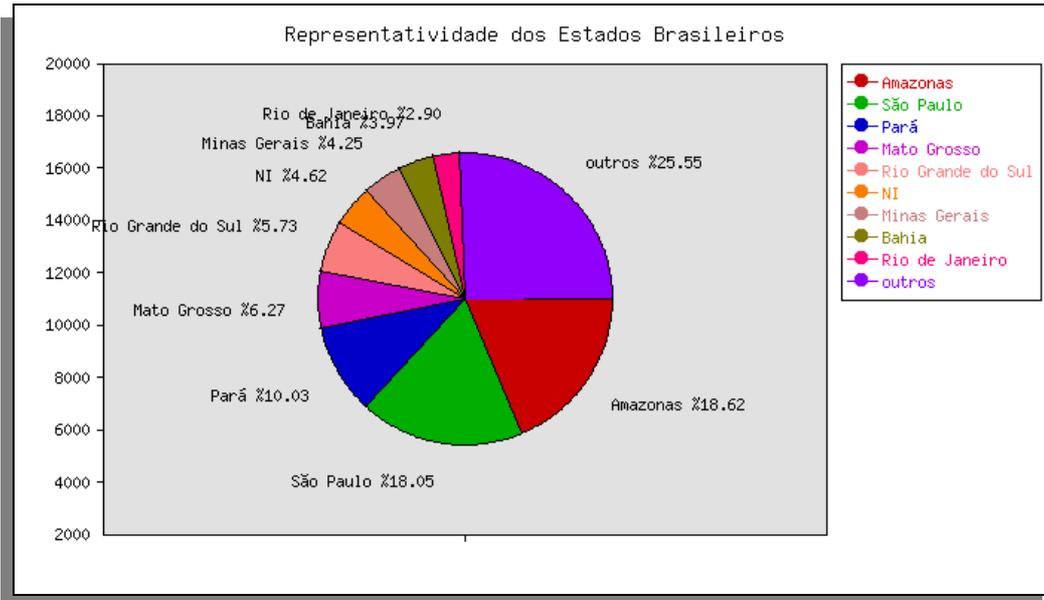


Figura 23. Representatividade dos Estados Brasileiros

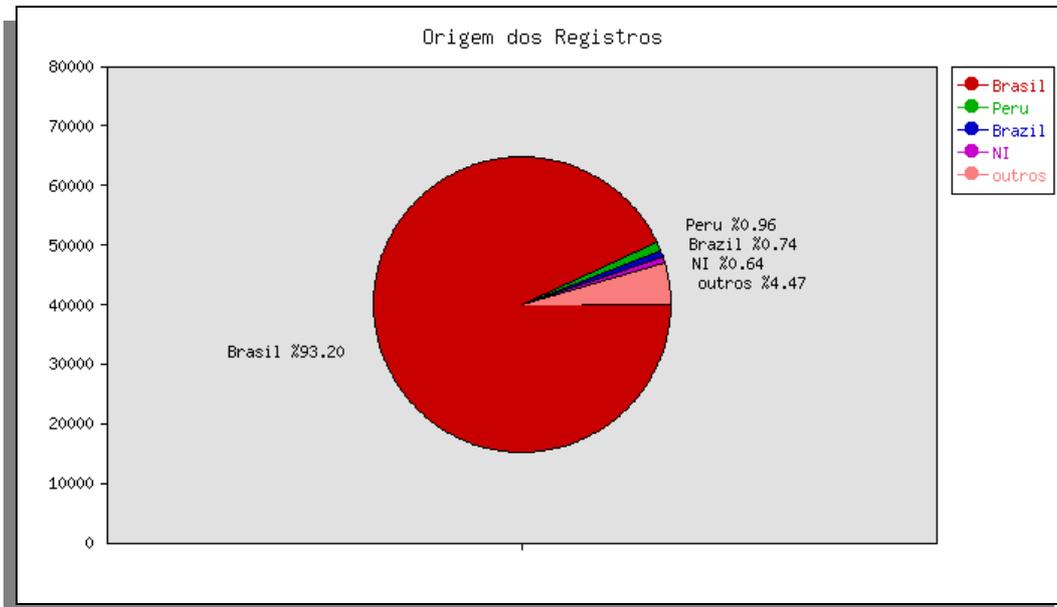


Figura 24. Países de origem dos registros

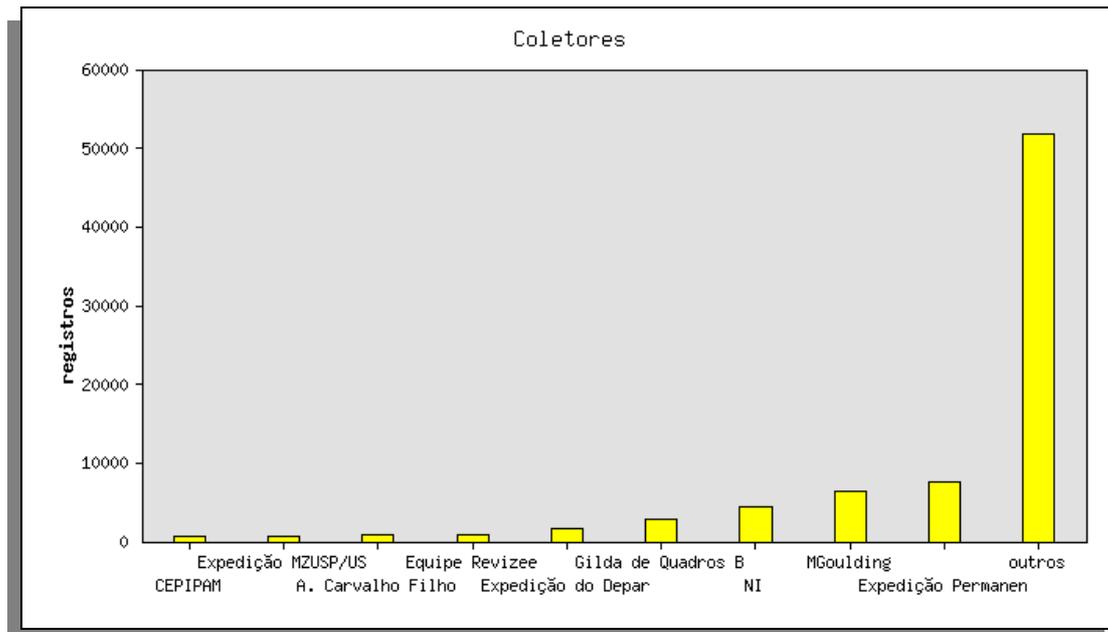


Figura 25. Coletores mais citados

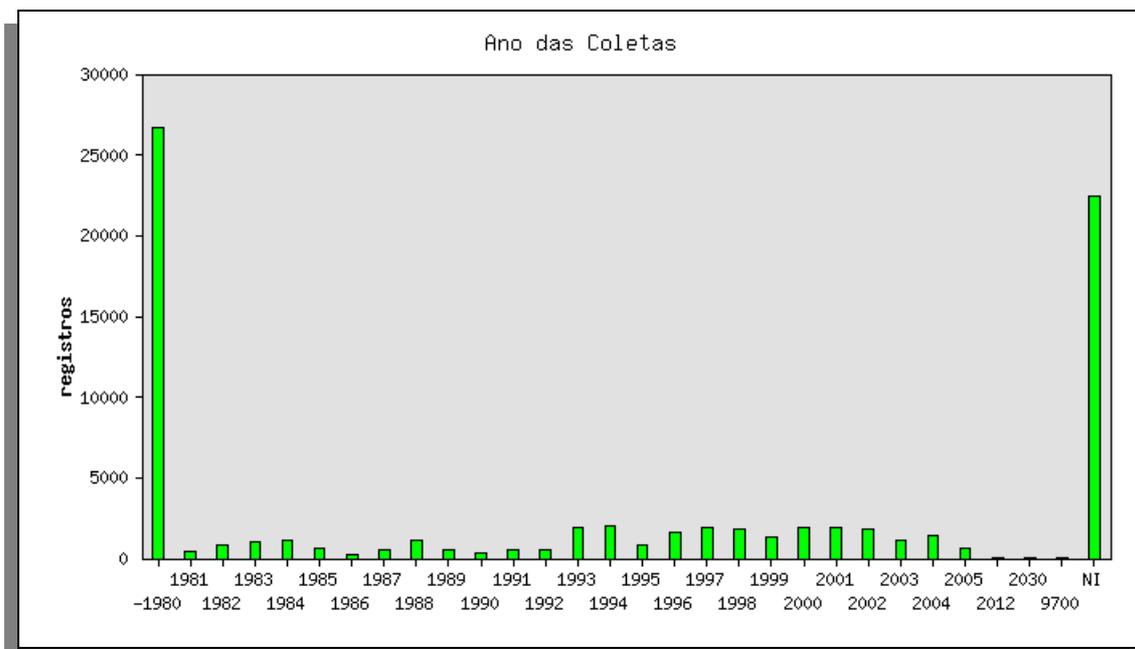


Figura 26. Ano das coletas

Além dos gráficos serem interessantes para avaliar o perfil do acervo, também podem auxiliar na padronização dos dados e na identificação de possíveis erros, como mostrado nesse último gráfico dos anos das coletas, onde podemos observar registros com anos 2012, 2030 e 9700.

A página de perfil da coleção traz ainda a lista de tipos por gênero, espécie e subespécie com o total de registros na coleção.

Total de registros com o campo "Tipo" preenchido: 1894

genus	species	subspecies	typestatus	ocor_col
SP Acestrocephalus	bohlkei		Parátipo	1
SP Acestrorhynchus	altus		Parátipos	1
SP Acestrorhynchus	apurensis		Holótipo	1
SP Acestrorhynchus	britskii		Holótipo	1
SP Acestrorhynchus	falcatus varius		Holótipo	1
SP Acestrorhynchus	falcatus varius		Parátipos	1
SP Acestrorhynchus	grandoculis		Parátipos	7
SP Acestrorhynchus	isalinae		Parátipos	1
SP Acestrorhynchus	maculipinna		Holótipo	1
SP Acestrorhynchus	minimus		Holótipo	1
SP Acestrorhynchus	pantaneiro		Holótipo	1
SP Acnodon	senai		Parátipos	1
SP Adenops	dissimilis		Paratipo	2
SP Adontosternarchus	clarkae		Parátipos	1
SP Aequidens	epae		Holótipo	1
...				

Figura 27. Relação de "tipos" da coleção

A figura a seguir mostra o esquema da ferramenta *data cleaning*. Toda noite uma rotina automática verifica se alguma coleção fez alguma atualização, e, em caso afirmativo, copia todos os dados necessários para um banco de dados local (PostgreSQL) utilizando o protocolo DiGIR, iniciando o processo de *data cleaning* e a análise do perfil. Essa rotina permite que uma coleção visualize possíveis registros suspeitos na manhã seguinte da atualização.

Essas ferramentas têm sido testadas pelos curadores de herbários e museus da rede *speciesLink* e estão em constante desenvolvimento.

Data Cleaning - speciesLink out/2004

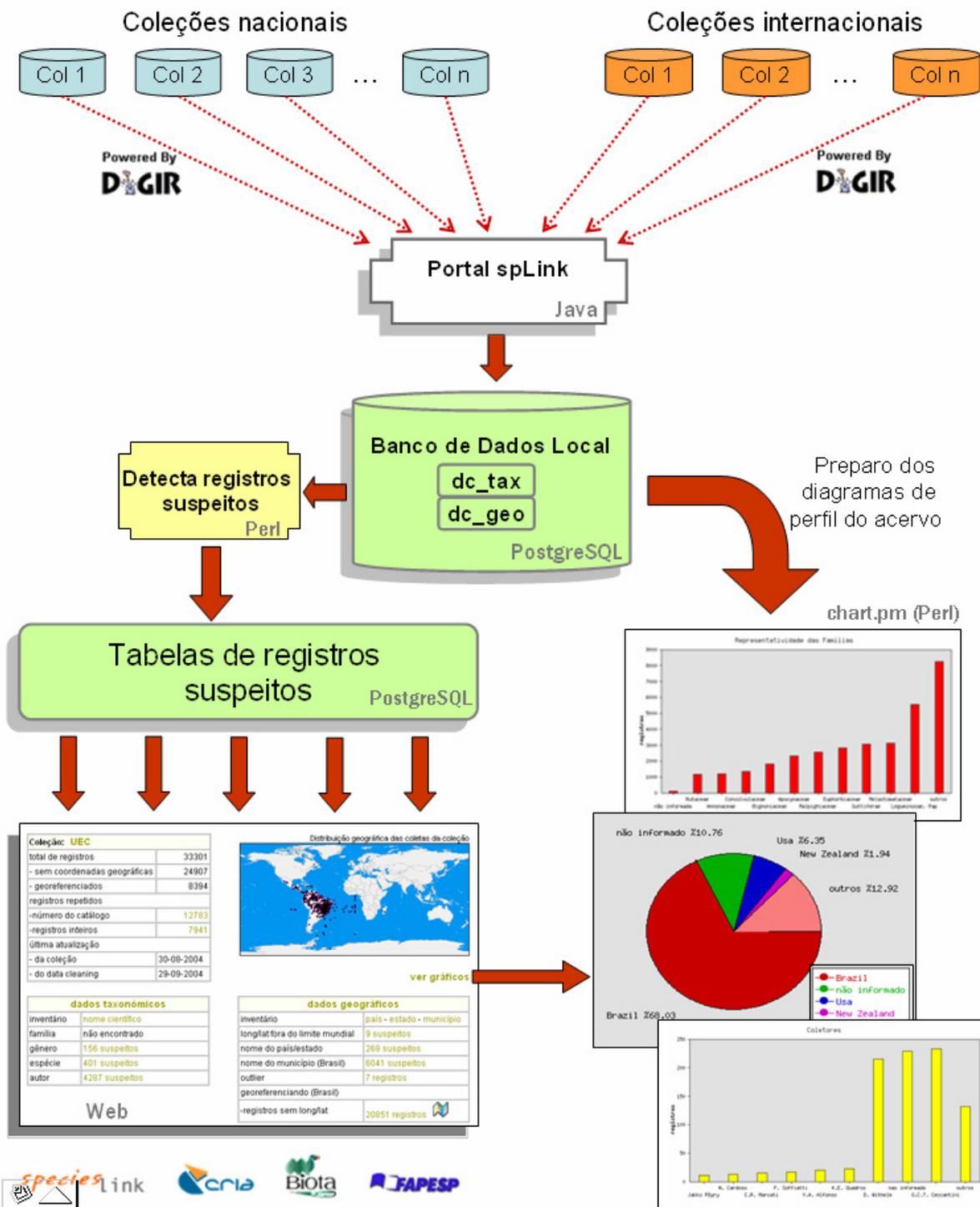


Figura 28. Esquema do processo de *data cleaning* e de visualização do perfil do acervo na rede speciesLink

Todas as ferramentas listadas acima, com exceção do “Conversor” dispõem de um visualizador de mapas dinâmico denominado *Mapcria web service* descrito no próximo item.

5. APLICATIVO PARA A PRODUÇÃO DINÂMICA DE MAPAS NA INTERNET

Desde o início do desenvolvimento do Sistema de Informação do Programa Biota/Fapesp, o *SinBiota*, foi detectada a necessidade de um aplicativo para a produção dinâmica de mapas na internet. A primeira versão do *SinBiota* utilizava como ferramenta de gerenciamento e produção dos mapas, o ArcInfo versão Unix, software proprietário de alto custo. Esta solução não era viável pela complexidade e alto custo anual das licenças. Uma solução alternativa foi implementada na época utilizando um outro software proprietário, porém de menor custo, o MapObjects (ambiente Microsoft Windows). Essa solução, apesar de muito mais barata que a anterior, apresentava problemas de performance e grande dificuldade de manutenção dos equipamentos onde as instâncias do MapObjects eram executadas. Assim, buscou-se uma solução que pudesse ser implementada inteiramente em ambiente Unix, que fosse de código aberto, seguindo a filosofia do CRIA, e que definitivamente resolvesse o problema de servir mapas para os diversos sistemas que estavam sendo implementados.

A solução encontrada foi o MapServer, um pacote de código aberto, originalmente desenvolvido pelo projeto ForNet⁴⁴, da Universidade de Minnesota (UMN) em cooperação com a NASA e o Departamento de Recursos Naturais (MNDNR) dessa mesma Universidade. Essa iniciativa atualmente é mantida pelo projeto TerraSIP⁴⁵, financiado pela NASA e a UMN.

O MapServer foi escolhido por ser de código aberto, ter desenvolvimento colaborativo, ser multi-plataforma e, pelo fato de também disponibilizar uma biblioteca que poderia ser utilizada como base para o desenvolvimento de aplicações desenhadas especificamente para as necessidades dos nossos projetos, o MapScript. Assim, ainda no escopo do projeto *SinBiota*, alguns estudos e protótipos foram implementados e testados mostrando bons resultados.

Com o início do projeto *speciesLink*, a necessidade de disponibilizar mapas para a visualização espacial dos dados providos pela rede se tornou ainda mais evidente. Tanto para a simples visualização espacial da abrangência geográfica do acervo das coleções quanto para viabilizar o desenvolvimento de ferramentas que permitissem o georreferenciamento de registros (geoLoc), ou a análise estatística de coordenadas geográficas (spOutlier) ou ainda para as diferentes ferramentas de correção de dados (*data cleaning*), era necessário dispor de um sistema mais genérico e robusto capaz de atender essas diferentes demandas.

O sistema deveria considerar a possibilidade de padronizar o uso da biblioteca para não depender das mudanças introduzidas a cada nova versão da biblioteca MapScript, e a viabilidade de desenvolver um visualizador genérico que pudesse ser utilizado pela maioria dos navegadores utilizados pelos usuários da rede (Microsoft Internet Explorer e navegadores baseados em Mozilla como o Mozilla FireFox).

Estudos concluíram que seria possível implementar um serviço web padronizado que fizesse a interface entre os diferentes aplicativos e a biblioteca MapScript. Definida a interface entre os aplicativos e o serviço, estes não mais precisariam ser alterados cada vez que uma nova versão da biblioteca fosse disponibilizada.

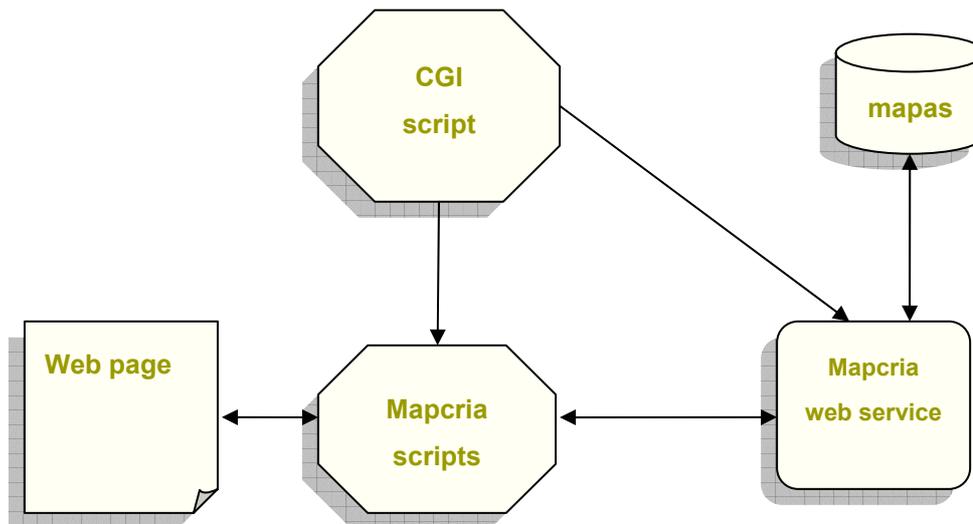
Estudos a respeito do visualizador concluíram que seria possível desenvolver uma aplicação padrão capaz de receber parâmetros de mapas previamente inicializados pelo serviço de mapas que tivesse autonomia para continuar a interação com o usuário provendo funções básicas como *zoom (in/out)*, ligar e desligar camadas (*layers*), etc. Além disso, o aplicativo deveria ser suficientemente leve na rede e ser compatível com os navegadores mais utilizados.

⁴⁴ <http://www.gis.umn.edu/fornet>

⁴⁵ <http://terrasip.gis.umn.edu>

Foram estudadas e testadas várias técnicas, quase todas ainda embrionárias, algumas utilizando Java Applets (como o Rosa Applet). Essas últimas se mostraram razoavelmente eficientes em termos de facilidades disponíveis, porém muito pesadas e, apesar de escritas em Java, apresentavam grandes problemas de compatibilidade com as máquinas Java dos navegadores. Foi decidido então que a interface deveria ser desenvolvida utilizando apenas DHTML e JavaScript, sem a utilização de Java.

Foram então implementados um serviço web, por nós chamado de **mapcria web service**, e um visualizador, batizado de **mapcria viewer**. A figura abaixo mostra a idéia de interação básica entre os módulos e as aplicações que os utilizam.



5.1. MAPCRIA WEB SERVICE

O serviço Mapcria para construção e manipulação de mapas, está na sua versão 3.0 e é baseado na biblioteca MapScript C com interface Perl SWIG para MapServer 4.2, utilizando ainda, SOAP::Lite. A documentação completa sobre o serviço pode ser encontrada em <http://www.cria.org.br/mapcria/doc/>

O serviço está disponível em duas máquinas mantidas no CRIA e várias instâncias do serviço podem estar disponíveis simultaneamente. O serviço pode ser utilizado por aplicações externas, de terceiros, mediante acordo com o CRIA.

A conexão ao serviço em Perl pode ser feita da seguinte maneira:

```

#!/usr/bin/perl
use SOAP::Lite;
my $soap = SOAP::Lite
-> uri("http://mapcria.cria.org.br/manager")
-> proxy("http://mapcria.cria.org.br:59000/",
timeout => 20)
-> on_fault( sub { return undef } );
  
```

As funções principais disponíveis pelo serviço são descritas a seguir.

a. DRAW

Função básica que permite a definição do mapa a ser criado. Essa função utiliza como parâmetro um documento XML que define as características do mapa a ser produzido. A função *draw* retorna um *ticket*, ou seja, um identificador para o mapa pedido, que é utilizado pelas aplicações nas outras funções de modificação do mapa, permitindo assim, a interatividade com o usuário.

O documento XML que deve ser passado como parâmetro para a função está definido pelo XML Schema em <http://www.cria.org.br/schema/mapcria3.0.xsd> cuja documentação completa encontra-se *on-line*⁴⁶.

Apenas para ilustração, apresentamos um exemplo de XML bastante simples:

```
<?xml version="1.0" encoding="UTF-8"?>
<mapcria xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://www.cria.org.br/schema/mapcria3.0.xsd">
  <!-- para ter acesso ao serviço a aplicação precisa ter uma identificação
  previamente acordada junto ao CRIA -->
  <user>USER_IDENTIFICACION</user>
  <!-- início da definição do mapa a ser produzido -->
  <map id="MAPA">
    <!-- definições básicas do mapa como abrangência geográfica, tamanho da imagem a ser
    produzida, transparência, cor do fundo, etc. -->
    <set_defaults>SP</set_defaults>
    <set_extent>-53.11,-19.78,-42.69,-25.30</set_extent>
    <set_extent_text>>false</set_extent_text>
    <set_image_background_color>0,204,255</set_image_background_color>
    <set_image_transparent>>false</set_image_transparent>
    <set_image_width>160</set_image_width>
    <set_image_height>120</set_image_height>
    <!-- definição das camadas (layers) que deverão estar disponíveis no mapa e suas
    aracterísticas como visibilidade inicial, cores de fundo e contorno, etc. -->
    <layer id="BRASIL_UF">
      <set_visibility>>true</set_visibility>
      <set_fill_color>255,255,255</set_fill_color>
    </layer>
    <layer id="BRASIL_MUN">
      <set_visibility>>false</set_visibility>
    </layer>
    <!-- definição de um polígono a ser "pintado" no mapa a partir de uma coordenada
    geográfica -->
    <query_by_coord id="municipio">
      <set_layer>BRASIL_MUN</set_layer>
      <set_outline_color>255,0,0</set_outline_color>
      <set_symbol_color>255,0,0</set_symbol_color>
      <set_symbol>point</set_symbol>
    </query_by_coord>
  </map>
</mapcria>
```

⁴⁶ <http://www.cria.org.br/mapcria/doc/doc/index.html>

```

        <set_symbol_size>8</set_symbol_size>
        <set_coords>-46.65,-23.55</set_coords>
    </query_by_coord>
    <!-- definição de conjunto de pontos a serem plotados e suas características que
    podem
    ser definidas para todos os pontos e modificadas para cada ponto, se necessário.
    -->

    <points id="servers">
        <set_title>Servidor Regional</set_title>
        <set_projection>proj=longlat</set_projection>
        <set_symbol>triangle</set_symbol>
        <set_symbol_size>10</set_symbol_size>
        <set_symbol_color>0,255,0</set_symbol_color>
        <!-- definição das características das etiquetas dos pontos -->
        <label>
            <set_font_color>0,0,255</set_font_color>
            <set_force>>true</set_force>
            <set_position>AUTO</set_position>
            <set_font>Arial</set_font>
            <set_font_size>8</set_font_size>
        </label>
        <!-- definição dos pontos propriamente ditos -->
        <point>
            <set_coords>-47.052164,-22.813446</set_coords>
            <set_symbol_color>0,0,255</set_symbol_color>
            <set_label>spLink</set_label>
        </point>
        <point>
            <set_coords>-46.65,-23.55</set_coords>
            <set_symbol_color>0,153,0</set_symbol_color>
            <set_label>SR São Paulo</set_label>
        </point>
        <point>
            <set_coords>-51.2425,-20.4328</set_coords>
            <set_symbol_color>0,153,0</set_symbol_color>
            <set_label>SR Ilha Solteira</set_label>
        </point>
        <point>
            <set_coords>-47.82,-21.2</set_coords>
            <set_symbol_color>0,153,0</set_symbol_color>
            <set_label>SR Ribeirão Preto</set_label>
        </point>
    </points>
</map>
</mapcria>

```

Uma vez inicializado, o mapa está disponível para ser manipulado pelas outras funções descritas a seguir (apenas as principais):

b. GET_METADATA

Retorna dados sobre os *layers* (camadas) disponíveis.

**c. GET_IMAGE_WIDTH
GET_IMAGE_HEIGHT**

Retornam a largura e altura, respectivamente, em *pixels*, da imagem definida para o mapa a ser produzido.

d. GET_IMAGE

Retorna a imagem produzida pelo serviço de acordo com as definições atuais para o mapa.

e. GET_EXTENT

Retorna a extensão (top, left, right, top) de trabalho do mapa nas coordenadas definidas no momento.

f. GET_SCALEBAR

Retorna a imagem produzida pelo serviço que representa a escala sendo vista.

g. GET_REFERENCE

Retorna o mapa de referência marcando a área sendo visualizada.

**h. ZOOM_POINT
ZOOM_RECTANGLE
ZOOM_COORDS
ZOOM_ALL**

Permite realizar operações de *zoom* no mapa através de um ponto e um fator, de um retângulo, de coordenadas geográficas ou de todo o mapa.

i. FINISH

Encerra o serviço. O mapa não estará mais disponível após a função *finish* ser utilizada.

j. TOGGLE_LAYER

Liga ou desliga uma camada, tornando-a visível ou não.

k. SET_VISIBILITY

Define se um objeto (camada, conjunto de pontos, etc.) deve estar visível ou não.

l. GET_FIELD_VALUE_BY_COORD

Recupera valores associados aos *layers* através da especificação de uma coordenada. Por exemplo, o nome de um município, dada uma coordenada geográfica, em um *layer* dos municípios do estado de São Paulo.

Algumas outras funções auxiliares também estão disponíveis. As funções têm sido implementadas quando de sua necessidade.

5.2. MAPCRIA VIEWER 3.0

O **mapcria viewer** pode ser entendido com um conjunto de aplicações (CGI scripts), escritas em linguagem Perl, capazes de interagir com o serviço mapcria utilizando as várias funções disponibilizadas pelo serviço através do navegador do usuário, permitindo assim a interatividade entre o usuário e o serviço. Com a utilização destas aplicações, não é necessário que o desenvolvedor se preocupe em re-escrever as interações com o servidor já que os viewers já disponibilizam as funções básicas necessárias para essa interação.

Um módulo principal é responsável por criar a página web onde o mapa será apresentado ao usuário, assim como gerar o JavaScript específico para aquele caso e ambiente. É importante ressaltar que todo o código necessário para criar a página apresentada ao usuário é gerado dinamicamente para se adaptar às características tanto do mapa quanto do ambiente utilizado pelo usuário (tipo de navegador, versão, sistema operacional, tamanho de tela, etc.)

A aplicação mãe inicializa o mapa no servidor através de um XML que define suas características e chama o visualizador, passando a ele os parâmetros necessários para que ele assuma o controle da interação. A partir daí, os módulos do **mapcria viewer** são capazes de manter a interação com o usuário, tornando-se assim, genérico.

O módulo principal utiliza vários outros pequenos scripts, disponíveis via HTTP, para interagir com o serviço, realizando tarefas simples como zoom (in/out), ligar e desligar layers, alternar entre diferentes cenários definidos, etc.

Duas versões estão disponíveis e sendo atualmente utilizadas pelas aplicações do projeto *speciesLink* e do *SinBiota*: versões 2.0 e 3.0.

A principal diferença entre as duas versões é que a 2.0 abre apenas uma janela no navegador do usuário e todas as operações e janelas auxiliares são mantidas no mesmo espaço enquanto que a versão 3.0 abre janelas auxiliares (pop-ups) permitindo uma flexibilidade muito maior na visualização das opções e escolha das operação desejadas pelo usuário.

A aplicação mais importante que faz uso da versão 3.0 é o novo Atlas do *SinBiota* para o qual foram escritas rotinas auxiliares ao viewer para permitir buscas específicas nos bancos de dados e criar a interação destes com os mapas.

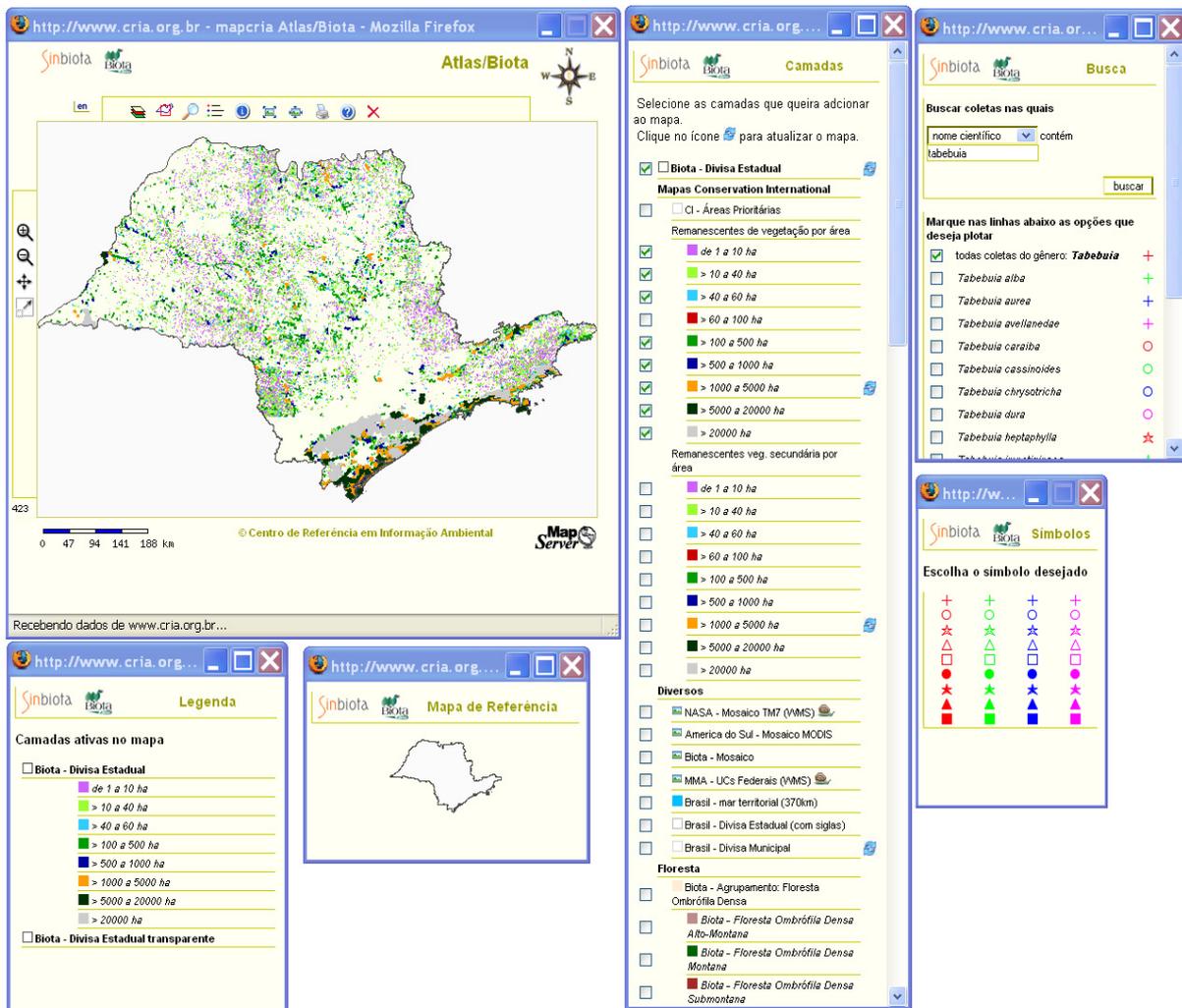


Figura 29. Aplicação do mapcria viewer no novo Atlas do SinBiota

Onde a visualização exige menos interação do usuário, a versão 2.0 ainda é utilizada.

São apresentadas a seguir algumas telas como exemplos de utilização do visualizador pelas diferentes ferramentas implementadas no âmbito do projeto speciesLink.

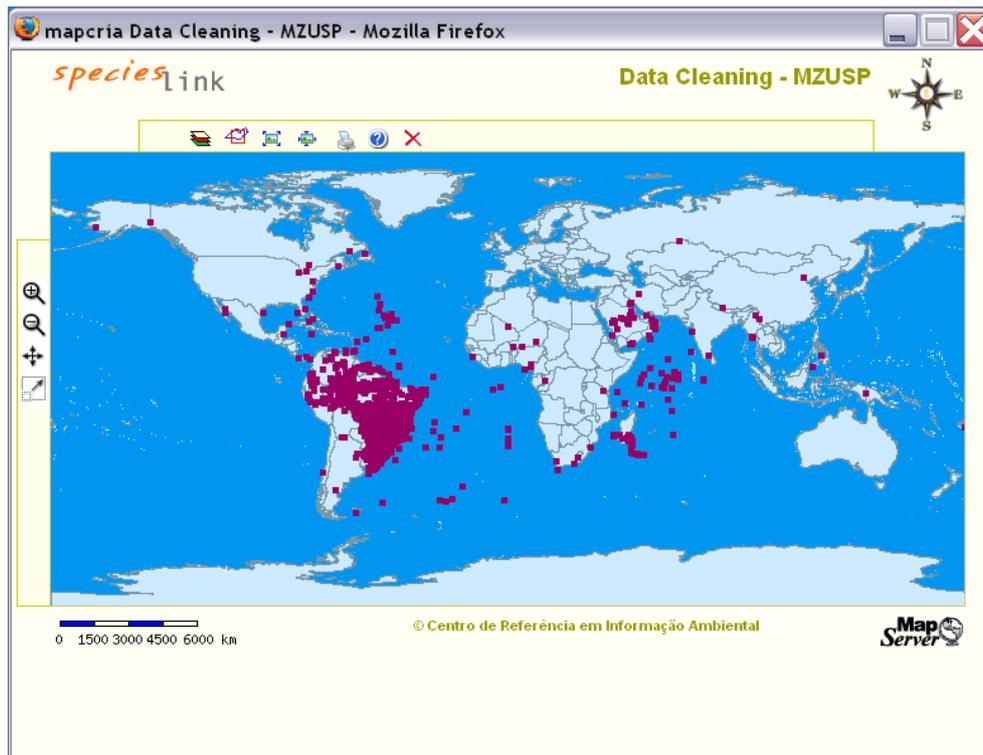


Figura 30. Distribuição geográfica do acervo da coleção de peixes do Museu de Zoologia da USP

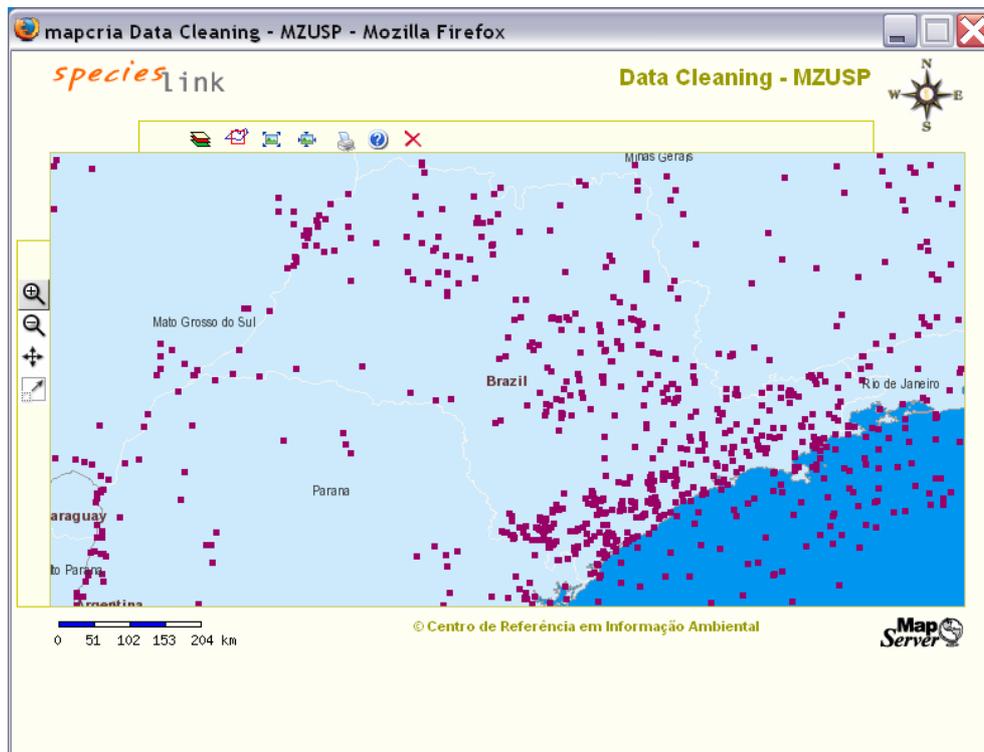


Figura 31. Ferramenta de zoom

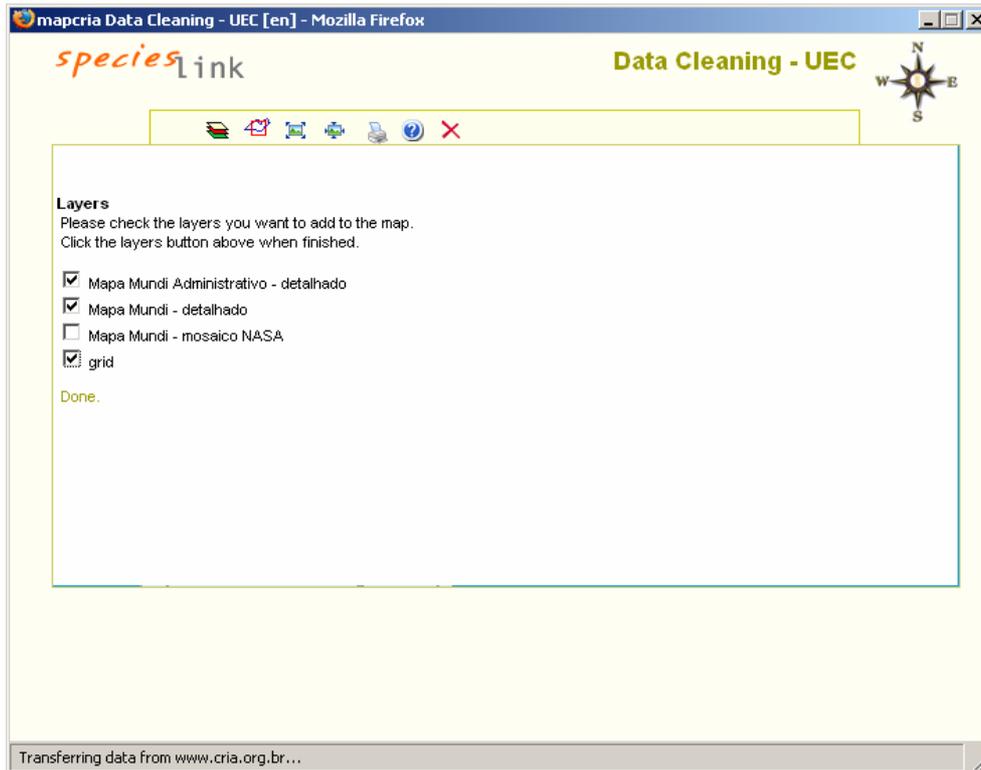


Figura 32. Exemplo de possibilidade de escolha dos layers (viewer 2.0)

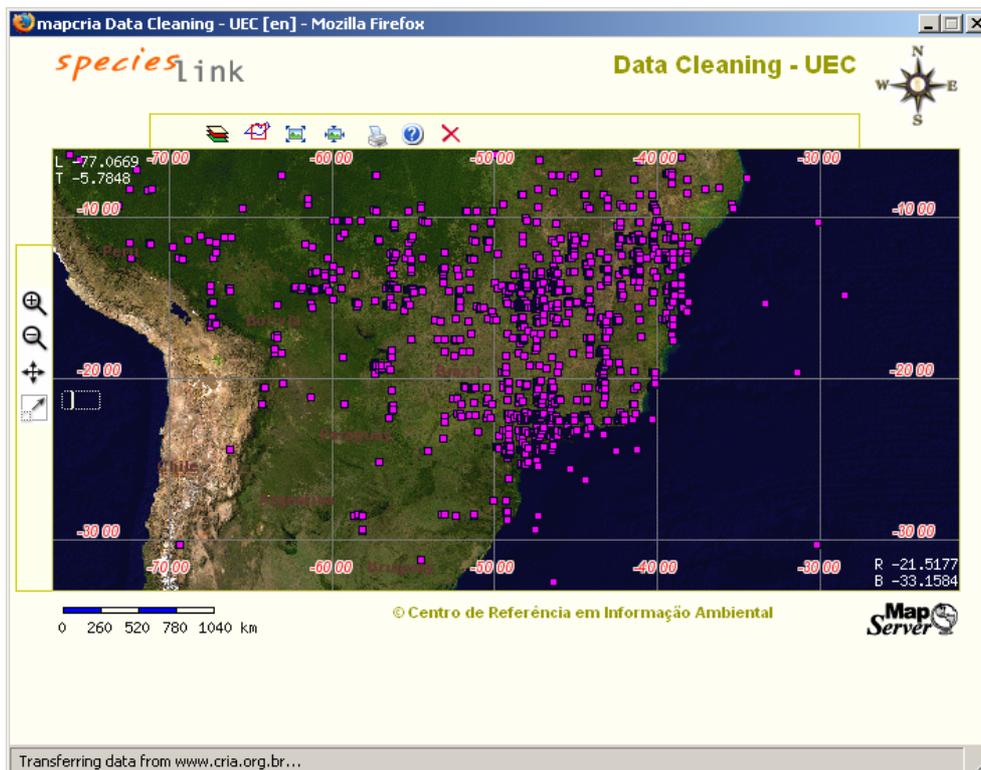


Figura 33. Resultado do zoom e seleção de layers

6. SISTEMA DE MONITORAMENTO DOS INDICADORES DO PROJETO ⁴⁷

Com o crescente número de coleções participantes da Rede *speciesLink*, principalmente a partir do início da segunda fase do projeto, tornou-se imprescindível desenvolver um sistema que permitisse o gerenciamento do *status* de cada uma das coleções. Foram implementados diferentes módulos, para monitorar a evolução de cada coleção, acompanhar a ação interna da equipe em relação a cada coleção e acompanhar a evolução da rede como um todo. É importante ressaltar que, apesar do sistema ter sido desenvolvido para uso interno do CRIA e do coordenador do projeto com a intenção de auxiliar no acompanhamento das atividades do projeto, passou a ser uma ferramenta de grande interesse para as próprias coleções, uma vez que permite que estas possam fazer um acompanhamento da evolução de sua participação no projeto.

Os módulos implementados são: **lista**, **monitor**, **mapa**, **estatísticas**, **diagrama**, **provedores** e **notas**.

6.1. LISTA

O sistema apresenta uma tabela com a lista de todas as coleções participantes do projeto, o nome oficial e sigla da coleção, a cidade onde está localizada, a que provedor de dados (servidor regional) está conectada, desde quando está ativa na rede (*on-line*), quantos computadores recebeu do projeto, que *software* de gerenciamento utiliza e em que fase passou a participar do projeto. A tabela apresentada pode ser ordenada por qualquer das colunas e o nome da coleção contém um *link* para uma página com detalhes específicos sobre ela. É possível ainda filtrar a lista através de busca em alguns campos, podendo-se ver, por exemplo, apenas a lista das coleções de fanerógamas ou de uma determinada cidade, ou ainda, as coleções que utilizam um determinado *software*

nome da coleção	cidade	provedor de dados	quando está ativa na rede	quantos computadores recebeu do projeto	software de gerenciamento	fase
FEIS Unesp (MEFEIS)	IhaSolteira	speciesLink Botucatu	15/08/2004	1	MS-Works	2
Herbário "Trina Delanova Gemtchújnicov" (BOTU)	Botucatu	speciesLink RJ	30/09/2004	0	Brahms	2
Herbário Dimitri Sucre Benjamin (JBRJ)	Rio de Janeiro	speciesLink SaoPaulo	06/07/2004	1	Brahms	2
Herbário Dom Bento Pickel (SPSF)	São Paulo	speciesLink Campinas	11/08/2004	1	Brahms	2
Herbário Rioclaresense (HRCB)	Rio Claro	speciesLink Campinas	25/06/2003	1	Brahms	2
Herbário da Escola Superior de Agricultura Luiz de Queiroz (ESA)	Piracicaba	speciesLink Campinas	01/06/2003	2	Brahms	1
Herbário da Universidade Estadual de Campinas (UEC)	Campinas	speciesLink IhaSolteira	24/08/2004	1	Brahms	2
Herbário de Ilha Solteira (HISA)	Ilha Solteira	speciesLink SJRP	17/11/2004	1	Brahms	2
Herbário de São José do Rio Preto (HSJRP)	São José do Rio Preto	speciesLink SaoPaulo	09/06/2003	1	Brahms	1
Herbário do Departamento de Botânica, SPF-IB USP (SPF)	São Paulo	speciesLink Campinas	01/07/2003	1	mySQL	1
Herbário do Instituto Agronômico de Campinas (IAC)	Campinas	speciesLink CRIA	09/06/2003	0	PostgreSQL	1
Sistema de Informação do Programa Biota Fapesp (SinBiota) Fapesp	Campinas	speciesLink SaoPaulo	08/10/2004	1	Excel	2
Xiloteca Calvino Mainieri (BCTw)	São Paulo	speciesLink SaoPaulo	07/10/2004	1	Brahms	2
Xiloteca do Instituto de Biociências da Universidade de São Paulo (SPFW)	São Paulo	speciesLink SaoPaulo				

Figura 34. Lista *on-line* do sistema Manager

⁴⁷ <http://splink.cria.org.br/manager>

6.2. MONITOR

Esse módulo apresenta uma tabela com o nome da coleção, a sigla, a cidade onde está localizada, número de registros disponíveis na rede, número total de registros no acervo, porcentagem de registros disponíveis em relação ao total, número de registros georreferenciados disponíveis e a porcentagem destes em relação ao número de registros disponíveis *on-line*. Além dessas informações, a disponibilidade de conexão com a coleção é sinalizada. Processos de verificação da conectividade e do número de registros *on-line* são executados a cada hora e alimentam o banco de dados central. O número de registros geo-referenciados na coleção é extraído do processo de *data cleaning* e pode não refletir a situação da coleção naquele momento. Mesmo assim, é um indicador suficientemente bom já que o processo de *data cleaning* é executado uma vez por dia, sempre que houver alterações nos dados disponibilizados.

The screenshot shows the 'Monitor' module in the SpeciesLink system. It displays a table with the following columns: Coleção, Sigla, Cidade, Registros Disponíveis, Total de Registros, % Disponíveis, Registros Georreferenciados, Total de Registros Georreferenciados, % Georreferenciados, and Status. The table lists various collections such as 'Coleção de Culturas de Fitobacterias do Instituto Biológico', 'Coleção Entomológica "Adolph Hempel" do Instituto Biológico', and 'Coleção Acarológica do Instituto Butantan'. A summary row at the bottom shows totals: 712.806 available records, 2.428.691 total records, 29% availability, 306.515 georeferenced records, and 43% georeferenced records.

Coleção	Sigla	Cidade	Registros Disponíveis	Total de Registros	% Disponíveis	Registros Georreferenciados	Total de Registros Georreferenciados	% Georreferenciados	Status
Coleção de Culturas de Fitobacterias do Instituto Biológico	IBSBF	Campinas	1.624	2.000	81%	0	0%		Concluído
Coleção Entomológica "Adolph Hempel" do Instituto Biológico	IBSP-IB	São Paulo	401	275.000	0%	0	0%		Concluído
Coleção Acarológica do Instituto Butantan	IBSP-Acari	São Paulo	2.700	9.201	29%	753	28%		Concluído
Coleção Herpetológica "Alphonse Richard Hoge"	IBSP-Herpeto	São Paulo	66.167	80.000	83%	0	0%		Concluído
Herbário Dimitri Sucre Benjamin	JBRJ	Rio de Janeiro	6.138	450.000	1%	0	0%		Concluído
Coleção de Peixes do Laboratório de Ictiologia de Ribeirão Preto	LIRP	Ribeirão Preto	5.167	30.000	17%	3.860	75%		Concluído
Coleção do Museu de Entomologia da FEIS/Unesp	MEFEIS	Ilha Solteira	3.519	28.000	13%	0	0%		Concluído
Coleção de Anfíbios do Museu de História Natural "Prof. Dr. Adão José Cardoso"	MHN-anfíbios	Campinas	12.231	12.231	100%	0	0%		Concluído
Coleção de Aves do Museu de História Natural "Prof. Dr. Adão José Cardoso"	MHN-aves	Campinas	2.224	2.224	100%	0	0%		Concluído
Coleção de Mamíferos do Museu de História Natural "Prof. Dr. Adão José Cardoso"	MHN-mamíferos	Campinas	2.385	2.385	100%	0	0%		Concluído
Coleção de Peixes do Museu de História Natural "Prof. Dr. Adão José Cardoso"	MHN-peixes	Campinas	6.138	6.138	100%	0	0%		Concluído
Coleção de Répteis do Museu de História Natural "Prof. Dr. Adão José Cardoso"	MHN-repteis	Campinas	2.832	2.832	100%	0	0%		Concluído
Coleção de Peixes do Museu de Zoologia da USP	MZUSP	São Paulo	77.876	84.000	93%	53.524	69%		Concluído
Coleção Camargo	RPSP	Ribeirão Preto	64.438	171.000	38%	48.957	76%		Concluído
Sistema de Informação do Programa Biota/Fapesp	SinBiota	Campinas	63.577	63.577	100%	63.600	100%		Concluído
Coleção de Algas do Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo"	SP-Algae	São Paulo	14.033	15.000	94%	0	0%		Concluído
Coleção de Fanerógamas do Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo"	SP	São Paulo	13.928	350.000	4%	3.494	25%		Concluído
Herbário do Departamento de Botânica, SPF-IB/USP	SPF	São Paulo	10.250	133.500	8%	4.064	40%		Concluído
Coleção de Algas do Departamento de Botânica, SPF-IB/USP	SPF-Algae	São Paulo	19.776	19.776	100%	0	0%		Concluído
Xiloteca do Instituto de Biociências da Universidade de São Paulo	SPFw	São Paulo	1.677	4.000	42%	80	5%		Concluído
Herbário Dom Bento Pickel	SPSF	São Paulo	31.210	34.000	92%	0	0%		Concluído
Coleção Científica de Aranhas (Araneae) do Depto. de Zoologia da Unesp, Campus Botucatu	UBTU	Botucatu	2.626	3.500	75%	0	0%		Concluído
Herbário da Universidade Estadual de Campinas	UEC	Campinas	35.699	134.000	27%	9.880	28%		Concluído
			712.806	2.428.691	29%	306.515	43%		

* O número de registros georreferenciados na coleção é extraído do processo de *data cleaning* e pode não refletir exatamente a situação da coleção no mesmo momento que os outros dados já que são coletados em diferentes momentos.

Figura 35. Exemplo do módulo monitor

6.3. MAPA

Apresenta o mapa com a distribuição geográfica das coleções participantes da rede, sinalizando sua disponibilidade para consulta de acordo com os dados coletados pelos processos de verificação executados a cada hora. Em verde são apresentados as coleções e servidores regionais acessíveis e em vermelho aqueles que apresentam algum problema de conexão ou que ainda não foram integrados à rede.

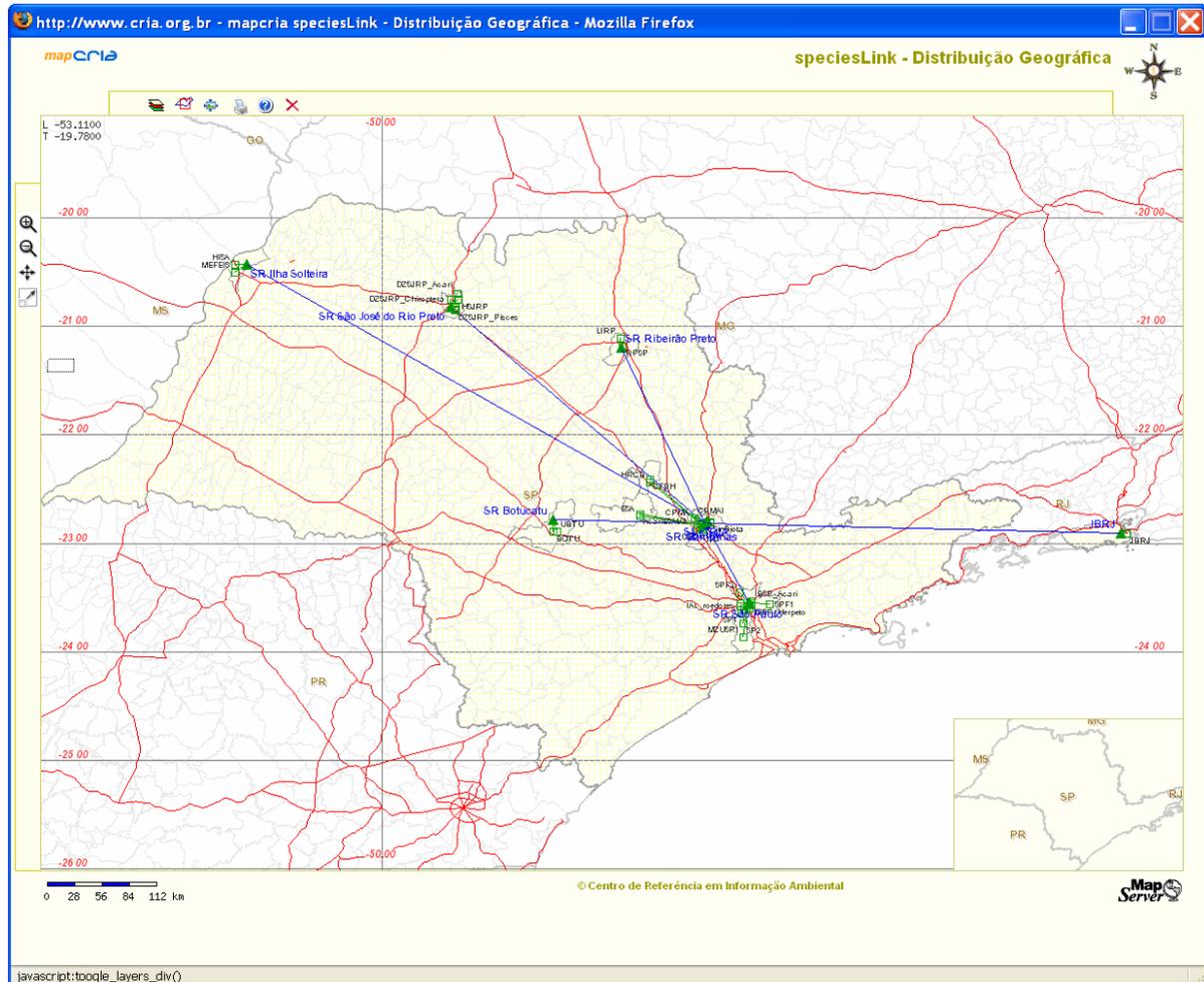
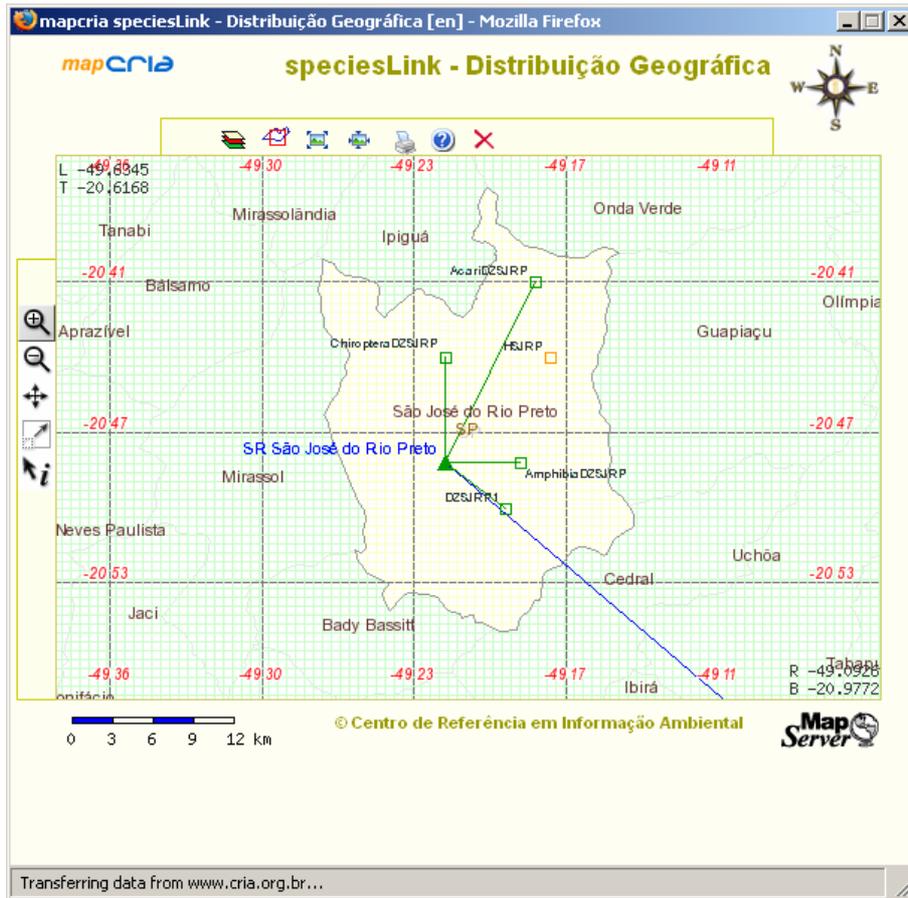


Figura 36. Distribuição geográfica das coleções participantes da Rede *speciesLink*

Para a criação do mapa é utilizado o serviço mapcria (item II.4. nesse relatório) o que torna o mapa interativo. É possível selecionar um detalhe específico, como um município, por exemplo, e clicar em uma das coleções para se obter informações detalhadas sobre ela.



Figuras 37 a e b. Uso do zoom no exemplo anterior e detalhes sobre a coleção

Mozilla Firefox

http://splink.cria.org.br/manager/index?action=detail&node=DZSJRP_Acari&level=des

speciesLink Detalhes sobre a Coleção

Identificação: DZSJRP-Acari

<p>Coleção de Ácaros UNESP, Campus São José do Rio Preto Departamento de Zoologia e Botânica http://www.ibilce.unesp.br</p> <p>Departamento de Zoologia e Botânica Rua Cristóvão Colombo, 2265 Jardim Nazareth 13034-000 São José do Rio Preto SP</p>	<p>país BRA região SE</p> <p>fase 1 comp 1 software Biota online desde 16-05-2003 total registros 7026</p>	<p>São José do Rio Preto SP</p> <p>6296 registros 10/10/05 10:15</p>
---	--	--

contato

<p>curador Dr. Reinaldo José Fazzio Feres e-mail: reinaldo@zoo.ibilce.unesp.br fone: (17) 3221-2368 fax: (17) 3221-2374</p>	<p>tecnico Eduardo Rodrigo Oliveira da Silva e-mail: edusilva_br@yahoo.com.br fone: (17) 3221-2368 fax: (17) 3221-2374</p>
--	---

classificação

localização AMS BRA SP	taxonomia ANIMAIS ARACNÍDEOS	fonte dos dados COL
----------------------------------	--	-------------------------------

menos detalhes ▲

descrição

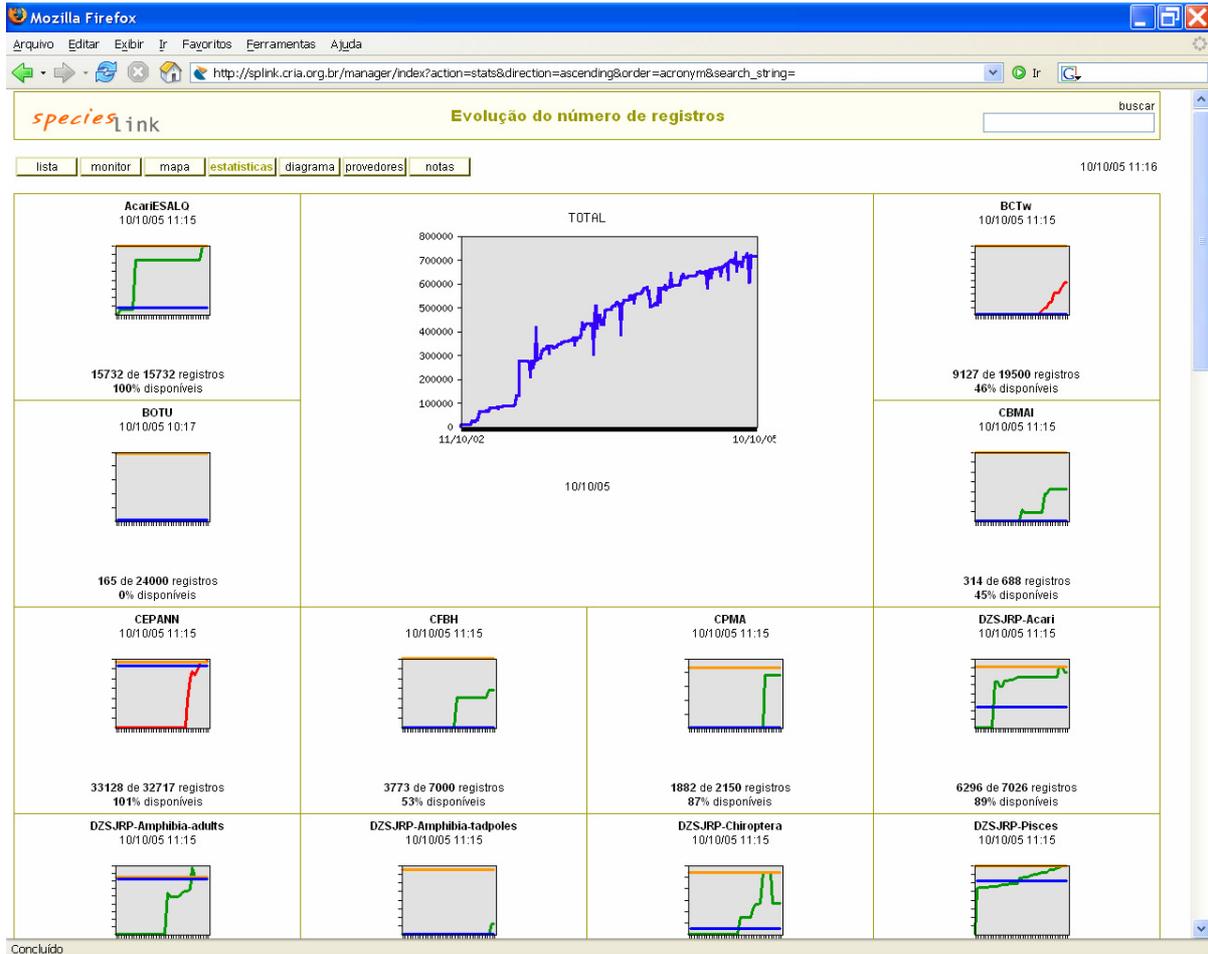
A coleção conta com cerca de 7.000 exemplares, digitalizados através de planilhas do MicroSoft Excel, e migrando para o gerenciador de banco de dados Biota (Colwell). Trata-se de uma coleção principalmente regional, do noroeste do Estado de São Paulo, com amostras de exemplares da Argentina, Colômbia e Indonésia, e parátipos doados (8) de 2 espécies africanas (8) e filipina (2).

mais detalhes ▼

Concluído

6.4. ESTATÍSTICAS

Esse módulo apresenta um gráfico geral sobre a evolução do número de registros disponíveis na rede desde o dia 11 de outubro de 2002, quando a versão beta foi lançada. Apresenta também gráficos individuais sobre a evolução da participação de cada coleção. Os gráficos são gerados dinamicamente de acordo com os dados coletados a cada hora pelos processos de verificação. Nos gráficos de cada coleção, em verde (ou vermelho se o servidor estiver fora do ar) é apresentado o gráfico da evolução dos registros. Em laranja, o número total estimado de registros na coleção e em azul, o número total de registros georeferenciados.



6.5. DIAGRAMA

Um diagrama detalhado das conexões das coleções à rede é apresentado. O diagrama indica, por exemplo, a que servidor regional cada coleção está conectada, que tipo de acervo existe na coleção (botânico, zoológico, microbiológico), onde estão localizadas, nome e instituição, etc.

Esse diagrama não é gerado dinamicamente e precisa ser atualizado cada vez que uma nova conexão é efetivada.

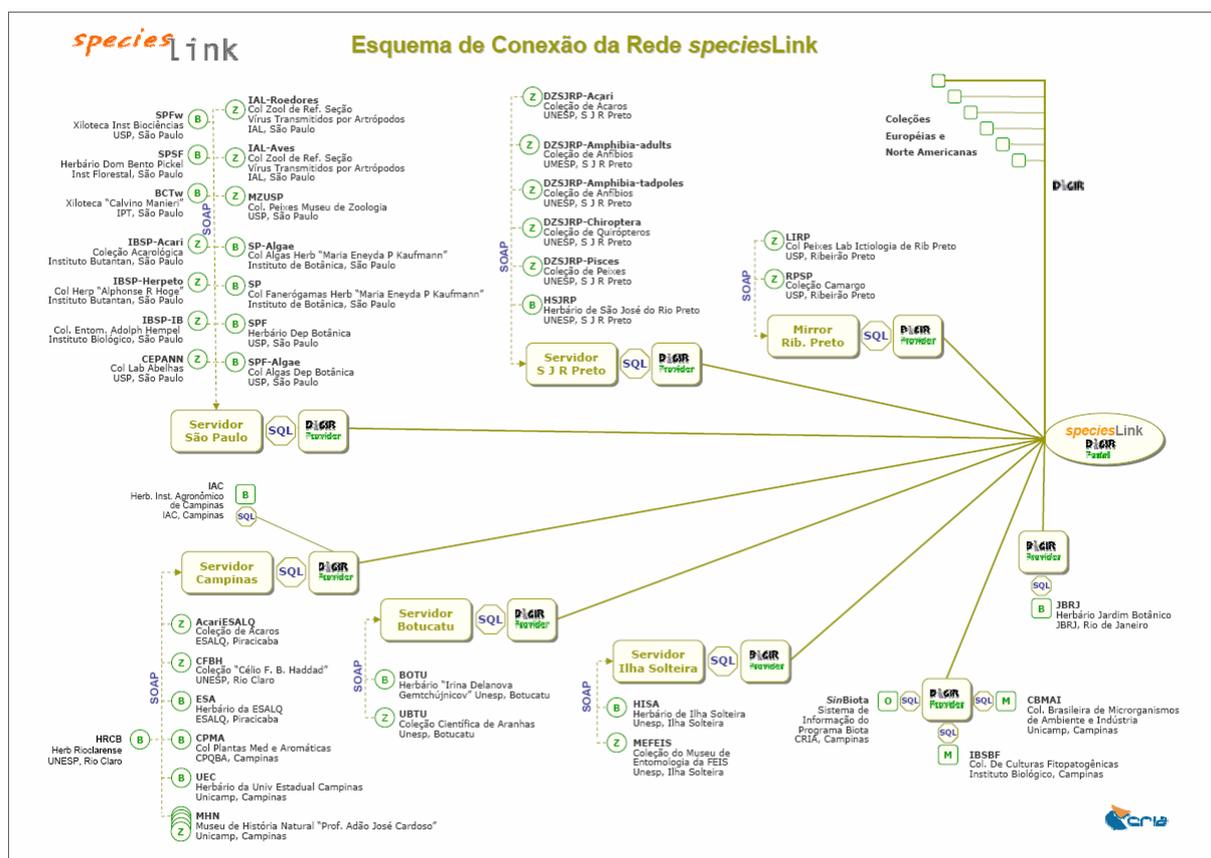


Figura 38. Esquema da Conexão da Rede speciesLink

6.6. PROVEDORES

Apresenta gráficos de disponibilidade dos servidores regionais nas últimas 96 horas, além de uma estatística dessa disponibilidade ao longo do tempo. A checagem dos servidores é feita a cada hora. É importante para análise da qualidade das conexões do portal com os servidores regionais.

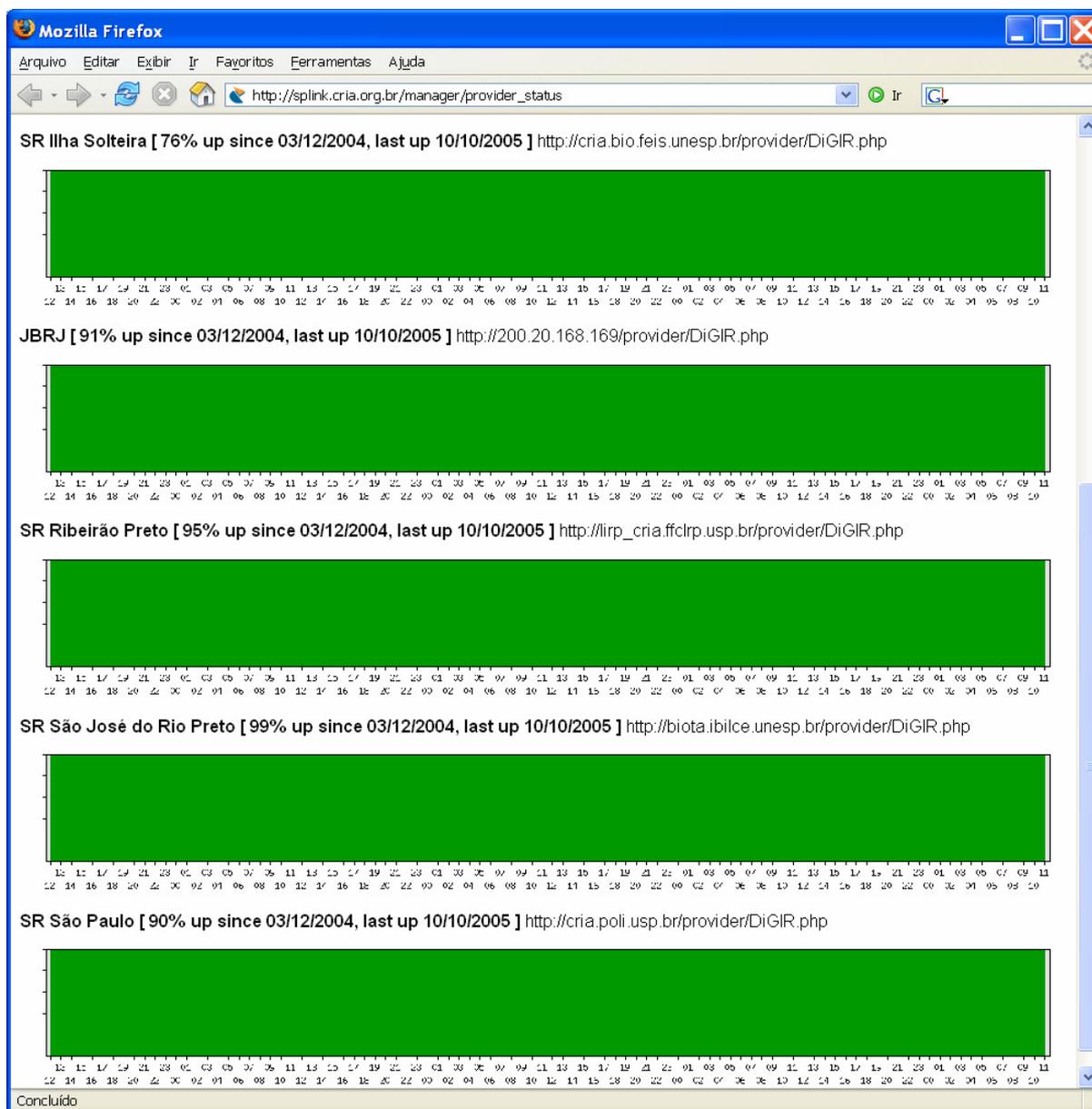


Figura 39. Estatísticas de disponibilidade dos servidores regionais

6.7. NOTAS

Para auxiliar no gerenciamento das atividades do projeto, foi criado um mecanismo de anotação de todas as ações feitas pela equipe do CRIA junto às coleções. Todas as visitas, instalação dos computadores e software, treinamento técnico no uso dos softwares de gerenciamento dos acervos, apoio para adaptação e importação de dados, etc. são anotadas nesse sistema permitindo o acesso fácil ao histórico dessas ações. Essa opção não está disponível ao público em geral, sendo de acesso exclusivo da equipe do CRIA.

7. INTEROPERABILIDADE E INTEGRAÇÃO DE DADOS COM OUTROS SISTEMAS

A proposta original do projeto tinha como meta a integração da rede de coleções biológicas do estado de São Paulo com o *SinBiota* e a rede *Species Analyst*.

7.1. SINBIOTA

SinBiota é o sistema responsável por integrar as informações geradas pelos pesquisadores vinculados ao programa BIOTA/FAPESP, O Instituto Virtual da Biodiversidade. A integração do *SinBiota* e a rede *speciesLink* foi feita através de uma tabela no banco de dados do primeiro, específica para esta finalidade, que foi conectada diretamente a um provedor DiGIR. A tabela criada no *SinBiota* para fazer a conexão é atualizada automaticamente de hora em hora. Isso significa que a inclusão de uma nova coleta no banco de dados do *SinBiota* é disponibilizada praticamente de forma instantânea no *speciesLink*.

Hoje, mais de 60 mil ocorrências de espécies contidas no *SinBiota* estão disponíveis para consulta no *speciesLink*. Estes dados podem ser integrados aos dados das coleções em tempo real, e podem ser obtidos em formatos de fácil utilização (html, planilhas ou visualizadas em mapas). A integração dos dados trouxe mais um benefício ao *SinBiota*, além de uma disseminação mais ampla e da integração de seus dados com as coleções, que é o uso da ferramenta *data cleaning*. Os registros “suspeitos” do *SinBiota* podem ser analisados via ferramenta do *data cleaning* no *speciesLink* (<http://splink.cria.org.br/dc>) ou através do site do *SinBiota* com acesso restrito aos coordenadores de projeto (http://SinBiota.cria.org.br/info/info_amb).

7.2. COLEÇÕES BIOLÓGICAS DO EXTERIOR

A rede *Species Analyst*, coordenada pela Universidade de Kansas, foi pioneira no desenvolvimento de sistemas distribuídos de informação sobre espécimes e sua localização geográfica, utilizando o protocolo Z39.50. Essa rede, com mais de 40 coleções integradas, evoluiu para redes temáticas como a rede de mamíferos *Manis* e a rede de peixes *Fishnet*. Aos poucos todos os sistemas estão adotando o protocolo DiGIR, o mesmo protocolo de comunicação utilizado pela rede *speciesLink*, tornando possível integrar dados da rede paulista com fontes de dados de outras redes regionais ou internacionais. Esta integração se dá em duas direções. A primeira delas é a integração dos dados de coleções internacionais, tornando-as visíveis nas buscas do *speciesLink*. De maneira simétrica, as coleções paulistas podem ser conectadas aos portais das redes internacionais de maneira simples e automática.

Para testar a interoperabilidade da rede foram incluídas as seguintes coleções do exterior na opção de busca do sistema:

Coleção	Acervo Disponível	Registros do Brasil
Swedish Museum of Natural History		
Bryophytes - mosses (NRM)	85.726	1.179
lichens specimens	52.995	1.604
Nordic Herbarium	108.035	
fish collection	40.479	2.113
Invertebrates	70.198	94
SG-BIRDS - bird observations in Sweden	1.818.065	
Herbarium of Oskarshamn, Sweden	85.998	
Kansas University		
KUNHM bird collection	94.915	1.061
KUNHM herp collection	292.038	
University of Kansas Insect Collection	494.565	6
KUNHM fish collection	36.028	29
KUFishTissues	7.386	
KUNHM specimen data for mammals	158.095	124
University of Kansas Plant Collection	162.517	33
Museum of Vertebrate Zoology, Terrestrial vertebrate specimens	630.562	5.743
Total	4.137.602	11.986

Uma outra iniciativa que evoluiu rapidamente após a concepção original da proposta é a Infra-estrutura Global de Informação sobre Biodiversidade - GBIF (*Global Biodiversity Information Facility*). O GBIF implementou um portal DiGIR com os acervos dos países membros e iniciativas globais que aderiram a este esforço internacional. A rede GBIF⁴⁸ conta hoje com 141 provedores de dados e disponibiliza mais de 75 milhões de registros *on-line*, todos de acesso público e gratuito.

O Brasil ainda não aderiu ao GBIF, mas alguns acervos de outros países que estão disponibilizando os seus dados possuem informações sobre coletas realizadas no Brasil. Em outubro de 2005 a rede GBIF possuía cerca de 190 mil registros do Brasil.

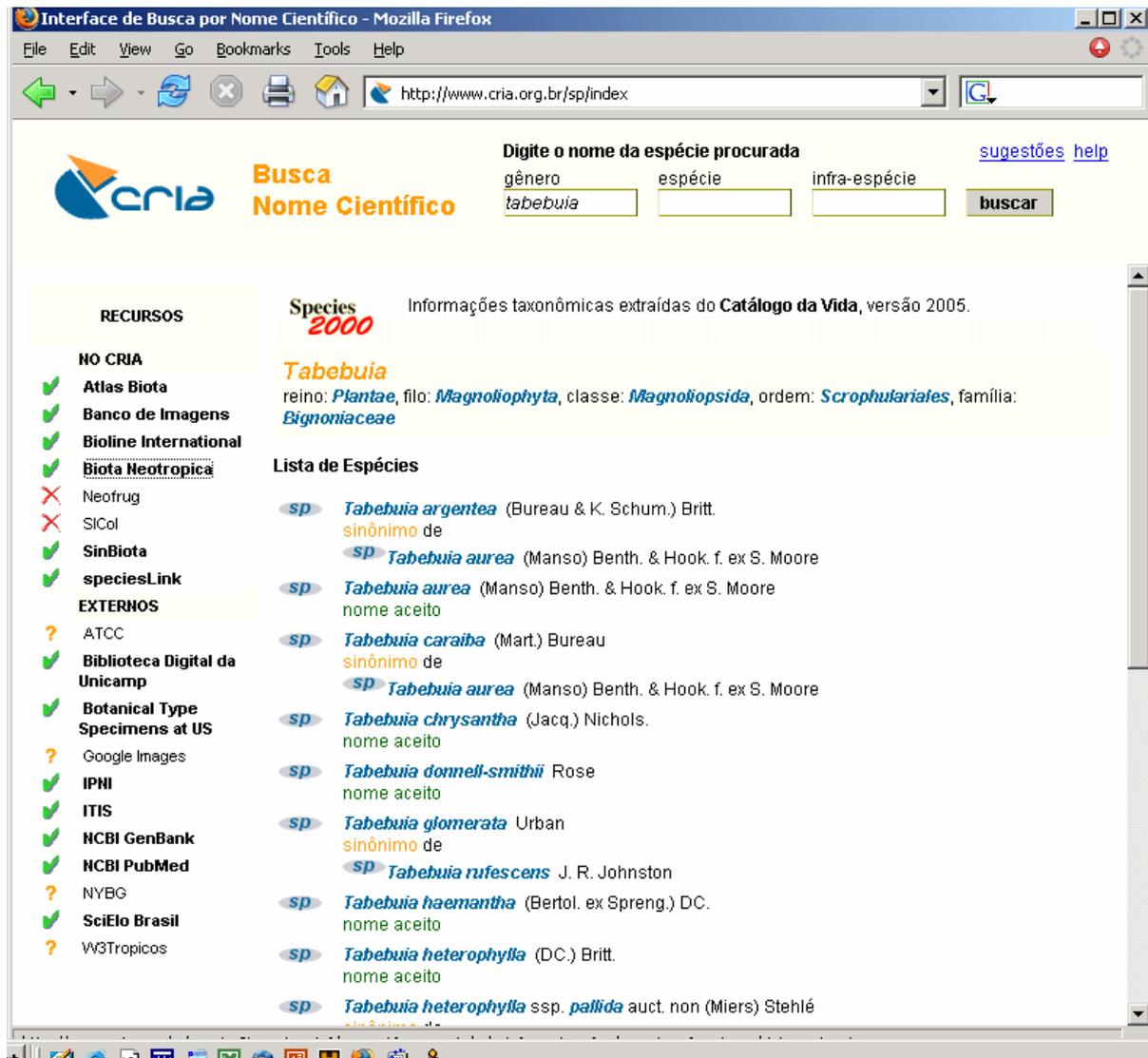
É importante ressaltar que, por utilizar o mesmo protocolo adotado pelo *speciesLink*, todas as coleções disponíveis na rede GBIF podem ser conectadas ao portal do *speciesLink* imediatamente.

7.3. OUTROS SISTEMAS

O CRIA trabalha com sistemas de informação nos quais o nome científico do organismo é o objeto principal. É o caso do *speciesLink*, do *SinBiota*, do SICol, do Neofrug, entre outros. Trabalhamos também com sistemas nos quais os nomes científicos, apesar de não serem o objeto central, são citados e representam uma rica fonte de informação complementar, como é o caso das revistas *Biota Neotropica* e *Bioline International* e do Banco de Imagens da Biodiversidade Brasileira.

Foram criados mecanismos de indexação e recuperação de dados de forma a permitir o cruzamento das informações associadas aos organismos existentes nos diferentes sistemas. Para os sistemas que têm uma estrutura de dados própria, como é o caso do

⁴⁸ <http://www.gbif.net>



Interface de Busca por Nome Científico - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.cria.org.br/sp/index

Busca Nome Científico

gênero: espécie: infra-espécie:

[sugestões](#) [help](#)

RECURSOS

NO CRIA

- Atlas Biota
- Banco de Imagens
- Bioline International
- Biota Neotropicala**
- Neofrug
- SICol
- SinBiota
- speciesLink

EXTERNOS

- ATCC
- Biblioteca Digital da Unicamp
- Botanical Type Specimens at US
- Google Images
- IPHI
- ITIS
- NCBI GenBank
- NCBI PubMed
- NYBG
- SciELO Brasil
- W3Tropicos

Species 2000 Informações taxonômicas extraídas do **Catálogo da Vida**, versão 2005.

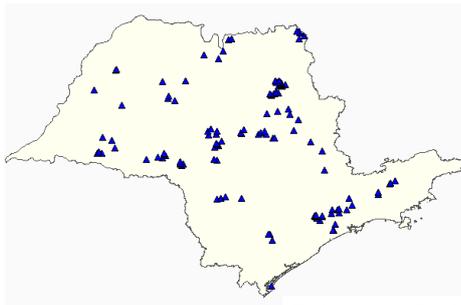
Tabebuia
reino: *Plantae*, filo: *Magnoliophyta*, classe: *Magnoliopsida*, ordem: *Scrophulariales*, família: *Bignoniaceae*

Lista de Espécies

- SP** *Tabebuia argentea* (Bureau & K. Schum.) Britt.
sinônimo de
- SP** *Tabebuia aurea* (Manso) Benth. & Hook. f. ex S. Moore
nome aceito
- SP** *Tabebuia aurea* (Manso) Benth. & Hook. f. ex S. Moore
nome aceito
- SP** *Tabebuia caraiiba* (Mart.) Bureau
sinônimo de
- SP** *Tabebuia aurea* (Manso) Benth. & Hook. f. ex S. Moore
- SP** *Tabebuia chrysantha* (Jacq.) Nichols.
nome aceito
- SP** *Tabebuia donnell-smithii* Rose
nome aceito
- SP** *Tabebuia glomerata* Urban
sinônimo de
- SP** *Tabebuia rufescens* J. R. Johnston
- SP** *Tabebuia haemantha* (Bertol. ex Spreng.) DC.
nome aceito
- SP** *Tabebuia heterophylla* (DC.) Britt.
nome aceito
- SP** *Tabebuia heterophylla* ssp. *pallida* auct. non (Miers) Stehlé
sinônimo de

Figura 41. Página integrando a busca pelo nome da espécie, no caso *Tabebuia*

Além da validação do nome pelo Catálogo da Vida (Species2000), o resultado da busca apresenta a ocorrência do nome em outros sistemas internos e externos ao CRIA (figuras a seguir).



Tabebuia (Dados Sistema CRIA)

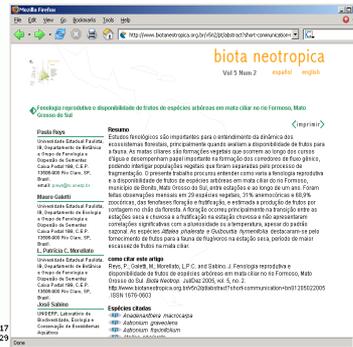
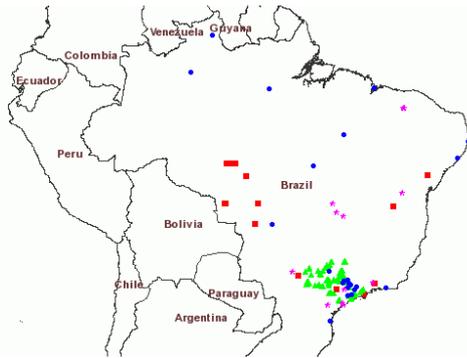


Figura 42. Exemplos de sistemas no CRIA com dados sobre o gênero Tabebuia

A figura acima mostra os dados da distribuição do gênero *Tabebuia* no Estado de São Paulo (*SinBiota*), uma foto da árvore no campo (banco de imagens), a distribuição no Brasil (*speciesLink*), e um artigo citando o gênero (revista *Biota Neotropica*). A figura a seguir mostra a ocorrência do gênero *Tabebuia* em sistemas externos ao CRIA.

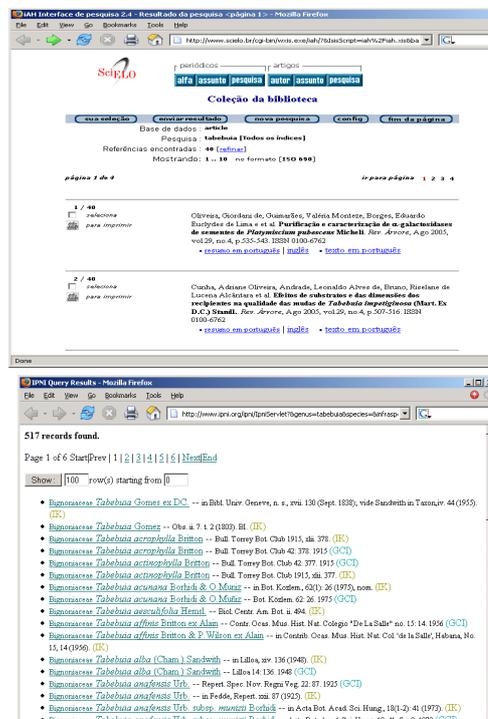
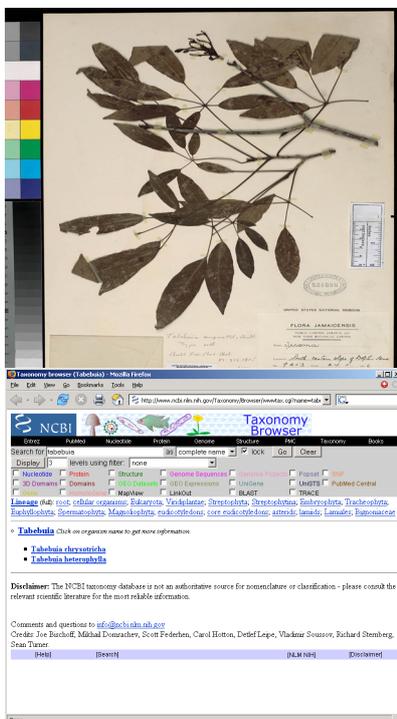


Figura 43. Exemplos de ocorrência do gênero Tabebuia em sistemas externos ao CRIA

8. MODELAGEM

8.1. HISTÓRICO

A idéia de se desenvolver um ambiente computacional para a realização de modelagem de nicho ecológico teve como principal objetivo promover a integração e o uso dos dados biológicos que estavam sendo disponibilizados pela rede *speciesLink*. Várias atividades foram realizadas no período, mas o que realmente sedimentou esta área no CRIA foi a vinda dos pesquisadores Dr. Andrew Townsend Peterson e Dr. Arthur David Chapman e a realização do Workshop Inter-Americano sobre Acesso a Dados Ambientais.

A vinda do Dr. Townsend, professor de Ecologia e Biologia Evolutiva e Curador da Coleção de Ornitologia do *Natural History Museum*, da Universidade de Kansas foi o marco inicial do desenvolvimento de atividades ligadas à modelagem de distribuição potencial de espécies no CRIA. Dr. Townsend trabalhou durante 2 meses no CRIA, estabelecendo parcerias e capacitando as pessoas para trabalharem com modelagem de distribuição potencial de espécies, culminando em várias publicações na área e no desenvolvimento de várias linhas de pesquisa. Durante sua estada o pesquisador ministrou um curso de treinamento intitulado “*Métodos e Ferramentas para Modelagem Preditiva de Espécies*”, com a presença de 52 participantes, incluindo pesquisadores renomados da área. O relatório de atividades do Dr. Townsend no Brasil está disponível na página de relatórios do projeto *speciesLink*⁴⁹.

A vinda do Dr. Chapman para o Brasil durante um período de 12 meses possibilitou o aprimoramento das técnicas de modelagem de nicho ecológico de espécies utilizadas. Dr. Chapman tem uma visão bastante crítica do GARP (principal algoritmo adotado pelo Dr. Townsend) o que fez com que a equipe trabalhasse com outros algoritmos fazendo análises comparativas tornando as ferramentas para modelagem muito mais ricas e interessantes. Além da modelagem, Dr. Chapman atuou junto à equipe no desenvolvimento de ferramentas para a identificação de erros em bancos de dados de biodiversidade. O seu relatório também está disponível na página de relatórios do projeto.

Oitenta pesquisadores do Brasil e do exterior participaram do Workshop Inter-Americano sobre Acesso a Dados Ambientais⁵⁰. O Workshop teve por objetivo reunir especialistas em biodiversidade com os de dados ambientais identificando possíveis recomendações que poderiam ser apresentadas ao Comitê Executivo do CODATA. Foi uma oportunidade para que essas duas comunidades pudessem iniciar possíveis colaborações e projetos comuns. A proposta de realização de um ambiente computacional de modelagem apresentada pelo CRIA, Poli e INPE foi um dos frutos desta reunião.

Quando a equipe do CRIA passou a ser usuária do algoritmo Desktop Garp, foram identificados vários problemas e dificuldades no manejo dos dados, que serviram de motivação para a realização da proposta para o desenvolvimento do ambiente computacional para modelagem. As principais dificuldades encontradas referiam-se à aquisição, preparo e utilização dos dados ambientais com os dados biológicos disponíveis. Percebeu-se que o pesquisador gastava mais tempo preparando os dados do que na análise dos resultados. Neste contexto, o ambiente de modelagem idealizado deveria ser capaz de:

- Ler mapas georeferenciados armazenados em diferentes formatos (GeoTiff, Arc/Info Grid, GXF, etc).
- Lidar com diferentes sistemas de coordenadas e de projeções para combinar diferentes mapas e pontos de ocorrência de espécies.
- Permitir a execução de diferentes algoritmos ao mesmo tempo, com os mesmos dados, possibilitando a comparação entre os resultados gerados.

⁴⁹ <http://splink.cria.org.br/reports>

⁵⁰ <http://www.cria.org.br/eventos/iaed/>

- Permitir que o pesquisador pudesse concentrar o seu tempo no desempenho dos algoritmos e nos resultados gerados.
- Lidar com inconsistências nas coordenadas geográficas (erros ou diferenças no datum utilizado).
- Identificação de erros com a máscara (para entrada de dados) utilizada, que nem sempre é adequada para a resolução disponível dos dados.

A realização de todos esses processos depende bastante esforço e tempo, uma vez que tinha que ser realizado pelo próprio pesquisador. O objetivo seria automatizar ao máximo esses processos.

Os recursos solicitados para dar continuidade ao desenvolvimento de ferramentas de modelagem não foram aprovados pela Fapesp no termo aditivo ao projeto. O CRIA então lançou o conceito e uma versão preliminar de um ambiente computacional para modelagem no sourceforge⁵¹ como uma biblioteca computacional de código aberto (*open source*). Em seguida, em parceria com a Escola Politécnica da USP e o Instituto Nacional de Pesquisas Espaciais propôs à Fapesp um projeto para o desenvolvimento desse ambiente que foi denominado *openModeller*. O projeto foi aprovado e teve início em abril de 2005. Todo o desenvolvimento continua sendo alimentado no *sourceforge*. Desta forma, usuários das diversas áreas poderão contribuir com o desenvolvimento do projeto e com a avaliação de seus resultados.

8.2. ESTUDOS DE CASO

A qualidade e precisão necessárias para um processo de modelagem dependem fundamentalmente da pergunta a ser feita. Alguns trabalhos desenvolvidos durante a primeira fase do projeto estudavam grandes biomas com perguntas como qual seria a distribuição futura de espécies arbóreas no Cerrado brasileiro diante de diferentes cenários de mudanças climáticas. Para perguntas mais locais como quais as melhores espécies arbóreas a serem utilizadas em um projeto de recuperação de uma área degradada na bacia do Médio Paranapanema, é necessário ter dados mais precisos para gerar modelos mais detalhados.

Além dos dados sobre espécies e distribuição geográfica, o CRIA está procurando criar um acervo de dados não bióticos. Todos os dados ambientais de acesso público estão relacionados na página do projeto no item “Bases ambientais disponíveis”⁵².

A seguir apresentamos alguns exemplos de estudos de caso desenvolvidos (ou em fase de desenvolvimento) pela equipe do CRIA neste período.

8.2.1 UTILIZAÇÃO DE DIFERENTES ESCALAS DE DADOS CLIMÁTICOS

Foi realizado um experimento comparando o resultado da modelagem utilizando diferentes escalas de dados climáticos (figura 44). O exemplo mostra a importância da escala do dado ambiental na produção do modelo. Mostra também a qualidade da modelagem quando são consideradas outras variáveis ambientais.

⁵¹ <http://openModeller.sf.net>

⁵² <http://splink.cria.org.br/bases>

***Aspidosperma cylindrocarpon* Müll. Arg. (Apocynaceae)**

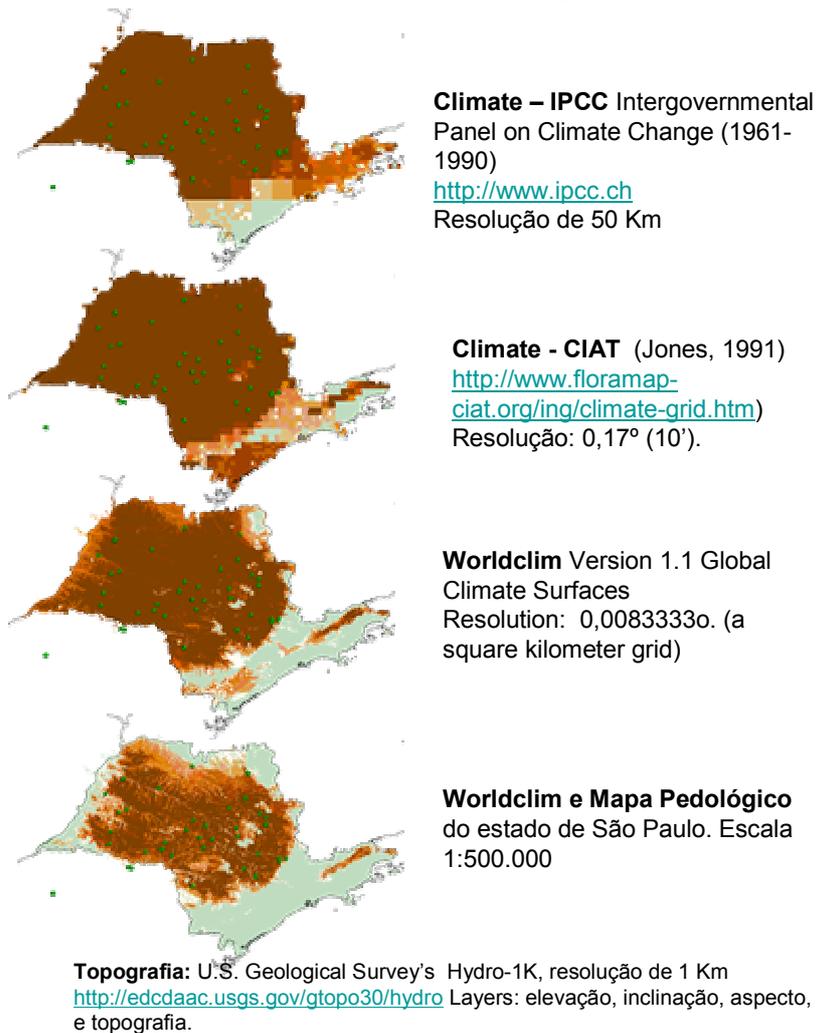


Figura 44. Mapas com resultados dos experimentos de modelagem feitos com resoluções distintas e com o acréscimo de dados de solo.

8.2.2 MODELAGEM DE ESPÉCIES LENHOSAS DE CERRADO EM TODO O ESTADO DE SÃO PAULO

As diferentes formações de cerrado recobriam originalmente cerca de 14% do território paulista, atualmente esta área é de menos de 1% da sua distribuição original (Kronka et al. 1998), ou seja, cerca de 23.790 ha extremamente fragmentados como podemos ver na Figura 45.

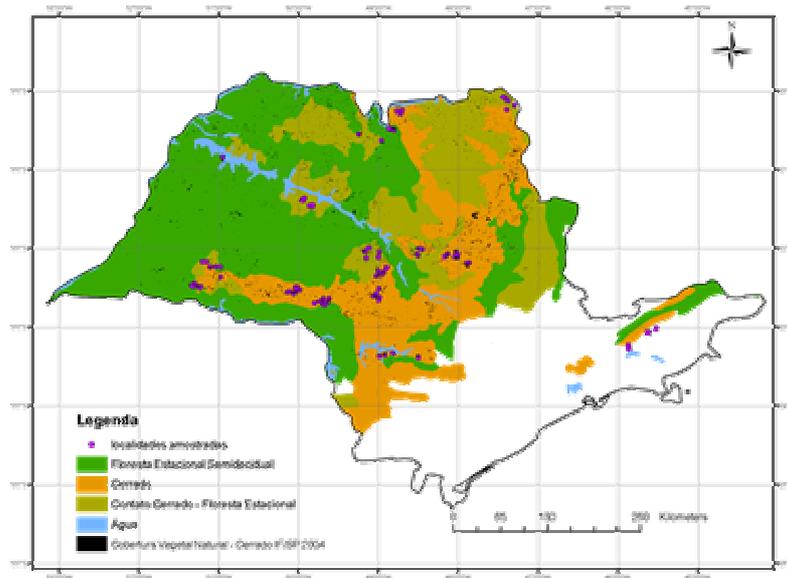


Figura 45: Região de ocorrência de Cerrado, vegetação ecotonal e da Floresta Estadual Semidecidual no estado de São Paulo

Utilizando técnicas de modelagem, foi desenvolvido um mapa com a distribuição potencial de cerrado, baseado em 17 espécies arbóreas, para o estado de São Paulo (Figura 46)⁵³.

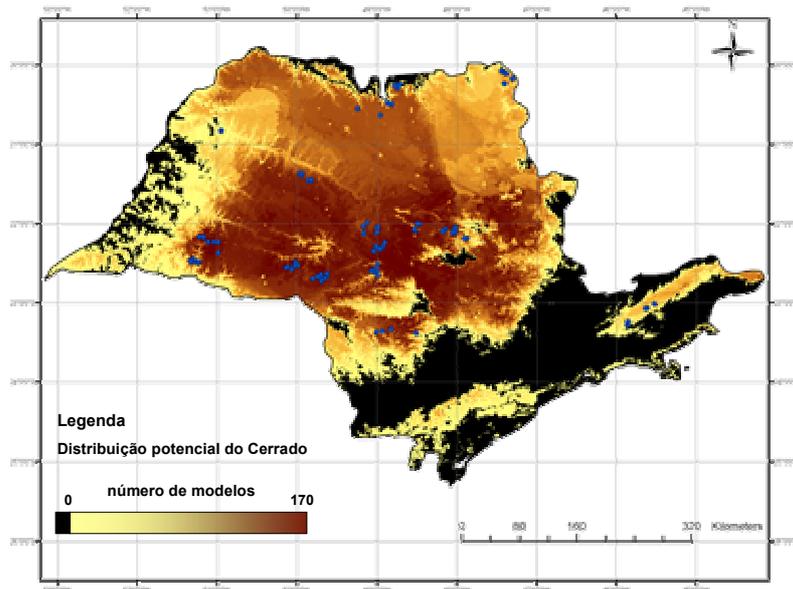


Figura 46. Distribuição potencial do Cerrado baseado em 17 espécies arbóreas

⁵³ KRONKA, F. J. N.; NALON, M. A.; MATSUKUMA, C. K.; PAVÃO, M.; GUILLAUMON, J. R.; CAVALLI, A. C.; GIANNOTTI, E.; IWANE, M. S. S.; LIMA, L. M. P. R.; MONTES, J.; DEL-CALI, I. H.; HAACK, P. G. (1998). *Áreas de domínio de cerrado no Estado de São Paulo*. Secretaria de Estado do Meio Ambiente, Instituto Florestal. São Paulo. 84p.

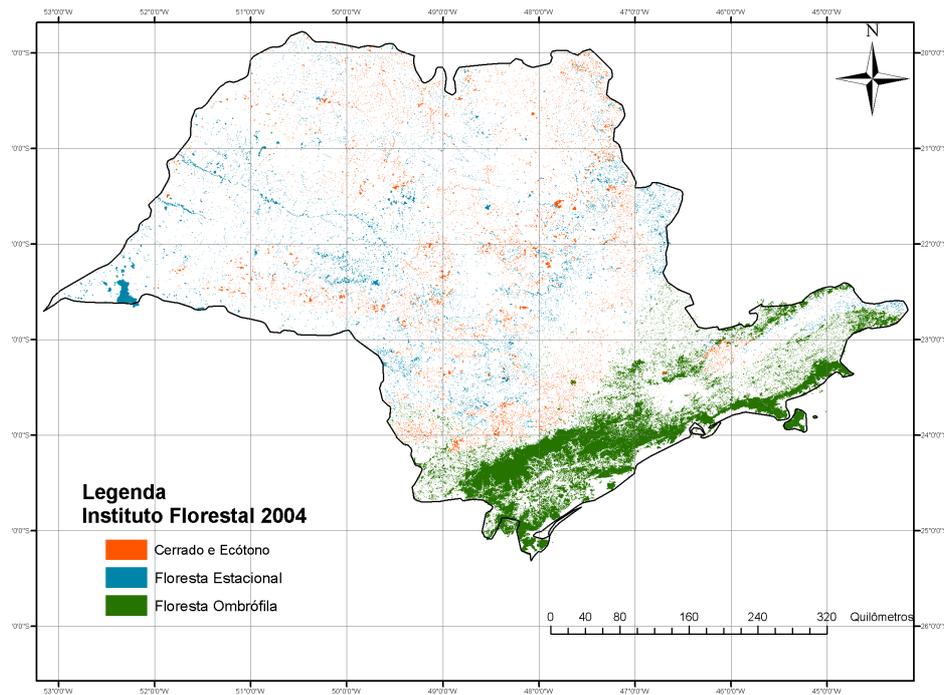


Figura 47: Distribuição potencial do Bioma Cerrado, obtida a partir de modelagem e distribuição dos fragmentos de vegetação natural no estado para o estado de São Paulo.

Neste mapa vemos a área potencial de cerrado, com relação às variáveis ambientais utilizadas (clima, topografia e solos) é bem maior que a distribuição original dada pelo IBGE. A área indicada como ecótono (zona de transição) entre o Cerrado e a Floresta Estacional deve ser muito mais extensa, englobando boa parte do oeste do estado, onde segundo o mapeamento do IBGE, seria uma área de Floresta Estacional Semidecidual. Isso também pode ser observado na distribuição geográfica dos fragmentos de Cerrado pelo mapeamento do Instituto Florestal (Kronka et. al. 1998).

8.2.3 MODELAGEM DE ESPÉCIES ARBÓREAS DO MÉDIO PARANAPANEMA: APLICAÇÕES PARA RECUPERAÇÃO DE ÁREAS DEGRADADAS

Com o objetivo de testar a modelagem para uma escala mais local, foram realizados alguns experimentos para a Bacia do Médio Paranapanema, sudoeste de São Paulo (Figura 48) a partir de dados de clima, topografia, solos e geologia da região.

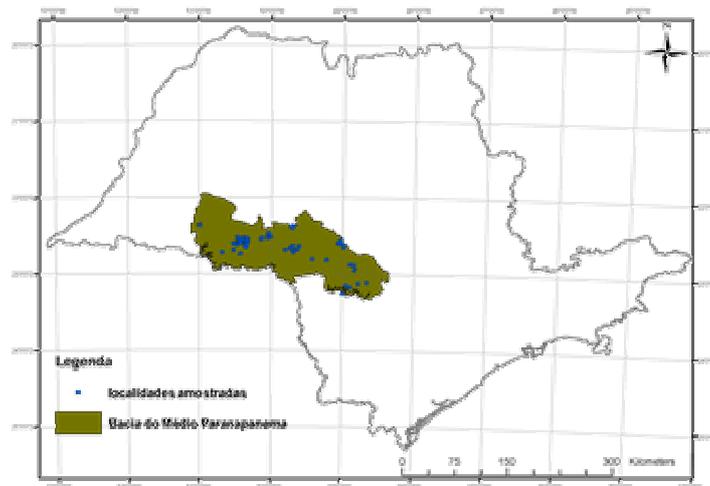


Figura 48: Localização geográfica da Bacia do Médio Paranapanema no estado de São Paulo com os pontos de amostragem da vegetação.

O modelo resultante gerado (Figura 49) representa o potencial de distribuição geográfica das 28 espécies arbóreas analisadas. As áreas mais escuras representam um maior número de espécies com potencial de ocorrência na área, baseado nos dados ambientais utilizados na modelagem.

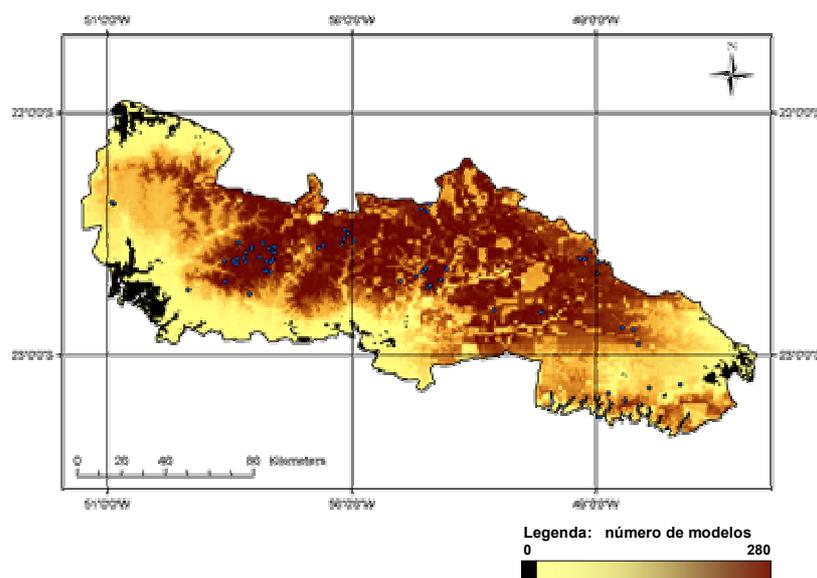


Figura 49: Mapa da riqueza potencial para 28 espécies arbóreas analisadas para a Bacia do Médio Paranapanema.

Através do uso de SIG (Sistemas de Informação Geográfica) o modelo gerado foi cruzado com a informação disponível em mapas do meio físico.

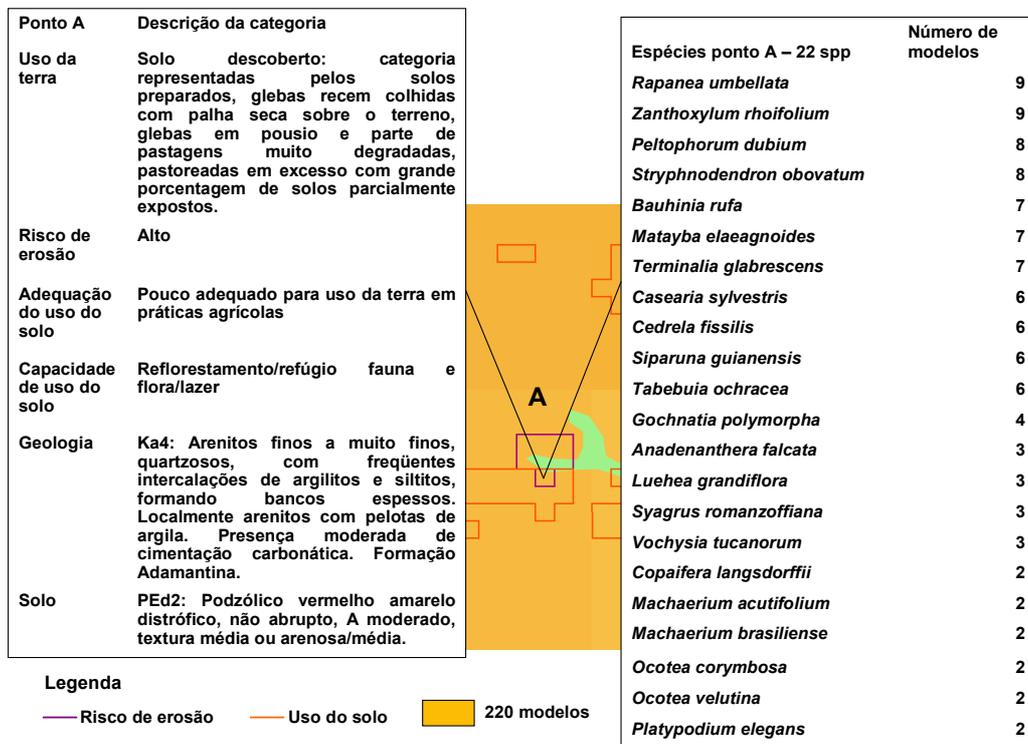
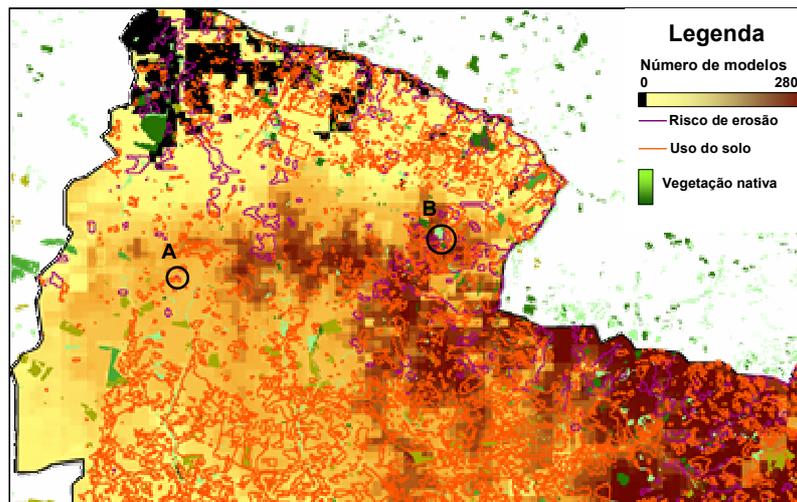


Figura 50: Cruzamento de dados entre o modelo de riqueza de espécies e as áreas de risco para a Bacia do Médio Paranapanema.

Com isso é possível saber quais áreas são potencialmente mais ricas em espécies, quais áreas tem potencial risco de erosão, quais apresentam solo descoberto, qual a proximidade dessas áreas com fragmentos de vegetação natural e assim, poder indicar, para a região da Bacia, quais áreas poderiam ser indicadas para recuperação ambiental, indicando também quais espécies seriam mais propícias. Com isso seria possível indicar, por exemplo, quais áreas seriam prioritárias para a criação de Unidades de Conservação para a Bacia.

8.2.4 MODELAGEM DE ESPÉCIES COM DISTRIBUIÇÃO RESTRITA OU COM POUCOS PONTOS DE AMOSTRAGEM

Para utilizar algoritmos de modelagem, geralmente é necessário um número mínimo de registros de ocorrência para cada espécie. Para o GARP, o ideal é ter pelo menos 20 pontos (STOCKWELL & PETERSON 2002⁵⁴). Obter este número de pontos de amostragem para cada espécie nem sempre é possível. No Cerrado, por exemplo, cerca de 30% das espécies arbóreas são conhecidas por apenas um único ponto de ocorrência (RATTER et al. 2000⁵⁵). Para estes casos existe a possibilidade de utilizar modelos mais simples baseados em similaridade ambiental. Este trabalho foi desenvolvido com o objetivo de orientar trabalhos de campo na busca por espécies consideradas endêmicas ou raras. A Figura 51 mostra os resultados para *Byrsonima subterranea*, espécie que foi considerada provavelmente extinta no estado (SMA-SP 1997⁵⁶).

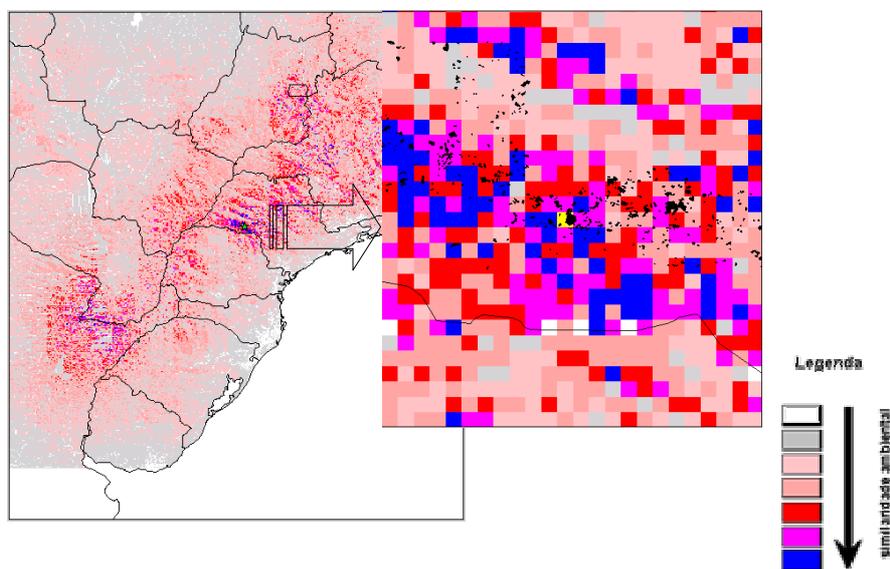


Figura 51: Modelo de distribuição potencial de *Byrsonima subterranea*.

Este modelo foi criado baseado em um único registro de ocorrência (Estação Ecológica de Assis, Município de Assis, SP) e utilizando apenas dados climáticos e topográficos. Os quadrados (*pixels*) azuis representam as áreas de maior similaridade ambiental e os brancos e cinza-claros representam as áreas de menor similaridade ambiental. Com este modelo foi possível selecionar novas áreas de coleta (figura 52) e como resultado do trabalho de campo, foram incluídos mais 6 registros de ocorrência para a espécie *Byrsonima subterranea* no estado de São Paulo, encontrados em áreas de grande similaridade ambiental em relação ao ponto de origem da modelagem.

⁵⁴ STOCKWELL, D. R. B.; PETERSON, A. T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling*. 148: 1-13.

⁵⁵ RATTER, J. A.; BRIDGEWATER, S.; RIBEIRO, J. F.; DIAS, T. A. B.; SILVA, M. R. (2000). Distribuição das espécies lenhosas da fitofisionomia cerrado sentido restrito nos estados compreendidos no bioma cerrado. *Boletim do Herbário Ezechias Paulo Heringer*. 5: 5-43.

⁵⁶ SMA-SP. (1997). Espécies da flora ameaçadas de extinção no estado de São Paulo: lista preliminar. Secretaria do Meio Ambiente/ Governo do Estado de São Paulo. São Paulo.

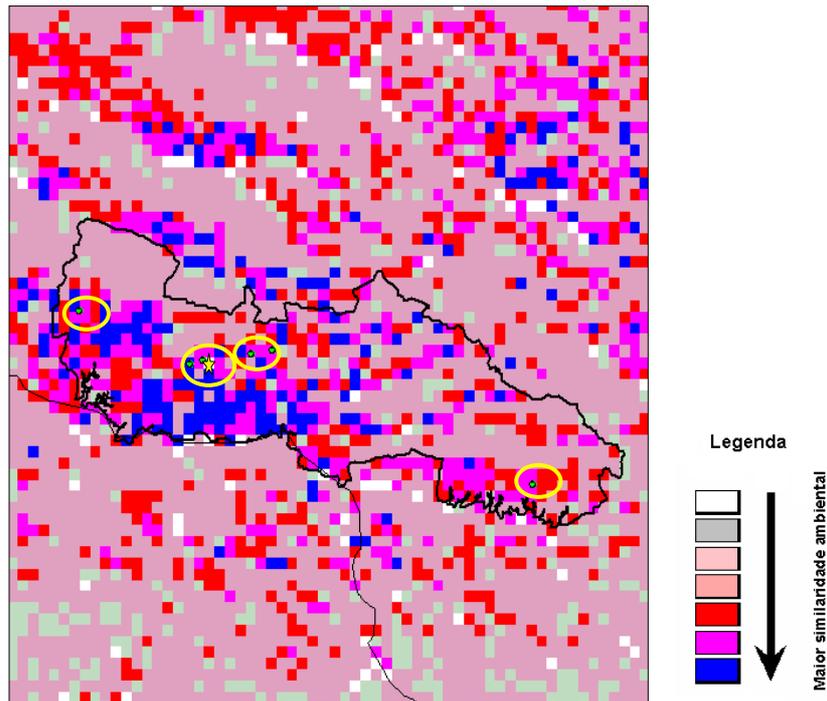


Figura 52: Seleção de locais de coleta e a inclusão de mais 6 pontos de ocorrência para *Byrsonima subterranea*

Esta abordagem mostrou ser bastante eficiente na busca de espécies no campo não só por tratar espécies com poucos pontos de amostragem mas também por otimizar o tempo envolvido no trabalho de campo para a busca de novos pontos de ocorrência. Esta metodologia é simples e pode ser replicada para qualquer grupo taxonômico que apresente deficiências ou dificuldades de coleta assim como para espécies raras, endêmicas ou ameaçadas de extinção.

8.2.5 MODELOS DE DISTRIBUIÇÃO GEOGRÁFICA DE ESPÉCIES VEGETAIS DISPERSAS PELA MEGAFUNA: OS EFEITOS DA PERDA DE DISPERSORES.

Em muitos casos, o movimento de sementes e seu padrão de dispersão podem ser associados ao movimento de vertebrados. Porém, algumas plantas possuem frutos que parecem estar mais adaptados para serem dispersos por animais de megafauna do período Pleistoceno, sendo que a sobrevivência dessas populações sem seus dispersores ainda não foi explicada satisfatoriamente. O estudo do padrão de distribuição de espécies vegetais em sua distribuição potencial pode auxiliar na análise de como a ausência ou diminuição do número de dispersores influencia a distribuição geográfica da planta.

Este trabalho tem por objetivo investigar os efeitos da ausência de dispersores em espécies vegetais do Cerrado brasileiro que eram provavelmente dispersas pela megafauna, através da análise de modelos de distribuição geográfica potencial gerados pelo software *DesktopGARP*. Os modelos gerados para as condições ambientais atuais serão comparados com projeções realizadas para as condições climáticas do Pleistoceno.

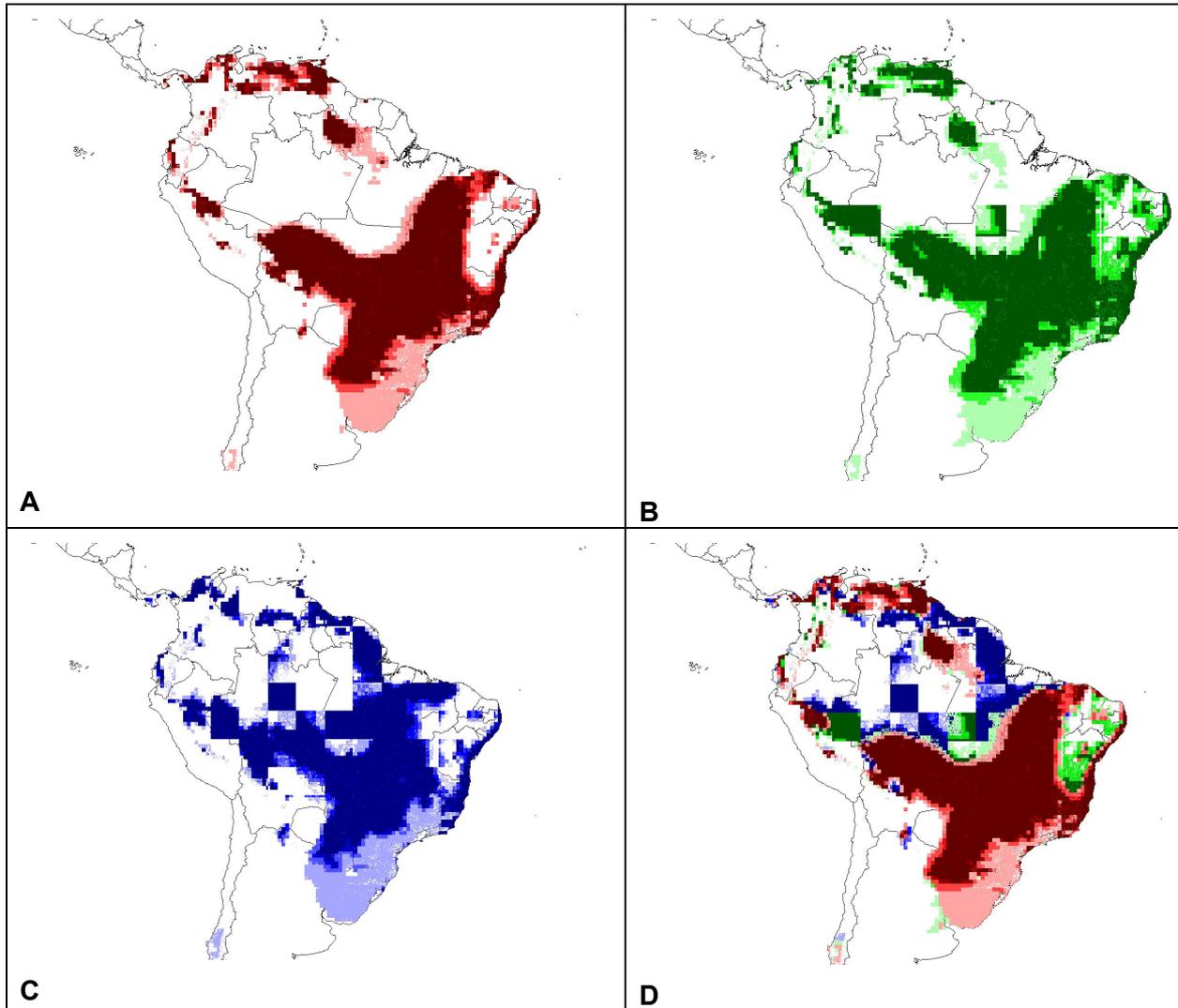


Figura 53. Modelagem da distribuição de *Hancornia speciosa* (Apocynaceae) no cenário atual (A), há 6000 anos (B), há 21.000 anos (C) e a sobreposição dos três modelos (D).

A figura 53 A mostra a distribuição atual de *Hancornia speciosa* (Apocynaceae). Na figura B o cenário projetado para 6000 anos atrás, utilizando como base dados do *Canadian Climate Center*. Em C pode-se observar a projeção para 21000 anos atrás. A sobreposição dos três cenários pode ser observada em D.

Esse trabalho faz parte de uma tese de mestrado em andamento.

8.2.6 MODELAGEM DE DISTRIBUIÇÃO POTENCIAL DAS ESPÉCIES DA SEÇÃO TÍPICA DO GÊNERO *RAUVOLFIA* (APOCYNACEAE) – APLICAÇÃO COMO FERRAMENTA AUXILIAR NA TAXONOMIA E ESTUDOS BIOGEOGRÁFICOS.

O gênero *Rauvolfia* pertence à família Apocynaceae, subfamília Rauvolfioideae (Endress & Bruyns, 2000⁵⁷). Suas espécies distribuem-se pelos trópicos e possuem caracteristicamente ramos e folhas verticiladas, flores hipocrateriformes, geralmente pequenas, e frutos drupáceos.

⁵⁷ Endress, M.E. & Bruyns, P. 2000. A revised Classification of the Apocynaceae s.l. Bot. Rev. 66: 1-56.

A distribuição do gênero é pantropical e Rao (1956)⁵⁸ sugeriu que dispersão por água corrente e pássaros seriam os responsáveis pela ampla distribuição das espécies.

Em estudos das relações entre as Américas do Norte e do Sul (Graham e Jarzen 1969⁵⁹, Graham 1972⁶⁰, Raven & Axelrod 1974⁶¹) elaborados a partir da análise da flora polínica do Oligoceno superior, da formação Porto Rico, foi sugerido que o gênero deveria ser Sul-americano e pode ter chegado na América Central e nas Antilhas, no Oligoceno superior, cruzando as barreiras d'água via Yucatan e Cuba.

A partir da distribuição atual das espécies é possível observar certos padrões que podem indicar o isolamento do gênero na América do Sul e a posterior dispersão para a América Central. Segundo Burnham & Graham⁶² (1999) um grupo de plantas isolado na América do Sul até o estabelecimento da "conexão de terra" na América Central deveria apresentar um grande número de espécies na América do Sul com poucos táxons, talvez amplamente distribuídos, se estendendo até a América Central. Entre as 37 espécies neotropicais de *Rauvolfia*, 26 são endêmicas à América do Sul, sete são endêmicas à América Central e Antilhas e quatro são amplamente distribuídas, ocorrendo na América do Sul, Central e México e Antilhas.

As quatro espécies, que ocorrem desde a América do Sul até a América Central, México e Antilhas, pertencem a uma seção do gênero definida por caracteres morfológicos peculiares, e, apesar de não haver ainda um estudo da filogenia do gênero usando caracteres moleculares, estas espécies são aparentemente relacionadas.

Existem alguns estudos comparando as características dos nichos de espécies irmãs, que sugerem que os nichos ecológicos tendem a permanecer relativamente constantes ao longo do tempo evolucionário e do espaço (e.g. Peterson *et al.* 1999⁶³). Peterson & Holt⁶⁴ (2003) usaram modelos de distribuição de nicho ecológico e distribuição geográfica de várias espécies e subespécies de pássaros para demonstrar que se podem identificar ambos: nichos ecológicos conservativos e nichos com evolução intra-específica. Eles usaram os modelos para testar a capacidade de previsão do nicho de uma espécie por outra espécie/subespécie relacionada. Alta predição representaria mínima diferenciação do nicho, enquanto baixa predição é consistente com a evolução das características do nicho dentro da espécie.

O trabalho atual tem por objetivo visualizar as similaridades e diferenças de cada espécie da seção típica de *Rauvolfia* no espaço ecológico, ou seja, a similaridade entre os nichos ecológicos das espécies e as relações de afinidade entre as mesmas.

⁵⁸ Rao, A. S. 1956. A revision of *Rauvolfia* with particular reference to the American species. *Ann. Missouri Bot. Gard.* 43(3): 253-355

⁵⁹ Graham, A. & Jarzen, D. J. 1969. Studies in Neotropical Paleobotany. I. The Oligocene communities of Puerto Rico. *Ann. Missouri Bot. Gard.* 56: 308-357

⁶⁰ Graham, A. 1972. Some aspects of tertiary vegetational history about Caribbean Basin. *Mem. Symp. I Congr. Latinoamer. Bot.* 97-117

⁶¹ Raven, P. H. & Axelrod, D. I. 1974. Angiosperm biogeography and past continental movements. *Ann. Missouri Bot. Gard.* 61(3): 539-673

⁶² Burnham, R. J. & Graham, A. 1999. The history of neotropical vegetation; new developments and status. *Ann. Missouri Bot. Gard.* 86(2): 546-589

⁶³ Peterson, A.T., Soberón, J. & Sánchez-Cordero, V. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285: 1265-1267

⁶⁴ Peterson, A.T. & Holt, R.D. 2003. Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. *Ecology Letters*, 6:774-782

a. ESTUDOS DE CASO I – AFINIDADES ENTRE AS ESPÉCIES DA SEÇÃO E ROTAS DE OCUPAÇÃO

A modelagem de distribuição potencial das espécies da seção típica do gênero *Rauvolfia*, e análises de similaridade ecológica entre as mesmas, poderiam nos dar pistas a respeito das relações de parentesco entre as mesmas? A projeção dos modelos para os cenários do Pleistoceno poderia auxiliar no entendimento das rotas de ocupação e fatores que contribuíram para o padrão de distribuição atual?

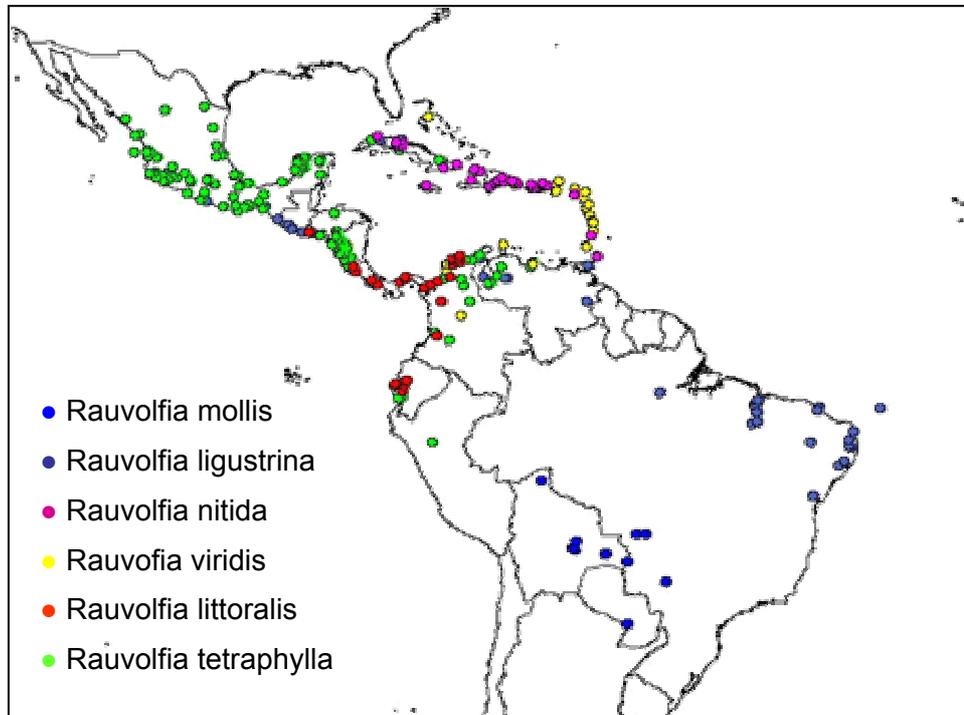


Figura 54. Dados de ocorrência das espécies da seção típica do gênero *Rauvolfia*.

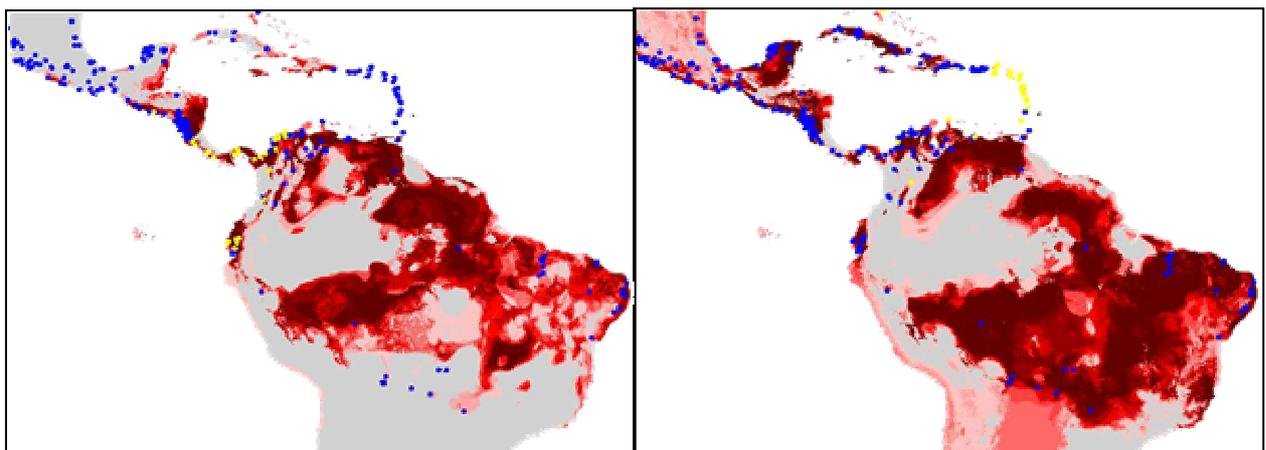


Figura 55. Resultados iniciais da modelagem de distribuição potencial das espécies da seção *Rauvolfia* mostrando a similaridade entre os nichos – a) *Rauvolfia littoralis* (pontos amarelos) prevê área de distribuição de parte das outras espécies (pontos azuis); b) *R. nitida* prevê área de distribuição de todas as outras espécies.

b. ESTUDO DE CASO II RAUVOLFIA LIGUSTRINA X R. MOLLIS

Rauvolfia ligustrina e *R. mollis* são separadas por características morfológicas pouco conservativas, como a presença de indumento na face abaxial das folhas. Analisando este caráter em um grande número de indivíduos de *R. ligustrina* foi possível detectar também diferenças entre os indivíduos que ocorrem na América Central e aqueles da região leste da América do Sul. *R. mollis* entretanto, apresenta-se distribuída em área disjunta do restante dos indivíduos de *R. ligustrina* e isso têm contribuído para a manutenção de seu status de espécies.

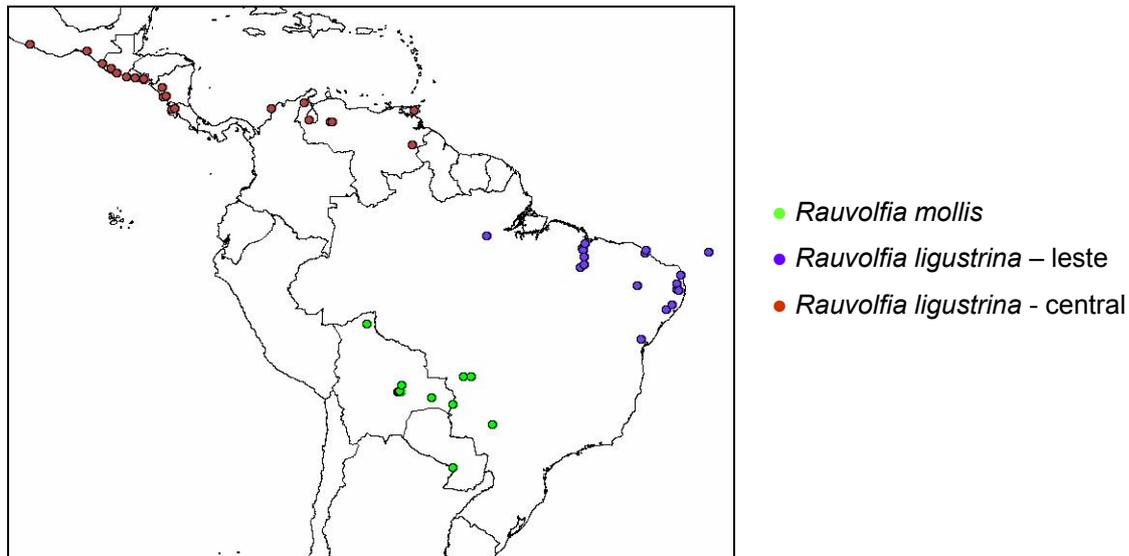
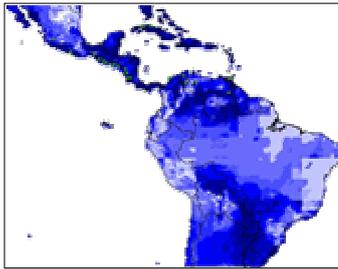


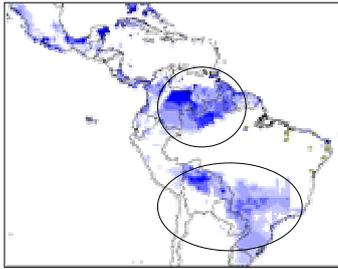
Figura 56. Pontos de ocorrência de subgrupos de *Rauvolfia ligustrina*.

Acreditamos que *R. mollis* é na verdade uma subespécie de *R. ligustrina* e que o padrão de distribuição disjunta se deve ao fato de esta ter tido uma distribuição mais ampla no passado, com posterior isolamento de suas populações, provavelmente ocasionadas por mudanças na disponibilização do nicho de ocupação. Tratamos também os indivíduos da região leste da América do Sul como uma possível subespécie considerando aspectos de distribuição e do indumento.

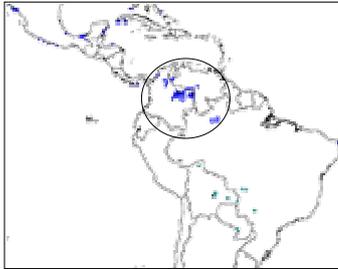
***Rauvolfia ligustrina* - Distribuição prevista para 21.000 anos atrás (Pleistoceno)**



Análise feita a partir de pontos de ocorrência da América Central



Análise feita a partir de pontos de ocorrência da região leste da América do Sul



Análise feita a partir de pontos de ocorrência de *R. mollis*

Figura 57. Resultados de modelagens dos subgrupos de *Rauvolfia ligustrina* com projeções para o período Pleistoceno (21.000 anos atrás).

A pergunta que queremos responder é se a projeção dos modelos destes subgrupos para cenários do Pleistoceno poderia auxiliar no entendimento das disjunções e no estabelecimento das relações taxonômicas entre os mesmos. Trata-se de um trabalho de pós-doc que será publicado em breve.

Os trabalhos realizados nos últimos três anos, com raras exceções, utilizaram o *DesktopGarp*. As próximas etapas deverão trabalhar com o *openModeller* e aprofundar as pré e pós-análises dos dados e modelos gerados e o teste e comparação de vários algoritmos, entre eles o próprio GARP.

8.2.7 CONTROLE BIOLÓGICO DE *TETRANYCHUS EVANSI* (ACARI: TETRANYCHIDAE) NA ÁFRICA.

O ácaro vermelho *Tetranychus evansi* (Baker and Pritchard) é uma espécie invasora que ataca plantações de tomate no Leste e Sueste Africano, causando grandes prejuízos econômicos. O ácaro é provavelmente nativo da América do Sul e foi introduzido no continente africano.

Esse trabalho de modelagem tem por objetivo identificar áreas na América do Sul similares às invadidas na África para busca de inimigos naturais do ácaro vermelho e possível aplicação de controle biológico na África.

Utilizamos o Desktop-GARP (*Genetic Algorithm for Rule-set Production*) e camadas ambientais de clima e topografia. Os dados de invasão foram coletados no Zimbábue e Kenya.

Os modelos determinaram áreas prioritárias no sul do Brasil, Argentina, Paraguai e Argentina. Isso causou mudanças no projeto de busca de inimigos naturais, pois os esforços estavam direcionados somente para o Nordeste brasileiro.

Foram encontradas 5 espécies de ácaros que controlam *Tetranychus evansi* nas áreas indicadas pelo modelo. Essas espécies atualmente estão sendo estudadas para testar sua eficiência na aplicação em lavouras.

Esse projeto foi desenvolvido através de uma parceria do CRIA com Kenyatta University – Nairobi/ Kenya, ESALQ/USP e *International Center of Insect Physiology and Ecology* (ICIPE)- Nairobi/ Kenya. O trabalho foi submetido à revista *Biological Control*.

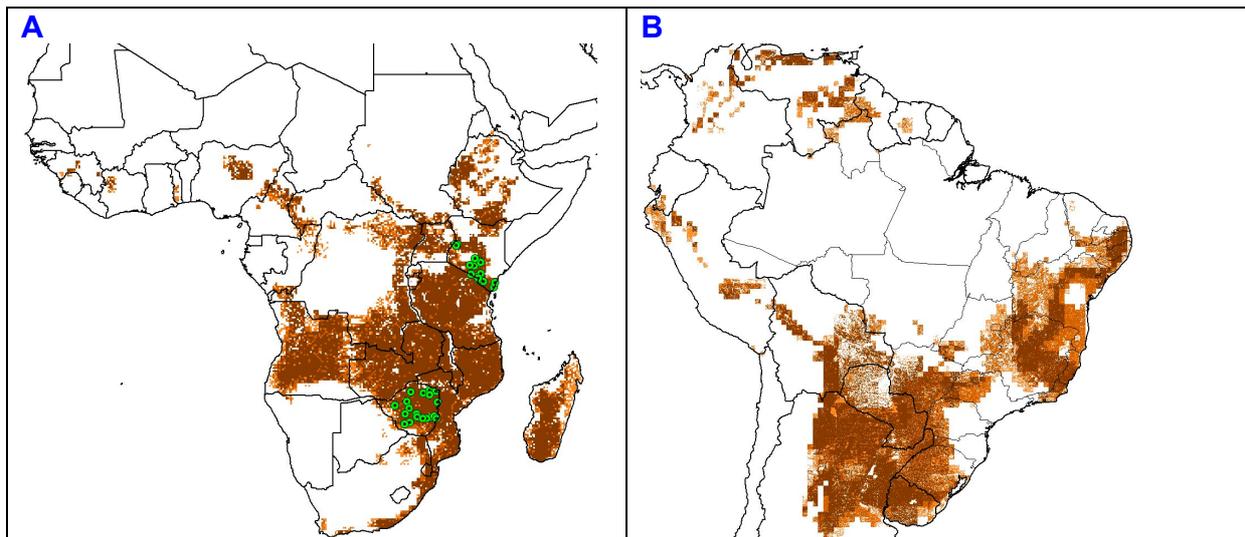


Figura 58. Resultado da modelagem de distribuição potencial da espécie invasora *Tetranychus evansi*. A) modelo de distribuição na região invadida e os pontos de ocorrência no Kenya e Zimbabwe. B) Modelo na área nativa, identificando uma grande área ao sul que não era esperada pelos pesquisadores. Nessa área foram encontradas 5 espécies controle.

8.2.8 DISTRIBUIÇÃO DE *PARIDES ASCANIUS* NO BRAZIL: URBANIZAÇÃO, ÁREAS PROTEGIDAS E CONSERVAÇÃO DE UMA ESPÉCIE AMEAÇADA DE BORBOLETA

Nesse trabalho foi estudada a distribuição geográfica potencial da espécie ameaçada Fluminense swallowtail, *Parides ascanius* (Cramer, 1775), relacionando a distribuição com a presença de urbanização e áreas protegidas. O modelo de distribuição potencial foi realizado utilizando GARP.

O resultado mostrou alta probabilidade de ocorrência para o Estado do Rio de Janeiro, com uma área contínua de 2.038.253 ha ao longo da costa. Porém, cerca de 17.8% da área prevista está dentro de manchas urbanas com alto índice de edificação.

Somente 8,74% (178.187 ha) do que restou da área prevista (sem urbanização) foi sobreposta com áreas de conservação (16 unidades ao todo). Algumas dessas áreas (5) se sobrepõem com a área urbana, pois são APAs (área de proteção ambiental) e permitem certo grau de ocupação (não são restritivas). Isso resulta em uma perda de 58.751 ha da área de proteção.

O modelo de distribuição potencial mostrou ser útil para identificar áreas onde a espécie pode ser re-introduzida e também para auxiliar na prevenção de perda de áreas naturais. Esse projeto foi uma parceria do CRIA com o Instituto de Biologia da Unicamp. O manuscrito foi submetido à revista *Diversity and Distribution*.

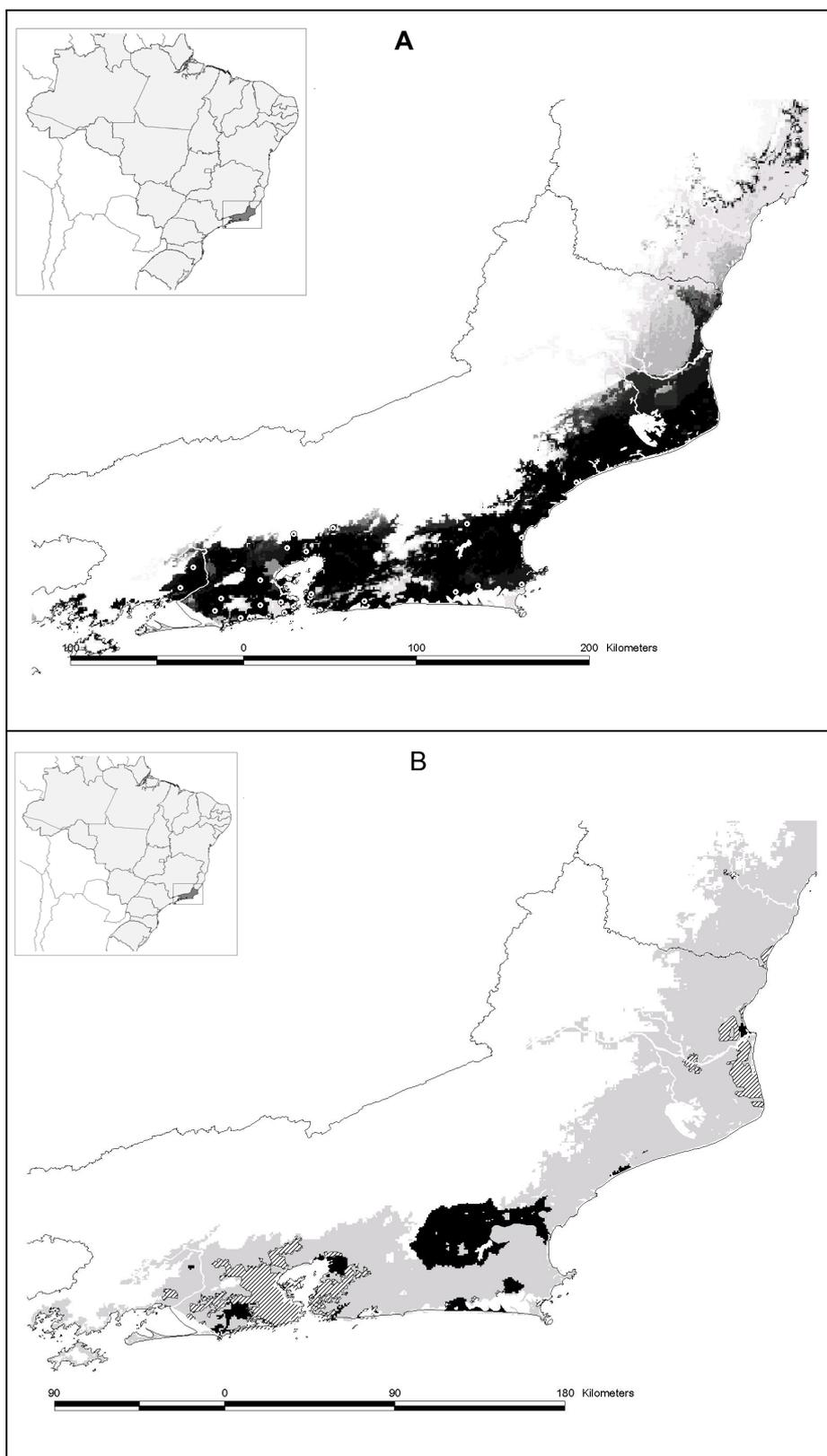


Figura 59. A) Distribuição potencial de *P. ascanius* e pontos de ocorrência utilizados na elaboração do modelo. B) A distribuição potencial se sobrepôs somente 16 unidades conservação (em preto), em um total de 46 estudadas. A área urbana dentro da área prevista está representada no mapa (área hachurada)

8.2.9 PREVENDO A INVASÃO DA ESPÉCIE INVASORA *CROTALARIA PALLIDA* NAS UNIDADES DE CONSERVAÇÃO DO BRASIL.

Espécies invasoras estão entre as ameaças mais importantes para a conservação da biodiversidade. Nesse trabalho foi elaborado o modelo de distribuição da espécie invasora *Crotalaria pallida* e analisado o grau de ameaça das unidades de conservação do Brasil.

O modelo de distribuição potencial foi desenvolvido utilizando GARP (Genetic Algorithm for Rule-set Prediction). As áreas previstas para a região nativa (África) e suas projeções na América do Sul foram testadas e tiveram grande sucesso ao prever pontos já conhecidos de invasão.

Nossos resultados mostraram que todas as unidades de conservação do Brasil possuem um alto risco de invasão por *C. pallida*. As regiões mais abertas, como cerrado e pantanal, similares à savana africana, possuem um risco maior de invasão. Também foi verificado que a Mata Atlântica, devido ao seu processo de fragmentação possui um alto risco. Algumas regiões do Brasil, como a região amazônica, possuem um baixo risco de invasão.

Manejo e pesquisas de áreas prioritárias serão essenciais para evitar uma grande invasão das áreas protegidas no Brasil.

Esse projeto foi realizado através de uma parceria do CRIA com UNICAMP e Universidade do Kansas. O trabalho foi aceito na revista *Weed Science*.

8.2.10 INVASÃO DE *CALLITHRIX JACCHUS* AND *C. PENICILLATA*: CONTAMINAÇÃO GENÉTICA DO GÊNERO *CALLITHRIX* E DECLÍNIO DE ESPÉCIES ENDÊMICAS DA MATA ATLÂNTICA NO BRASIL

Duas espécies de micos (*Callithrix jacchus* e *C. penicillata*) têm sido intencionalmente introduzidas em todo sudeste brasileiro. Ambas espécies de micos são nativas do nordeste brasileiro e são ilegalmente comercializadas como animais de estimação.

As duas espécies hibridizam e competem por recursos com outras 4 espécies endêmicas e ameaçadas. Nesse trabalho foi utilizado Desktop GARP para gerar modelos de distribuição potencial para *C. jacchus* e *C. penicillata* e prever a invasão na América do Sul.

Os resultados indicam que a ocupação dessas espécies pode ter efeito catastrófico nas espécies nativas do sudeste (através de contaminação genética), em outros primatas endêmicos (competição e doenças) e em suas presas (pássaros e ovos).

Esse projeto foi realizado através de uma parceria do CRIA com UNICAMP e UNESP- Rio Claro.

9. MANUTENÇÃO DO SISTEMA

9.1. INFRA-ESTRUTURA COMPUTACIONAL

O Centro de Processamento de Dados do CRIA dá suporte a toda a rede *speciesLink*. A rede interna é composta por dois servidores DELL modelo PowerEdge 6600, cada um com 4 processadores Intel Pentium III Xeon, 8GB de memória RAM e um array de discos Dell PowerEdge. Essa infra-estrutura tem uma capacidade total de armazenamento em disco rígido de 1.5 TeraBytes. Todos os volumes do disco são montados usando a tecnologia RAID-5, o que permite total confiabilidade dos dados e proteção completa contra eventuais defeitos em algum dos discos rígidos.

Com o objetivo de tornar o armazenamento dos dados ainda mais confiável, o CRIA possui uma unidade de *backup* em fita modelo Dell/EMC com capacidade para 20 fitas DLT com até 200GB de capacidade cada uma.

Além dos dois servidores principais, o CRIA possui mais quatro servidores menores, utilizados para geração de mapas, recuperação de informação, gerenciamento de serviços

de impressão e testes da equipe de desenvolvimento. Todo o CRIA é interligado por cabeamento de dados estruturado, suportando uma capacidade de transferência de até 100 Mbps. Os servidores principais são interligados por uma sub-rede de fibra óptica que suporta transferências de até 1 Gbps

O CRIA possui uma conexão de dados com o nó de Campinas da rede ANSP (*Academic Network of São Paulo*) formada por um *link* de fibra óptica entre o prédio do CRIA e o Centro de Computação da UNICAMP com a velocidade de 1 Gbps. O nó da UNICAMP está conectado à FAPESP por uma conexão de 155 Mbps. Essa conexão entre o CRIA e a rede ANSP é controlada por dois roteadores Foundry de última geração. A rede interna do CRIA é protegida por um *firewall* instalado nos roteadores e por *firewalls* locais em cada um dos servidores. Todo o tráfego de rede para dentro e para fora do CRIA é examinado para prevenir a entrada de *vírus* de computador e de invasões por agentes externos (*hackers*).

A rede interna é composta por 25 computadores (entre *desktops* e *notebooks*) executando o sistema operacional Microsoft Windows e/ou Linux.

O CPD do CRIA também possui um aparelho de ar condicionado AirSplit 24000BTU para que os servidores sejam mantidos na temperatura ideal de funcionamento e um no-break modelo Prestige 6000, que suporta quedas de tensão elétrica por até 2 horas sem que os servidores saiam do ar.

Todos os servidores do CRIA possuem o sistema Linux instalado.

9.2. INFRA-ESTRUTURA DE SOFTWARE

Todo o software desenvolvido no CRIA funciona sobre o sistema operacional Linux, distribuição RedHat versão 8.0. Os dados do projeto *speciesLink* localizados no CRIA são armazenados em um sistema gerenciador de bancos de dados PostgreSQL e o software foi desenvolvido usando-se as linguagens PHP, Perl e Java. Como protocolos de transferência de dados são utilizados HTTP, SOAP e XML. Todo o sistema faz uso do software Apache como servidor de páginas web.

O *software* dos servidores do CRIA (inclusive sistema operacional) é atualizado periodicamente pela equipe de suporte e por vezes de forma urgente quando uma nova versão possui uma característica útil ao projeto em desenvolvimento, quando a equipe de suporte considera que a instalação de uma nova versão tornará o sistema mais eficiente e principalmente quando é descoberta alguma falha de segurança em algum programa ou parte do sistema operacional. Para este fim, a equipe de suporte monitora diariamente sites de fabricantes de *software* e listas de discussão de falhas de segurança. Em todo o período de execução do projeto *speciesLink*, o CRIA não teve sua segurança computacional ameaçada e não tivemos nenhuma invasão do sistema por elementos mal-intencionados.

O sistema de verificação de vírus e invasões de *hackers* é atualizado diariamente. O *software* dos *desktops* dos pesquisadores é atualizado semanalmente, de acordo com as recomendações dos fabricantes, com o objetivo de eliminar falhas de execução e brechas de segurança.

9.3. ROTINA DE BACKUP E PROTEÇÃO DOS DADOS DO PROJETO

Costuma-se dizer que a principal ferramenta de proteção de dados importantes é uma boa rotina de *backup*. A equipe de suporte do CRIA realiza *backups* diários em todos os servidores do projeto, tanto os servidores localizados no CPD do CRIA como também os servidores localizados em instituições externas (*servidores regionais*).

O *backup* dos servidores centrais é feito através do software ArcServe e através de uma série de *scripts* desenvolvidos pela equipe de suporte. Diariamente é feito um backup diferencial do dia anterior de todos os dados provenientes das bases de dados, arquivos da equipe de desenvolvimento e arquivos do sistema operacional, o qual é guardado em um

disco protegido. O mesmo procedimento é feito com os *servidores regionais*. O procedimento é automático para todos os servidores, tanto locais quanto remotos.

Semanalmente é feito um *backup* completo de todos os servidores, que é armazenado em fitas DLT de alta confiabilidade. O armazenamento é feito em um formato padronizado, que torna a fita passível de ser lida em qualquer equipamento compatível. Os dados remotos dos *servidores regionais* também são guardados em fita, sendo que os mesmos são transmitidos ao CRIA através de um canal de dados seguro que utiliza a conexão Internet.

Mensalmente é realizado um *full backup*, armazenado em fita DLT e mantido fisicamente nas instalações da unidade da Embrapa Informática Agropecuária (CNPTIA) no campus da Unicamp.

9.4. INSTALAÇÃO E MANUTENÇÃO DOS SERVIDORES REGIONAIS

Durante a execução do projeto foram instalados seis servidores regionais, a saber:

- Servidor Regional de Campinas – localizado no CPD do CRIA
- Servidor Regional de Botucatu – Localizado no CPD da UNESP/Botucatu
- Servidor Regional de São Paulo – Localizado no CPD da Poli/USP
- Servidor Regional de Ilha Solteira – Localizado no CPD da UNESP/Ilha Solteira
- Servidor Regional de São José do Rio Preto – Localizado no CPD da UNESP/SJRP
- Servidor Regional de Ribeirão Preto - – Localizado no CPD da USP/Ribeirão

Todos os servidores remotos executam o sistema operacional Linux. Esses servidores remotos são checados quanto à sua integridade, sua conectibilidade à rede e seu tempo de resposta a conexões do projeto. Para essa checagem é utilizado o software de código aberto Nagios e *scripts* desenvolvidos pelo CRIA. O período de verificação é de quinze em quinze minutos.

Quando necessário, o software desses servidores regionais é atualizado remotamente pela equipe do CRIA. As instituições participantes necessitam apenas garantir o acesso remoto a esses computadores e seu funcionamento ininterrupto.

II. INDICADORES

O projeto estabeleceu como indicadores para avaliação e acompanhamento técnico da implementação do projeto: volume, uso e qualidade.

O indicador **volume** refere-se ao aumento progressivo de registros disponibilizados pelas coleções na rede *speciesLink*. Além do número de registros *on-line* o sistema monitora o número de registros geo-referenciados.

O indicador **uso** é o acesso aos servidores web. Os parâmetros que estão sendo monitorados são acessos (*hits*), visitas (*visits*), páginas (*pages*) e uso de banda (*bandwidth*).

O índice **qualidade** foi definido como sendo a capacidade de estabelecer parcerias com os provedores de informação, grupos de pesquisa, instituições e indivíduos que estejam trabalhando com sistemas de informação ou com o desenvolvimento de aplicativos. No relatório final apresentamos também uma breve análise da qualidade dos dados, tendo como base o resultado da análise da ferramenta *data cleaning*.

Por fim, o relatório final traz mais dois indicadores que são **treinamento e divulgação dos resultados do projeto**, que inclui publicações e a participação em eventos científicos.

1. VOLUME

Além de dar continuidade à informatização dos dados de acervos das 12 coleções iniciais, o projeto teve como meta ampliar a rede com novas coleções, de forma a aumentar a abrangência geográfica e taxonômica do sistema de informação *speciesLink*. Nesta última fase foram integrados dados de coleções botânicas (herbários e xilotecas), coleções zoológicas (aracnídeos, insetos, peixes, répteis e anfíbios, mamíferos, aves) e de microrganismos, de forma a fomentar a formação de redes temáticas colaborativas com base na competência taxonômica consolidada no Estado de São Paulo.

A tabela 6 a seguir apresenta os dados sobre os acervos, a evolução do número de registros *on-line* ao longo dos últimos 3 anos, incluindo o número de registros geo-referenciados.

Tabela 6. Evolução do acervo on-line da rede speciesLink

COLEÇÕES BIOLÓGICAS	ACRÔNIMO	ACERVO	REGISTROS ON-LINE				GEO-REF	VARIÇÃO ANUAL	VARIÇÃO 2A. FASE	% ON-LINE	% GEO-REF
			out-03	out-04	out-05	out-05					
HERBÁRIOS	BOTU	24.000			165	161			1%	98%	
	CPMA	2.150			1.882	0			88%	0%	
	ESA	120.000	730	40.356	74.549	19.606	85%	10112%	62%	26%	
	HISA	10.235			718	62			7%	9%	
	HRCB	40.000		2.593	6.541	0	152%		16%	0%	
	HSJRP	28.000			19.377	0			69%	0%	
	IAC	45.000	25.296	31.548	39.770	31.825	26%	57%	88%	80%	
	SP	350.000	12.901	18.302	13.928	3.494	-24%	8%	4%	25%	
	SP - Algae	15.000			14.033	0			94%	0%	
	SPF	133.500	22.769	22.762	10.250	4.064	32%	32%	8%	40%	
SPF-Algae	19.776			19.776	0			100%	0%		
SPSF	34.000			31.168	2	0%		92%	0%		
UEC	134.000	12.860	33.301	36.081	9.957	8%	181%	27%	28%		
JBRJ	450.000		1.383	6.138	0	344%		1%	0%		
Subtotal		1.405.661	74.556	181.413	274.488	69.171	51%	268%	20%	25%	
XILOTECAS	BCTw	19.500	9.127	908	9.127	0	905%	0%	47%	0%	
	SPFw	4.000		908	1.677	80	85%		42%	5%	
	Subtotal	23.500	9.127	1.816	10.804	80	495%	18%	46%	1%	
PEIXES	DZSJRP-Pisces	7.885	5.725	6.888	7.885	6.068	14%	38%	100%	77%	
	LIRP	30.000	4.454	4.454	5.167	3.860	16%	16%	17%	75%	
	MZUSP	84.000		72.714	77.613	53.319	7%		92%	69%	
	MHN-peixes	6.138			6.138	0			100%		
	Subtotal	128.023	10.179	84.056	96.803	63.247	15%	851%	76%	65%	
ARACNÍDEOS											
	DZSJRPAcari	7.026	5.382	5.753	6.296	2.344	9%	17%	90%	37%	
	AcariESALQ	15.732	12.392	12.392	15.732	1.417	27%	27%	100%	9%	
	IBSP-Acari	9.201		4.210	2.771	1.143	-34%		30%	41%	
	UBTU	3.500		2.623	2.626	0	0%		75%	0%	
Subtotal	35.459	17.774	24.978	27.425	4.904	10%	54%	77%	18%		

COLEÇÕES BIOLÓGICAS RÉPTEIS E ANFÍBIOS	ACRÔNIMO	ACERVO	REGISTROS ON-LINE			GEO-REF	VARIÇÃO ANUAL	VARIÇÃO 2A. FASE	% ON-LINE	% GEO-REF
			out-03	out-04	out-05					
Anfíbios	DZS.JRP-Amphibia-adults	7.358		4.823	7.358	7.124	53%		100%	97%
Anfíbios	DZS.JRP-Amphibia-tadpoles	1.758			1.158	0			66%	0%
Anfíbios	CFBH	7.000		2.935	3.773	0	29%		54%	0%
Anfíbios	MHN-anfíbios	12.231			12.231	0			100%	0%
Répteis	MHN-repteis	2.832			2.832	0			100%	0%
Répteis e Anfíbios	IBSP-Herpeto	80.000			66.167	0			83%	0%
Subtotal		111.179	0	7.758	93.519	7.124	1105%		84%	8%
MAMÍFEROS										
Roedores	IAL-Roedores	21.000			12.029	0			57%	0%
Mamíferos	MHN-mamíferos	2.385			2.385	0	-100%			
Morcegos	DZS.JRPChiroptera	10.678		2.829	5.339	815	89%		50%	15%
Subtotal		34.063	0	2.829	19.753	815	598%		58%	4%
INSETOS										
Insetos	MEFEIS	28.000		3.519	3.519	0	0%		13%	0%
Insetos	IBSP-IB	275.000			401	0				
Abelhas	CEPANN	34.171			33.636	31.650			98%	94%
Abelhas	RPSP	171.000			68.599	50.467				
Subtotal		508.171	0	3.519	106.155	82.117			21%	77%
AVES										
	IAL-Aves	110.000			21.157	12.488			19%	59%
	MHN-Aves	2.224			2.224	0			100%	0%
Subtotal		112.224	0	0	23.381	12.488			21%	53%
MICROORGANISMOS										
	CBMAI	688	110	78	314	0	303%	185%	46%	0%
	IBSBF	2.000	929	929	1.750	0	88%	88%	88%	0%
Subtotal		2.688	1.039	1.007	2.064	0	105%	99%	77%	0%
TOTAL ACERVOS		2.360.968	112.675	307.376	654.392	239.946	113%	481%	28%	37%
OBSERVAÇÃO	<i>Sim</i> Biota	63.586	39.402	51.606	63.586	63.584	23%	61%	100%	100%
TOTAL GERAL		2.424.554	152.077	358.982	717.978	303.530	100%	372%	30%	42%

A tabela permite comparar a evolução das diferentes coleções ao longo do projeto. No caso dos herbários e xilotecas podemos destacar o aumento do número de coleções participantes, que passou de 6 na fase 1, para 16 no final do projeto. Do final dos dois primeiros anos até o final do projeto houve um crescimento de 270% no número de registros *on-line*.

No caso de peixes, a rede com 2 coleções na primeira fase, passou a ter 4 na fase final, aumentando o número de registros *on-line* em cerca de 850%. O mesmo acontecendo com os aracnídeos, com um aumento de 54% do número de registros *on-line*.

Répteis, anfíbios, mamíferos, insetos e aves não tinham dados no sistema no final da primeira fase. No caso de microrganismos não foi acrescentada nenhuma coleção no período, mas o acervo disponível *on-line* aumentou 185%.

A tabela 7 apresenta os mesmos dados por grupo taxonômico.

Tabela 7. Porcentagem de registros *on-line* e geo-referenciados por grupo taxonômico

Coleções Biológicas	Acervo Total	Registros <i>On-line</i>	% <i>on-line</i>	Registros geo-refe	% geo-ref
Herbários e Xilotecas	1.429.161	285.292	20%	69.251	24%
Peixes	128.023	96.803	76%	63.247	65%
Aracnídeos	35.459	27.425	77%	4.904	18%
Répteis e Anfíbios	111.179	93.519	84%	7.124	8%
Mamíferos	34.063	19.753	58%	815	4%
Insetos	508.171	106.155	21%	82.117	77%
Aves	112.224	23.381	21%	12.488	53%
Microrganismos	2.688	2.064	77%	0	0%
Total dos Acervos	2.360.968	654.392	28%	239.946	37%
Dados de Observação	63.586	63.586	100%	63.584	100%
Total Geral	2.424.554	717.978	30%	303.530	42%

Estima-se que o Brasil tenha aproximadamente 5 milhões de exsicatas depositadas em seus herbários. Um acervo potencial de cerca de 1,5 milhão de registros na rede *speciesLink* é, portanto, bastante significativo, representando cerca de 30% do acervo nacional. Destacamos o trabalho realizado pelos herbários ESA e IAC. O ESA na digitalização de seu acervo e o IAC no uso das ferramentas disponíveis, atingindo um valor de 80% do total de registros *on-line* geo-referenciados. Esse fato é muito importante porque aumenta a utilidade dos dados. O geo-referenciamento permite, por exemplo, o uso dos dados do acervo em trabalhos de modelagem da distribuição de espécies, na definição de políticas de conservação ou de recuperação de áreas degradadas, entre outros inúmeros exemplos.

Destaque também deve ser dado às duas coleções de abelhas, das quais cerca de 50% dos acervos está *on-line* após apenas um ano de participação no projeto, e com elevado índice de geo-referenciamento (80%).

A figura a seguir procura mostrar graficamente o número total de registros nos acervos, *on-line* e geo-referenciados por grupo taxonômico.

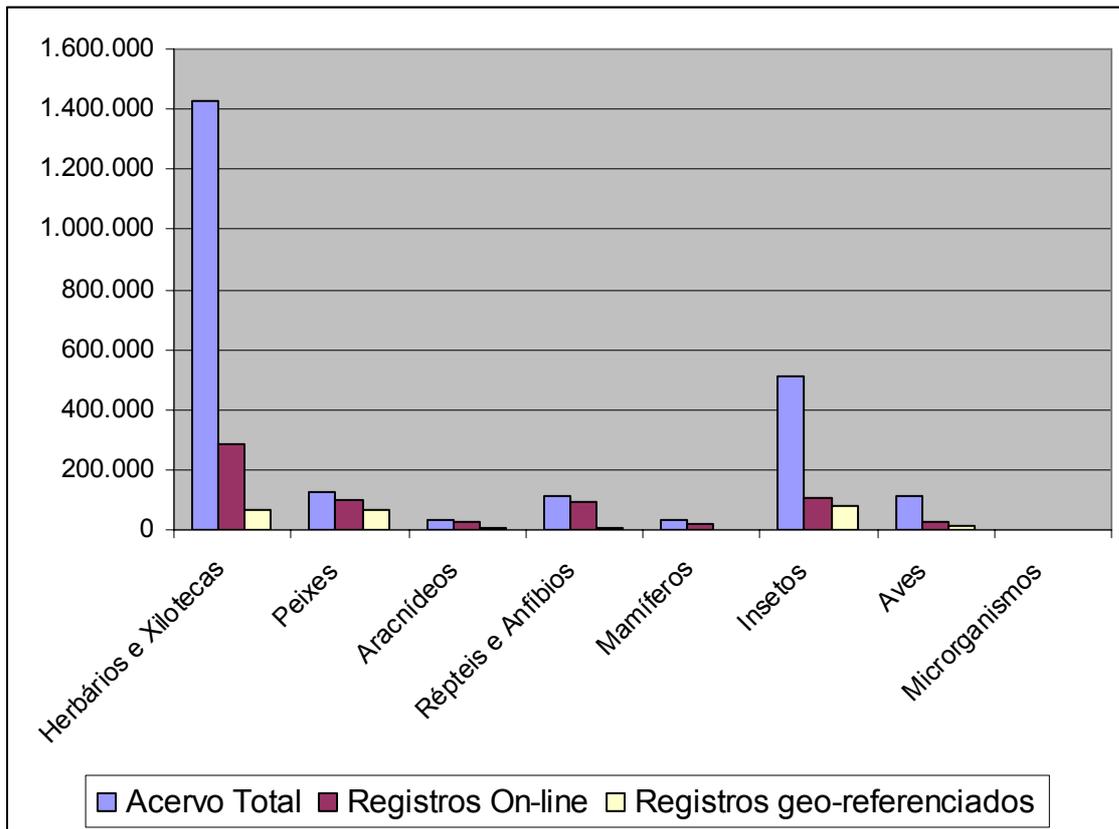


Figura 60. Número total dos acervos, registros *on-line* e geo-referenciados por grupo taxonômico

Podemos observar que o grupo de peixes está comparativamente bem informatizado, 76% com 65% dos registros *on-line* geo-referenciados. Podemos também verificar que os grupos com acervos maiores, como plantas e insetos, têm pela frente um extenso trabalho de digitalização de seus acervos. Também deve-se destacar a inclusão do Jardim Botânico do Rio de Janeiro à rede. Esse fato mostra a escalabilidade do sistema desenvolvido e permite supor como sendo provável a integração de mais coleções de outras regiões do país à rede *speciesLink*.

Em números gerais os registros *on-line* dos acervos das coleções biológicas do Estado de São Paulo aumentaram 113% no último ano e quase quintuplicaram do segundo (final da fase 1) ao último ano do projeto. No mesmo período, o sistema *SinBiota* aumentou 23% e 61% respectivamente.

Estamos fechando o projeto com cerca de 720 mil registros *on-line*, disponíveis para qualquer pessoa com acesso à Internet de forma livre e aberta. 42% desses registros estão geo-referenciados e o número total de registros *on-line* representa cerca de 30% do acervo das coleções participantes da rede.

Esse dado significa que as coleções ainda precisam de apoio para a digitalização de seus acervos.

Recomenda-se que a Fapesp mantenha uma linha de apoio para a digitalização dos acervos das coleções biológicas do Estado de São Paulo.

O volume de dados e informações é monitorado em tempo real através da ferramenta *manager* (item 1.6. desse relatório), importante instrumento de gestão do projeto. Esse sistema apresenta um gráfico que mostra a evolução do número de registros *on-line* das coleções brasileiras.

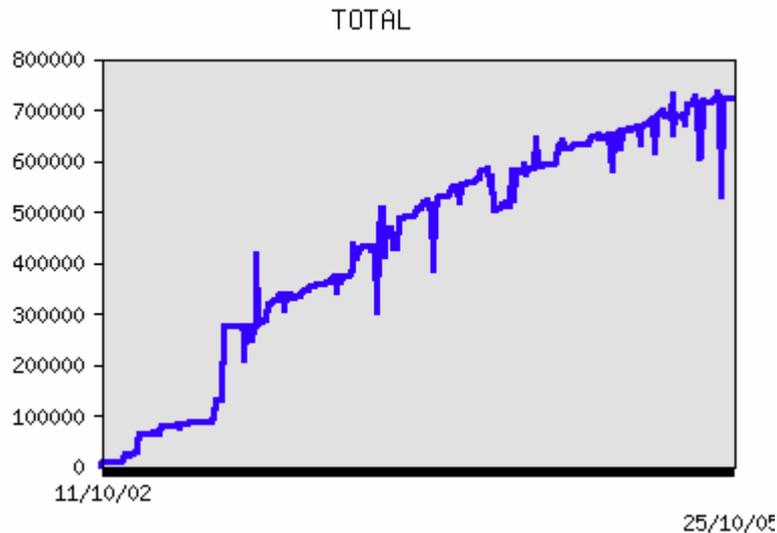


Figura 61. Evolução do número de registros de coleções do país disponíveis na rede *speciesLink*

O gráfico reflete tanto a inserção como a retirada de dados da rede. Isso mostra uma relação dinâmica entre o provedor do dado (as coleções) e o sistema de integração dos dados, a rede *speciesLink*, gerenciada pelo CRIA.

2. Uso

Como o objetivo fundamental do projeto é contribuir para a construção de uma infraestrutura compartilhada de dados de acesso aberto e livre, disponível na Internet, o *uso* é medido pelo acesso ao sistema *on-line*.

Existem vários parâmetros para medir o acesso ao sistema. O parâmetro mais usual, porém com distorções, é *hits*. Por ter sido um dos primeiros parâmetros usados na Internet, é um número que o usuário da rede normalmente compreende a sua magnitude. Estamos também adotando os indicadores “páginas” “visitas” e “uso da banda” para indicar o acesso ao sistema e o volume de troca de dados.

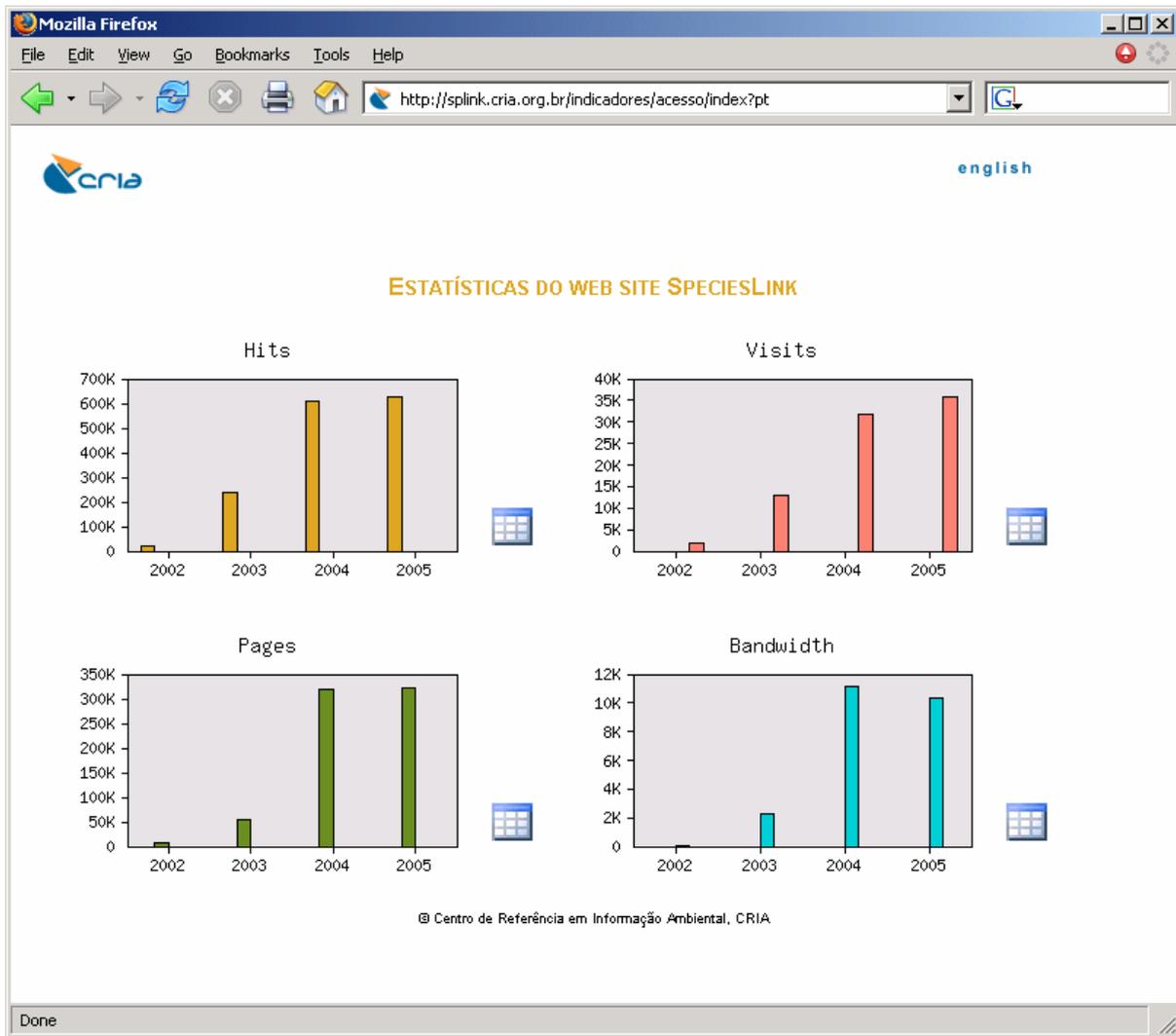


Figura 62. Estatística de acesso ao website *speciesLink*

Como o CRIA é responsável pela manutenção do site do Programa Biota/Fapesp - O Instituto Virtual da Biodiversidade, pela Revista Biota Neotropica, além da rede *speciesLink* integrada ao *SinBiota*, é interessante comparar o acesso a cada sistema de informação do programa.

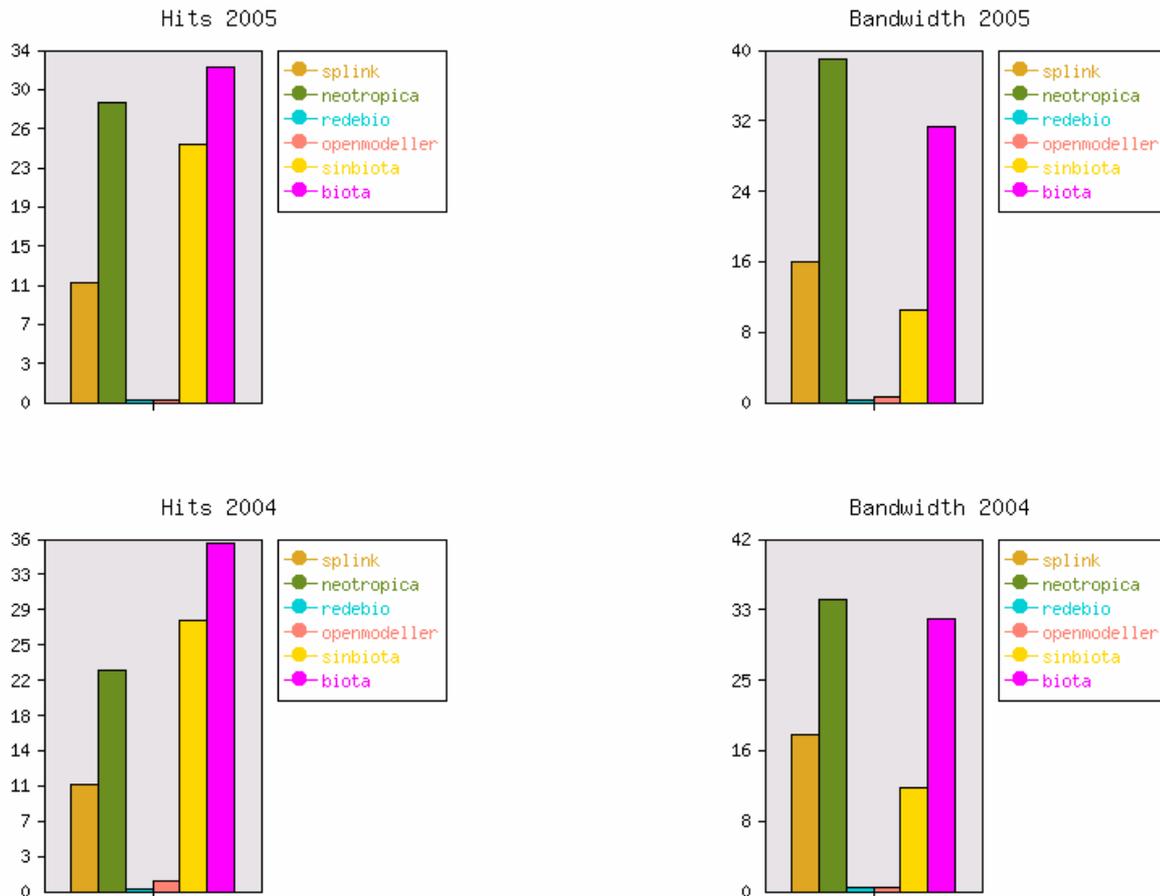


Figura 63. Acesso a cada sistema de informação hospedado no CRIA referente ao Programa Biota/Fapesp nos anos 2004 e 2005 para os critérios *hits* e *bandwidth*

O gráfico mostra que o site do programa Biota/Fapesp é o mais acessado seguindo o critério *hits*, mas em transmissão de dados na rede (*bandwidth*) é a Revista Biota Neotropica. A rede *speciesLink*, na realidade somente lançada em outubro de 2004, começa a ter destaque principalmente no critério *bandwidth*. O uso da banda é um parâmetro importante uma vez que temos por objetivo criar uma infra-estrutura de dados para uso amplo.

3. QUALIDADE

No início do desenvolvimento dos trabalhos, a capacidade de articulação, de estabelecimento de parcerias, foi definido como sendo um dos indicadores de qualidade do projeto. O CRIA é um mecanismo facilitador na disseminação de dados *on-line*. A parceria com o provedor de dados é essencial. O CRIA é também uma instituição pequena, que se mantém através de projetos e não tem apoio governamental para a sua manutenção. Depende exclusivamente da contratação de projetos e da venda de serviços especializados. Não pode, portanto, ter uma equipe grande e novamente depende do estabelecimento de parcerias para expandir a sua capacidade de desenvolvimento.

Um segundo fator de qualidade foi incluído neste relatório: a qualidade dos dados *on-line*.

3.1. DESENVOLVIMENTO DE PARCERIAS

Como o CRIA está desenvolvendo um mecanismo facilitador para disseminar dados primários dos acervos de coleções e museus do Estado de São Paulo acreditamos que um indicador de qualidade é a capacidade de interagir com grupos de excelência. Podemos

dividir esses grupos em duas categorias: o provedor de dados (as coleções biológicas) e os parceiros no desenvolvimento do sistema.

3.1.1 PROVEDORES DE DADOS

Em setembro de 2003 concluímos a primeira fase com 12 coleções participantes. Em apenas um ano dobramos este valor para o Estado de São Paulo e iniciamos o processo de ampliação da rede para outros estados, com a participação do Jardim Botânico do Rio de Janeiro. Concluímos o projeto com parcerias estabelecidas com 41 coleções pertencentes às 3 Universidades estaduais, a sete institutos de pesquisa do Estado de São Paulo e ao Jardim Botânico do Rio de Janeiro. São elas:

1. Universidade Estadual Paulista "Julio de Mesquita Filho", Unesp

- Campus de Rio Claro
 - Coleção "Célio F. B. Haddad" (CFBH), Instituto de Biociências
 - Herbário Rioclarense (HRCB) - Instituto de Biociências, UNESP, Campus de Rio Claro, SP
- Campus de Botucatu
 - Coleção Científica de Aranhas (Araneae) do Depto. de Zoologia da Unesp, Campus Botucatu (UBTU)
 - Herbário "Irina Delanova Gemtchújnicov" (BOTU) - Instituto de Biociências, UNESP, Campus de Botucatu, SP
- Campus São José do Rio Preto
 - Coleção de Peixes (DZSJRP-Pisces) - Unesp, IBILCE/Unesp
 - Coleção de Quirópteros (DZSJRP-Chiroptera) - Unesp, IBILCE/Unesp
 - Coleção de Anfíbios (DZSJRP-Amphibia-adults)
 - Coleção de Girinos (DZSJRP-Amphibia-tadpoles)
 - Herbário de São José do Rio Preto (HSJRP)
 - Coleção de Ácaros (DZSJRP-Acari)
- Campus Ilha Solteira
 - Coleção do Museu de Entomologia da FEIS/Unesp (MEFEIS) - FEIS/Unesp
 - Herbário de Ilha Solteira (HISA)

2. Unicamp

- Museu de História Natural "Prof. Dr. Adão José Cardoso"
 - Coleção de Anfíbios (MHN-anfibios)
 - Coleção de Aves (MHN-aves)
 - Coleção de Mamíferos (MHN-mamiferos)
 - Coleção de Peixes (MHN-peixes)
 - Coleção de Répteis (MHN-repteis)
- Herbário da Universidade Estadual de Campinas (UEC)
- Centro Pluri-disciplinar de Pesquisas Químicas, Biológicas e Agrícolas CPQBA
 - Coleção de plantas medicinais e aromáticas (CPMA)
 - Coleção Brasileira de Microrganismos de Ambiente e Indústria (CBMAI)

3. Universidade de São Paulo

- Campus de Ribeirão Preto
 - Coleção Camargo (RPSP)
 - Coleção de Peixes do Laboratório de Ictiologia de Ribeirão Preto (LIRP)
- Campus São Paulo

- Coleção de Peixes do Museu de Zoologia da USP (MZUSP)
 - Coleção do Laboratório de Abelhas do IB/USP (CEPANN)
 - Herbário do Departamento de Botânica, SPF-IB/USP (SPF)
 - Coleção de Algas Departamento de Botânica (SPF-Algae)
 - Xiloteca do Instituto de Biociências da Universidade de São Paulo (SPFw)
 - Campus Piracicaba (ESALQ)
 - Herbário da Escola Superior de Agricultura Luiz de Queiroz (ESA)
 - Coleção de Ácaros do Departamento de Entomologia, Fitopatologia e Zoologia (AcariESALQ) - ESALQ/USP
- 4. Institutos de Pesquisa do Estado de São Paulo**
- Instituto Butantan:
 - Coleção Acarológica (IBSP-Acari)
 - Coleção Herpetológica "Alphonse Richard Hoge" (IBSP-Herpeto)
 - Instituto Biológico
 - Coleção Entomológica "Adolph Hempel" (IBSP-IB)
 - Coleção de Culturas de Fitobactérias (IBSBF)
 - Instituto Adolfo Lutz
 - Coleção Zoológica de Referência da Seção de Vírus Transmitidos por Artrópodos (IAL-roedores)
 - Coleção Zoológica de Referência da Seção de Vírus Transmitidos por Artrópodos - Banco de Aves (IAL-aves)
 - Instituto de Botânica, IBt
 - Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo"
 - Coleção de Fanerógamas (SP)
 - Coleção de Algas (SP-Algae)
 - Instituto Florestal de São Paulo
 - Herbário Dom Bento Pickel (SPSF) -
 - Instituto Agrônomo de Campinas
 - Herbário do Instituto Agrônomo de Campinas (IAC) -
 - Instituto de Pesquisas Tecnológicas, IPT
 - Xiloteca Calvino Mainieri (BCTw)
- 5. Institutos de Pesquisa do Estado do Rio de Janeiro**
- Instituto de Pesquisas Jardim Botânico do Rio de Janeiro
 - Herbário Dimitri Sucre Benjamin (JBRJ)

3.1.2 PARCEIROS TÉCNICOS

O fato de trabalhar com software livre e open source propicia o estabelecimento de parcerias técnicas para o desenvolvimento conjunto de aplicativos. Talvez o principal componente nesse projeto foi o protocolo de comunicação DiGIR. Foram parceiros no desenvolvimento dos trabalhos as seguintes instituições:

- Global Biodiversity Information Facility - GBIF
- Taxonomic Database Working Group – TDWG
- The University of Kansas
- California Academy of Sciences
- Biological Collection Access Service for Europe – BioCASE

- Botanic Garden and Botanical Museum Berlin-Dahlem

No caso do desenvolvimento do openModeller são parceiros:

- The University of Reading (BDWorld project)
- The University of Kansas (BTRA project)
- QGIS project (Tim Sutton, Jens Oberender, Tom Elwertowski)
- Escola Politécnica, USP
- Instituto Nacional de Pesquisas Espaciais (INPE)

3.2. QUALIDADE DOS DADOS

Não era objetivo do projeto trabalhar com a questão da qualidade dos dados. No entanto, com a estada do Dr. Arthur Chapman durante um ano no CRIA, dada a sua competência na área, foram desenvolvidas ferramentas para facilitar a identificação e visualização de erros.

Como não existe um Catálogo de Nomes do Brasil, possíveis erros de digitação, são identificados comparando os nomes com catálogos internacionais (p.ex. o Catálogo da Vida)⁶⁵ e através de buscas fonéticas, apresentando nomes foneticamente iguais com grafias diferentes.

A tabela 8 a seguir mostra o resultado do *data cleaning* para todos os registros das coleções do estado de São Paulo, o Jardim Botânico do Rio de Janeiro e o SinBiota juntos. Nessa tabela o número já supera 720 mil registros por que a coleta de dados foi realizada em data diferente das demais análises apresentadas nesse relatório.

Tabela 8. Relação de alguns dos critérios analisados no *data cleaning* para o conjunto de dados de todas as coleções do país participantes da rede speciesLink

CRITÉRIO	NÚMERO	%
CARACTERIZAÇÃO DOS DADOS		
Número Total de Registros <i>On-line</i>	721.300	100%
Número Total de Registros com Geo-referenciamento	316.357	44%
Número Total de Registros sem Geo-referenciamento	404.943	56%
Geo-referenciamento sugerido para registros brasileiros sem geo-ref (precisão = Município)	247.803	61%
registros no mar	15.888	5%
total de tipos	23.783	3%
REGISTROS SUSPEITOS		
Possível duplicação do registro		
registros inteiramente duplicados	49.762	7%
registros com número de catálogo duplicado	132.132	18%
Possível erro de digitação do nome ou falta de padronização (autor)		
autores suspeitos	68.648	10%
famílias suspeitas	9.179	1%
generos suspeitos	24.539	3%
espécies suspeitas	48.551	7%
subespecies suspeitas	3.683	1%
Total (sem considerar "autor")	85.952	12%
Problemas com a localização geográfica da coleta		
nome de países suspeitos	3.549	0%
sugestões para nomes de países	334	0%
nome de municípios suspeitos (Brasil)	75.269	10%

⁶⁵ <http://www.species2000.org/2005/search.php>

CRITÉRIO	NÚMERO	%
sugestões para nomes de municípios (Brasil)	12.752	2%
long/lat iguais	3	0%
long ou lat igual a zero	289	0%
fora do limite geográfico mundial (long -180/180 e lat -90/90)	7	0%

Na caracterização dos dados podemos observar que 44% do acervo *on-line* está geo-referenciado. Tem-se ainda que para 56% (405 mil) dos registros sem coordenadas geográficas, o sistema pode oferecer valores de lat/long com precisão do município para 61% (247 mil). Isso elevaria o número de registros geo-referenciados para cerca de 78%, garantindo a precisão no nível de município. As características dos dados também apresentam valores como 5% dos registros geo-referenciados são do mar. Esse número indica que ou as coleções com acervos do mar não estão bem representadas na rede ou existem poucas coletas nesse meio ambiente. Um outro dado interessante é o número de tipos, 3% do total de registros *on-line*. Esse número teria que ser melhor analisado uma vez que o sistema apenas verifica o número de registros onde esse campo está preenchido.

Com relação à nomenclatura, observa-se que um grande problema é a falta de padronização do nome do autor. 10% dos registros apresentam o nome do autor como sendo "suspeito". Somando todos os registros suspeitos para os nomes de família, gênero, espécie e subespécie tem-se que 12% do acervo *on-line* precisariam ser melhor analisados. É importante deixar claro que não se trata de nenhum mecanismo de validação taxonômica. O sistema apenas está procurando por erros de digitação.

Por fim, com respeito aos dados da localização geográfica, aparentemente o maior problema está no nome do município. Mais uma vez seria importante analisar os dados antes de tirar qualquer conclusão. Pode ser de fato erro no registro do dado ou problema, por exemplo, do datum.

4. FORMAÇÃO DE RECURSOS HUMANOS

O projeto contou com um número total de 25 bolsistas para trabalhar na organização e digitação dos acervos, a saber:

- Alan Rosendo, Bolsa: TT1, Coleção: BCTw
- Alex Franceschet, Bolsa: TT3, Coleção: MZUSP
- Anna Carolina Pastorek, Bolsa: TT3, Coleção: IAC
- Bruno Leite Tacarambí, Bolsa: TT2, Coleção: UEC
- Carla Fernanda Nardin, Bolsa: TT1, Coleção: IAC
- Carolina Soares latesta, Bolsa: TT1, Coleção: IAC
- Cesar Augusto Falcetti, Bolsa: TT3, Coleção: IBSP-Herpeto
- Denise de Araujo Alves, Bolsa: TT3, Coleção: CEPANN
- Eduardo Rodrigo Oliveira da Silva, Bolsa: TT3, Coleção: DZSJRP-Acari
- Eliana Yuriko Miyada, Bolsa: TT3, Coleção: MEFEIS
- Ingrid Koch, Bolsa: Pós-Doc, Coleção: UEC
- Janete Moscardi, Bolsa: TT3, Coleção: UEC
- Juliana Venâncio Oliveira, Bolsa: TT1, Coleção: RPSP
- Kátia Freire da Silva, Bolsa: TT3, Coleção: SP
- Luciana Bastos de Queiroz Lima, Bolsa: TT3, Coleção: IAL
- Luciano Luna Rodrigues, Bolsa: TT3, Coleção: BCTw
- Luiz Felipe Mania, Bolsa: TT1, Coleção: HRCB

- Marcela Firens da Silveira, Bolsa: TT1, Coleção: ESA
- Marcelo Rodrigo Pace, Bolsa: TT3, Coleção: SPFW
- Marcos Yoshihiro Yokoi, Bolsa: TT3, Coleção: MZUSP
- Mariana Mitiko Suzuki, Bolsa: TT3, Coleção: MEFEIS
- Priscila Bongiorno, Bolsa: TT3, Coleção: SPSF
- Reginaldo Almazan, Bolsa: TT3, Coleção: IAL
- Sérgio Morbiolo, Bolsa: TT3, Coleção: UEC
- Thaís Maestri Rodrigues, Bolsa: TT3, Coleção: RPSP
- Thiago Motta de Oliveira, Bolsa: TT3, Coleção: DZSJRP-Amphibia
- Viviane Yuri Jono, Bolsa: TT1, Coleção: SPFW

Contou também com uma bolsista pós-doc, Dra. Ingrid Kock, que colaborou no desenvolvimento das ferramentas de *data cleaning* e no uso dos aplicativos de modelagem.

Destaque deve ser dado a Marinez Ferreira de Siqueira, do CRIA, que no período fez o seu doutorado em Ciências da Engenharia Ambiental junto ao Centro de Recursos Hídricos e Ecologia Aplicada – Escola de Engenharia de São Carlos – USP defendendo a tese *Uso de modelagem de nicho fundamental na avaliação do padrão de distribuição geográfica de espécies vegetais*.

Por fim, ainda dentro do critério **treinamento** a equipe do CRIA organizou vários workshops onde procurou promover a necessidade de se estruturar no país uma infra-estrutura compartilhada de dados sobre espécies e espécimes de acesso livre e aberto. Esses eventos reuniram mais de 400 especialistas e técnicos.

1. Workshop: Diretrizes e Estratégias para a Modernização de Coleções Biológicas Brasileiras e a Consolidação de Sistemas Integrados de Informação sobre Biodiversidade⁶⁶

Realizado nos dias 05 e 06 de julho de 2005, em Brasília, DF, no Centro de Gestão e Estudos Estratégicos (CGEE) teve como objetivo central contribuir para o estabelecimento de um plano estratégico para a consolidação de uma rede integrada de coleções biológicas no Brasil.

Os objetivos específicos:

- Fazer uma análise crítica das profundas transformações que estão ocorrendo nas áreas de conhecimento associadas ao gerenciamento e manutenção de coleções biológicas, sistemática em biologia, e informática para biodiversidade.
- Fazer recomendações que levem à ampliação da capacidade do governo e sociedade em responder rapidamente aos desafios associados ao uso dos recursos naturais e seus impactos à biodiversidade.
- Recomendar diretrizes e estratégias para a modernização e consolidação de uma rede integrada de coleções biológicas, associada a uma infra-estrutura compartilhada de dados e informações sobre biodiversidade

Foram produzidos documentos setoriais e notas técnicas⁶⁷ para subsidiar os trabalhos nas áreas de botânica, zoológica, microbiologia e informação. Tendo os documentos como base, foi realizado um seminário com especialistas para discutir possíveis diretrizes e estratégias. O documento consolidado foi apresentado e discutido no workshop com a participação de 77 especialistas do país e do exterior.

⁶⁶ <http://www.cria.org.br/cgee/col/>

⁶⁷ <http://www.cria.org.br/cgee/col/documentos>

2. Workshop: Centros de Recursos Biológicos e Avaliação da Conformidade de Material Biológico⁶⁸

Realizado no período de 16 a 20 de maio de 2005 em Campinas, SP, no Instituto de Tecnologia de Alimentos – ITAL, contou com o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPQ, Centro Brasileiro Argentino de Biotecnologia – CBAB e do Ministério de Ciência e Tecnologia.

O workshop teve como objetivos apresentar os conceitos básicos de sistemas de gestão da qualidade em organizações ligadas à preservação e distribuição de material biológico para atividades e aplicações industrial, bem como dos conceitos de conformidade de material biológico, sistemas de propagação/produção e preservação, controles do processo e distribuição, aspectos de conformidade, segregação, rastreabilidade e biossegurança.

Apresentações da equipe do CRIA:

- Vanderlei P. Canhos: "Coleções Microbiológicas, Centros de Recursos Biológicos e Conformidade do Material Biológico" e "De Coleções de Culturas a Centros de Recursos Biológicos, Iniciativas Internacionais"
- Dora A. L. Canhos: "Sistemas de Informação sobre Recursos Biológicos: CABRI, SICol e speciesLink"
- Renato De Giovanni: "Padrões e Protocolos em Informática para Biodiversidade"

Participaram 16 alunos e 6 ouvintes de várias coleções biológicas do Brasil.

3. Inter-American Workshop on Environmental Data Access⁶⁹

Esse workshop foi realizado nos dias 03 a 06 de março de 2004, em Campinas, SP e foi fruto de uma colaboração entre a Academia Brasileira de Ciências (ABC), Codata (*Committee on Data for Science and Technology*), Academia Americana de Ciências e o CRIA com apoio da Fapesp, da Academia Brasileira de Ciência (ABC), ICSU (*International Council for Science*) e NSF (*National Science Foundation*).

O workshop teve por objetivo reunir especialistas em dados ambientais, biodiversidade, bancos de dados e sistemas de informação *on-line* com formuladores de políticas e agências de fomento para identificar e discutir:

- Aspectos científicos, técnicos e institucionais a respeito a práticas e políticas de acesso a dados ambientais;
- Oportunidades e barreiras para o acesso a dados ambientais; e
- Ações específicas que podem ser implementadas para melhorar as políticas e a cooperação regional no que tange o acesso a dados ambientais.

O relatório final do evento está disponível no endereço:

<http://www.cria.org.br/eventos/iaed/WorkshopFinalReport.pdf>.

A equipe do CRIA apresentou os seguintes trabalhos:

- *speciesLink: a system for integrating distributed primary biodiversity data.* Vanderlei P. Canhos, Renato De Giovanni e Sidnei de Souza.
- *GeoLoc and spOutlier: on-line tools for geo-coding and validating biological data.* Alexandre Marino, Fabrício Pavarin, Sidnei de Souza e Arthur Chapmann.
- *OpenModeller – a Framework for Biological/Environmental Modelling.* Mauro E. S. Muñoz, Renato De Giovanni e Sidnei de Souza.

⁶⁸ <http://www.cria.org.br/eventos/confmt/>

⁶⁹ <http://www.cria.org.br/eventos/iaed/>

- *Use of biodiversity modeling in environmental conservation – a case study.* Marinez Ferreira de Siqueira, Gisela Durigan, Mauro E. S. Muñoz e Fabrício Pavarin.

O workshop ofereceu uma oportunidade significativa para indivíduos e instituições envolvidas com o gerenciamento de dados e o desenvolvimento de sistemas de informação identificarem necessidades comuns e oportunidades para colaboração. Reuniu mais de 70 pesquisadores, representando 48 instituições e 14 países. Deu início a um movimento de articulação entre sistemas de informação sobre biodiversidade com dados ambientais. No caso específico do estado de São Paulo o workshop certamente colaborou para estreitar as relações entre as instituições INPE, Poli e CRIA e resultou na formulação de uma proposta que foi submetida para a análise da Fapesp.

4. **Trends and Developments in Biodiversity Informatics Forum**⁷⁰

Um evento importante realizado no período foi o Fórum internacional "*Trends and Developments in Biodiversity Informatics*" realizado em Indaiatuba, SP, em outubro de 2002, pois apresentou o CRIA para instituições nacionais e internacionais, bem como possibilitou a divulgação da proposta do speciesLink aos participantes. Foram 12 dias onde foram realizados 8 eventos reunindo mais de 100 especialistas. Os tópicos que foram discutidos incluíram: padrões e protocolos, flora, polinizadores e informática para biodiversidade.

5. **Métodos e Ferramentas para Modelagem Preditiva de Espécies**⁷¹

Curso realizado no período de 19 e 20 de Junho de 2002, na Embrapa Informática Agropecuária (CNPTIA/EMBRAPA), ministrado pelo Prof. Dr. A. Townsend Peterson da Universidade de Kansas e Ricardo Scachetti Pereira do CRIA. Teve a participação de 53 alunos de várias regiões do Estado de São Paulo e do Brasil.

6. **Primeira Reunião de curadores das coleções participantes da rede speciesLink**

A primeira reunião com os curadores das coleções foi realizada nos dias 25 e 26 de agosto de 2004 no CNPTIA, Embrapa Informática Agropecuária, em Campinas, SP. A reunião foi coordenada por Vanderlei Perez Canhos e contou com a participação de 19 curadores e 34 técnicos e bolsistas representando 29 coleções científicas, 28 do estado de São Paulo e uma do Rio de Janeiro.

A reunião teve por objetivo aumentar a interação do CRIA com os curadores das coleções. Na oportunidade foram apresentados os resultados do projeto speciesLink e foi dado início a uma discussão mais ampla sobre o futuro da rede. Na ocasião foram apresentadas as seguintes palestras:

- Direito à informação: Marco Legal – Dora Ann Lange Canhos
- Panorama Internacional: Convenção sobre Diversidade Biológica (CBD), Global Biodiversity Information Facility (GBIF), Inter-American Biodiversity Information Network (IABIN) - Vanderlei Perez Canhos
- Sistemas de Informação de Apoio ao Programa Biota: do SinBiota ao speciesLink, Dora Ann Lange Canhos
- A Rede speciesLink: Aspectos técnicos e arquitetura do sistema, Renato De Giovanni
- Ferramentas: *data cleaning*, geo-referenciamento, mapeamento, Alexandre Marino
- Uso de modelagem na avaliação do padrão de distribuição geográfica de espécies arbóreas, Marinez Ferreira de Siqueira

⁷⁰ <http://www.cria.org.br/eventos/tdbi/>

⁷¹ <http://www.cria.org.br/eventos/mfmpe/>

- Uso dos dados: Análises e geração de modelos de biodiversidade, Ingrid Koch

No final do encontro foi discutido o planejamento do lançamento oficial da rede *speciesLink*.

7. Segunda Reunião de curadores das coleções participantes da rede *speciesLink*

Em 2005, no dia 22 de setembro, a Reunião Técnica da Rede Species Link foi também realizada no CNPTIA, Embrapa Informática Agropecuária, em Campinas, SP (<http://www.cria.org.br/eventos/reuniaosp/>).

O objetivo da reunião foi discutir os resultados do projeto e desenvolvimentos futuros. Foram apresentadas as seguintes palestras:

- Apresentação dos objetivos e expectativas da reunião – Vanderlei Perez Canhos, coordenador do projeto *speciesLink*
- Objetivos originais e resultados alcançados – Dora Ann Lange Canhos
- Demonstração das principais ferramentas desenvolvidas – Equipe de desenvolvimento do CRIA

Foram discutidos com os participantes os seguintes itens:

- O que deu certo, o que deu errado, o que faltou: relato das coleções – coordenação dos trabalhos Vanderlei Perez Canhos
- Desenvolvimentos Futuros – Coordenação dos trabalhos Vanderlei Perez Canhos

Participaram da reunião 43 pessoas, incluindo convidados e equipe do CRIA, sendo 21 curadores, 8 representantes de coleções, no total de 29 coleções representadas.

5. DIVULGAÇÃO DOS RESULTADOS DO PROJETO

A divulgação dos resultados do projeto se dá através

- do próprio web site, onde existe um setor específico para assuntos relacionados ao projeto (<http://splink.cria.org.br/project>);
- de uma lista de discussão estruturada para este fim (<http://www.cria.org.br/mailman/listinfo/splink-l>) com 91 membros;
- da participação de eventos no país e no exterior, ministrando palestras sobre componentes do projeto ou participando de reuniões científicas;
- da participação em comitês técnicos;
- da publicação de trabalhos e relatórios técnicos; e através
- da divulgação do projeto e de eventos na mídia especializada.

5.1. PARTICIPAÇÃO EM REUNIÕES CIENTÍFICAS

A participação da equipe em reuniões científicas talvez seja um dos métodos mais eficazes de disseminar o trabalho que está sendo realizado e estimular a participação de provedores e usuários no sistema. O lançamento oficial da Rede *speciesLink* foi realizado no dia 05 de outubro de 2004 na sede da FAPESP, São Paulo. A rede foi apresentada pelo coordenador do projeto Dr. Vanderlei Perez Canhos. O CRIA aproveitou o momento para divulgar o resultado dos trabalhos desenvolvidos pelo projeto e discutir com os participantes necessidades e ações futuras.

Durante o desenvolvimento dos trabalhos, além do lançamento da rede, a equipe do CRIA participou de 47 eventos no país e 28 no exterior (anexo 2), apresentando resultados do projeto *speciesLink* ou participando de reuniões técnicas de interesse para o projeto.

Graças aos resultados e repercussão dos trabalhos realizados, membros da equipe do CRIA foram convidados a participar de importantes comitês técnicos. Alguns grupos são

importantes pelo contato e articulação política, outros pela definição de padrões e protocolos. A seguir estão relacionados aqueles considerados mais importantes para o projeto e o nome do técnico envolvido.

- Comitê de Coordenação do Programa Biota-FAPESP. Vanderlei Perez Canhos
- *Global Biodiversity Information Facility Sub-Committee on Digitization of Biological Collections* (DIGIT). Vanderlei Perez Canhos
- *Task Force on Biological Resource Centers. Organization for Economic Cooperation and Development-OECD*. Vanderlei Perez Canhos
- Grupo de Trabalho da Mata Atlântica, Brasília, DF. Vanderlei Perez Canhos
- *Subcommittee on Capacity Building and Outreach of the Global Biodiversity Information Facility* – GBIF – <http://www.gbif.org> . Vanderlei Perez Canhos
- *Committee on Networking and Interoperability - World Federation for Culture Collections* - <http://wdcm.nig.ac.jp/wfcc/>. Vanderlei Perez Canhos
- *Species 2000 Board of Directors e Project Management Team* - <http://www.sp2000.org/>. Vanderlei Perez Canhos
- Board of Directors do Expert Taxonomy Information Center - ETI, da Universidade de Amsterdam- <http://www.eti.uva.nl/>. Vanderlei Perez Canhos
- *Board of Trustees of the Electronic Publishing Trust for Development* - <http://dspace.dial.pipex.com/bioline/>. Vanderlei Perez Canhos
- GBIF (*Global Biodiversity Information Facility*) *Subcommittee for Data Access and Data Interoperability*. Renato de Giovanni.
- Comitê de Avaliação e Acompanhamento do Programa Plantas do Nordeste, - PNE (Portaria 318/2000, CNPq, 21/12/2000), na área de Informação e Difusão do Conhecimento. Dora A. L. Canhos
- *Conservation Commons Interim Steering Committee*: Dora A. L. Canhos
- *Codata Task Group on Preservation and Archiving of Scientific and Technical Data in Developing Countries* (<http://www.tgdc-codata.org.cn/>). Dora A. L. Canhos.
- *Liaison Working Group do Biosafety Clearing-House* da Convenção sobre Diversidade Biológica (<http://bch.biodiv.org/>). Dora A. L. Canhos
- *Clearing-House Mechanism Informal Advisory Committee* da Convenção sobre Diversidade Biológica (<http://www.biodiv.org/>). Dora A. L. Canhos

5.2. PUBLICAÇÕES E RELATÓRIOS TÉCNICOS

Como grande parte da equipe do projeto não é da área acadêmica, consideramos importante incluir na relação de publicações (anexo 3) os relatórios técnicos apresentados no período e que têm relevância para os trabalhos desenvolvidos. Foram 41 trabalhos, entre publicações, teses, monografias e relatórios sendo que um dos trabalhos, de autoria da Dra. Marinez Ferreira de Siqueira em colaboração com pesquisadores do exterior, foi publicado na revista Nature.

5.3. O PROJETO SPECIESLINK NA MÍDIA

Apesar da rede speciesLink ter sido oficialmente lançada somente em outubro de 2004, alguns desenvolvimentos do projeto já têm merecido a atenção da imprensa. O Anexo 4 traz a relação de algumas notícias arquivadas no CRIA.

III. RESULTADOS

1. PROCESSO DE DIGITALIZAÇÃO DOS ACERVOS

O projeto adotou como estratégia dar autonomia plena às coleções com relação à escolha do software de gerenciamento e à disponibilização dos dados. Toda a entrada, correção, saída e controle de dados sensíveis seriam feitos pela coleção.

O pressuposto inicial era que cada coleção dispunha de um software de gerenciamento e o papel do CRIA seria o de entender o sistema e, com a mínima interferência possível, conectar a coleção à rede. No entanto, constatou-se que a grande maioria das coleções não tinha sistema de gerenciamento implementado e nem equipe de apoio para dar suporte a esse trabalho. Por esse motivo foi necessário oferecer suporte com relação à escolha do software de gerenciamento do acervo e realizar serviços como importação dos dados, configuração do computador, entre outros.

A proposta original também previa a implementação de bolsas para a digitalização de acervos no exterior visando a repatriação de dados de interesse para o país. Foi feito um trabalho de avaliação no Jardim Botânico de Edimburgo e em Kew Gardens, mas na época a tendência das grandes coleções era de acelerarem o processo de digitalização e disponibilização de seus dados na Internet. Optou-se então por investir mais no desenvolvimento de tecnologia de integração de dados ao invés de "trazer" esses dados para um servidor central, dando prioridade a bolsas para auxiliar as coleções do Estado de São Paulo na digitalização de seus acervos.

O apoio dado às coleções foram hardware e software, orientação técnica e ferramentas de *data cleaning* e de geo-referenciamento para auxiliar no processo de digitalização de seus acervos:

Os resultados estão apresentados na tabela 9 a seguir.

Tabela 9. Resultados do processo de digitalização dos acervos

sigla	registros on-line				média de novos registros por mês	Bolsistas		média / bolsista / mês	Custo total bolsistas	custo por registro	Resultados			
	início	Out/2005	evolução	meses		No	meses				total	%	georef	%geoe
ácaros														
AcariESALQ	888	15.732	14.844	36	412						15.732	100%	1.417	9%
DZSJRP-Acari	4.734	6.296	1.562	30	52	1	8	195	4.900	R\$ 3,14	7.026	90%	2.344	37%
IBSP-Acari	1.270	2.771	1.501	12	125						9.201	30%	1.143	41%
Total	6.892	24.799	17.907	78	230	1	8	195			31.959	78%	4.904	20%
aranhas														
UBTU	2.623	2.626	3	14	0						3.500	75%	0	0%
Total	2.623	2.626	3	14	0						3.500	75%	0	0%
entomologica														
IBSP-IB	0	401	401		0						275.000	0%	0	0%
MEFEIS	0	3.519	0	14	0	2	16	0	11.200	R\$ 3,18	28.000	13%	0	0%
Total	3.519	3.519	0	14	0	2	16	0			303.000	1%	0	0%
abelhas														
CEPANN	10.180	34.477	24.297	9	2.700	1	12	2.025	8.400	R\$ 0,35	34.477	100%	31.891	92%
RPSP	0	68.599	68.599	8	8.575	1	12	5.717	7.500	R\$ 0,11	171.000	40%	50.467	74%
Total	10.180	103.076	92.896	17	5.464	2	24	3.871			205.477	50%	82.358	80%
mamíferos														
DZSJRP-Chiroptera	2.829	5.339	2.510	16	157	1	8	314	5.600	R\$ 2,23	10.678	50%	815	15%
IAL-roedores	4.067	12.029	7.962	9	885	1	7	1.137	4.900	R\$ 0,62	21.000	57%	0	0%
MHN-mamíferos	2.359	2.385	26	12	2						2.385	100%	0	0%
Total	9.255	19.753	10.498	25	420	2	15	700			34.063	58%	815	4%
aves														
IAL-aves	889	21.157	20.268	6	3.378	1	8	2.534	5.600	R\$ 0,28	110.000	19%	12.488	59%
MHN-aves	2.197	2.224	27	12	2						2.224	100%	0	0%
Total	3.086	23.381	20.295	6	3.383	1	8	2.534			112.224	21%	12.488	53%

sigla	registros on-line			média de novos registros por mês	Bolsistas		média / bolsista / mês	Custo total bolsistas	custo por registro	Resultados				
	início	Out/2005	evolução		meses	No				meses	total	%	georef	%geo
peixes														
DZSJRP-Pisces	5.278	7.885	2.607	37			70				7.885	100%	6.068	77%
LIRP	4.314	5.167	853	29			29				30.000	17%	3.860	75%
MHN-peixes	7.767	6.138	-1.629	12			-136				6.138	100%	0	0%
MZUSP	72.714	77.613	4.899	20	2	31	245	21.700	R\$ 4,43		84.000	92%	53.319	69%
Total	90.073	96.803	6.730	86	2	31	78				128.023	76%	63.247	65%
repteis e anfíbios														
CFBH	2.935	3.773	838	16			52				7.000	54%	0	0%
DZSJRP-Amphibia-adults	4.823	7.499	2.676	16	1	12	167	8.400	R\$ 3,14		7.499	100%	7.124	95%
DZSJRP-Amphibio-tadpoles		1.158	1.158	16			72				1.758	66%	0	0%
IBSP-Herpeto	57.397	66.167	8.770	11	1	8	797	5.600	R\$ 0,64		80.000	83%	0	0%
MHN-anfíbios	16.062	12.231	-3.831	12			-319				12.231	100%	0	0%
MHN-repteis	2.439	2.832	393	12			33				2.832	100%	0	0%
Total	83.656	93.660	10.004	43	2	20	233				111.320	84%	7.124	8%
herbários														
BOTU	2	165	163	5			33				24.000	1%	161	98%
CPMA	1.882	1.882	0	6			0				2.150	88%	0	0%
ESA	730	74.549	73.819	28	1	12	2.636	3.000			120.000	62%	19.606	26%
HISA	4	718	714	6			119				10.235	7%	62	9%
HRCB	2.593	6.541	3.948	15	1	8	263	2.000	R\$ 0,51		40.000	16%	0	0%
HSJRP	19.462	19.377	-85	11	3		-8				28.000	69%	0	0%
IAC	26.136	39.770	13.634	18	1	28	757	10.600	R\$ 0,78		45.000	88%	31.825	80%
JBRJ	1.383	6.138	4.755	13			366				450.000	1%	0	0%
SP*	1.088	13.928	12.840	27	1	12	476	8.400	R\$ 1,28		350.000	4%	3.494	25%
SPF	1.121	10.250	9.129	29			315				133.500	8%	4.064	40%
SPSF	31.168	31.280	112	16	1	8	7	5.600	R\$ 50,00		34.000	92%	2	0%
UEC	11.800	36.081	24.281	30	4	66	809	44.600	R\$ 1,84		134.000	27%	9.957	28%
Total	97.369	240.679	143.310	204	12	134	703				1.370.885	18%	69.171	29%

sigla	registros on-line			média de novos registros por mês	Bolsistas		média / bolsista / mês	Custo total bolsistas	custo por registro	Resultados				
	início	Out/2005	evolução		meses	No				meses	total	%	georef	%geo
algas														
SP-Algae	11.088	14.033	2.945	27							15.000	94%	0	0%
SPF-Algae	19.776	19.776	0	29							19.776	100%	0	0%
Total	30.864	33.809	2.945	27	0	109					34.776	97%	0	0%
xiloteca														
BCTW	0	9.127	9.127	11	2	24	380	11.400	R\$ 1,25		19.500	47%	0	0%
SPFW	908	1.677	769	13	2	20	38	8.600	R\$ 11,18		4.000	42%	80	5%
Total	908	10.804	9.896	24	4	44	225				23.500	46%	80	1%
microorganismos														
CBMAI	78	314	236	18							688	46%	0	0%
IBSBF	929	1.750	821	18	1	4	205	2.800	R\$ 3,41		2.000	88%	0	0%
Total	1.007	2.064	1.057	36	1	4	264				2.688	77%	0	0%

SP* - informação prestada pelo herbário é que o bolsista digitou 6585 registros no período, o restante foi digitalizado pela equipe (email de 11/08/05)

A avaliação do custo por registro tendo por base o valor do bolsista e a evolução do número de registros *on-line* é, na realidade, uma avaliação bastante grosseira. Os valores variam entre R\$ 0,11 e R\$ 50,00. Se não considerarmos os valores discrepantes, obtemos um valor médio de R\$ 1,70 por registro digitado. É importante ter em mente que o valor refere-se tão somente ao tempo do bolsista, Não estão sendo consideradas questões como a limpeza e validação dos dados, nem os custos institucionais, que poderiam ser objeto de estudos futuros.

Foi também elaborado um questionário perguntando às coleções informações sobre o processo de digitalização. Trata-se de uma experiência muito rica que pode ajudar na orientação de trabalhos futuros. Foi perguntado se a coleção tinha ou não um plano de digitalização e algumas estratégias de digitalização apresentadas foram:

- Primeiro organizar os fichários e iniciar os trabalhos com os espécimes já organizados e identificados. Incluir também todos os organismos em estudo na pós-graduação na conclusão dos trabalhos.
- inclusão seguindo a ordem crescente da numeração
- treinamento prévio do bolsista na rotina da coleção
- Iniciar pela digitalização do acervo já tombado, trabalhar na correção e refinamento. Escolher a próxima coleção, tombar e iniciar o processo de digitalização, correção e refinamento. E assim, sucessivamente.
- A digitalização está sendo feita em 3 etapas: (1) digitalização dos campos mínimos; (2) imagens digitais de cada espécime; (3) digitação de todos os campos.
- Digitalização seguindo a ordem alfabética por famílias com verificação de “ruídos”. Durante o processo a planilha foi simplificada e os dados serão completados em uma segunda etapa.
- Dar prioridade aos materiais recém incluídos e àqueles solicitados para empréstimo (informatização antes de sair do acervo)
- Contratação de empresa de digitação.
- Designar um funcionário para participar do processo de digitação junto com alunos e bolsistas para assegurar uma continuidade mínima

É interessante comparar as estratégias com os resultados obtidos. Sem dúvida alguma, as coleções que indicaram ter estratégias claras, são muito mais organizadas e obtiveram resultados muito bons. Um destaque deve ser dado às coleções de abelhas.

Todas as coleções que responderam que não tinham uma estratégia para a digitalização do acervo tiveram um processo bem lento e, via-de-regra, com muitos erros.

Recomenda-se à Fapesp exigir na apresentação de toda proposta solicitando apoio à informatização de coleções, uma estratégia para a digitalização de acervos.

O questionário também solicitou às coleções que fizessem sugestões quanto a desenvolvimentos futuros. Algumas sugestões foram:

- Incluir a associação parasito – hospedeiro *on-line*;
- Desenvolver uma ferramenta para gerenciar referências bibliográficas;
- Desenvolver uma ferramenta que permite pesquisar organismos relacionados a determinadas espécies;
- Desenvolver uma ferramenta de distribuição temporal dos espécimes;

- Elaborar um manual explicando “passo a passo” como configurar o programa que filtra os dados;
- Aprimorar o Gazetteer;
- Ter listas de referência para a atualização de nomes e autores de espécies;
- Incluir outros layers no speciesMapper (hidrografia, vegetação, ...); e,
- Incluir um sistema de ‘help’ e incluir alguns exemplos de uso das ferramentas.

Recomendação à Fapesp apoiar a produção de catálogos de nomes e *checklists* online para servir de referência para a atualização da nomenclatura dos acervos das coleções.

Das 24 coleções que responderam ao questionário, tivemos a grata surpresa de saber que 20 estão usando a ferramenta *data cleaning*, 15 o *speciesMapper*, 13 o *geoLoc*, 13 o *spOutlier*, 10 o *Conversor* e ainda uma coleção está utilizando o Banco de coletores e *infoxy*. No caso do *infoxy* trata-se de uma resposta espontânea uma vez que esta ferramenta, por erro nosso, não foi incluída como opção no questionário.

O projeto não previu a realização de treinamento, uma vez que pressupunha que as coleções já tivessem iniciado o processo de digitalização de seus acervos. O questionário levantou essa questão e as respostas indicam que seria oportuno realizar alguns cursos como:

- Uso dos programas de gerenciamento das coleções: Biota, Brahms, Access
- Estrutura dos dados e a padronização da nomenclatura (p.ex. localidades)
- Temas relacionados ao CRIA e às coleções, visando o crescimento profissional dos bolsistas
- Treinamento nas ferramentas desenvolvidas no projeto *speciesLink* (incluindo o *openModeller*)

Por fim, foram solicitadas recomendações gerais que poderiam auxiliar o processo de estabelecimento de uma infra-estrutura de dados de coleções biológicas de acesso livre e aberto na Internet. São elas:

- Incluir técnicos no quadro funcional da instituição para dar suporte aos curadores e pesquisadores
- Instituir nas universidades e instituições de pesquisa a função de “curador”
- Estudar junto à Fapesp a possibilidade da criação de bolsas para digitação de acervos para dar continuidade aos trabalhos
- Falta visibilidade ao projeto (comunidade científica e sociedade): ampliar a divulgação da rede *speciesLink*
- continuar com o suporte do CRIA via extensão do projeto ou realização de outro projeto
- extensão do prazo para a digitalização do acervo
- Em projetos futuros, prever uma quantidade maior de equipamento de informática para as coleções
- à Fapesp: a bolsa deveria ser concedida ao orientador ou coordenador e não ao candidato. Hoje, cada substituição significa reiniciar o processo;
- Para bolsas que visam a digitalização dos dados dos acervos seria importante não descontar o tempo da bolsa TT caso o aluno faça pós-graduação.

Destacamos, portanto, algumas recomendações à Fapesp.

Recomendações:

Criação de bolsas associadas ao Programa Biota/Fapesp para digitalização de acervos visando dar continuidade aos trabalhos, evitando que cada curador tenha que apresentar um projeto para esse fim.

Conceder a bolsa ao orientador e não ao candidato e desvincular a bolsa do programa de pós-graduação, ou seja, para a digitalização de acervos não se descontaria o tempo da bolsa caso o aluno faça pós-graduação.

Estudar uma forma de dar suporte à manutenção da rede.

2. DESENVOLVIMENTO DA REDE

A proposta apresentada à Fapesp em março de 2001 tinha por objetivo implementar uma rede distribuída de informação sobre biodiversidade para o Estado de São Paulo, explorando as novas tecnologias e uma melhor conectividade (Internet 2) para a integração de bases de dados heterogêneas e independentes. As bases de dados a serem integradas incluíam 12 coleções biológicas do Estado de São Paulo, a rede *Species Analyst* e o *SinBiota*, sistema que armazena os dados de coletas realizadas no âmbito do programa Biota/Fapesp.

Em junho de 2001, portanto antes da aprovação da proposta (outubro de 2001), foi realizada uma reunião envolvendo as principais redes de dados de acervos biológicos para discutir a possibilidade de desenvolver um novo protocolo de comunicação. As seguintes redes estavam representadas:

- Australian Virtual Herbarium (AVH)
- Bio-resource network in Japan (BRnet)
- European Natural History Specimen Information Network (ENHSIN)
- Red Mundial de Información sobre Biodiversidad (REMIB)
- The Species Analyst (TSA)

Quando o projeto *speciesLink* foi aprovado, como estávamos em contato com alguns coordenadores da rede TSA e da AVH, resolvemos colaborar com essa iniciativa ao invés de desenvolver um protocolo próprio para integrar as coleções. Desse trabalho nasceu o protocolo DiGIR e o modelo de dados DarwinCore, hoje adotados internacionalmente.

Na proposta, os requisitos esperados da rede eram:

- **Disponibilidade:** a rede deveria estar disponível, em funcionamento sempre;
- **Tolerância a falhas:** a rede deveria ser capaz de operar quando alguns nós não estivessem funcionando;
- **Desempenho e Robustez:** a rede deveria ser capaz de processar e responder em um tempo razoável um grande número de solicitações simultâneas;
- **Descoberta de novos nós:** a rede deveria ter mecanismos automáticos ou manuais para inclusão de novos nós;
- **Independência de plataforma:** o software deveria operar em ambientes computacionais heterogêneos; e,
- **Escalabilidade:** a rede deveria ter capacidade para aumentar o número de nós, sem que fosse necessário desenvolver novos componentes de *software*.

Todos esses requisitos foram atendidos.

Adotou-se como estratégia

- sempre que possível utilizar software livre ou de código aberto;
- utilizar equipamentos de boa performance e baixo custo relativo, tanto em termos do custo do equipamento como também no custo de manutenção;
- participar do desenvolvimento *open source* do protocolo de comunicação (DiGIR);
- participar da discussão e adotar o modelo de dados DarwinCore;
- desenvolver um sistema *on-line* de acesso livre e aberto;
- garantir autonomia plena às coleções com relação à escolha do software de gerenciamento;
- garantir autonomia plena às coleções com relação à disponibilização dos dados: a entrada, correção, saída e controle de dados sensíveis teriam que ser feitos pela coleção; e,
- garantir a participação plena de coleções com conectividade lenta ou instável.

Como resultado temos:

1. A rede é baseada em equipamento Intel, portanto de baixo custo relativo, tanto do equipamento em si como também na sua manutenção.
2. Todo o sistema, inclusive as ferramentas e aplicativos foram desenvolvidos usando software livre e de código aberto.
3. O protocolo de comunicação desenvolvido (DiGIR) em base colaborativa, foi adotado internacionalmente, o que permite a integração da rede *speciesLink* com as mais importantes redes internacionais.
4. As coleções têm total autonomia em relação a seus dados, definindo o que pode ou não estar disponível, podendo inserir, retirar, alterar ou atualizar dados, sempre que necessário. Somente a coleção tem este poder.
5. As coleções com conectividade lenta ou instável ou sem capacitação local para a manutenção de um servidor, estão participando da rede plenamente através dos servidores regionais.

3. CONTEÚDO

O produto final do projeto é o número de dados disponíveis de forma livre e aberta na Internet. A tabela a seguir mostra o grau de informatização e o percentual de registros geo-referenciados, em cada acervo, por grupo taxonômico e os números totais (tabela 10).

Tabela 10. Grau de informatização e de geo-referenciamento por coleção e por grupo taxonômico.

COLEÇÕES BIOLÓGICAS	ACRÔNIMO	ACERVO	REGISTROS ON-LINE	GEO-REF	% ON-LINE	% GEO-REF
			out-05			
Herbários	BOTU	24.000	165	161	1%	98%
	CPMA	2.150	1.882	0	88%	0%
	ESA	120.000	74.549	19.606	62%	26%
	HISA	10.235	718	62	7%	9%
	HRCB	40.000	6.541	0	16%	0%
	HSJRP	28.000	19.377	0	69%	0%
	IAC	45.000	39.770	31.825	88%	80%
	SP	350.000	13.928	3.494	4%	25%
	SP - Algae	15.000	14.033	0	94%	0%
	SPF	133.500	10.250	4.064	8%	40%

COLEÇÕES BIOLÓGICAS	ACRÔNIMO	ACERVO	REGISTROS ON-LINE	GEO-REF	% ON-LINE	% GEO-REF
			out-05			
	SPF-Algae	19.776	19.776	0	100%	0%
	SPSF	34.000	31.280	2	92%	0%
	UEC	134.000	36.081	9.957	27%	28%
	JBRJ	450.000	6.138	0	1%	0%
Subtotal		1.405.661	274.488	69.171	20%	25%
Xilotecas	BCTw	19.500	9.127	0	47%	0%
	SPFw	4.000	1.677	80	42%	5%
Subtotal		23.500	10.804	80	46%	1%
Peixes	DZSJRP-Pisces	7.885	7.885	6.068	100%	77%
	LIRP	30.000	5.167	3.860	17%	75%
	MZUSP	84.000	77.613	53.319	92%	69%
	MHN-peixes	6.138	6.138	0	100%	
Subtotal		128.023	96.803	63.247	76%	65%
Aracnídeos						
	Ácaros DZSJRPacari	7.026	6.296	2.344	90%	37%
	Ácaros AcariESALQ	15.732	15.732	1.417	100%	9%
	Ácaros IBSP-Acari	9.201	2.771	1.143	30%	41%
	Aranhas UBTU	3.500	2.626	0	75%	0%
Subtotal		35.459	27.425	4.904	77%	18%
Répteis e Anfíbios						
	Anfíbios DZSJRP-Amphibia-adults	7.358	7.358	7.124	100%	97%
	Anfíbios DZSJRP-Amphibia-tadpoles	1.758	1.158	0	66%	0%
	Anfíbios CFBH	7.000	3.773	0	54%	0%
	Anfíbios MHN-anfíbios	12.231	12.231	0	100%	0%
	Répteis MHN-repteis	2.832	2.832	0	100%	0%
	Répteis e Anfíbios IBSP-Herpeto	80.000	66.167	0	83%	0%
Subtotal		111.179	93.519	7.124	84%	8%
Mamíferos						
	Roedores IAL-Roedores	21.000	12.029	0	57%	0%
	Mamíferos MHN-mamíferos	2.385	2.385	0		
	Morcegos DZSJRPChiroptera	10.678	5.339	815	50%	15%
Subtotal		34.063	19.753	815	58%	4%
Insetos						
	Insetos MEFEIS	28.000	3.519	0	13%	0%
	Insetos IBSP-IB	275.000	401	0		
	Abelhas CEPANN	34.171	33.636	31.650	98%	94%
	Abelhas RPSP	171.000	68.599	50.467		
Subtotal		508.171	106.155	82.117	21%	77%
Aves						
	IAL-Aves	110.000	21.157	12.488	19%	59%
	MHN-Aves	2.224	2.224	0	100%	0%
Subtotal		112.224	23.381	12.488	21%	53%
Microrganismos						
	CBMAI	688	314	0	46%	0%
	IBSBF	2.000	1.750	0	88%	0%
Subtotal		2.688	2.064	0	77%	0%
Total dos Acervos		2.360.968	654.392	239.946	28%	37%
Dados de Observação	SinBiota	63.586	63.586	63.584	100%	100%
Total Geral		2.424.554	717.978	303.530	30%	42%

Analisando os dados das coleções que hoje têm informações *on-line* na rede *speciesLink* temos que no 30% dos registros estão digitalizados. Esse índice era de 20% em outubro de 2004 e 10% no final da primeira fase. Do total de registros *on-line*, 42% estão geo-referenciados, contra 33% em outubro de 2004. Infelizmente não foi medido o grau de geo-referenciamento no final da primeira fase do projeto. Um outro dado significativo obtido da ferramenta *data cleaning* é que se as coleções incluíssem a coordenada geográfica do município para os dados que não tem registro da latitude e longitude do ponto da coleta, o percentual de registros geo-referenciados saltaria para 78%.

4. PARCERIA COM OUTRAS INICIATIVAS

No decorrer do projeto, o CRIA recebeu diversas visitas de instituições nacionais e internacionais. A equipe procurou, sempre que possível, mostrar todos os desenvolvimentos do projeto. No caso do *data cleaning* nos foi solicitado testar dados de uma expedição nas Guianas do Smithsonian de Washington, indexamos também uma coleção de Mamíferos do Espírito Santo, e dados de Madagascar, do Jardim Botânico de Missouri a pedido do diretor do GBIF. Acreditamos que durante estes anos adquirimos “*know how*” nessa área e esse trabalho está se tornando modelo para outras instituições de várias regiões do Brasil.

As seguintes parcerias foram firmadas envolvendo cooperação internacional:

O GBIF (<http://www.gbif.org/>) se mostrou interessado em vários trabalhos desenvolvidos pela equipe do CRIA. O CRIA participou de um processo de concorrência e foi contratado para trabalhar no desenvolvimento do portal, de uma novo protocolo de comunicação (TAPIR) e para desenvolver uma ferramenta de validação de dados para o portal do GBIF para servir de ferramenta para todas as coleções conectadas ao portal. Todo o código desenvolvido em Java está disponível no Source Forge (<http://sourceforge.net>).

OBIS - Ocean Biogeographic Information System (<http://www.iobis.org>): O CRIA é responsável por abrigar o portal OBISSA (<http://obissa.cria.org.br>) que representa o nó Sul-americano composto por Brasil, Chile e Argentina. O CRIA é também responsável por criar e manter o provedor dos dados brasileiros. Nesse projeto temos como parceiro a Universidade de São Paulo - USP (Departamento de Zoologia e Instituto Oceanográfico). O projeto está sendo patrocinado pela *Alfred P. Sloan Foundation*.

O projeto Biogeomancer (<http://www.biogeomancer.org>) foi proposto à Moore Foundation e tem por objetivo o desenvolvimento de uma ferramenta de georeferenciamento automático (o *biogeomancer*). Trata-se de uma iniciativa conjunta das seguintes instituições lideradas pelo *Museum of Vertebrate Zoology* da Universidade da Califórnia, Berkeley:

- Tulane Museum of Natural History
- Yale Peabody Museum
- California Academy of Sciences
- Centro de Referência em Informação Ambiental (CRIA)
- Berkeley Natural History Museums
- Open GIS Consortium
- Center for Biodiversity and Conservation Research, Australian Museum
- Institute of Natural Resource Analysis and Management (INRAM)
- University of Colorado
- University of Illinois
- University of Kansas Biodiversity Research Center

Nesse projeto o CRIA tem a participação no desenvolvimento do módulo de validação, baseado na idéia de identificação de *outliers*.

5. DEFINIÇÃO DE UM MODELO

O projeto *speciesLink* desenvolveu um modelo de rede para integrar os acervos de coleções biológicas, disponibilizando os dados de maneira livre e aberta *on-line*. O modelo inclui as seguintes características próprias da rede *speciesLink*:

- Possibilidade da coleção se conectar ao servidor regional ao invés de manter um servidor dedicado à rede;
- Capacitação dada a cada coleção de criar filtros a dados sensíveis, atualizar ou apagar todos os seus dados de forma autônoma através do software *spLinker*;
- Possibilidade de visualizar possíveis erros e o perfil de cada coleção; e,
- Sistema *on-line* de monitoramento da rede.

As demais características como o portal, provider, e o modelo de dados DarwinCore são comuns às demais redes que adotaram o protocolo DiGIR, como o GBIF.

A adoção desse modelo está em fase de discussão ou contratação com as seguintes iniciativas:

- Dois projetos selecionados pelo Edital CT-BIOTEC/MCT/CNPq – Nº 021/2005 “Seleção Pública de Projetos para Organização, Consolidação e Dinamização de Programas Institucionais para Manutenção, Melhoria e Ampliação das Coleções Biológicas de Interesse para a Pesquisa Biotecnológica e para a Conservação da Biodiversidade Brasileira.”
- Uma rede de coleções a ser implementada no Espírito Santo com recursos da Fapes (Fundação de Amparo à Pesquisa do Espírito Santo) e o MCT.
- Uma rede de coleções a ser implementada na Bahia com recursos da Fapesb (Fundação de Amparo à Pesquisa da Bahia).
- A rede PPBio que inclui o Museu Goeldi e o Instituto Nacional de Pesquisa na Amazônia (INPA) e conta com recursos do MCT.

IV. COMENTÁRIOS FINAIS

A construção da base do conhecimento sobre biodiversidade é necessariamente coletiva. O estudo, a conservação e o uso sustentável da biodiversidade requer um tratamento multi e inter disciplinar. Requer também um ambiente de colaboração global. A identificação de uma espécie depende de uma base comparativa de dados primários e informações associadas. Por este motivo, no âmbito da comunidade acadêmica, há poucas restrições ao compartilhamento de dados científicos. A resistência que existe hoje por parte de alguns especialistas em disponibilizar os dados na Internet, se deve mais ao fato da Internet ser aberta a qualquer pessoa, do que em uma eventual resistência em compartilhar dados com seus pares. Muitos pesquisadores têm receio do mau uso, da cópia não autorizada ou de interpretações erradas de seus dados por não especialistas. Aos poucos, porém, com as evidentes vantagens do compartilhamento de dados on-line, essa resistência vem diminuindo.

Os acervos científicos e dados associados devem ser vistos como infra-estrutura de pesquisa. Durante séculos cientistas vêm sistematicamente registrando suas observações de pesquisa e publicando os resultados obtidos. Mas não havia, e em muitos casos ainda não há, qualquer preocupação em manter os dados primários de observação e coletas ou de disseminá-los a pesquisadores de outras áreas de conhecimento. Nem tampouco há a preocupação em preservá-los para gerações futuras. Com o avanço da tecnologia de informação e comunicação, houve também uma evolução do pensamento e da organização científica, criando diferentes demandas por diferentes tipos de dados ou ainda, graças à integração e a interoperabilidade de sistemas, por diferentes conjuntos de dados.

Com os avanços tecnológicos e a demanda por dados e informações científicas em constante evolução, é fundamental promover o livre acesso a dados e informações visando o fortalecimento da ciência para o benefício da sociedade. O progresso científico depende do acesso pleno a dados e da divulgação científica aberta dos resultados da pesquisa na literatura. Um forte componente de dados e informações acessíveis e em domínio público, promove um maior retorno do investimento público, estimulando a inovação e a decisão informada. A integração de dados de diferentes disciplinas, de diferentes espaços geográficos, analisados por pesquisadores também de diferentes disciplinas e de origem cultural diversa, abre o caminho para novas perguntas e para a inovação.

Publicações são importantes e até vitais para o progresso científico, mas refletem a visão dos autores, a sua leitura dos fatos (dos dados) em um determinado momento. Uma releitura pelo próprio autor, pelos seus pares ou por gerações futuras dependerá da disponibilidade permanente de dados primários.

Os trabalhos desenvolvidos no decorrer do projeto *speciesLink* alcançaram resultados muito melhores do que a própria equipe esperava. O entusiasmo das coleções participantes no esforço e a mudança cultural de grande parte dos curadores e técnicos das coleções certamente contribuíram para o sucesso e a visibilidade que a rede *speciesLink* está tendo no cenário nacional e internacional.

No entanto, é necessário definir uma estratégia de sustentabilidade da rede, definindo o seu modelo de gestão. Uma grande discussão sempre recorrente diz respeito à natureza jurídica da principal instituição gestora, se deve ser pública ou privada de interesse público.

Algumas características essenciais com relação à instituição gestora são:

- Ter como missão o desenvolvimento e manutenção de uma infra-estrutura de dados sobre coleções científicas de acesso livre e aberto
- Ser sustentável, com apoio permanente ou de longo prazo garantido;
- Ter flexibilidade para poder contratar e demitir mão-de-obra necessária, alterar tecnologias, incluir novas prioridades, e manter um plano estratégico dinâmico;

- Ter alta capacidade de articulação e de estabelecimento de parcerias
- Ter transparência para garantir a sua credibilidade e,
- Contar com equipe altamente especializada.

A opção por uma instituição pública tem como aspecto positivo uma relativa estabilidade financeira com a possibilidade de manutenção de longo prazo. Por outro lado normalmente é uma administração mais “engessada”, com possíveis problemas de contratação e menor facilidade para a articulação para o estabelecimento de parcerias. Outro risco é a susceptibilidade às mudanças políticas com a eleição de novos governos dificultando a implementação de políticas de longo prazo. No caso de universidades, seria muito difícil o compromisso institucional de ter na rede a sua missão.

Uma instituição privada de interesse público, que pode ser uma Organização Social (OS) ou uma Organização da Sociedade Civil de Interesse Público (OSCIP), apresenta maior facilidade para mobilizar recursos, sinergizar iniciativas e promover parcerias em prol de seus objetivos e metas. Com objetivos e metas claras é também mais simples desenvolver um planejamento de longo prazo com indicadores claros e mensuráveis. Para manter a certificação como OS ou OSCIP elas precisam

- ter transparência,
- prestar contas ao Poder Executivo; e
- executar as atividades estatutárias.

Acreditamos que seria importante realizar um estudo junto à Fapesp com relação ao melhor modelo para garantir a manutenção dessa infra-estrutura de dados de acesso livre e aberto. Sugerimos que o apoio ao sistema seja permanente e à instituição responsável pela manutenção desse sistema, no caso da rede *speciesLink*, seja de longo prazo com avaliações periódicas.

Acreditamos que a Fapesp deverá ter um papel central na viabilização de um projeto nacional de criação de uma infra-estrutura compartilhada de dados biológicos de apoio ao desenvolvimento científico e tecnológico do país.

V. EQUIPE

Coordenador Geral: Vanderlei Perez Canhos

Desenvolvimento do Projeto:

Sidnei de Souza (coordenador técnico)

Renato de Giovanni (desenvolvimento do DiGIR, *openModeller*)

Alexandre Marino (desenvolvimento de ferramentas, data cleaning)

Ricardo Scachetti Pereira (desenvolvimento, fase 1)

Mauro Muñoz (desenvolvimento do *openModeller*, DiGIR)

Rosely Aurea Lopes Coelho (teste de sistemas)

Fabrcio Pavarin (desenvolvimento de ferramentas, apoio às coleções)

Rafael L. Fonseca (apoio às coleções, modelagem)

Ana Paula de Souza Albano (apoio às coleções, suporte)

Benedito Cruz (suporte)

Marinez F. Siqueira (modelagem)

Ingrid Koch (modelagem, data cleaning)

Dora Ann Lange Canhos (relatório, análise de indicadores, elaboração de estratégia)

Cristina Umino (coordenação administrativa)

Silvia Beltrane Cintra (auxiliar administrativa)

Curadores da rede speciesLink:

Carlos Flechtmann

MEFEIS, Coleção do Museu de Entomologia da FEIS, Unesp, Ilha Solteira

Célia SantAnna

SP-Algae, Coleção de Algas do Herbário do Estado, Instituto de Botânica

Célio F.B. Haddad

CFBH, Coleção - IB/UNESP, Rio Claro

Darci Moraes Barros Battesti

IBSP-Acari, Instituto Butantan

Denise de Cerqueira Rossa Feres

DZSJRP-Amphibia adults e DZSJRP-Amphibia tadpoles, Coleção de Anfíbios, UNESP, São José do Rio Preto

Eliana Morielle Versute

DZSJRP-Chiroptera, Coleção de Quirópteros do Departamento de Zoologia e Botânica - IBILCE/UNESP, São José do Rio Preto

Flávio A. Bockmann

LIRP, Laboratório de Ictiologia, USP, Ribeirão Preto

Flavio Augusto de Souza Berchez

SPF-Algae, Coleção de Algas Vasculares, Herbário, Depto. Botânica - IB/USP

Francisco Langeani

DZSJRP-Pisces, Coleção de Peixes do Departamento de Zoologia e Botânica, IBILCE/UNESP, São José do Rio Preto

Francisco Luís Franco

IBSP, Coleção Herpetológica "Alphonse Richard Hoge", Inst. Butantan

Glyn Mara Figueira

CPMA, Coleção de Plantas Medicinais e Aromáticas

Gregório Ceccantini

SPFw, Coleção de Madeiras (Xiloteca) , Herbário, Depto. Botânica - IB, USP

Isabela M P Rinaldi

UBTU, Coleção de Aranhas, Unesp, Botucatu

João Aurélio Pastore

SPSF, Herbário D. Bento Pickel, Instituto Florestal

João Maria Franco de Camargo

RPSP, Coleção Camargo do Departamento de Biologia da FFCLRP, Universidade de São Paulo, Ribeirão Preto

Júlio Rodrigues Neto

IBSBF, Coleção de Culturas de Fitobactérias do Laboratório de Bacteriologia Vegetal - Instituto Biológico, Campinas

Lara Durães Sette

CBMAI, Coleção Brasileira de Microrganismos de Ambiente e Indústria, CBQBA/Unicamp

Lasaro Vanderlei F. Silva

ESALQ-Acari, Coleção de Referencia de Ácaros ESALQ, Piracicaba

Lúcia Garcez Lohmann

SPF, Coleção de Plantas Vasculares, Herbário, Depto. Botânica - IB/USP

Luiz Eloy Pereira

IAL, Instituto Adolpho Lutz (aves e roedores)

Marco Antonio de Assis

HRCB, Herbarium Rioclarense, UNESP, Rio Claro

Maria Candida Henrique Mamede

SP, Coleção de Fanerógamas do Herbário do Estado, Instituto de Botânica

Maria José de Andrade C. Miranda

BCTw, Xiloteca Calvino Mainieri, Divisão de Produtos Florestais, IPT

Marilia Rodrigues Pereira de Noronha

HISA, Herbario de Ilha Solteira, UNESP

Naércio Aquino Menezes

MZUSP, Coleção de Peixes do Museu de Zoologia, IB/USP

Neusa Taroda Ranga

HSJRP, Herbário, UNESP, São José do Rio Preto

Paulo José Fernandes Guimarães

JBRJ, Jardim Botânico do Rio de Janeiro

Reinaldo José Fazzio Feres

DZSJRP-Acari, Coleção de Ácaros/UNESP, São José do Rio Preto

Sergio Ide

IBSP-IB, Coleção Entomológica Adolph Hempel, Inst. Biológico

Sigrid Luiza Jung Mendaçolli

IAC, Herbário IAC, Instituto Agronômico

Vera Lúcia Imperatriz Fonseca

CEPANN, Coleção Entomológica Paulo Nogueira-Neto, Lab. de Abelhas, USP

Vinicius Castro Souza

ESA, Herbário da Escola Superior de Agricultura Luiz de Queiroz, ESALQ, Piracicaba

Washington Marcondes Ferreira Neto

UEC, Herbário IB, Unicamp



Wesley Rodrigues Silva

MHN, Museu de História Natural, Unicamp (Coleções de anfíbios, aves, mamíferos, peixes e répteis)

Yuriko Nogueira Pinto

BOTU, Herbário Irina Delanova Gemtchújnicov, UNESP, Botucatu

ANEXO I

QUESTIONÁRIO PARA AVALIAÇÃO DA INFRA-ESTRUTURA DA COLEÇÃO PARA SUA PARTICIPAÇÃO NO PROJETO SPECIESLINK

I. Dados Institucionais

Nome da Instituição:

Nome e email do diretor:

Nome da coleção:

Sigla:

Nome e email do curador:

Nome e email do responsável técnico⁷²:

Endereço para correspondência:

Fone:

Fax:

Endereço na Internet (URL): http://

II. Descrição do Acervo

Descreva o acervo, indicando qual o seu escopo, relevância, principais serviços oferecidos (empréstimos, troca, identificação de amostras, etc) e qual(is) o(s) público(s)-alvo a que a coleção se destina.

III. Informatização do Acervo

Total de amostras no acervo⁷³:

Porcentagem de amostras informatizadas⁷⁴:

Porcentagem de amostras georeferenciadas⁷⁵:

Qual o sistema de informação utilizado?

- Sistema de gerenciamento de coleções biológicas⁷⁶ (qual):
- Banco de dados (qual):
- planilha eletrônica (qual):
- editor de texto (qual):
- outro (qual):

Existe algum catálogo das amostras da coleção (em papel ou em formato digital)?

Descreva os campos que estão sendo registrados, indicando os campos mínimos obrigatórios e as informações complementares:

⁷² Contato para discussão de assuntos técnicos em informática e Internet.

⁷³ Por amostra subentende-se qualquer amostra de material coletado que faz parte do acervo da coleção, tais como: exsiccatas, fruto, semente, pele, esqueleto, espécime, linhagem. Caso a coleção tenha mais de um tipo de material no acervo, informe o total de amostras para cada tipo de material no acervo.

⁷⁴ Número de registro de amostras do acervo que estão armazenados em algum formato digital.

⁷⁵ Número de registros (informatizados ou não) que possuem uma coordenada geográfica associada (em latitude/longitude, UTM, ou outro sistema de coordenadas).

⁷⁶ Software específico para gerenciamento de coleções biológicas tais como: Biota (Robert Colwell), Specify, Biótica (CONABIO, México), Brahms, entre outros.

Existe estratégia para informatização e georeferenciamento das amostras do acervo?
Descreva-as sucintamente:

É utilizado algum vocabulário controlado nos dados da sua coleção? Qual a sua fonte (IBGE, ITIS, Sp2000, etc)? Em quais campos é utilizado (taxonomia, localidades, municípios, autores, coletores, etc)?

Indique outra informação que achar relevante:

IV. Acesso à Internet

Qual o tipo de acesso à internet:

- Nenhum
- Linha discada (via modem)
- Linha dedicada (indique a velocidade)

Os dados do acervo estão disponíveis via internet (quais)?

Qual a URL (endereço web da coleção)? http://

Responsável pela informação:

Data:

Favor enviar o questionário preenchido para vcanhos@cria.org.br

ANEXO II

PARTICIPAÇÃO EM EVENTOS CIENTÍFICOS

EVENTOS NO PAÍS

1. I Seminário de Monitoramento da Biodiversidade em UCs da Amazônia, 17 a 18 de Outubro de 2005, Manaus, AM. Participação de Alexandre Marino com a apresentação da palestra Banco de dados em Biodiversidade
2. 56º. Congresso Brasileiro de Botânica, 09 a 15 de outubro de 2005, Curitiba, PR.
 - Participação de Dora A. L. Canhos na Mesa Redonda "Avanços na Integração das coleções biológicas brasileiras" com a palestra: "Sistemas de informação abertos e as coleções biológicas"
 - Participação de Ingrid Koch com a apresentação do trabalho "Distribuição potencial de espécies de Rauvolfia (apocynaceae) e projeções para cenários climáticos do passado"
3. Reunião Técnica da Rede *speciesLink*, 22 de setembro de 2005, CNPTIA/EMBRAPA, Campinas, SP. Participação de Vanderlei P. Canhos, Dora A. L. Canhos, Sidnei de Souza, Marinez F. Siquiera, Ingrid Koch, Rafael L. Fonseca, Alexandre Marino, Renato De Giovanni
4. IBAMA: Seminário de Infra-estrutura de Dados sobre Biodiversidade e Sistemas de Informação (ênfase aos projetos *SinBiota* e *Species Link*), 03 de agosto de 2005, Brasília, DF. Participação de Vanderlei P. Canhos e Sidnei de Souza
5. Workshop Preparatório Programa Biota Bahia, 28 a 29 de julho de 2005, Salvador, BA. Apresentação do Programa Biota Fapesp por Vanderlei P. Canhos
6. The Annual Meeting of the Association for Tropical Biology and Conservation, 25 a 28 de Julho de 2005, Uberlândia, MG. Apresentação dos trabalhos "Using GIS and complex network to describe biological diversity: the distribution of tree species in Cerrado of São Paulo state" e "Pleistocene projections for megafauna dispersed-fruits in South America"
7. 19th Annual Meeting of the Society for Conservation Biology na UNB. (<http://www.scb2005.unb.br/symposiaschedule.htm#principles>), 15 a 19 de julho de 2005, Brasília, DF.
 - Participação de Dora A. L. Canhos na Mesa Redonda "Principles and tools for freely sharing conservation information" com a apresentação da palestra "Helping scientific collections share their information *on-line*".
 - Participação de Rafael L. Fonseca com a apresentação do painel: Predicting the potential the invasion of exotic marmosets (*Callithrix* spp.) in Southeastern Atlantic Forest: genetic contamination of endemic marmoset
8. Workshop: Diretrizes e Estratégias para a Modernização de Coleções Biológicas Brasileiras e a Consolidação de Sistemas Integrados de informação sobre Biodiversidade, 05 e 06 de julho de 2005, Brasília, DF. Participação de Vanderlei P. Canhos e Dora A. L. Canhos
9. Oficina de Planejamento - Biodiversidade na Secretaria de Estado de Ciência e Tecnologia do Espírito Santo, 14 de junho de 2005, Vitória, ES. Participação de Vanderlei P. Canhos
10. Workshop Centros de Recursos Biológicos e Avaliação da Conformidade de Material Biológico, 16 a 20 de maio de 2005, Campinas, SP. Apresentação das palestras:

- Vanderlei P. Canhos: "Coleções Microbiológicas, Centros de Recursos Biológicos e Conformidade do Material Biológico" e "De Coleções de Culturas à Centros de Recursos Biológicos, Iniciativas Internacionais"
 - Dora A. L. Canhos: Sistemas de Informação sobre Recursos Biológicos: CABRI, SICol e *speciesLink*
 - Renato De Giovanni: Padrões e Protocolos em Informática para Biodiversidade
11. 6º Workshop RNP, 10 de maio de 2005, Fortaleza, CE. Participação de Vanderlei P. Canhos
 12. Summer School on Environmental Modeling of Amazonia (Modelagem), 10 a 15 de abril de 2005, Angra dos Reis, RJ. Participação de Marinez F. de Siqueira
 13. Oficina de Modelagem no INPE: Rede GEOMA - Biodiversidade: Modelos de Distribuição de Espécies Amazônicas, 23 a 25 de fevereiro de 2005, Centro de Treinamento do INPE, São José dos Campos, SP. Participação de AMrinez F. Siqueira e Renato De Giovanni.
 14. Workshop Rede Fluminense, 08 de dezembro de 2004, Rio de Janeiro, RJ. Apresentação do "Species Link - Jardim Botânico" por Vanderlei P. Canhos e Sidnei de Souza
 15. Ciclo de Palestras de Atualização em Botânica, 10 de novembro de 2004, Instituto de Biociências em Botucatu – Unesp, Botucatu, SP. Participação de Marinez F. Siqueira e Ingrid Koch com as palestras:
 - O projeto *speciesLink* e *SinBiota*
 - Ferramentas de georeferenciamento e limpeza de dados
 - Uso dos dados: Modelagem de Biodiversidade
 16. Apresentação da Palestra "O Site CRIA e o uso do "Programa Species Link" aos pesquisadores do Instituto Agrônomo de Campinas (IAC), em 04 de novembro de 2004, Campinas, SP por Alexandre Marino.
 17. III Workshop sobre a biodiversidade na Bacia do Lago Tupé, 13 a 15 de outubro de 2004, Manaus, AM. Participação de Vanderlei P. Canhos
 18. Lançamento da Rede Species Link, 05 de outubro de 2004, São Paulo, SP. Participação de Vanderlei P. Canhos, Dora. A. L. Canhos, Sidnei de Souza, Renato De Giovanni, Alexandre Marino, Rafael L. Fonseca, Marinez F. Siqueira, Ingrid Koch, Alexandre Marino
 19. II FIAM - Feira Internacional da Amazônia, 15 a 18 de setembro de 2004, Manaus, AM. Participação de Vanderlei P. Canhos.
 20. Workshop Lista Oficial das Espécies da Flora Ameaçada de Extinção do Estado de São Paulo, 13 a 14 de setembro de 2004, São Paulo, SP. Participação de Rafael L. Fonseca em Mesa Redonda: "Importância das Listas Oficiais de Espécies da Flora Ameaçadas para a Conservação da Biodiversidade" com a palestra: *Mapeamento das espécies da flora ameaçadas de extinção no Estado de São Paulo como subsídio para a conservação.*
 21. XX Congresso Brasileiro de Entomologia, 09 de setembro de 2004, Gramado, SC. Participação de Ingrid Koch em Mesa Redonda "Coleções entomológicas brasileiras: responsabilidade científica versus responsabilidade legal" com a apresentação da palestra *A Rede speciesLink*
 22. I Encontro Nacional de Usuários de MAPESERVER, 03 a 05 de setembro de 2004, Itajaí, SC. Participação de Alexandre Marino

23. Reunião de Curadores do Species Link, 25 e 26 de agosto de 2004, CNPTIA/EMBRAPA, Campinas, SP. Participação de Vanderlei P. Canhos, Dora A. L. Canhos, Sidnei de Souza, Marinez. F. Siqueira. Ingrid Koch, Rafael L. Fonseca, Alexandre Marino, Renato De Giovanni
24. 56ª Reunião Anual SBPC, 18 a 23 de julho de 2004, Cuiabá, MT. Participação de Vanderlei Perez Canhos como Coordenador do Simpósio *Infomática na Fronteira da Ciência* e Coordenador da Mesa Redonda "Computadores nas Ciências da Vida"
25. 1º Simpósio de Frugivoria e Biologia da Conservação, UNESP, Rio Claro, SP 15 e 16 de junho de 2004. Participação de Rafael L. Fonseca com a palestra: *Recriando a megafauna: padrões de distribuição geográfica de frutos de megafauna no cerrado*
26. NAE: Mudança do Clima - Dia Internacional da Diversidade Biológica, 22 de maio de 2004, Brasília, DF. Participação de Vanderlei P. Canhos
27. Palestra Vanderlei Perez Canhos *Organização institucional de coleções científicas*. 21 de maio de 2004 – Sede Fiocruz, Rio de Janeiro, RJ.
28. Brasil - USA Workshop on Cooperation in Climate Science and Technology, 20 de maio de 2004, São José dos Campos, SP. Participação de Vanderlei P. Canhos
29. Inter-American Workshop on Environmental Data Access, 04 a 06 de março de 2003, Campinas, SP. Participação de Vanderlei P. Canhos, Dora A. L. Canhos. Sidnei de Souza, Alexandre Marino, Fabrício Pavarin, Marinez F. Siqueira, Ingrid Koch, Rafael L. Fonseca, Renato De Giovanni.
30. II Simpósio da Sociedade Brasileira de Entomologia, 09 e 10 de fevereiro de 2004, Brasília, DF. Participação de Marinez Ferreira de Siqueira com apresentação da palestra: *Modelagem da Biodiversidade - A Experiência do CRIA*
31. Congresso de Zoologia, 08 a 13 de fevereiro de 2004, Brasília, DF. Participação de Marinez Ferreira de Siqueira em mesa redonda "Banco de dados em biodiversidade" com a palestra: *O Projeto speciesLink*
32. IV Reunião de Avaliação do Programa BIOTA/FAPESP 12 e 13 de dezembro de 2003 –, Águas de Lindóia, SP. Participação de Vanderlei Perez Canhos, Sidnei de Souza e Renato de Giovanni
33. IV Simpósio BIOTA/FAPESP, 08 a 13 de dezembro de 2003, Águas de Lindóia, SP. Participação de Vanderlei Perez Canhos e equipe do CRIA na realização do Mini-Curso *Informatização de Coleções* e a apresentação dos seguintes trabalhos:
 - Sistema de Informação Distribuído para Coleções Biológicas: A integração do Species Analyst e *Sin*Biota, Renato de Giovanni e Vanderlei Perez Canhos
 - Modelagem para auxílio em estudos de controle biológico de *Lantana camara* L. (VERBENACEAE), Rafael L. Fonseca e Sérgio R. Morbiolo
 - Sistema de Análise de Similaridade de Áreas, Rafael L. Fonseca e J. Zullo Jr.
 - *Coletores de Plantas do Brasil*, Ingrid Koch, George Shepherd e Alexandre Marino
 - Previsão de distribuição geográfica de espécies de plantas tropicais lenhosas, Ingrid Koch, Marinez Ferreira de Siqueira e Andrew Townsend Peterson
 - *Sistema de Informação do Programa Biota/FAPESP: possibilidade de interação*, Alexandre Marino, Erica Speglish, Dora A. L. Canhos, Benedito A. Cruz, Sidnei de Souza e Carlos Alfredo Joly.
 - Ambiente computacional para modelagem de espécies. Mauro E. S. Muñoz.

- Species Mapper: uma ferramenta para visualização de pontos de ocorrência. Fabrício Pavarin, Alexandre Marino e Sidnei de Souza
34. XVIII Simpósio de Biologia Marinha, 28 a 30 de novembro de 2003., São Sebastião, SP. Participação de Dora A. L. Canhos com apresentação da palestra: *Distributed network based on DiGIR implemented for the speciesLink project in Brazil as a model for Ocean Biogeographic Information System of South America (OBISSA)*.
 35. SIPAM - Aplicações de Modelagem Ambiental, 06 a 08 de agosto de 2003, Manaus, AM. Apresentação da palestra: “Modelagem de Biodiversidade” por Marinez Ferreira de Siqueira e Mauro Muñoz.
 36. 54º. Congresso Nacional de Botânica, 13 a 18 de julho de 2003, Belém, PA.. Participação de Dora A. L. Canhos na Mesa-Redonda Nº 21 – Informática como Ferramenta para o Avanço da Taxonomia.
 37. Workshop – Modelagem de Biodiversidade 10 a 12 de fevereiro de 2003, Belém, PA. Apresentação dos trabalhos:
 - Ricardo Schachetti Pereira, GARP Genetic Algorithm for Rule-set Production
 - Marinez Ferreira de Siqueira, *Algoritmos Genéticos (GARP)*
 38. Curso de Especialização, Modalidade Extensão Universitária em Geoprocessamento, Universidade Estadual de Campinas – UNICAMP, Departamento de Geociências, 01 de fevereiro a 25 de novembro de 2003. Apresentação das seguintes monografias:
 - Alexandre Marino, *Ferramenta de classificação da qualidade da água da Bacia do Rio Piracicaba - PCJ* (<http://hydro.cria.org.br/iqa>).
 - Fabrício Pavarin, *Uso de SIG na identificação de áreas para recuperação ambiental da Bacia do Médio Paranapanema*.
 - Rafael L. Fonseca, *Sistema de Análise de Similaridade de Áreas*, <http://www.cria.org.br/similar/>
 39. I Workshop de Síntese e III Simpósio do Programa Biota / FAPESP, 24 a 28 de novembro de 2002, São Carlos, SP. Participação de Vanderlei P. Canhos, Dora A. L. Canhos, Érica Speglich, e Paula Drummond
 40. Fórum “Tendências e Desenvolvimentos em Informática para a Biodiversidade”, outubro de 2002, Indaiatuba, SP. Participação de Vanderlei P. Canhos, Dora A. L. Canhos, Fabrício Pavarin, Mauro Muñoz, Renato De Giovanni, Ricardo Scachetti Pereira, Rosely A. Coelho Lopes, Sidnei de Souza, Érica Speglich, Paula Drummond, Rafael. L. Fonseca, Daniela La Selva.
 41. Workshop Coleções Biológicas Desafios e Perspectivas no Brasil, Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Escola Nacional de Botânica Tropical, 05 a 07 de novembro de 2002, Rio de Janeiro, RJ. Apresentação da palestra “A experiência do CRIA com bancos de dados e sistemas de informação” por Dora A. L. Canhos
 42. 1a. Feira Internacional da Amazônia - Seminário sobre Biodiversidade, Biotecnologia e Bionegócios na Amazonia, 12 de setembro de 2002, Manaus, AM. Participação no painel Coleções e Sistemas de Informação para a Biodiversidade no encontro da Rede de Inventário, coleta e cultivo com a apresentação do trabalho “A experiência do CRIA com bancos de dados e sistemas de informação” por Dora A. L. Canhos
 43. International Colloquium on Invertebrate Pathology and Microbial Control, agosto de 2002. Apresentação da palestra “Electronic Publishing: open Access, integration, and interoperability” por Dora A. L. Canhos

44. 54º Reunião Anual da SBPC, 11 de Julho de 2002, Univ. Federal de Goiás, Goiânia, GO. Apresentação da palestra "Disseminação do Conhecimento Científico: a experiência do CRIA com publicações eletrônicas" por Sidnei de Souza.
45. 53o. Congresso Nacional de Botânica. Seminário sobre Banco de Dados botânicos no Brasil, 23 de julho de 2002, Recife, PE. Apresentação do trabalho "Sistemas de Informação" por Dora A. L. Canhos
46. Workshop sobre Geoprocessamento e Modelagem Ambiental - Desafios para a Amazônia no período 2002-2006, 5 a 7 de março de 2002, Petrópolis, RJ. Participação de Alexandre Marino
47. XIV Congresso da Sociedade Botânica de São Paulo, 2 a 5 de setembro de 2002, Rio Claro, SP. Apresentação da palestra "Modelagem de Nicho Ecológico para Predição de Distribuição Geográfica de Espécies Arbóreas de Cerrado" por Marinez Ferreira de Siqueira.

EVENTOS NO EXTERIOR

1. Species 2000, 8 a 12 de outubro de 2005, Londres, Inglaterra. Participação de Vanderlei P. Canhos
2. INCOFISH start-up workshop, 4 a 7 de outubro de 2005, Kiel, Alemanha. Participação de Dora A. L. Canhos
3. TDWG 2005 Meeting, 11 a 17 de setembro de 2005, St. Petersburg, Rússia. Participação de Renato De Giovanni.
4. BiodiversityWorld GRID Workshop, 30 de junho e 1 de julho de 2005, Edimburgo, Escócia. Participação de Renato De Giovanni
5. ETI BioInformatics sobre o Species 2000 (www.eti.uva.nl), 24 a 26 de abril de 2005, Amsterdam, Holanda. Participação de Vanderlei P. Canhos
6. GBIF 10th Governing Board Meeting e 3rd. Science Symposium, 17 a 23 de abril de 2005, Bruxelas, Bélgica. Participação de Vanderlei Perez Canhos e Renato De Giovanni
7. Fourth Inter-American Biodiversity Information Network (IABIN) Council, 6 a 8 de abril de 2005, Panamá. Participação de Vanderlei P. Canhos
8. Species 2000, 14 a 19 de março de 2005, Malta, Espanha. Participação de Vanderlei P. Canhos
9. GBIF SpeciesBank Workshop, 2 a 4 de março de 2005, Amsterdam, Holanda. Participação de Dora A. L. Canhos
10. First BioGeomancer Project Functional Requirements Workshop, no Museum of Vertebrate Zoology, University of Califórnia, 2 a 5 de março de 2005, Berkeley, EUA, Alexandre Marino e Sidnei de Souza.
11. 3rd IUCN - World Conservation Congress, 17 a 25 de dezembro de 2004, Bangkok, Tailândia. Participação de Dora A. L. Canhos, com a apresentação da palestra *Data Sharing: The Experience of SinBiota and speciesLink in Brazil*
12. 19ª Conferência Internacional - The Information Society New Horizons For Science e Assembléia Geral do ICSU's Committee on Data on Science and Technology - Codata www.codata.org. 7 a 10 de novembro de 2004, Berlim, Alemanha. Participação de Vanderlei P. Canhos na mesa redonda: "Interoperability – Biodiversity Data Systems" com apresentação da palestra: "Data Sharing: the experience of the speciesLink network in Brazil" sobre os resultados do "Workshop Inter-Americano sobre Acesso a Dados Ambientais" www.cria.org.br/eventos/iaed

13. Taxonomic Databases Working Group. Annual Meeting, 11 a 17 de outubro de 2004 - Biodiversity Informatics 2004, University of Canterbury, Christchurch, New Zealand. Participação de Renato De Giovanni com apresentação dos trabalhos
 - *The integration of DiGIR and BioCAsE*, em conjunto com Döring, M.; Hobern, D.; Vieglais, D.; Güntsch, A.; Blum, S.; Wieczorek, J. & de la Torre, J.
 - *openModeller: An open, collaborative environmental niche modelling toolkit* em conjunto com Tim Sutton (BDWorld), Mauro E. S. Muñoz (Cria), Ricardo Pereira Scachetti (University of Kansas), Peter Brewer (BDWorld) e Neil Caithness (BDWorld).
http://www.tdwg.org/2004meet/paperabstracts/TDWG_2004_Papers_Sutton_1.htm
14. 1st Meeting of the OBIS Regional Nodes, 16 a 18 de setembro de 2004, Dartmouth, Nova Scotia, (Atlantic) Canada. Participação de Alexandre Marino com a palestra *The Brazilian experience in setting up a distributed network on biological data*
15. GBIF Data Portal, 16 de agosto a 3 de setembro de 2004, Copenhagen, Dinamarca. Participação de Mauro H. Muñoz
16. General Meeting about Unification of DiGIR and BioCase, Museu Botânico, 24 de Julho a 12 de agosto de 2004, Berlim, Alemanha. Participação de Renato De Giovanni.
17. ELPUB 2004 - 8ª ICCO International Conference on Electronic Publishing, 23 a 26 de junho de 2004, Brasília, DF. Participação de Vanderlei Perez Canhos com a palestra *Networking Biological Diversity Information: Linking Local to Global Efforts* e na mesa redonda: *Digital Technologies and Media for Publishing Purposes*.
18. International Workshop on Strategies for Preservation and Open Access to Scientific Data, 22 a 24 de junho de 2004 -. Beijing, China. Participação de Dora A. L. Canhos no CODATA Task Group on *Preservation and Archiving of Scientific and Technical Data in Developing Countries*
19. Biodiversity Commons Meeting 25 a 26 de maio de 2004, IUCN, Genebra, Suíça. Apresentação da Palestra: *The experience with data sharing at CRIA (Reference Center on Environmental Information)* por Dora A. L. Canhos
20. Species 2000 Annual Meeting, 25 a 30 de abril de 2004, Washington, USA. Participação de Vanderlei P. Canhos.
21. GBIF Science Symposium – Biodiversity Data Users Serving Science and Society. 27 a 28 de abril de 2004 –, Oaxaca, México. Apresentação das seguintes palestras:
 - The brazilian Biota/FAPESP Virtual Institute of Biodiversity, Vanderlei Perez Canhos
 - Biodiversity data users serving science and society Renato De Giovanni
22. Seminário sobre Bioinformática 17 a 19 de novembro de 2003 -, San Jose, Costa Rica. Participação de Mauro E. S. Muñoz com a apresentação das seguintes palestras: "*speciesLink - La experiencia brasileña en montar una red*" e "*ACME - Un ambiente computacional para modelagem de especies*".
23. Reunião do Liaison Group on the Biosafety Clearing House, Secretariado da Convenção sobre Diversidade Biológica, 10 e 11 de abril de 2003, Montreal, Canadá. Participação de Dora A. L. Canhos
24. Workshop: Promoting Best Practices for Conservation Sustainable Use of Biodiversity of Global Significance in Arid and Semiarid Zones, 18 a 22 de Março de 2002, Santiago, Chile. Apresentação do trabalho "BIOTA/FAPESP – The Virtual Institute of Biodiversity: a Brazilian experience that can be used as model for research program on conservation and sustainable use of the biodiversity" por Érica Speglich.

25. Sp2000 Team meeting; Catalogue of Life Workshop; Outreach and GBIF Capacity Building Committee and GBIF Governing Board, 13 a 22 de março de 2002, Austrália. Participação de Vanderlei P. Canhos e Ricardo S. Pereira.
26. "Workshop on Exchange Format for Species Related Data" , co-financiado pela American Type Culture Collection (ATCC) US National Science Foundation (NSF) e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), 16 a 18 de janeiro de 2002, Miami, FL, USA. Participação de Vanderlei Canhos, Sidnei de Souza e Ricardo Scachetti Pereira.
27. Workshop: "Testing Alternative Methodologies for Modeling Species Ecological Niches and Predicting Geographic Distributions" promovido pelo University of Kansas Natural History Museum and Biodiversity Research Center e Museum of Vertebrate Zoology, University of California at Berkeley e financiado pelo National Center for Ecological Analysis and Synthesis (NCEAS), 27 de maio a 02 de junho de 2002, Santa Barbara, CA, USA. Participação de Ricardo Schachetti Pereira
28. Workshop "Data Access and Database Interoperability" do Global Biodiversity Information Facility (GBIF), 27 e 28 de junho de 2002, San Diego, CA, USA. Participação de Ricardo Schachetti Pereira

ANEXO II PUBLICAÇÕES E RELATÓRIOS TÉCNICOS

- Bonaccorso, E., **Koch, I.**, Peterson, A.T. Pleistocene fragmentation of Amazon species' ranges. Diversity and Distributions. Accepted. 2005 (*Science*)
- Camaeron, A., Thomas, C.D., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., **Siqueira, M.F.**, Grainger, A., Hannah, L., Hughes, L., Huntley, B., Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O. & Williams, S.E. 2004. Will climate change catch us off guard? 2004. *Conservation In Practice* V5(2):28-30.
- Canhos, D. A. L. ; Canhos, V. P.; Souza, S.; Siqueira, M. F.; Muñoz, M.; De Giovanni, R.; Marino, A.; Koch, I.; Fonseca, R. L.; Umino, C. Y.; Lopes, R. A. C.; Cruz, B. A. & Albano, A. P. S. Relatório Técnico Anual: Sistema de Informação Distribuído para Coleções Biológicas: a Inegração do Species Analyst e *Sin*Biota. FAPESP, Outubro, 2004
- Canhos, D. A. L.; Canhos, V. P. & Souza, S. S.** Coleções Biológicas e Sistemas de Informação. In: Diretrizes e Estratégias para a Modernização de Coleções Biológicas Brasileiras e a Consolidação de Sistemas Integrados de Informação sobre a Biodiversidade. Centro de Estudos Estratégicos (CGEE) e Ministério da Ciência e Tecnologia (<http://www.cria.org.br/cgee/col/documentos>), 2005-10-18
- Canhos, D. A. L.; Souza, S.; De Giovanni, R.; Marino, A.; Siqueira, M. F.; Cruz, B. A. & Canhos, V. P.** Estudo de Caso: Sistemas de Informação *On-line* - A experiência do CRIA. In: Diretrizes e Estratégias para a Modernização de Coleções Biológicas Brasileiras e a Consolidação de Sistemas Integrados de Informação sobre a Biodiversidade. Centro de Estudos Estratégicos (CGEE) e Ministério da Ciência e Tecnologia (<http://www.cria.org.br/cgee/col/documentos>), 2005.
- Canhos, D.A.L., S. Souza e V.P.Canhos** Sistemas de Informação e Centros de Recursos Biológicos. III Simpósio de Recursos Genéticos para a América Latina e Caribe – SIRGEALC. Londrina, Paraná, Brasil. 19-22 de novembro de 2001
- Canhos, D.A.L., S. Souza e V.P.Canhos.** Electronic Publishing: open Access, integration, and interoperability. Em: International Colloquium on Invertebrate Pathology and Microbial Control (8.: 2002: Foz do Iguaçu, PR). ISSN 1516-781X; n. 184.
- Canhos, D.A.L.,** Uhlir, P.F., Esanu, J.M. (Report editors). Access to Environmental Data: Summary of an Inter-American Workshop. Codata. Julho, 2004.
<http://www.codata.org/FINAL%20Summary%20Report%20on%20Inter%20American%20Workshop.pdf>
- Canhos, D.A.L.; Chapman, A.; Canhos, V.P.** Study on Data-sharing with Countries of Origin. Relatório preparado para o Global Biodiversity Information Facility (GBIF), Contract Report GBIFS/2003/04. 2004. <http://www.gbif.org/News/NEWS1082472796>.
- Canhos, D.L, S. Souza e V.P.Canhos.** O uso de redes eletrônicas em biodiversidade, Em: Biodiversidade: valor econômico e social, Com Ciência, Junho de 2001
<http://www.comciencia.br/reportagens/biodiversidade/frameset/vogt.htm>
- Canhos, V.P., Canhos D.A.L., Souza, S.** Construindo a Rede Interamericana de Informação sobre Biodiversidade – IABIN, Sub-região 2 – Brasil, Fase II. Relatório para a Organização dos Estados Americanos. Julho, 2004.
- Canhos, V.P., Canhos D.A.L., Souza, S.** Informação Ambiental e Práticas de Cidadania em "Práticas de Cidadania". 89-100. organizador: J. Pinsky. Editora Contexto. ISBN: 85-7244-265-0. 2004.
- Canhos, V.P., Canhos D.A.L., Souza, S.** Informática para Biodiversidade no Brasil. Relatório para a Organização dos Estados Americanos. Julho, 2004.

- Canhos, V.P., Souza, S., De Giovanni, R., Canhos, D.A.L.** Global Biodiversity Informatics: Setting the Scene for a “New World” of Ecological Modeling. *Biodiversity Informatics*, 1, 2004, pp. 1-13, 2004
- Canhos, V.P.** Centros de Recursos Biológicos: suporte ao desenvolvimento científico e inovação tecnológica. *Ciência e Cultura* 55:01, p. 27-29, 2003
- Chapman, A. D.; **Muñoz, M. E. S. & Koch, I.** Environmental information: placing biodiversity phenomena in an ecological and environmental context. "Biodiversity Informatics, 2, 2005, pp.24-41 (<http://jbi.nhm.ku.edu>)
- Chapman, A.D.** Sistema de Informação Distribuído para Coleções Biológicas: Project Report. Fapesp/Biota process no. 2001/02175-5. March 2003 - March 2004. <http://splink.cria.org.br/docs/report.pdf>
- Doring M. & **De Giovanni R.** A unified protocol for search and retrieval of distributed data. Preparado para o Global Biodiversity Information Facility (GBIF), Contract Report GBIFS – CRIA. September, 2004. <http://www.cria.org.br/protocols/newprotocol.pdf>
- Ferraz, K.; **Scachetti-Pereira, R.**; Peterson, A. T. Distribuição da capivara na bacia do Rio Piracicaba no Estado de São Paulo. Aceito para publicação nos anais do IALE 2003 (Simpósio sobre landscape ecology).
- Fiaboe, K. K.M., **Fonseca, R. L.**, Moraes, G. J., Ogot, C. K.P.O. and Knapp, M. Identification of priority areas in South America for exploration of natural enemies for classical biological control of *Tetranychus evansi* (Acari: Tetranychidae) in Africa. *Biological Control*. (submitted), 2005
- Fonseca, R. L.** *Sistema de Análise de Similaridade de Áreas*, <http://www.cria.org.br/similar/>. Monografia apresentada no Curso de Especialização, Modalidade Extensão Universitária em Geoprocessamento, Universidade Estadual de Campinas – UNICAMP, Departamento de Geociências, novembro/2003
- Fonseca, R. L.**, Guimarães, P.R. & Galetti, M. Pleistocene projections for megafauna dispersed-fruits in South America. The Annual Meeting of the Association for Tropical Biology and Conservation, Uberlândia, MG, Brasil, 2005
- Fonseca, R. L.**; Bueno, R.; Guimarães, P.R. & Galetti, M. Predicting the potential the invasion of exotic marmosets (*callithrix* spp.) in southeastern atlantic forest: genetic contamination of endemic marmoset. XIX Annual Meeting of the Society for Conservation Biology, Brasília, DF, Brasil, 2005.
- Fonseca, R. L.**; Guimarães, B. R.; Morbiolo, S. R.; Pereira, R. S. & Peterson, A. T. Vulnerability of brazilian national parks to invasion by the alien weed *Crotalaria pallida* Ait. (Fabaceae). *Weed Science*. In press, 2005
- Guimarães, P.R., **Fonseca, R. L.** & Galetti, M. Using GIS and complex network to describe biological diversity: the distribution of tree species in Cerrado of São Paulo state. The Annual Meeting of the Association for Tropical Biology and Conservation, Uberlândia, MG, Brasil, 2005
- Koch, I.**; Peterson, A. T. & Shepherd, G. J. Distribuição Potencial de espécies de *Rauvolfia* (Apocynaceae) e projeções para cenários climáticos do passado. In: Resumos do 56o. Congresso Nacional de Botânica, Curitiba, PR, 2005
- Marino, A.** *Ferramenta de classificação da qualidade da água da Bacia do Rio Piracicaba (PCJ) na WEB* <http://hydro.cria.org.br/iqa>. Monografia apresentada no Curso de Especialização, Modalidade Extensão Universitária em Geoprocessamento, Universidade Estadual de Campinas – UNICAMP, Departamento de Geociências, novembro/2003

- Mauro Galetti, Camila I. Donatti, Marco Aurélio Pizo, Paulo R. Guimarães Jr., **Rafael Luís Fonseca** and Pedro Jordano. *Em revisão*. Living in the land of the megafauna: fruit traits in the Pantanal of Brazil. *Frugivores and Seed Dispersal*
- Meirelles, L. D.; Sheperd, G. J.; **Koch, I. & Siqueira, M. F.** Modelagem da distribuição geográfica de *Araucaria angustifolia* com projeções para cenários climáticos do passado. In: Resumos do 56o. Congresso Nacional de Botânica, Curitiba, PR, 2005
- Pavarin, F.** *Uso de SIG na identificação de áreas para recuperação ambiental da Bacia do Médio Paranapanema*. Monografia apresentada no Curso de Especialização, Modalidade Extensão Universitária em Geoprocessamento, Universidade Estadual de Campinas – UNICAMP, Departamento de Geociências, novembro/2003
- Peterson, A. T.; **Koch, I.; Scachetti-Pereira, R.**; Navarro-Sigüenza, A. G. Detectando problemas de identificação em conjuntos de dados sobre biodiversidade baseado em modelagem de nichos ecológicos. "Diversity and Distributions". 2003.
- Peterson, A. T.; Navarro-Sigüenza, A. G. & **Scachetti-Pereira, R.** Detecting errors in biodiversity data: Collectors' itineraries flag mislabeled specimens. "Bulletin of the British Ornithologists' Club". 124(2), 2004
- Peterson, A. T.; **Scachetti, R. P.** & Kluza, D. A. Avaliação do potencial de invasão de *Homoladisca coagulata* na costa oeste da América do Norte e na América do Sul. *Rev. Biota Neotropica*, v.3, n.1, 2003.
<http://www.biotaneotropica.org.br/v3n1/pt/abstract?article+BN00703012003>
- Rafael Luís Fonseca**, Juliana Gastaldello Rando e Vinicius Castro Souza. *Em revisão*. Mapeamento das espécies da flora ameaçadas de extinção no Estado de São Paulo como subsídio para a conservação. *Livro vermelho das espécies da flora ameaçadas de extinção no Estado de São Paulo*.
- Siqueira, M. F.** Uso de Modelagem de nicho fundamental na avaliação do padrão de distribuição geográfica de espécies vegetais. Tese de Doutorado. Ciências da Engenharia Ambiental, Escola de Engenharia de São Carlos, USP, Brasil, 2005
- Siqueira, M. F.**, Durigan, G., Peterson, A. T. Modelagem de distribuição geográfica para espécies com distribuição restrita In: 54 Congresso Nacional de Botânica, 2003, Belém, PA (<http://www.adaltech.com.br/evento/museugoeldi/resumoshtm/resumos/R0884-1.htm>)
- Siqueira, M. F.**, Peterson, A. T. Consequences of global climate change for geographic distributions of cerrado tree species. *Biota Neotropica*, v.3, n.2, 2003
(<http://www.biotaneotropica.org.br/v3n2/pt/download?article+BN00803022003+item>)
- Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., **Siqueira, M. F.**, Grainger, A., Hannah, L. Extinction risk from climate change. *Nature*, v.427, n.8, p.145 - 148, 2004.
<http://www.nature.com/nature/journal/v427/n6970/abs/nature02121.html;jsessionid=E578E12B630926FB406940E88F4B07F3>
- Uehara-Prado, M. & **Fonseca, R. L.** Distribution of the Fluminense Swallowtail in Brazil: Urbanization, protected areas and conservation of a threatened butterfly (submitted), 2005

ANEXO III. O PROJETO NA MÍDIA

- *Principles and Methods of Data Cleaning* Acrobat PDF 2.34 MB, Julho/2005
http://www.gbif.org/prog/digit/data_quality
- *Principles of Data Quality* Acrobat PDF 0.73 MB, Julho/2005
http://www.gbif.org/prog/digit/data_quality
- *Uses of Primary Species Occurrence Data* Acrobat PDF 2.38 MB, Julho/2005
http://www.gbif.org/prog/digit/data_quality
- Unesp/Rio Preto tem coleção científica de ácaros, 22/06/2005
http://www.universia.com.br/html/noticia/noticia_dentrodocampus_caiji.html
- Unesp/Rio Preto tem coleção científica de ácaros, 21/06/2005
<http://www.unesp.br/proex/universia/acaro.php>
- *Sharing and Accessing Biodiversity Data Globally through GBIF* (apresentado no paper ESRI User Conference 2005-07-25 in San Diego), 03/05/2005
http://circa.gbif.net/Public/irc/gbif/ict/library?l=/papers/esriuserconference/_EN_0.2_&a=d
- Rede óptica da RNP, Redecomep e Clara são assuntos do segundo dia do WRNP, 11/05/2005 <http://www.rnp.br/noticias/2005/not-050512.html>
- *Resources for Ecology, Evolutionary Biology, Systematics, and Conservation Biology*, 08/02/2005 <http://darwin.eeb.uconn.edu/links/link.php?id=414>
- BIOTA/FAPESP - um modelo para programas de pesquisa em caracterização, conservação e uso sustentável da biodiversidade, 01/02/2005
http://www.jornalcana.com.br/conteudo/noticia.asp?area=AcaoSocial%26Meio+Ambiente&secao=Meio+Ambiente&id_materia=15252
- Coleções do Instituto Biológico integram base de dados financiada pela FAPESP - Parte I, No. 11, Jan/05 http://www.biologico.sp.gov.br/bioin_janeiro05.htm
- Natureza sob novo olhar: Programa Biota completa 5 anos – 26/12/2004
<http://www.ambientebrasil.com.br/noticias/index.php3?action=ler&id=17389>
- Biota/Fapesp: Natureza sob novo olhar – 20/12/2004
<http://www.jornaldaciencia.org.br/Detail.jsp?id=24210>
- Natureza sob novo olhar, 20/12/2004 - Edição 10, Dez/04
http://www.agencia.Fapesp.br/boletim_dentro.php?data%5Bid_materia_boletim%5D=3030
- Natureza sob novo olhar, 17/12/2004 -
http://www.fundacaoboticario.org.br/site/br/educacao/noticias_34.htm
- *Clipping*, 1/12/2004 - http://www.ecolnews.com.br/clipping/dezembro_04.htm
- Pesquisa quer aumentar acesso a informações ambientais. Aventura Brasil.com.br, 20/10/2004 - <http://www.zone.com.br/aventurabrasil/>
- Fórum "Tendência e Desenvolvimentos em Informática para Biodiversidade". Jornal da Ciência, 11/10/2004
- Arca de Noé Digital. Revista Pesquisa FAPESP, ed. 104 - 10/2004
- Biodiversidade ganha rede. Agência de Notícias Fapesp, 6/10/2004
- Biodiversidade sai dos Museus para a Internet. O Estados de São Paulo, 6/10/2004
- Lançada em outubro, na sede da FAPESP, a Rede species link, um sistema de informação distribuído para a integração de dados de coleções biológicas, notícia na Revista *on-line* da Sociedade Botânica do Brasil, 04/11/2004 -
<http://www.botanica.org.br/noticias.php?noticia=77>

- Portal da Biodiversidade. Agência de Notícias Fapesp, 22/7/2004
- Pesquisas buscam aumentar acesso a informações ambientais. Eco Agência Informações em rede, 16/3/2004
- Painel da Semana: Botânica. Jornal da Unicamp, ed.242, de 01 a 07/03/04
- A terra mais quente. Revista Pesquisa Fapesp, ed. 96 - fev/2004
- Rede da Vida, 03/10/2003 -
<http://www.portaldomeioambiente.com.br/noticias.asp?tarefa=mostra&id=401>
- A Mata Atlântica e o aquecimento global, 10/8/2002,
<http://www.comciencia.br/reportagens/clima/clima15.htm>
- Rede da Vida. Revista Pesquisa FAPESP, ed. 82 - 12/2002
(http://www.revistapesquisa.Fapesp.br/show.php?id=revistas1.Fapesp1..20021202.20021282..SEC3_5)
- Modelagem distribui espécies com precisão. Revista Ciência e Cultura, out/nov/dez/2002.
- Simuladores de Futuro. Revista Pesquisa FAPESP, 10/2002
- PROSSIGA/IBICT Bases de Dados Brasileiras na Internet -
<http://www5.prossiga.br/basesdedados/asp/ListaRegCNPq.asp?cod=2.05.00.00-9&id=port>
- *Workshop on Strategies for the Preservation of and Open Access to Digital Scientific Data* - http://www7.nationalacademies.org/usnc-codata/China_Bios.html
- Programa BIOTA/FAPESP o modelo brasileiro para gestão de um recurso estratégico: a biodiversidade
http://www.desafios.org.br/conferencia/files/cncti3_artigos-5-Artigo.doc
- Citação na *Newsletter of Systematic Ichthyology* No. 25, last issue from C.A.S. -
http://www.calacademy.org/research/ichthyology/NLSI/newsletter_pdfs/NLSI_03.pdf
- Citação no artigo: *Comments on: Consultation Paper: Developing a National Biodiversity and Climate Change Action Plan* -
<http://www.deh.gov.au/biodiversity/publications/nbccap-consultation/submissions/pubs/chapman-a.pdf>