

# Rede speciesLink: avaliação 2006

## Índice

Sumário Executivo .....	i
Introdução .....	1
Volume e Qualidade .....	1
Herbários e Xilotecas .....	6
Evolução dos dados on-line.....	6
Repatriação de dados.....	8
Flora brasiliensis.....	8
Abelhas .....	9
Dados on-line.....	9
Catálogo de Abelhas Moure (moure.cria.org.br) .....	9
Tecnologia .....	10
Ferramentas de limpeza de dados ( <i>data cleaning</i> – splink.cria.org.br/dc).....	10
speciesBase (splink.cria.org.br/speciesbase) .....	11
spLinker (splink.cria.org.br/splinker).....	11
openModeller (openmodeller.sourceforge.net/) .....	11
Uso.....	11

**Equipe speciesLink**

Vanderlei Perez Canhos (coordenação)

Dora Ann Lange Canhos

**Desenvolvimento**

Sidnei de Souza

Alexandre Marino

Renato de Giovanni

Giscar Fernandes Vicente de Paiva

Tim Sutton

**Suporte**

Benedito A. Cruz

Luiz Henrique Pires Barione

**Apoio**

Cristina Y. Umino

Rosely Coelho

Leandro Luis Filippi

Vanessa Barbosa Villar

## Sumário Executivo

No ano de 2006 ocorreram avanços significativos na rede speciesLink. Durante a fase de implantação do projeto, com apoio da Fapesp (outubro de 2001 a 2005), foram desenvolvidos alguns indicadores de acompanhamento (*volume, qualidade, tecnologia, uso*) e a presente análise baseia-se na evolução desses indicadores.

O primeiro indicador diz respeito ao **volume** de dados e número de coleções participantes da rede. Em 2006 o volume de dados aumentou 63%, passando de cerca de 700 mil registros para cerca de 1,16 milhão. Esse aumento deve-se principalmente a dois fatores: (i) ao esforço de digitação do Jardim Botânico do Rio de Janeiro, que saltou de 6 mil para mais de 275 mil registros on-line; e (ii) à integração dos dados das coleções associadas à Rede Paranaense de Coleções Biológicas, a *Taxonline*.

O número de coleções e subcoleções participantes da rede speciesLink também aumentou no período. Em outubro de 2005 a rede incluía os acervos de 41 coleções e subcoleções pertencentes a três universidades estaduais e sete institutos de pesquisa do Estado de São Paulo e ao Jardim Botânico do Rio de Janeiro, além dos dados de observação do sistema *SinBiota* do programa Biota/Fapesp. A rede hoje conta com a participação de coleções do Estado de São Paulo, da Rede Paranaense, da Rede Fluminense, da rede SICol (coleções microbianas) e do SinBiota. São 68 coleções e subcoleções integradas em rede, sendo que dessas, cinco no momento não estão disponibilizando dados.

Com o aumento do número de coleções participantes, o número de registros nos acervos das coleções associadas à rede speciesLink saltou de 2,4 milhões para 6,1 milhões, o que representa um aumento de mais de 150%. Esse aumento é devido, principalmente à entrada das coleções de insetos da Universidade Federal do Paraná com mais de 3 milhões de exemplares. Como essas coleções têm pouco material digitado, a porcentagem de registros on-line caiu de 30% dos registros de todos os acervos para 19%. No entanto, o índice de geo-referenciamento aumentou, de 42% dos registros on-line georeferenciados em 2005 para 46% em 2006.

Em termos de registros on-line, 60% dos dados são de herbários seguido por insetos (13%), principalmente abelhas, e peixes (11%). O *SinBiota* contribui com 8% do total. Um outro dado importante diz respeito ao local das coletas. 88% dos registros são de coletas realizadas no país sendo que desse total, menos de 20% é do Estado de São Paulo.

Ainda dentro do indicador *volume*, tem-se a **repatriação de dados**. Durante o projeto, foram integradas algumas coleções do exterior, com o intuito de testar o protocolo e a viabilidade técnica do país em participar de uma rede internacional. Foram integrados os acervos do Museu de História Natural e o Herbário da Suécia e os acervos da Universidade de Kansas. Os acervos disponíveis somavam mais de 4 milhões de registros, mas com apenas 12 mil registros de material coletado no Brasil. Em 2006 foi estabelecida uma parceria com os Jardins Botânicos de Nova Iorque e de Missouri que passaram a integrar os seus dados à rede speciesLink. Os registros foram filtrados para somente incluir as coletas realizadas no Brasil. Assim, foram integrados mais 188 mil registros sendo 60 mil georeferenciados. Esses números representam um aumento 30% dos registros de herbários on-line.

Visando melhorar a qualidade dos dados disponíveis na rede speciesLink, foram desenvolvidas duas linhas de ação: (i) o desenvolvimento de ferramentas para auxiliar as coleções no trabalho de detecção de possíveis erros no registro de seus dados, e (ii) o desenvolvimento de listas de nomes que poderão ser usadas como dicionários.

Ainda com respeito à qualidade dos dados, o Brasil precisa de listas de nomes científicos para validar os nomes dos espécimes de seus acervos. Em 2006 o CRIA desenvolveu trabalhos com plantas e abelhas.

Com o apoio da Fapesp, Fundação Vitae e Natura, e em colaboração com o Jardim Botânico de Missouri e a Unicamp, foi desenvolvido o sistema Flora brasiliensis on-line ([florabrasiliensis.cria.org.br](http://florabrasiliensis.cria.org.br)). O sistema integra imagens de quase 4 mil pranchas e mais de 10 mil páginas (textos) digitalizadas em alta resolução, on-line, com nomes e descrições de 210 famílias, 2.298 gêneros e 22.550 espécies. Para a atualização dos nomes foi desenvolvido o sistema *Flora*

*brasiliensis revisitada* ([flora.cria.org.br](http://flora.cria.org.br)) cuja proposta é servir de base para a construção colaborativa de uma lista de nomes de plantas brasileiras. No momento o sistema contém 186 gêneros e 2.271 espécies, sendo que a família *Bignoniaceae* está completa. Foi feito um estudo junto com o Jardim Botânico de Missouri e foram desenvolvidos aplicativos para importar os dados do sistema Trópicos para o sistema *Flora brasiliensis revisitada*. As primeiras 2 famílias cujos dados foram importados são *Rutaceae* e *Simaroubaceae* que serão revisados pelo Prof. José Rubens Pirani do Departamento de Botânica, IB/USP.

Em 2006 foi também lançado o sistema de informação on-line com o Catálogo de Abelhas Moure ([moure.cria.org.br](http://moure.cria.org.br)). O trabalho de revisão taxonômica foi coordenado pela Universidade Federal do Paraná (UFPR) contando com a colaboração de pesquisadores da USP-Ribeirão Preto e da Universidade Federal de Minas Gerais, além da própria UFPR. São mais de 2.500 nomes de abelhas, validados por especialistas, com os respectivos autores, dados sobre tipos, distribuição geográfica e referências bibliográficas.

Um terceiro indicador *tecnologia*, mede o desenvolvimento de ferramentas e aplicativos. Em 2006 foram desenvolvidas ferramentas para limpeza de dados e para modelagem de nicho ecológico de espécies (*openModeller*). O software para espelhamento dos dados das coleções em servidores regionais (*spLinker*) foi aprimorado e foi desenvolvido em software para gerenciamento de coleções (*speciesBase*).

O último indicador *uso* traz dados muito importantes. Somando o acesso à rede speciesLink e ao sistema *Flora brasiliensis* tivemos em 2006 mais de 170 mil visitantes únicos, 250 mil visitas, mais de 3 milhões de páginas acessadas, mais de 19 milhões de hits e um tráfego de dados de cerca de 180 Gbytes. Uma análise dos domínios e países que acessaram os sistemas tem-se que mais de 70% dos usuários são do Brasil e menos que 1% é proveniente da rede comercial. Os dados indicam que os sistemas estão sendo utilizados prioritariamente por usuários do Brasil, que são o nosso público alvo.

## Introdução

O CRIA tem como missão institucional disseminar conhecimentos científicos e tecnológicos e promover a educação, visando a conservação e a utilização sustentável dos recursos naturais e a formação da cidadania. Tem como meta e estratégia a disseminação de informação eletrônica, como ferramenta na organização da comunidade científica e tecnológica do país. Assim, a rede speciesLink atualmente é uma das principais ações institucionais em andamento.

A rede speciesLink dissemina dados de espécimes existentes em acervos do país e do exterior (com foco em acervos coletados no Brasil) de forma integrada com o banco de dados de observação do programa Biota/Fapesp o *SinBiota* ([sinbiota.cria.org.br/atlas](http://sinbiota.cria.org.br/atlas)) e o sistema de informação sobre coleções microbianas *SICol* ([sicol.cria.org.br](http://sicol.cria.org.br)). Nasceu de um projeto financiado pela Fapesp (de outubro de 2001 a 2005) e desenvolvido em parceria com coleções biológicas do país (no início restrito ao estado de São Paulo) e com grupos de informática para biodiversidade do país e do exterior.

O presente relatório tem por objetivo divulgar junto aos nossos parceiros a evolução da rede ao longo do ano de 2006. A avaliação parte dos indicadores estabelecidos no projeto que são: *volume, qualidade, tecnologia e uso*.

## Volume e Qualidade

Por *volume* entendemos o número de registros disponíveis de forma livre e aberta a qualquer interessado na Internet. A tabela 1 apresenta a evolução da rede com os dados de outubro de 2003, 2004 e 2005 e dezembro de 2006.

Com base nesses dados podemos apresentar a evolução da rede speciesLink como sendo algo com crescimento constante (figura 1).

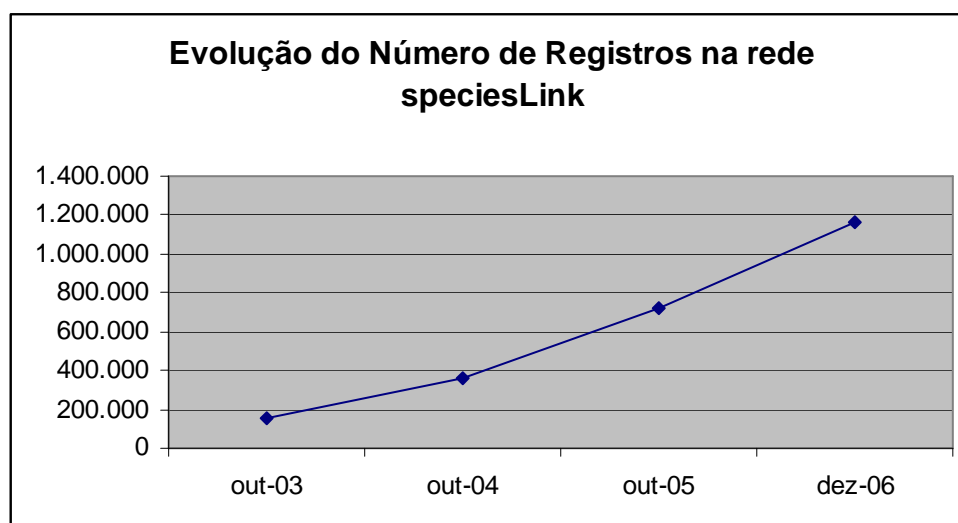
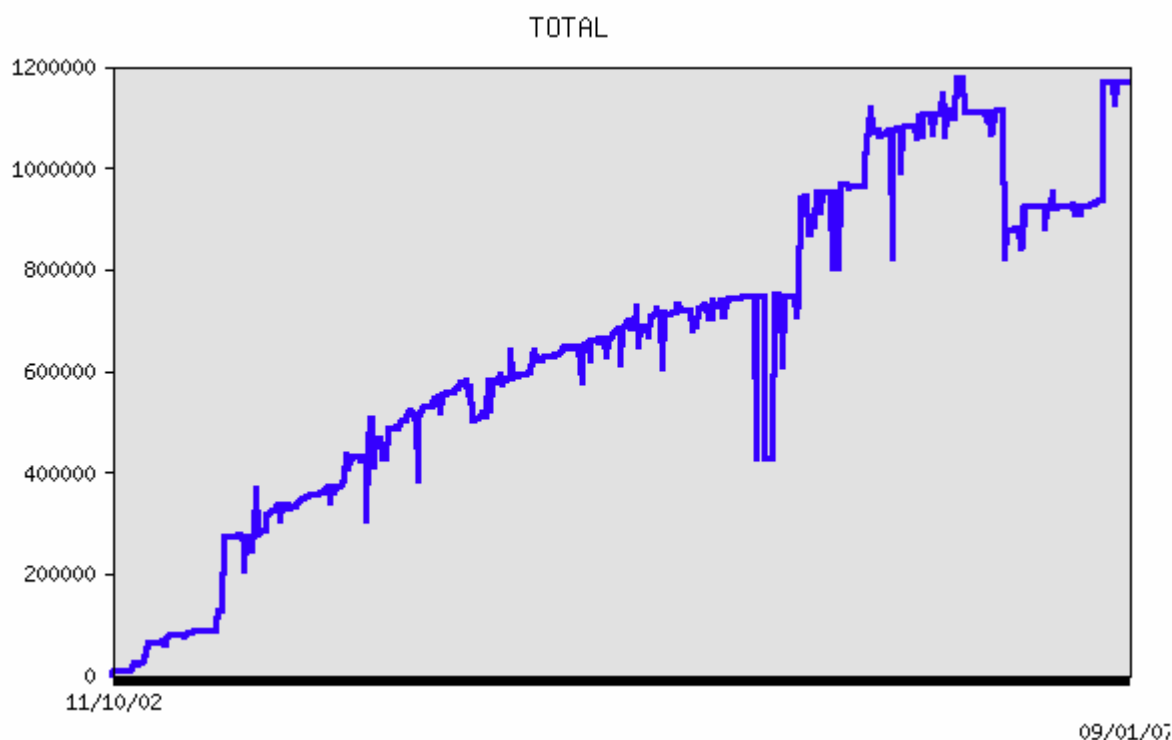


Figura 1. Evolução da rede speciesLink

No entanto a figura 2 é um retrato fiel da dinâmica da rede.



**Figura 2. Gráfico demonstrativo da entrada e saída de dados da rede speciesLink**

Podemos perceber que existem momentos onde há uma elevação do patamar dos números. A entrada da Coleção de Peixes do Museu de Zoologia da USP, a inclusão da rede Paranaense de Coleções Biológicas *Taxonline* ([www.taxonline.ufpr.br/](http://www.taxonline.ufpr.br/)) e do Jardim Botânico do Rio de Janeiro são exemplos desses “saltos” que a rede teve ao longo do tempo.

Ao todo, a rede possui hoje 59 coleções e sub coleções, além do SICol (com 9 coleções microbianas) e o sistema SinBiota do programa Biota/Fapesp. São, portanto 68 coleções ou subcoleções integradas à rede. 20 novas coleções foram integradas em 2006, sendo que 5 ainda não disponibilizaram seus dados.

Se separarmos os dados por grupo taxonômico, temos a seguinte fotografia para o final de 2006:

Coleções Biológicas	Acervo Total		Evolução do Acervo	Registros On-line	% on-line	Registros geo-ref	% geo-ref	Evolução dos registros on-line
	2005	2006						
Herbários e Xilotecas	1.429.161	1.854.517	30%	638.915	34%	234.908	37%	124%
Peixes	128.023	125.371	-2%	112.343	90%	66.717	59%	16%
Aracnídeos	35.459	33.697	-5%	26.444	78%	3.777	14%	-4%
Répteis e Anfíbios	111.179	121.787	10%	101.049	83%	8.227	8%	8%
Mamíferos	34.063	34.476	1%	19.766	57%	815	4%	0%
Insetos	508.171	3.744.061	637%	135.955	4%	113.316	83%	27%
Ascidiacea		600		408	68%	322	79%	%
Aves	112.224	112.223	0%	23.179	21%	12.396	53%	-1%
Microrganismos	2.688	8.724	225%	8.724	100%	0	0%	0%
<b>Acervos</b>	<b>2.360.968</b>	<b>6.035.456</b>	<b>156%</b>	<b>1.066.783</b>	<b>18%</b>	<b>440.478</b>	<b>41%</b>	<b>63%</b>
Dados de Observação	63.586	98.758	55%	98.758	100%	98.758	100%	55%
Coleção de Sons		7.000		577	8%	457	79%	
<b>Total Geral</b>	<b>2.424.554</b>	<b>6.141.214</b>	<b>153%</b>	<b>1.166.118</b>	<b>19%</b>	<b>539.693</b>	<b>46%</b>	<b>63%</b>

O potencial da rede (número de registros nos acervos) cresceu 156%, principalmente devido à entrada das coleções de insetos da Universidade Federal do Paraná. Em termos do acervo o grupo mais numeroso é o dos insetos representando 61% do total seguido pelos herbários e xilotecas com 31% do total (figura 3).

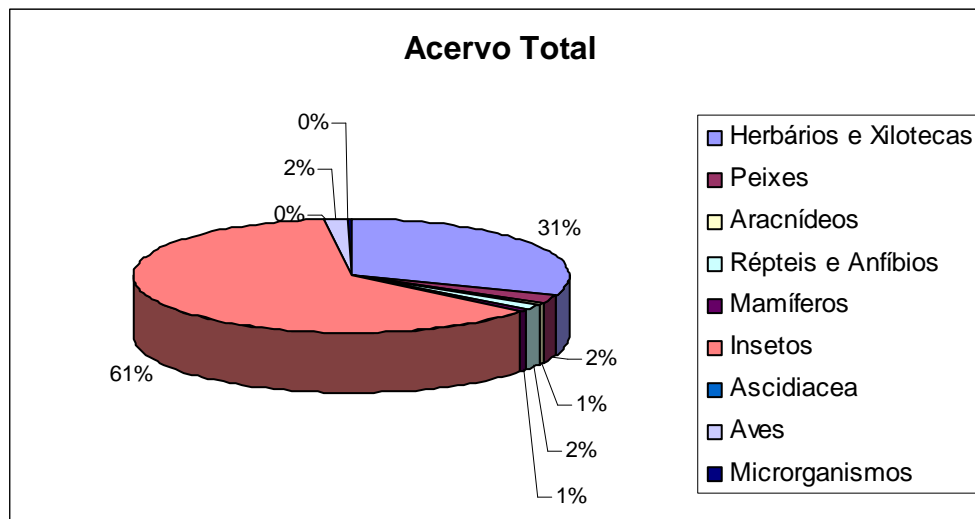


Figura 3. Porcentagem do número de registros nas coleções por grupo taxonômico

No entanto, se analisarmos somente os registros disponíveis on-line (figura 4) temos que herbários e xilotecas representam 60% do total, seguido por peixes (13%) e insetos (11%). Podemos supor que o desafio para informatizar os acervos de insetos certamente é maior.

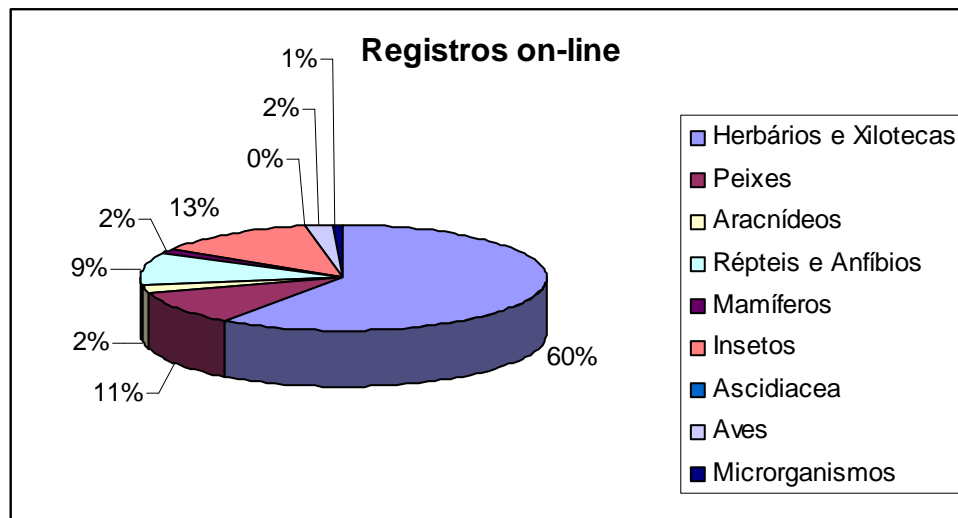
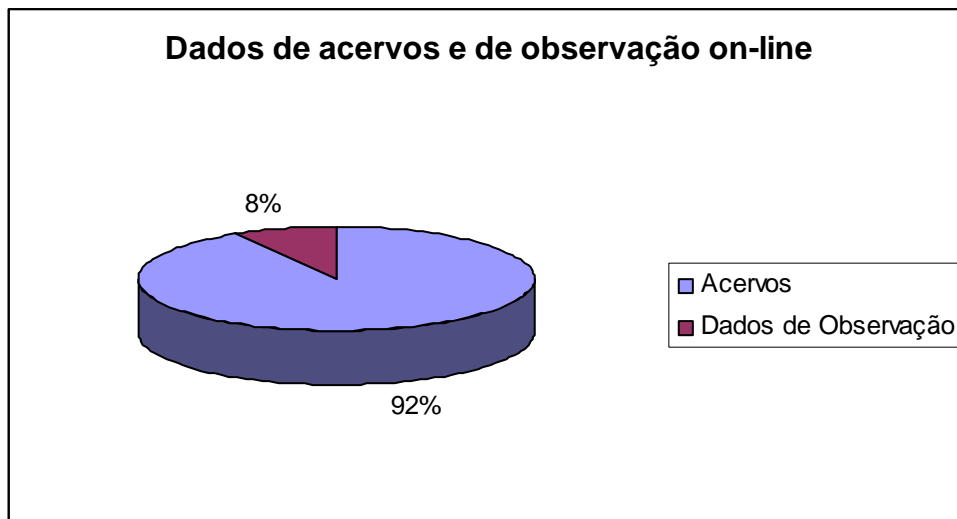


Figura 4. Porcentagem do número de registros on-line das coleções por grupo taxonômico

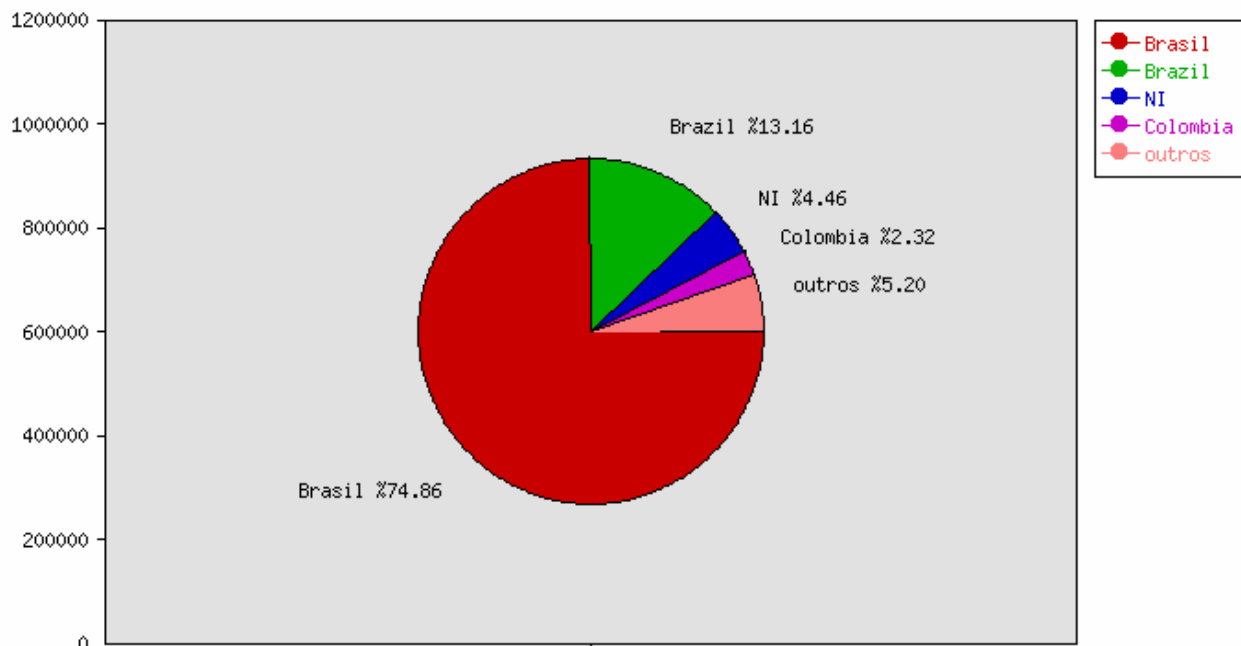
Se compararmos os dados de observação (exclusivamente do SinBiota) com a dos acervos, vimos que apesar do avanço do SinBiota que nesse ano cresceu 55%, dados de observação ainda são pouco representativos na rede, apesar das inúmeras iniciativas existentes no país. Esse número talvez também reflète o fato das grandes instituições conservacionistas não disponibilizarem os seus dados de forma que possam ser compartilhados com outras redes.



**Figura 5. Comparação do número de registros on-line de coleções biológicas com dados de observação**

Consideramos que é importante ter uma ação visando a estruturação de um sistema para receber dados de observação além do programa Biota/Fapesp. As principais instituições de conservação atuantes no país que realizam inventários baseados na observação da ocorrência de espécies devem ser contatadas para viabilizar a sua participação na rede.

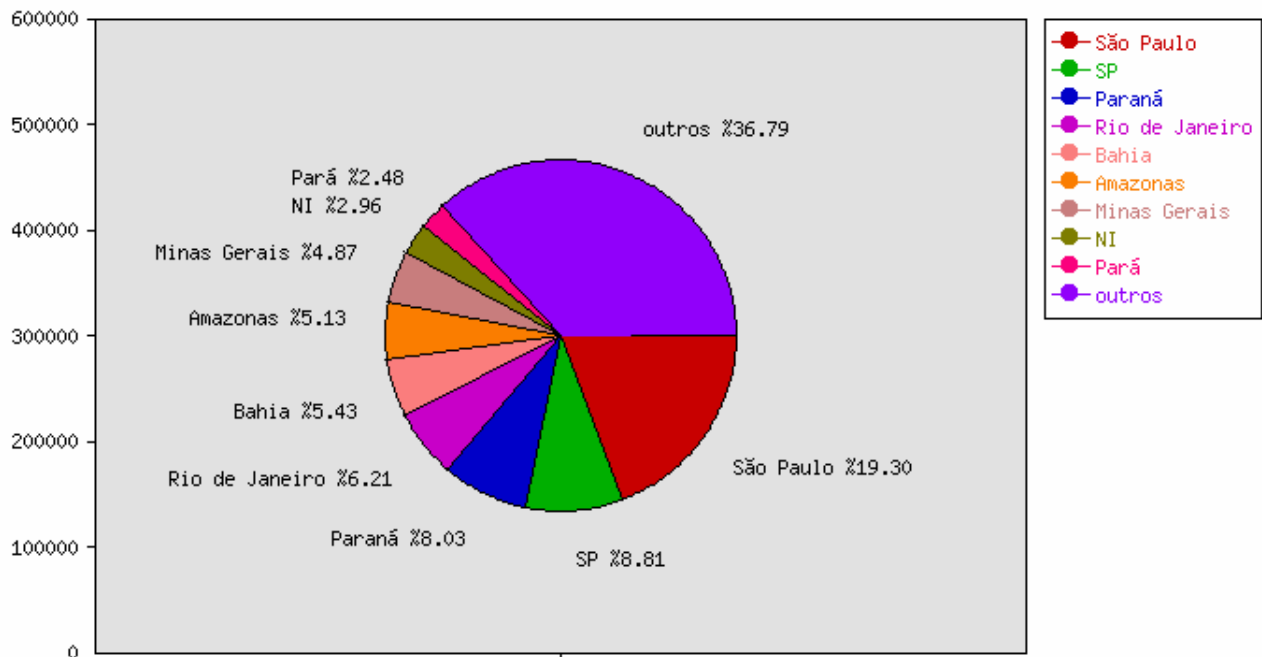
O perfil dos dados das coleções nacionais participantes da rede speciesLink também revela dados interessantes. 88% dos dados são de espécies coletadas no Brasil, como mostra a figura 6.



**Figura 6. País de origem da coleta**

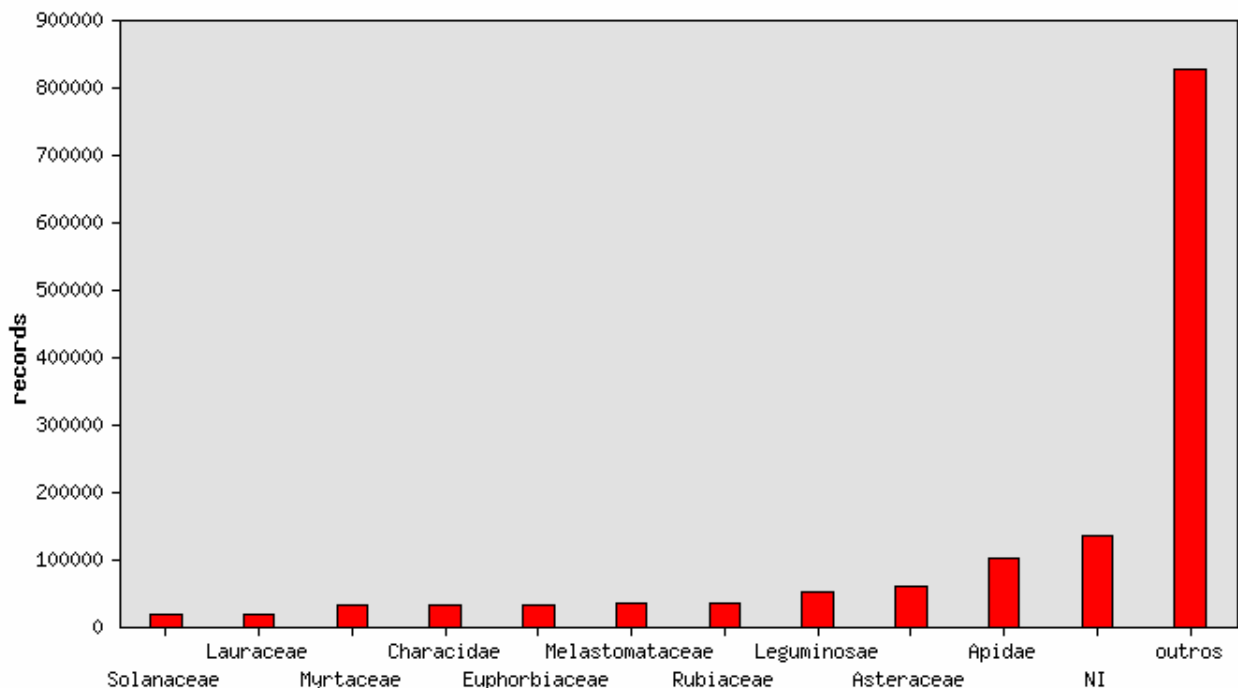
Com relação à distribuição estadual e apesar da rede speciesLink ter nascido como um projeto da Fapesp somente com coleções do estado de São Paulo, podemos observar na figura 7 que apenas cerca de 28% dos dados são de coletas de São Paulo. Para uma avaliação mais precisa é necessário realizar um trabalho de padronização dos nomes dos estados.





**Figura 7. Estado de origem da coleta**

Ainda, como último elemento do perfil da rede tem-se as 10 famílias mais representadas (figura 8). Nove são plantas e uma abelha.



**Figura 8. As 10 famílias mais representadas na rede speciesLink**

A evolução anual dos trabalhos de cada coleção está registrada na tabela 1 (anexo). Nesse relatório somente as redes de herbários e as coleções de abelhas serão avaliadas com mais detalhes por terem sido os grupos mais trabalhados pela equipe do CRIA e instituições parceiras em 2006.

## Herbários e Xilotecas

### Evolução dos dados on-line

Os herbários são as coleções com maior volume de dados na rede speciesLink e que, por esse motivo, têm merecido atenção especial do CRIA em relação ao desenvolvimento de ferramentas para a limpeza e repatriação de dados. Ao final de 2005 haviam 16 coleções on-line representando um acervo de cerca de 1,4 milhão de registros, dos quais cerca de 275 mil estavam disponíveis na Internet (cerca de 20%) sendo cerca de 25% georeferenciados. Os números hoje indicam 1,85 milhão de registros no acervo com cerca de 640 mil on-line (cerca de 34%) com 37% dos dados georeferenciados. A rede de herbários teve um crescimento de 123% com relação ao número de registros on-line, sendo o Jardim Botânico do Rio de Janeiro e as coleções da rede paranaense os principais responsáveis por este crescimento. A tabela a seguir mostra a relação dos herbários e xilotecas participantes da rede ordenadas pelo número de registros on-line.

Acrônimo	Acervo	on-line	Georef.	% on-line	% Georef.
JBRJ_RB	450.000	275.523	164.622	61%	60%
ESA	120.000	71.977	18.086	60%	25%
IAC	45.000	41.865	31.825	93%	76%
UEC	134.000	38.861	10.387	29%	27%
FUEL	40.000	36.523	0	91%	0%
UPCB	52.067	27.977	0	54%	0%
SPSF	35.089	25.050	690	71%	3%
SPF-Algae	19.776	19.776	0	100%	0%
HSJRP	28.000	19.377	0	69%	0%
BCTw	19.500	16.823	0	86%	0%
MBM	320.000	14.837	0	5%	0%
SP - Algae	15.000	14.539	1.408	97%	10%
SP	350.000	11.580	2.176	3%	19%
JBRJ_RBw	8.200	7.597	1.841	93%	24%
HRCB	40.000	6.541	0	16%	0%
JBRJ_RBdna	2.500	2.504	2.081	100%	83%
SPFw	4.000	2.098	157	52%	7%
CPMA	2.150	2.044	990	95%	48%
BOTw	1.500	1.416	0	94%	0%
HISA	10.235	1.242	411	12%	33%
SPF	133.500	600	73	0%	12%
BOTU	24.000	165	161	1%	98%
<b>Total</b>	<b>1.854.517</b>	<b>638.915</b>	<b>234.908</b>	<b>34%</b>	<b>37%</b>

Se agruparmos as coleções de uma mesma instituição como sendo uma coleção única, temos os seguintes gráficos comparativos dos acervos totais e on-line (figuras 9 e 10) dos herbários que participam da rede speciesLink.

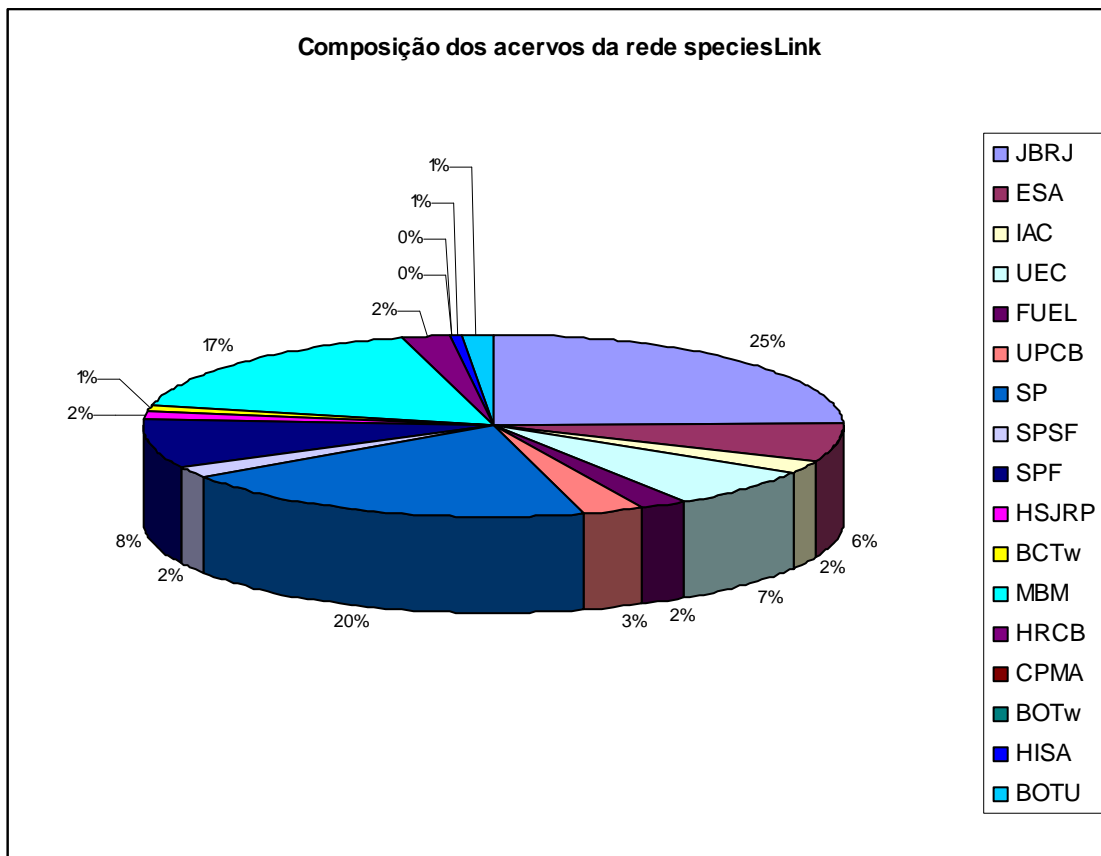


Figura 9. Composição dos acervos que participam da rede speciesLink

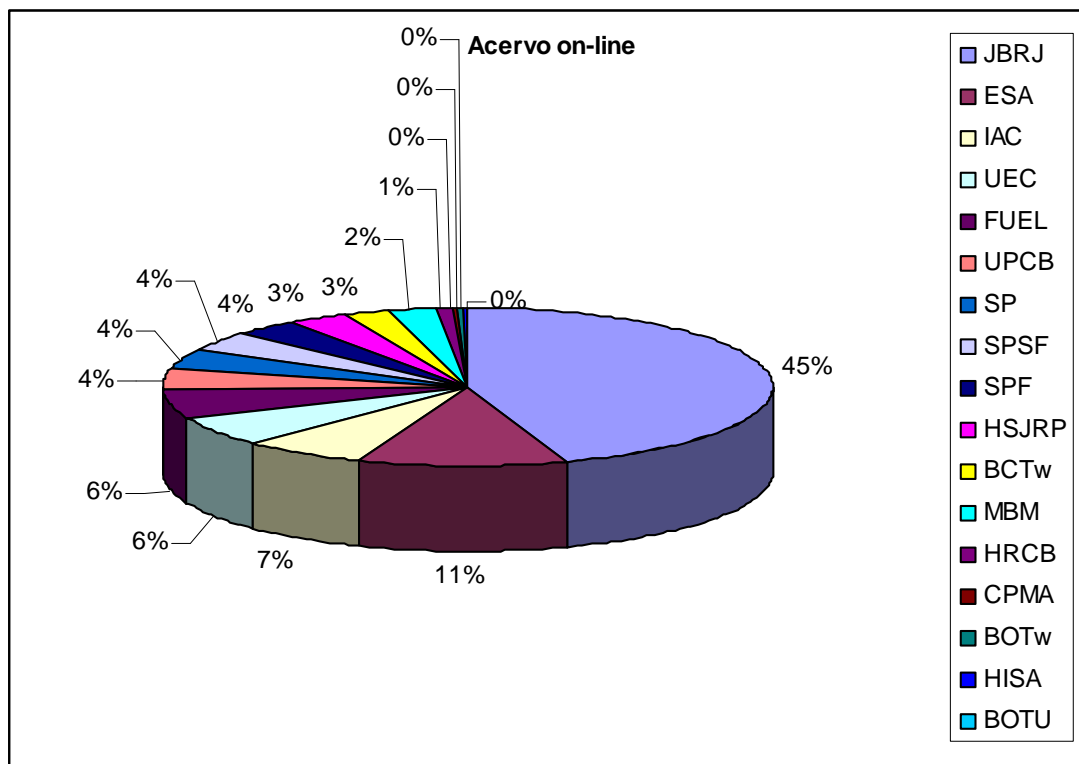


Figura 10. Contribuição dos herbários à rede speciesLink

Para as agências de fomento, esses dados são muito importantes. Mostram a necessidade de apoio específico para a digitalização dos acervos de coleções maiores como o SP, com um acervo bastante significativo (350 mil registros ou 20% do total dos acervos participantes da rede), mas com apenas 3% dos dados on-line. Os dados mostram também a importância de uma política de

fomento para coleções menores como o IAC e FUEL com mais de 90% de seus acervo disponíveis on-line. A disponibilidade dos dados on-line potencializa o uso dos resultados de anos de pesquisa, dando um retorno maior para o país. Esses gráficos também mostram a importância dos acervos menores. O IAC, por exemplo representa apenas 2% do total dos acervos da rede, mas representa 7% do número total de registros on-line, e 14% dos registros georeferenciados (figura 11), o que viabiliza o uso desses dados em trabalhos de modelagem do nicho ecológico. Acervos menores também são importantes por suas especificidades biogeográficas ou taxonômicas, o que muitas vezes tornam os seus acervos únicos. O grande destaque sem dúvida alguma é o Jardim Botânico do Rio de Janeiro cujo acervo representa 25% do total, 45% dos registros on-line e 72% dos registros georeferenciados.

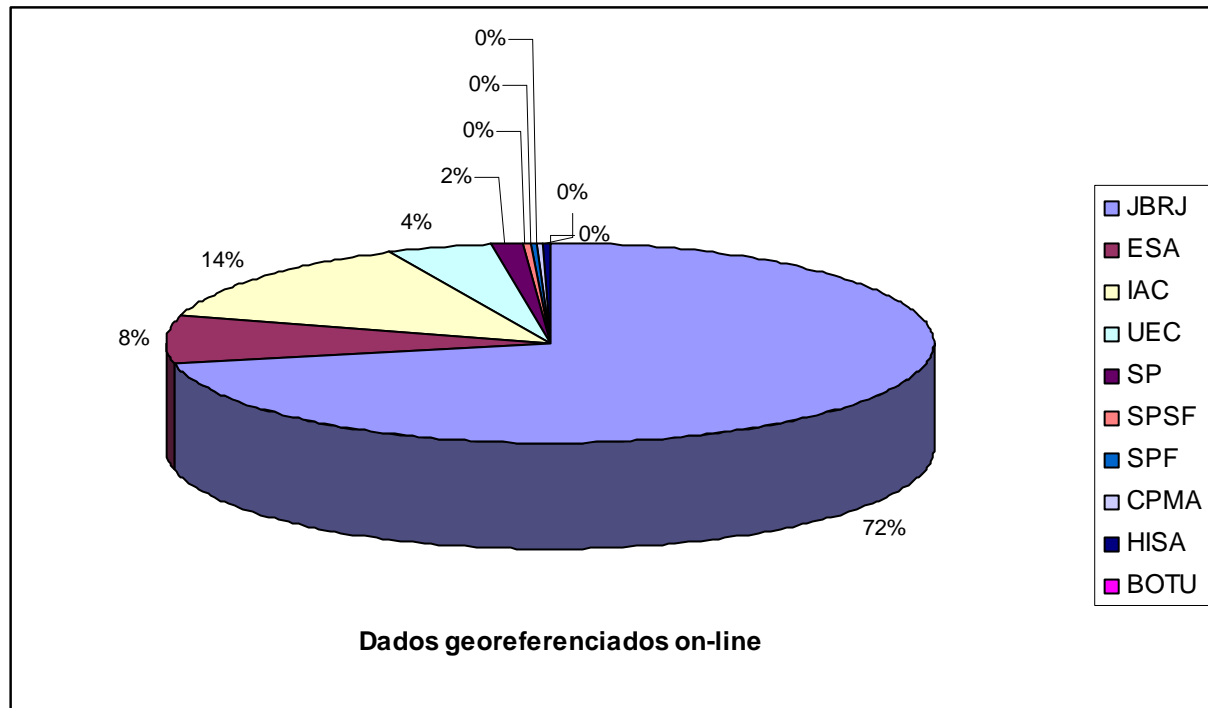


Figura 11. Proporção de dados georeferenciados on-line por herbário

Como último ponto de análise dos herbários, das 22 coleções e subcoleções temos apenas três que não alteraram os seus dados no ano de 2006. As demais, mesmo sem o apoio do projeto da Fapesp, continuaram atualizando e/ou melhorando a qualidade de seus dados na rede speciesLink.

## Repatriação de dados

Graças aos contatos realizados durante o evento *Revisão da Flora Brasileira: Desafios e Oportunidades* ([www.cria.org.br/eventos/flora](http://www.cria.org.br/eventos/flora)) realizado em Florianópolis, SC (Julho, 2006), no final de 2006, foram enviados ao CRIA os dados dos acervos dos jardins botânicos de Nova Iorque (NYBG) e Missouri (MOBOT) filtrados para os espécimes coletados no Brasil. São mais de 140 mil registros de NYBG, sendo 46 mil georeferenciados e 46 mil registros de MOBOT, sendo cerca de 14 mil georeferenciados. A inclusão desses acervos representa um aumento de 30% dos registros de herbários on-line (passa a ser de 827.706) com 26% dos registros georeferenciados (294.904). Ao mesmo tempo é gratificante constatar que ambas as coleções estão sendo beneficiadas por essa parceria uma vez que estão usando as ferramentas de *data cleaning*. Trata-se, portanto de uma parceria verdadeira, onde todos ganham.

## Flora brasiliensis

Ainda em relação a dados sobre plantas, durante o ano de 2006, graças ao apoio da Fapesp, Natura e Fundação Vitae, e em parceria com o Jardim Botânico de Missouri e a Unicamp, foi desenvolvido o sistema de informação *Flora brasiliensis on-line* ([florabrasiliensis.cria.org.br](http://florabrasiliensis.cria.org.br)). Todos os nomes citados na obra foram digitados e cadastrados em um banco de dados assim

como os metadados de cada prancha e página textual digitalizada em alta resolução por Missouri. São 3.849 pranchas, 10.213 páginas textuais, 210 famílias, 2.298 gêneros e 22.550 nomes de espécies disponíveis de forma livre e aberta a qualquer interessado através da Internet.

Deu-se início também ao trabalho de atualização dos nomes, sendo desenvolvido um segundo sistema, integrado à obra, denominado *Flora brasiliensis revisitada* ([flora.cria.org.br/](http://flora.cria.org.br/)). A proposta do sistema é servir de base para a construção colaborativa de uma lista de nomes de plantas brasileiras que deverá ser mantida e atualizada dinamicamente pelos diferentes especialistas. Cabe à comunidade de botânicos a articulação e o desenvolvimento dessa política colaborativa. Coube ao CRIA desenvolver o sistema para alimentação remota e validação dos dados através de um browser, além do sistema aberto à consulta pública, integrado ao sistema *Flora brasiliensis on-line*. Em 2006 foram trabalhadas algumas famílias que ajudaram a equipe do CRIA no desenvolvimento do sistema. São elas: *Alismataceae* (coordenadora: Maria do Carmo E Amaral); *Apocynaceae* (coordenadora: Ingrid Koch); *Bignoniaceae* (coordenadora: Lúcia G Lohmann) *Bonnetiaceae*, *Clusiaceae* e *Hypericaceae* (coordenador: Volker Bittrich). O sistema inclui hoje 185 gêneros, sendo 118 validados pelos coordenadores, e 2.266 espécies, sendo 1.067 validados pelos coordenadores. 464 nomes fazem referência a espécies citadas na obra *Flora brasiliensis*.

## Abelhas

### Dados on-line

A evolução de um outro grupo taxonômico merece destaque no ano de 2006, é o de abelhas, pelo trabalho de digitação e limpeza de dados das coleções de São Paulo e pela entrada da Coleção Entomológica Pe. Jesus Santiago Moure na rede, apesar de ter poucos dados digitalizados.

As seguintes coleções fazem parte da rede:

- RPSP, Coleção de Abelhas do Departamento de Biologia - FFCLRP/USP Ribeirão Preto SP com um acervo de 171.000 registros, 40% (68.598) on-line e desses 55.281 (81%) georeferenciados.
- CEPANN, Coleção Entomológica Paulo Nogueira-Neto - IB/USP São Paulo SP com 38.613 registros todos on-line, com 90% do acervo georeferenciado (34.596)
- DZUP-Hymenoptera, Coleção Entomológica Pe. Jesus Santiago Moure (Hymenoptera) Curitiba PR, com um acervo de 500.000 registros, mas apenas 1% (5.010) on-line e 96% dos registros on-line (4.813) georeferenciados.
- MZUEL-Abelhas, Museu de Zoologia da Universidade Estadual de Londrina - Coleção de Abelhas Londrina PR com 1.500 registros e apenas 5 registros georeferenciados on-line.

Os acervos das 4 coleções somam cerca de 710 mil registros, com 112.226 registros on-line (16%) e desses, 94.695 (84%) georeferenciados. Nenhuma coleção tem qualquer registro identificado como "suspeito" pela ferramenta *data cleaning* com relação ao nome da espécie. Trata-se, portanto, de acervos muito bem curados.

### Catálogo de Abelhas Moure ([moure.cria.org.br](http://moure.cria.org.br))

O projeto de elaboração do Catálogo de Abelhas Moure obteve financiamento junto à Finep e foi coordenado pela Universidade Federal do Paraná em parceria com o CRIA, contando com a colaboração dos seguintes especialistas:

- Prof. Gabriel Augusto Rodrigues de Melo, coordenador
- M. Sc. Anamaria Dal Molin, revisão de dados catalogados
- Prof. João Maria Franco de Camargo (USP - Ribeirão Preto): Meliponini
- Dra. Sílvia Regina Menezes Pedro (USP - Ribeirão Preto): Meliponini
- José Amílcar Tavares (USP - Ribeirão Preto): Meliponini
- Prof. Fernando Amaral da Silveira (UFMG): Exomalopsini e Teratognathini

- Prof. Vinalto Graf (UFPR): Paracolletini e Oxaeini
- Dr. Antônio José Camillo de Aguiar (UFPR): Tapinotaspidini
- M. Sc. Anamaria Dal Molin (UFPR): Megachilini

São mais de 2.500 nomes de abelhas, validados por especialistas, com os respectivos autores, dados sobre os tipos, distribuição geográfica e referências.

## Tecnologia

O indicador *tecnologia* é outro item importante de avaliação das atividades da rede speciesLink. Em 2006 a equipe do CRIA trabalhou nas seguintes ferramentas e aplicativos:

- Ferramentas de limpeza de dados (*data cleaning*)
- Software para gerenciamento de coleções (*speciesBase*)
- Desenvolvimento do aplicativo que viabiliza o espelhamento dos dados das coleções biológicas participantes da rede speciesLink em servidores regionais (spLinker)
- Desenvolvimento do openModeller, uma ferramenta de código aberto para modelagem estática de distribuição espacial utilizada para prever a distribuição de espécies.

## Ferramentas de limpeza de dados (*data cleaning* – [splink.cria.org.br/dc](http://splink.cria.org.br/dc))

Além da manutenção e do aprimoramento constante do conjunto de ferramentas para detectar possíveis erros de registro de dados, em 2006 novos aplicativos foram desenvolvidos como:

**Duplicatas:** de especial interesse para herbários que distribuem duplicatas de suas coletas a outros herbários essa ferramenta compara registros que têm o mesmo nome e número do coletor, destacando diferenças no nome da espécie. O objetivo é o de permitir que todos os herbários sejam beneficiados quando um especialista identifica uma exsicata. Essa nova identificação poderia ser propagada pela rede. Um grande problema nesse caso é a falta de padronização do nome do coletor, que deverá ser tema de discussão entre os curadores de herbário. Mas o sistema não está mais considerando espaços e pontos e está também desconsiderando letras duplas. Assim, nomes como "G M Felipe" e "G. M. Felipe" são considerados iguais. É evidente que essa não é a solução desejada, mas pelo menos podemos agora identificar duplicatas e comparar os seus dados. Temos hoje mais de 700 mil registros de exsicatas que podem ser comparados.

**data da coleta:** o sistema está listando todos os registros que têm data de coleta com mais de 75 anos (podem não ser erros) e datas de coleta cujo ano é maior que a última atualização (certamente são erros).

**registros suspeitos em relação aos nomes:** quando o número de registros suspeitos das diferentes tabelas (família, gênero, espécie, ...) é menor ou igual a 5, a tabela mostra também quais os números de catálogo para facilitar a correção dos dados.

**Planilha Excel:** cada tabela agora também traz no final, um link para uma planilha Excel que poderá ser aberta ou salva pelo usuário. Esperamos que isso facilite a correção dos dados na coleção.

**Mapa do data cleaning:** abrindo o mapa e clicando no ícone "i" de informação, o usuário poderá selecionar uma área e retornar todos os registros georeferenciados ou pontos que estão contidos no retângulo.

O sistema prepara um relatório on-line dinâmico por coleção que apresenta os registros tidos como *suspeitos* além de apresentar o perfil da coleção e o monitoramento de sua atividade de limpeza de dados. Com isso pode-se avaliar o indicador *qualidade* por coleção.

## speciesBase ([splink.cria.org.br/speciesbase](http://splink.cria.org.br/speciesbase))

speciesBase é uma base de dados genérica em Access para registros taxonômicos desenvolvida em Visual Basic for Applications (VBA). Seu desenvolvimento foi motivado pelo crescente número de coleções que desejam ingressar na rede speciesLink mas que não estão informatizadas. Em linhas gerais o que se procurou desenvolver é um sistema para o registro de dados taxonômicos associado às coletas, que tenha como apoio tabelas secundárias. O código fonte do speciesBase está disponível junto com o programa, sendo permitido a qualquer um alterá-lo e redistribuí-lo gratuitamente. A licença do software está disponível no próprio programa e no manual.

## spLinker ([splink.cria.org.br/splinker](http://splink.cria.org.br/splinker))

O spLinker é um aplicativo desenvolvido pela equipe do CRIA em Java para coleções biológicas que não possuem os pré-requisitos necessários para servir dados diretamente à rede speciesLink. Esses pré-requisitos são: uma boa conexão com a Internet, equipamento adequado funcionando 24h por dia com softwares específicos responsáveis por servir os dados, além de ter uma equipe qualificada responsável pela manutenção do sistema.

## openModeller ([openmodeller.sourceforge.net/](http://openmodeller.sourceforge.net/))

openModeller é uma ferramenta de código aberto para modelagem estática de distribuição espacial desenvolvido pela equipe do CRIA inicialmente como parte do projeto speciesLink (financiado pela Fapesp) e atualmente com projeto próprio, ainda contando com recursos da Fapesp, em parceria com a Escola Politécnica da USP (Poli) e o Instituto Nacional de Pesquisas Espaciais (INPE). Os modelos são gerados por um algoritmo que recebe como parâmetros um conjunto de pontos de ocorrência de espécies e um conjunto de mapas de variáveis ambientais. O desenvolvimento está sendo feito em C++ ANSI visando a independência de plataforma. Vários algoritmos podem ser utilizados. Utiliza a biblioteca GDAL para ler vários formatos de mapas de variáveis ambientais e proj4 para converter diferentes projeções e sistemas de coordenadas. Inicialmente está sendo usada uma arquitetura cliente-servidor, possibilitando a existência de diferentes interfaces (desktop, linha de comando e interface web). No futuro algumas tarefas poderão ser realizadas de forma distribuída, incluindo a possibilidade de processar os algoritmos separadamente em clusters remotos. O código fonte está disponível no sourceforge para todos os desenvolvedores que estiverem interessados.

## Uso

Analisando somente os 2 principais sistemas da rede (speciesLink e Flora brasiliensis) tem-se as seguintes estatísticas para o ano de 2006.

Ano de 2006	Visitantes únicos	Número de visitas	Páginas	Hits	GBytes
rede speciesLink	73.317	123.219	578.644	1.026.313	19,79
Flora brasiliensis	97.954	127.035	2.754.747	18.178.232	158,66
<b>Total</b>	<b>171.271</b>	<b>250.254</b>	<b>3.333.391</b>	<b>19.204.545</b>	<b>178,45</b>

São números bastante expressivos quando considerarmos que se trata de um sistema científico.

As figuras 12 e 13 apresentam o acesso aos sistemas speciesLink e Flora brasiliensis por país ou domínio.
























Visitas domínios/países (Primeiros 10) - Lista completa						
	Domínios/Países		Páginas	Hits	Bytes	
	Brazil	br	289453	603117	15.38 GB	
	Desconhecido	ip	182878	287247	2.47 GB	
	USA Educational	edu	28781	30992	841.33 MB	
	Commercial	com	18717	21387	183.63 MB	
	Network	net	17850	22316	176.26 MB	
	Turkey	tr	5211	5240	28.99 MB	
	Russian Federation	ru	5169	5222	29.07 MB	
	Japan	jp	4208	4442	26.78 MB	
	Mexico	mx	2527	3267	32.38 MB	
	Peru	pe	1642	1933	22.52 MB	
	Outros visitantes		22208	41150	649.30 MB	

Figura 12. Visitas à rede speciesLink





Visitas domínios/países (Primeiros 10) - Lista completa						
	Domínios/Países		Páginas	Hits	Bytes	
	Brazil	br	1903338	12800392	109.59 GB	
	Desconhecido	ip	631017	3969300	35.76 GB	
	Network	net	46422	330892	2.78 GB	
	Portugal	pt	37852	126069	1.19 GB	
	France	fr	32703	98323	1.73 GB	
	Commercial	com	28534	181798	1.27 GB	
	Germany	de	7622	68054	660.64 MB	
	Argentina	ar	7522	68892	909.47 MB	
	Canada	ca	7132	23275	429.34 MB	
	USA Educational	edu	6700	60281	496.73 MB	
	Outros visitantes		45905	450956	3.90 GB	

Figura 13. Visitas ao sistema Flora brasiliensis

Podemos ver que de longe o principal usuário é o Brasil, 78% do volume de dados em GBytes da rede speciesLink e cerca de 70% da rede Flora brasiliensis. A rede comercial (.com) representa menos que 1% em ambos os casos. Trata-se de um indicador muito importante, que justifica o investimento do governo tanto na ampliação, digitação e melhora da qualidade dos acervos como também no sistema de informação e nas redes de comunicação de dados, só pelo fato de ser de interesse público e de beneficiar, principalmente, o público nacional.



Tabela 1. Evolução da rede speciesLink

COLEÇÕES BIOLÓGICAS	ACRÔNIMO	ACERVO	REGISTROS ON-LINE				GEO-REF	VARIÇÃO ANUAL	% ON-LINE	% GEO-REF	última atualização
			out-03	out-04	out-05	dez-06					
HERBÁRIOS	BOTU	24.000			165	165	161	0%	1%	98%	23/8/2005
	CPMA	2.150			1.882	2.044	990	9%	95%	48%	5/9/2006
	ESA	120.000	730	40.356	74.549	71.977	18.086	-3%	60%	25%	29/9/2006
	HISA	10.235			718	1.242	411	73%	12%	33%	7/11/2006
	HRCB	40.000		2.593	6.541	6.541	0	0%	16%	0%	20/9/2006
	HSJRP	28.000			19.377	19.377	0	0%	69%	0%	4/8/2005
	IAC	45.000	25.296	31.548	39.770	41.865	31.825	5%	93%	76%	5/1/2007
	SP	350.000	12.901	18.302	13.928	11.580	2.176	-17%	3%	19%	23/11/2006
	SP - Algae	15.000			14.033	14.539	1.408	4%	97%	10%	8/2/2006
	SPF	133.500	22.769	22.762	10.250	600	73	-94%	0%	12%	13/9/2006
	SPF-Algae	19.776			19.776	19.776	0	0%	100%	0%	18/2/2005
	SPSF	35.089		31.168	31.280	25.050	690	-20%	71%	3%	27/4/2006
	UEC	134.000	12.860	33.301	36.081	38.861	10.387	8%	29%	27%	21/12/2006
	JBRJ_RB	450.000		1.383	6.138	275.523	164.622	4389%	61%	60%	4/1/2007
	JBRJ_RBdna	2.500				2.504	2.081		100%	83%	4/1/2007
	FUEL	40.000				36.523	0		91%	0%	31/10/2006
MBM	320.000				14.837	0		5%	0%	17/5/2006	
UPCB	52.067				27.977	0		54%	0%	16/5/2006	
<b>Subtotal</b>		<b>1.821.317</b>	<b>74.556</b>	<b>181.413</b>	<b>274.488</b>	<b>610.981</b>	<b>232.910</b>	<b>123%</b>	<b>34%</b>	<b>38%</b>	
XILOTECAS	BCTw	19.500	9.127	908	9.127	16.823	0	84%	86%	0%	10/11/2006
	SPFw	4.000		908	1.677	2.098	157	25%	52%	7%	7/3/2006
	BOTw	1.500				1.416	0		94%	0%	6/12/2006
	JBRJ_RBw	8.200				7.597	1.841		93%	24%	19/12/2006
<b>Subtotal</b>		<b>33.200</b>	<b>9.127</b>	<b>1.816</b>	<b>10.804</b>	<b>27.934</b>	<b>1.998</b>	<b>159%</b>	<b>84%</b>	<b>7%</b>	
PEIXES	DZSJRP-Pisces	8.496	5.725	6.888	7.885	8.496	6.475	8%	100%	76%	6/7/2006
	LIRP	5.744	4.454	4.454	5.167	5.744	4.364	11%	100%	76%	19/12/2006
	MZUSP	84.000		72.714	77.613	79.278	54.323	2%	94%	69%	24/3/2006
	MHN-peixes	6.304			6.138	6.304	0	3%	100%	0%	17/11/2006
	MHNCI	11.537				7.997	336		69%	4%	9/10/2006
	MZUEL-Peixes	9.290				4.524	1.219		49%	27%	4/12/2006

COLEÇÕES BIOLÓGICAS	ACRÔNIMO	ACERVO	REGISTROS ON-LINE				GEO-REF	VARIÇÃO ANUAL	% ON-LINE	% GEO-REF	última atualização	
			out-03	out-04	out-05	dez-06						
	Subtotal	125.371	10.179	84.056	96.803	112.343	66.717	16%	90%	59%		
<b>ARACNÍDEOS</b>	Ácaros	DZSJRPAcari	7.225	5.382	5.753	6.296	7.225	583	15%	100%	8%	20/10/2006
	Ácaros	AcariESALQ	13.771	12.392	12.392	15.732	13.771	2.004	-12%	100%	15%	8/11/2006
	Ácaros	IBSP-Acari	9.201		4.210	2.771	2.771	1.143	0%	30%	41%	20/10/2005
	Aranhas	UBTU	3.500		2.623	2.626	2.677	47	2%	76%	2%	22/11/2006
	<b>Subtotal</b>	<b>33.697</b>	<b>17.774</b>	<b>24.978</b>	<b>27.425</b>	<b>26.444</b>	<b>3.777</b>	<b>-4%</b>	<b>78%</b>	<b>14%</b>		
<b>RÉPTEIS E ANFÍBIOS</b>												
	Anfíbios	DZSJRP-Amphibia-adults	8.324		4.823	7.499	8.324	8.227	11%	100%	99%	21/12/2005
	Anfíbios	DZSJRP-Amphibia-tadpoles	1.758			1.158	1.158	0	0%	66%	0%	23/9/2005
	Anfíbios	CFBH	15.000		2.935	3.773	8.695	0	130%	58%	0%	20/9/2006
	Anfíbios	MHN-anfíbios	13.657			12.231	13.657	0	12%	100%	0%	16/11/2006
	Répteis	MHN-repteis	3.048			2.832	3.048	0	8%	100%	0%	16/11/2006
	Répteis e Anfíbios	IBSP-Herpeto	80.000			66.167	66.167	0	0%	83%	0%	11/10/2005
	<b>Subtotal</b>	<b>121.787</b>	<b>0</b>	<b>7.758</b>	<b>93.660</b>	<b>101.049</b>	<b>8.227</b>	<b>8%</b>	<b>83%</b>	<b>8%</b>		
<b>MAMÍFEROS</b>	Roedores	IAL-Roedores	21.000			12.029	12.029	0	0%	57%	0%	31/10/2005
	Mamíferos	MHN-mamíferos	2.398			2.385	2.398	0	1%	100%	0%	16/11/2006
	Morcegos	DZSJRPChiroptera	10.678		2.829	5.339	5.339	815	0%	50%	15%	8/7/2005
	Mastozoologia	DZUP-Mammalia	400				0	0				
	<b>Subtotal</b>	<b>34.476</b>	<b>0</b>	<b>2.829</b>	<b>19.753</b>	<b>19.766</b>	<b>815</b>	<b>0%</b>	<b>57%</b>	<b>4%</b>		
<b>INSETOS</b>	Insetos	MEFEIS	28.000		3.519	3.519	4.699	0	34%	17%	0%	14/12/2005
	Insetos	IBSP-IB	275.000			401	401	0	0%	0%	0%	27/9/2005
	Abelhas	CEPANN	38.613			34.477	38.613	34.596	12%	100%	90%	17/10/2006
	Abelhas	RPSP	171.000			68.599	68.598	55.281	0%	40%	81%	10/11/2005
	Abelhas	MZUEL-Abelhas	1.500				5	5		0%	100%	26/9/2006
	Insetos	DZUP-Aphidoidea	9.400				0	0		0%		9/6/2006
	Insetos	DZUP-Coleoptera	700.000				13.947	13.944		2%	100%	24/5/2006
	Insetos	DZUP-Dermaptera	1.389				430	425		31%	99%	16/11/2006
	Insetos	DZUP-Diptera	1.500.000				0	0		0%		17/11/2006
	Insetos	DZUP-Hymenoptera	500.000				5.010	4.813		1%	96%	2/9/2006
	Insetos	DZUP-Lepidoptera	268.005				0	0		0%		
	Insetos	DZUP-Trichoptera	151.154				4.252	4.252		3%	100%	29/8/2006
	Insetos	FSP-USP	100.000				0	0		0%		
	<b>Subtotal</b>	<b>3.744.061</b>	<b>0</b>	<b>3.519</b>	<b>106.996</b>	<b>135.955</b>	<b>113.316</b>	<b>27%</b>	<b>4%</b>	<b>83%</b>		

COLEÇÕES BIOLÓGICAS	ACRÔNIMO	ACERVO	REGISTROS ON-LINE				GEO-REF	VARIÇÃO ANUAL	% ON-LINE	% GEO-REF	última atualização
			out-03	out-04	out-05	dez-06					
<b>ASCIDIACEA</b>											
	DZUP-Ascidiacea	600				408	322		68%	79%	9/6/2006
	<b>Subtotal</b>	<b>600</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>408</b>	<b>322</b>		<b>68%</b>	<b>79%</b>	
<b>AVES</b>											
	IAL-Aves	110.000			21.157	20.956	12.396	-1%	19%	59%	31/10/2005
	MHN-Aves	2.223			2.224	2.223	0	0%	100%	0%	16/11/2006
	<b>Subtotal</b>	<b>112.223</b>	<b>0</b>	<b>0</b>	<b>23.381</b>	<b>23.179</b>	<b>12.396</b>	<b>-1%</b>	<b>21%</b>	<b>53%</b>	
<b>MICROORGANISMOS</b>											
	SICol	8.724				8.724	0		100%	0%	5/1/2007
	<b>Subtotal</b>	<b>8.724</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>8.724</b>	<b>0</b>		<b>100%</b>	<b>0%</b>	
<b>TOTAL ACERVOS</b>		<b>6.034.856</b>	<b>111.636</b>	<b>306.369</b>	<b>653.310</b>	<b>1.066.783</b>	<b>440.156</b>	<b>63%</b>	<b>18%</b>	<b>41%</b>	
<b>DADOS DE OBSERVAÇÃO</b>											
	SinBiota	98.758	39.402	51.606	63.586	98.758	98.758	55%	100%	100%	20/12/2006
	<b>Subtotal</b>	<b>98.758</b>	<b>39.402</b>	<b>51.606</b>	<b>63.586</b>	<b>98.758</b>	<b>98.758</b>	<b>55%</b>	<b>100%</b>	<b>100%</b>	<b>0%</b>
<b>COLEÇÃO DE SONS</b>											
	CSUEL	7.000				577	457		8%	79%	18/12/2006%
	<b>Subtotal</b>	<b>7.000</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>577</b>	<b>457</b>		<b>8%</b>	<b>79%</b>	
<b>TOTAL GERAL</b>		<b>6.140.614</b>	<b>151.038</b>	<b>357.975</b>	<b>716.896</b>	<b>1.166.118</b>	<b>539.371</b>	<b>63%</b>	<b>19%</b>	<b>46%</b>	