ICT, Social Networks and the Brazilian Biodiversity Infrastructure

Vanderlei Canhos

vcanhos@cria.org.br

Codata Conference 2010

Biocomplexity: data and information Biosphere – The world we live in Ecosystem – The set of communities of all domains of life that interact with one another and the abiotic environment to form a unit **Community** – Interacting populations of organisms **Population** – All individuals of a species or phylotype within a community Organism – A single individual **Organ system**– a specialized functional system of an organism Organ – a set of tissues that function as a unit **Tissue** A set of interacting cells Cell – the functional unit of all living organisms Organelle a specialized subunit within a cell

Molecule – biochemical constituents of cells

Source: Committee on A New Biology for the 21st Century

Data, information & knowledge

To predict and control the activities of biological systems we need data and information on:

- about each component
- about how components work together as systems

Complexity of biodiversity data

Requirements:

- Assemble virtual collaborations at different scales
- Integration (data, information, people ...)
- Advanced communication and informatics infrastructure

Information is the fundamental currency of the new biology



http://splink.cria.org.br

Key elements

Biological collections in Brazil

- A small number of "large" collections
- A large number of small key collections

Informatics profile

- Human resources: lack of expertise in informatics
- Equipments and installations (not adequate)
- Connectivity (normally slow or unstable)

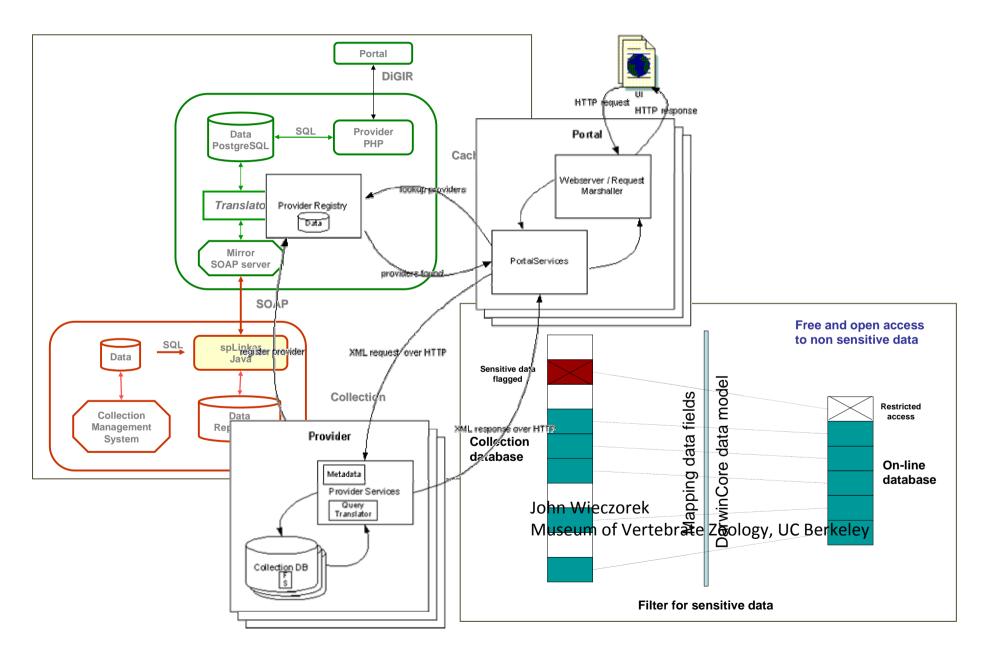
Challenges

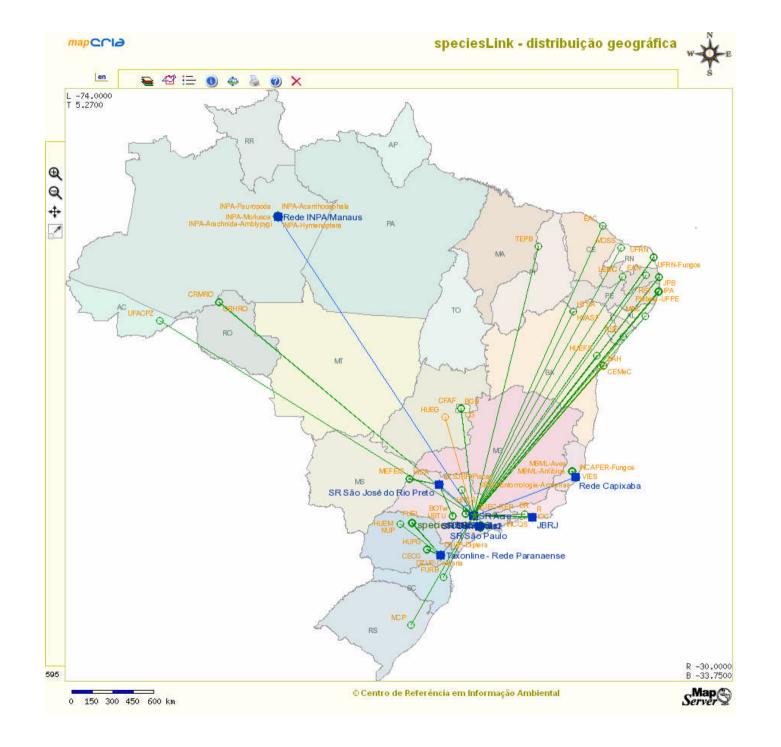
- Integration of primary data from all taxa, from distributed collections, using different software in diverse environments
- Integrating data from collections with low and/or unstable internet connectivity, using basic hardware and lack of computer experts
- Data providers with full control over the data served to the network

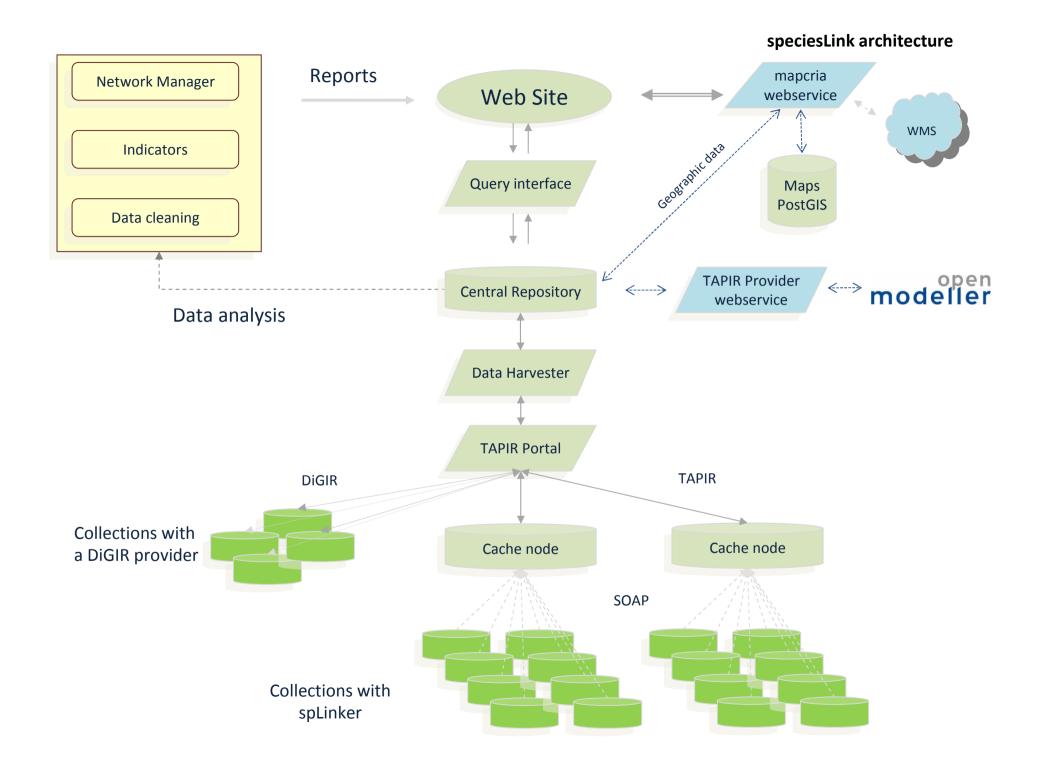
Development parameters - architecture

- Maintenance of collection's routine
 - Practically any software is accepted (Excel, Access, Specify, Biota, Brahms, PostgreSQL, MySQL, ...)
- Data provider with full control over the data
 - What is sensitive data, what is open and free
 - Digitization strategy, data cleaning strategy
- Data providers fully acknowledged
- Connectivity problems addressed
- Network interoperable with international initiatives

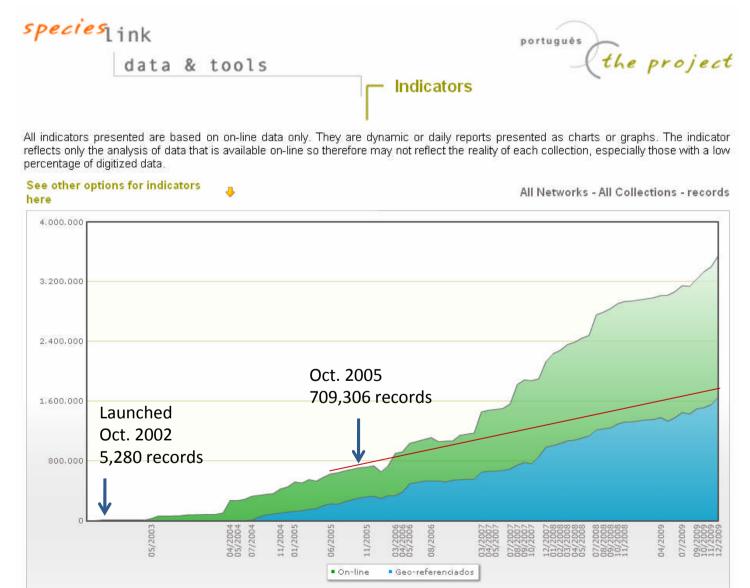
Network architecture



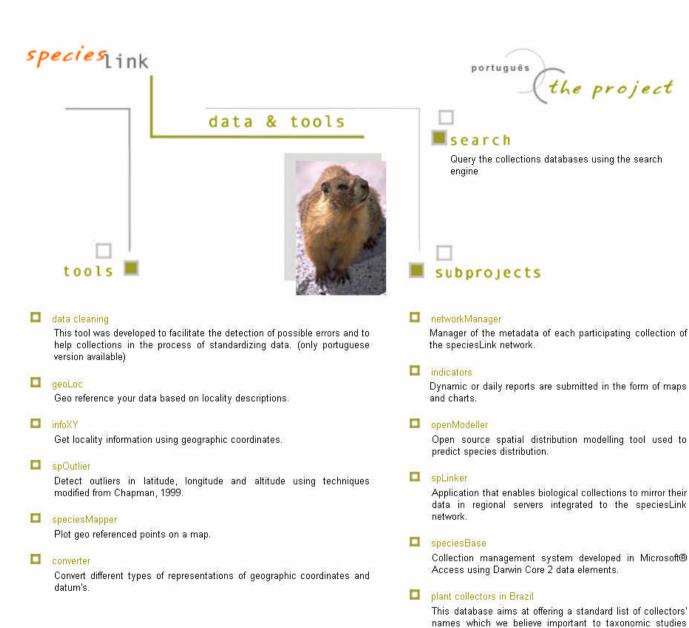




The development of the speciesLink network



3.8 million



search indicators data cleaning networkManager and herbaria digitization projects. email Discussion List

Centro de Referência em Informação Ambiental Fundação de Amparo à Pesquisa do Estado de São Paulo

portugués

engine

the project

Query the collections databases using the search

Data Cleaning Tools

- □ Detects "possible errors"
 □ Flags "suspect records" to facilitate the work of data providers and users
 □ Records are not modified by CRIA
- ☐ Main features include:
 - Checking names using phonetic comparisons: names of the network, the Catalogue of Life (CoL), and selected authority files
 - checking dates indicating earliest collection date and latest update
 - checking geographic data inconsistencies: country/state/municipality versus geographic coordinates



email



data & tools

data cleaning

Select a collection NYBG_BR

collection: NYBG_BR	
total number of records on-line	241281
- without coordinates	142083
- georeferenced	99198
- access to georeferenced data denied	0
- in the sea	1995
repeated records	
catalog number	30
all fields	2
collector's name and number	30348
last update	
of the collection	11-09-2009
of data cleaning	24-09-2009

	taxonomic data
inventory	scientific name - collector - types
family	not found
genus	244 suspect records
species	484 suspect records
subspecies	not found
author	4194 suspect records
duplicate	1271 suspect records

date collected	
collect before 1767	15 suspect records
last update previous to date collected	1 suspect records



collection profile data cleaning statistics geographic coordinates analysis

locality data		
inventory	country - state - municipality	
name of the country/state	419 suspect records	
outlier	not found	
long/lat outside the world limit	not found	
equal long/lat	not found	
long or lat equal to zero	ero 33 suspect records	
long/lat in the sea (Brazil) 71 suspect record		
municipality name (Brazil)	21305 suspect records	
coordinate unit analysis (Brazil)	459 suspect records	

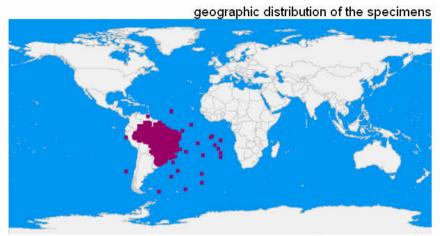
suggestions for blank fields		
long/lat (Brazil)	84090 suggestions	
country/state name	52 suggestions	
municipality name (Brazil)	22360 suggestions	

Data cleaning reports for each collection based on online data

collection: NYBG_BR	
total number of records on-line	241281
- without coordinates	142083
- georeferenced	99198
- access to georeferenced data denied	0
- in the sea	1995
repeated records	
catalog number	30
all fields	2
collector's name and number	30348
last update	
of the collection	11-09-2009
of data cleaning	24-09-2009

	taxonomic data
in∨entory	scientific name - collector - types
family	not found
genus	244 suspect records
species	484 suspect records
subspecies	not found
author	4194 suspect records
duplicate	1491 suspect records

date collected	
collect before 1767	15 suspect records
last update previous to date collected	1 suspect records



collection profile data cleaning statistics geographic coordinates analysis

locality data		
in∨entory	country - state - municipality	
name of the country/state	419 suspect records	
outlier	not found	
long/lat outside the world limit	not found	
equal long/lat	not found	
long or lat equal to zero	33 suspect records	
long/lat in the sea (Brazil)	71 suspect records	
municipality name (Brazil)	21305 suspect records	
coordinate unit analysis (Brazil)	459 suspect records	

suggestions for blank fields				
long/lat (Brazil) 84090 suggestions				
country/state name	52 suggestions			
municipality name (Brazil)	22360 suggestions			





collection: NYBG_BR Suspect genus names

List of records that have the same family name with a phonetic variation of the genus

Title: The names in red are not included in the dictionaries available, those in green are.

family	genus	species	subspecies	re	cords	speciesLink	status CoL	catalog number
[Rhamnaceae]	🥨 [Zyziphus]	[joazeiro]	[]	0		1		
[Rhamnaceae]	<pre>\$P [Zyzyphus]</pre>	[joazeiro]	[]	0		3		
[Rhamnaceae]	SP [Ziziphus]	[joazeiro]	[]	30	see	331		
[Rhamnaceae]	💴 [Zizyphus]	[joazeiro]	[]	1	see	24		1034302
[Rubiaceae]	SP [Emmeorhiza]	[umbelata]	[]	0		1		
[Rubiaceae]	<pre>SP [Emmeorrhiza]</pre>	[umbellatta]	[]	0		2		
[Rubiaceae]	<pre>SP [Emmeorriza]</pre>	[umbellata]	[]	0		1		
[Rubiaceae]	<pre>SP [Emmeorrhiza]</pre>	[umbellata]	[]	92	see	109		
[Rubiaceae]	SP [Emmeorhiza]	[umbellata]	[]	17	see	477	accepted name	
[Euphorbiaceae]	[]	[hieronyma]	[]	0		1		
[Euphorbiaceae]	SD [Hieronyma]			5	see	38		1064903 1064904 1064905 1064906 1064907
[Euphorbiaceae]	SP [Hyeronima]	[]	[]	3	see	24		195423 40920 896135

collection: ESA	
total number of records on-line	94179
- without coordinates	77149
- georeferenced	17030
- access to georeferenced data denied	643
- in the sea	2177
repeated records	
catalog number	1273
all fields	42
collector's name and number	3394
last update	
of the collection	08-11-2007
of data cleaning	07-04-2008

	taxonomic data
in∨entory	scientific name - collector - types
family	3 suspect records
genus	1389 suspect records
species	2769 suspect records
subspecies	not found
author	41477 suspect records
duplicate	1348 suspect records

date collected					
collect before 1930	220 suspect records				
last update previous to date collected	not found				

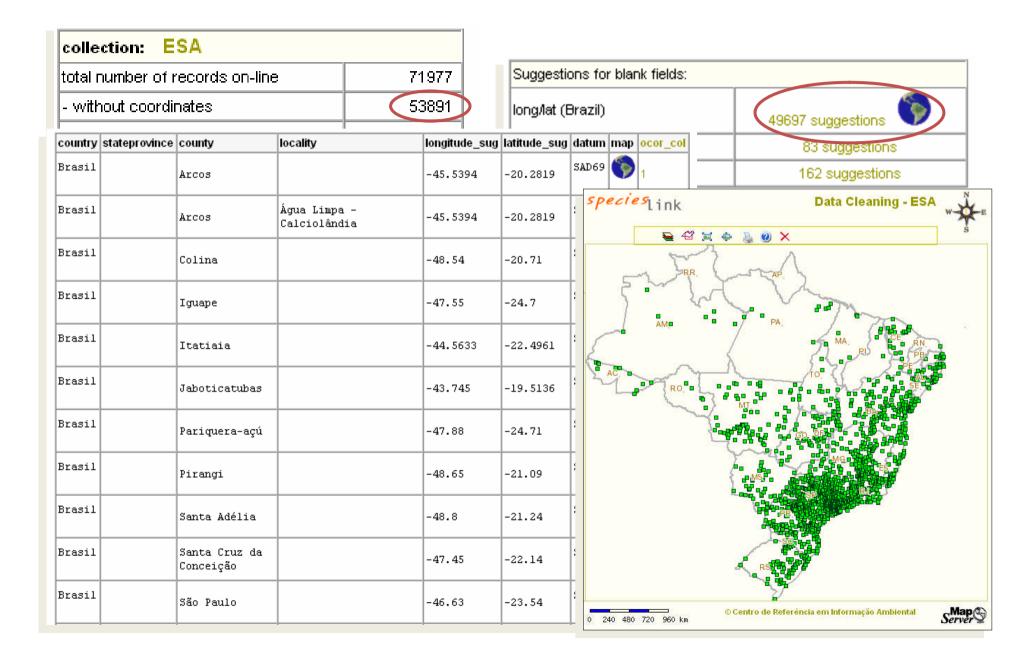


collection profile data cleaning statistics geographic coordinates analysis

locality data						
inventory	country - state - municipality					
name of the country/state	3 suspect records					
outlier	1110 suspect records					
long/lat outside the world limit	3 suspect records					
equal long/lat	not found					
long or lat equal to zero	1498 suspect records					
long/lat in the sea (Brazil)	1738 suspect records					
municipality name (Brazil)	6128 suspect records					
coordinate unit analysis (Brazil)	2770 suspect records					

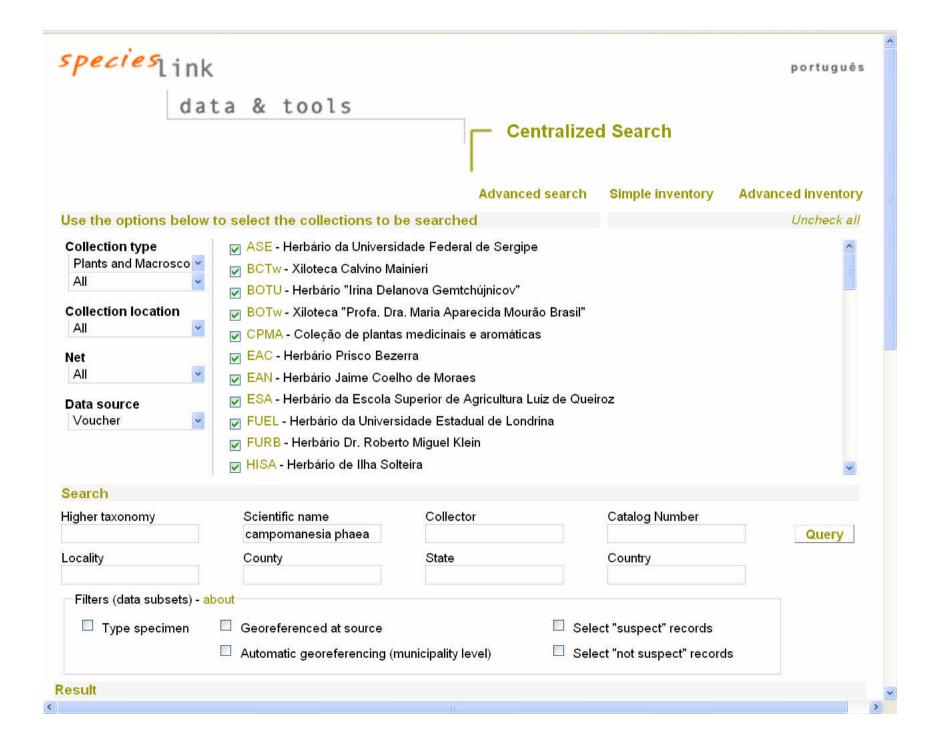
suggestions	s for blank fields
long/lat (Brazil)	67066 suggestions
country/state name	2 suggestions
municipality name (Brazil)	36 suggestions

Automatic Georeferencing



Centralized query

- Classification system to select data providers
- Data outputs: files (html, xml, Excel), mapCRIA,
 Google Maps
- Data filters: type specimen, georeferencing at source or automatic, inconsistent records
- Results: presented per collection, integrated with red lists, show geographic inconsistencies, smart links to CRIA's name server and GBIF



Result							
		Georeferenced					
Source	Records	At source	Automatic	Content	Format	W	-
ESA	4	2	2	All	html 💌	see	₩ G
FUEL	9	0	1	Locality 🔽	excel 💌	see	
HUMC	1	1	0	Summary 💆	xml	see	
IAC	5	5	0	All	html 💌	see	
IPA	1	0	1	All	html	see	
JBRJ_RB	6	6	0	All	html 💌	see	
МВМ	6	0	6	All	html 💌	see	
MBML-Herbario	2	2	0	All	html 💌	see	G
MOBOT_BR	5	1	0	All	html 💌	see	
NYBG_BR	5	1	3	All	html 💌	see	
SP	22	2	18	All	html 💌	see	
SPSF	8	1	7	All 💌	html 💌	see	
UEC	3	2	0	All	html	see	♀ G
UPCB	1	0	1	All	html 💌	see	
Total	70	23	39	All	html 💌	see	G



data & tools

Centralized Search



Escola Superior de Agricultura Luiz de Queiroz, ESALQ

How to cite e Use conditions



spLink		Collection	Catalog number	Scientific name	Basis of record	Kingdom	Phylum	Class	Ord
sp 🛂 🏲	ESALQ	ESA	75352	Campomanesia phaea	S	Plantae	Magnoliophyta	₹ 7	70
sp 🐇	ESALQ	ESA	17676	Campomanesia phaea	S	Plantae	Magnoliophyta	æ	=
sp 🖔	ESALQ	ESA	13138	Campomanesia phaea	S	Plantae	Magnoliophyta		8
sp 🖔 🟲	ESALQ	ESA	68271	Campomanesia phaea	S	Plantae	Magnoliophyta	3.5	

FUEL - Herbário da Universidade Estadual de Londrina

Universidade Estadual de Londrina, UEL

How to cite e Use conditions



spLink	Institution	Collection	Catalog number	Scientific name	Basis of record	Kingdom	Phylum	Class	Order	Farr
sp 🖔	UEL	FUEL	39902	Campomanesia phaea	S	(-)		-	-	Myrta

HUMC - Herbário Mogiense

Universidade de Mogi das Cruzes, UMC

How to cite e Use conditions





spLink	Institution	Collection	Catalog number	Scientific name	Basis of record	Kingdom	Phylum	Class	Order	F
sp 🖔 🏲	UMC	HUMC	¥.	Campomanesia phaea	(#V	1 = 0	120	-	-	N



Scientific Name Search

Genus Species subspecies Kingdom

[Campomanesia phaea in All v

português

search

HELP SAVE PECIES IOW!

ferms of use

Help Dictionaries Resources at CRIA

Flora brasiliensis speciesLink

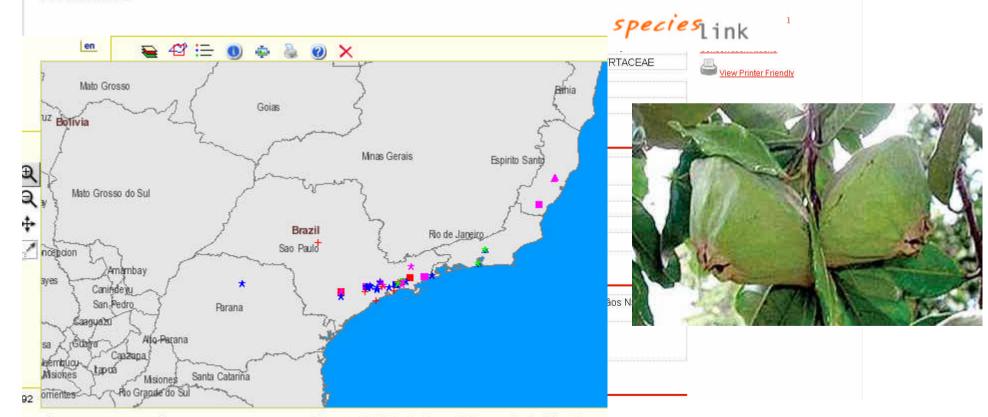
speciesLink

speciesLink is a distributed information system that integrates primary data from biological collections of the State of São Paulo, Parana's network Taxonline, the SICol network, Rio de Janeiro Botanical Garden Herbarium and observation data from the Biota/Fapesp program, and some collections outside of Brazil. The development was funded by Fapesp (Process 2001/02175-5).

Centro de Referência em Informação Ambiental, CRIA

See matches

0 110 220 330 440 km



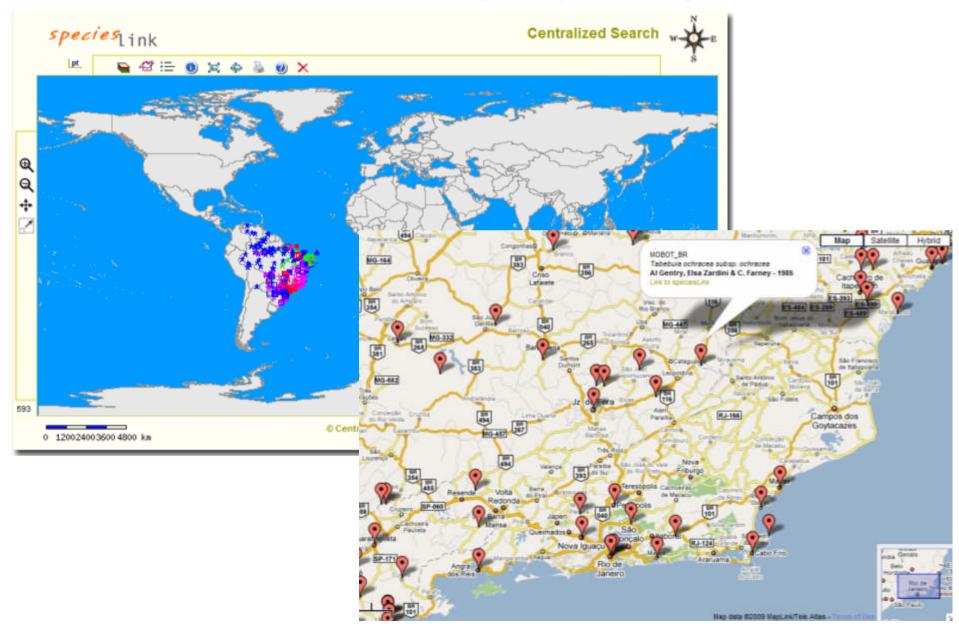
© Centro de Referência em Informação Ambiental

External Resources

Centralized query: Tabebuia

Result								
Source	Records	Georeferenced At source	records Automatic	Con	tent	Format		
ASE	33	2	27	All	<u>~</u>	html 💟	see	 G
воти	4	4	0	All	~	html 🔽	see	€ G
CPMA	10	8	2	All	~	html 🔽	see	₩ G
EAC	12	2	10	All	V	html 💟	see	€ G
ESA	192	12	172	All	V	html 💟	see	€ G
FUEL	39	2	32	All	~	html 💟	see	₩ G
FURB	11	7	4	All	~	html 💟	see	₩ G
HISA	5	4	1	All	V	html 💟	see	© G
HRCB	11	0	11	All	V	html 🔛	see	₩ G
HST	79	6	54	All	V	html 💟	see	
нимс	24	12	12	All	V	html 🔽	see	₩ G
HUPG	6	0	6	All	V	html 💟	see	
IAC	57	51	0	All	V	html 💟	see	₩ G
INPA-Herbario	295	46	100	All	V	html 💟	see	
IPA .	17	0	13	All	~	html 💟	see	
JBRJ_RB	747	747	0	All	~	html 💟	see	
JPB	117	17	91	All	$\overline{\mathbf{v}}$	html 💟	see	
MAC	47	18	27	All	V	html 💟	see	G
мвм	294	0	214	All	V	html 💟	see	
MBML-Herbario	52	43	7	All	V	html 🔽	see	
MOSS	22	0	20	All	$\overline{\mathbf{v}}$	html 🐷	see	
SP	41	2	37	All	V	html 💟	see	G
SPSF	133	6	119	All	V	html 🔛	see	₩ G
UEC	60	19	39	All	V	html 💟	see	
UFRN	13	0	13	All	V	html 💟	see	€ G
UPCB	59	2	53	All	V	html 🔛	see	
VIES	31	0	29	All	V	html 💟	see	
Total	2411	1010	1093	All	~	html 🔛	see	

Centralized query – maps



Centralized query – retrieved record



Indicators

"On the fly" indicators based on online data ☐ Dynamic daily reports presented as charts or graphs ☐ Only reflects the analysis of data that is available online and may not reflect the reality of each collection or group of collections, especially those with a low percentage of digitized data ☐ With the growth of the network (more collections and more digital data) the indicators will support the identification of information gaps (geographic and taxonomic)



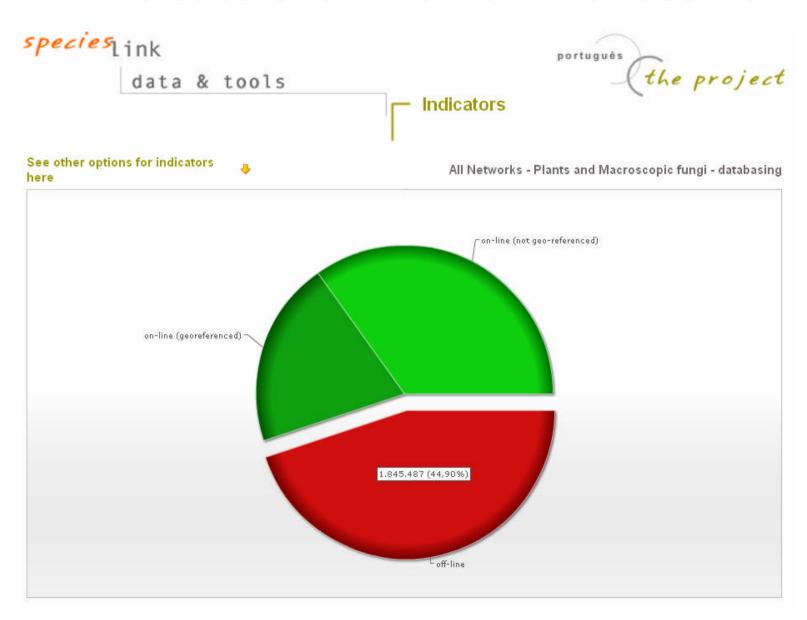
All indicators presented are based on on-line data only. They are dynamic or daily reports presented as charts or graphs. The indicator reflects only the analysis of data that is available on-line so therefore may not reflect the reality of each collection, especially those with a low percentage of digitized data.



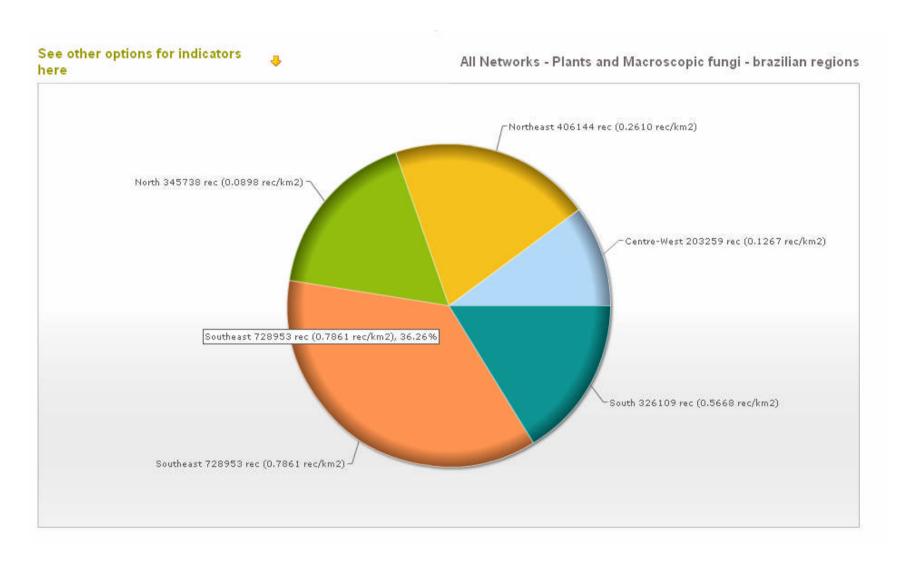
This graph reflects the movement of data (entry and removal) in the network. Monthly averages are presented for both total on-line records and total georeferenced records.



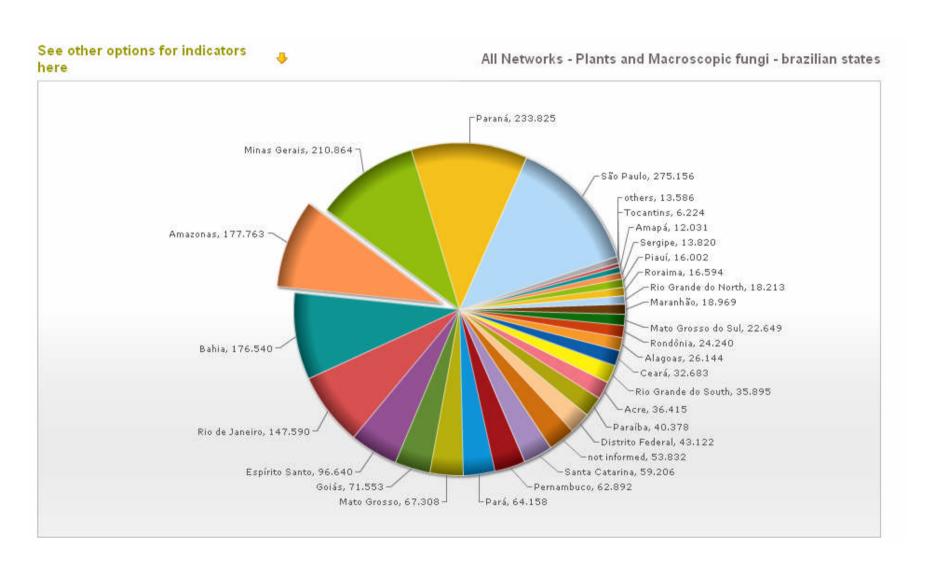
Indicators: online x offline records



Indicators: records per region, per km2



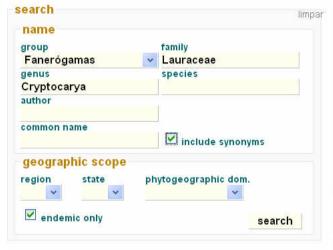
Indicators: records per state











Open online system

List of Species of the Brazilian Flora

Brazil, as a party of the Convention on Biological Diversity, assumed a series of commitments to be delivered by 2010. To prepare a list of Brazilian formally described species of plants, animals and microorganisms is among them. The Rio de Janeiro Botanic Garden was designated by the Ministry of the Environment to coordinate the elaboration of the List of Species of the Brazilian Flora, Moreover, besides the political obligation, such list represents an old dream of the Brazilian botanical community.

In order to achieve it, is necessary that taxonomists who actively study the Brazilian flora work together on a purpose built database that will allow not just the elaboration of the list, but will also serve as a baseline to gather further information regarding Brazilian biodiversity.

Welcome to the List of Species of the Brazilian Flora system.

Coordination

Rafaela Campostrini Forzza, Jardim Botânico do Rio de Janeiro.

Organizer Committee

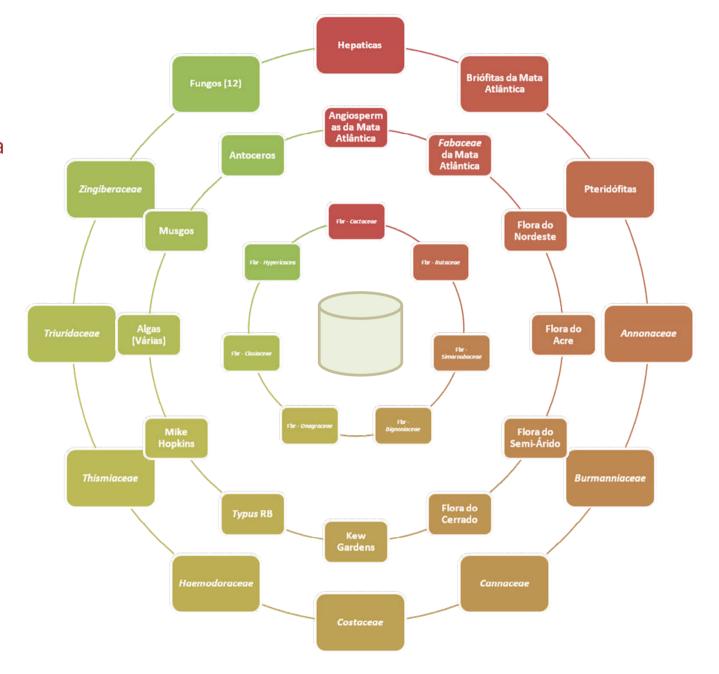
Andrea Costa (Museu Nacional): Ariane Luna Peixoto (JBRJ): Bruno Machado Teles Walter (CENARGEN); Daniela Zappi (KEW); Eduardo Lleras (CENARGEN); Gustavo Martinelli (JBRJ); Haroldo Cavalcante de Lima (JBRJ); João Renato Stehmann (UFMG); José Fernando Baumgratz (JBRJ); José Rubens Pirani (USP); Lucia G. Lohmann (USP); Luciano Paganucci (UEFS): Marcos Silveira (UFAC): Marcus Nadruz (JBRJ): Maria Cândida Mamede (IBt-SP); Maria Nazaré C. Bastos (Museu Goeldi); Maria Regina Barbosa (UFPB); Marli Pires Morim (JBRJ); Mike Hopkins (INPA); Ricardo Secco (Museu Goeldi); Taciana Cavalcanti (CENARGEN); Vinícius de Castro Souza (ESALQ/USP); Denise Pinheiro da Costa (JBRJ) (Briofitas): Lana Sylvestre (UFRRJ) & Jefferson Prado (IBt-SP) (Pteridofitas): Leonor Costa Maia (UFPE) & Anibal Alves de Carvalho Jr. (JBRJ) (Fungos); Mariângela Menezes (Museu Nacional) & Carlos Bicudo (IBt-SP) (Algas).

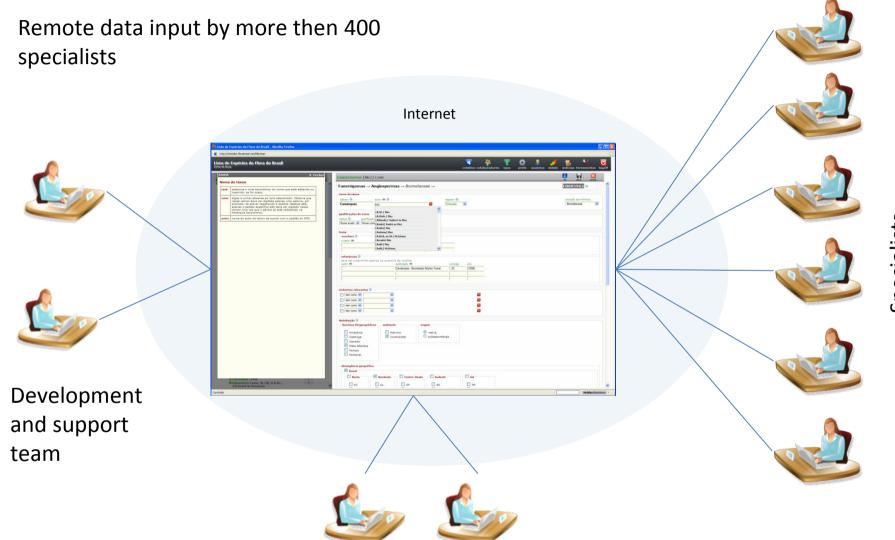
Information System

Sidnei de Souza, Dora A. L. Canhos, Centro de Referência em Informação Ambiental, CRIA

Contact

Paula Leitman, Assistant Email: listadobrasil@jbrj.gov.br Uploading and integrating data from available lists





Coordination





















Users	
Adriana de Mello Gugliotta Adriana Guglieri	Polyporales
Adriana Lobão	Annonaceae
Alain Chautems	Gesneriaceae
Alessandro Rapini	Apocynaceae
Alexa Araujo de Oliveira Paes Coelho	Portulacaceae
Alexandre Indriunas	Acanthaceae
Alexandre Quinet	Lauraceae
Alexandre Salino	Ctenitis Cyclodium Dryopteris Equisetaceae Megalastrum Salviniaceae Stigmatopteris Thelypteridaceae
Allan Carlos Pscheidt Anabela Silveira de Oliveira-Deble	Euphorbiaceae
Ana Claudia Araújo	Cyperaceae Juncaginaceae Thurniaceae
Ana Flora de Novaes Pereira	Anemiaceae Lomariopsidaceae Lygodiaceae Schizaeaceae
Ana Luiza Andrade Côrtes	Acanthaceae
Ana Maria Giulietti	Eriocaulaceae
Ana M.G.A. Tozzi	Dahlstedtia Deguelia

Derris Desmodium Lonchocarpus Millettia Mucuna Muellera Tephrosia Zornia Ana Odete Santos Vieira Campanulaceae Onagraceae Ana Paula Fortuna Perez Zornia Ana Paula Santos-Alvimia Anomochloa Gonçalves Apoclada Arthrostylidium Athroostachys Atractantha

users activity log

users	logins	creates	edit	moves	delete	last logii
Flávio França	33	37	590		29	12-12-2009 17:3:
Jose M. Valls	21	48	413		10	12-12-2009 17:23
Maria do Carmo Estanislau do Amaral	13	1	28		1	12-12-2009 17:20
Mara Angelina Galvão Magenta	23	35	206			12-12-2009 17:02
Volker Bittrich	19	37	625		98	12-12-2009 16:42
Valquíria Dutra	11	90	360		29	12-12-2009 16:33
Haroldo Cavalcante Lima	116	213	1251		63	12-12-2009 15:23
Leila Carvalho da Costa	16	4	105		1	12-12-2009 15:23
Jomar Gomes Jardim	24		402		1	12-12-2009 14:36
Maurício Watanabe	25		58		4	12-12-2009 12:58
José Rubens Pirani	10	18	200		4	12-12-2009 12:4
Julie H. A. Dutilh	14		101			12-12-2009 12:4:
Bruno Walnoffer	6	6	18			12-12-2009 12:1
Cássia Mônica Sakuragui	36	4	364		16	12-12-2009 11:5
Ana Maria Giulietti	8		188			12-12-2009 11:5
Andréa Pozetti Spina	17		215		3	12-12-2009 11:4
José Fernando Baumgratz	48	109	695		24	12-12-2009 11:2
Raymond Harley	14	1	124			12-12-2009 11:2
Lilian Auler Mentz	25	47	586		12	12-12-2009 10:5
Maria Silvia Ferruci	34	29	585		5	12-12-2009 10:4
Andréa de Araújo	1					12-12-2009 10:3
Alessandro Rapini	17	65	788		27	12-12-2009 10:0
Rose Bortoluzzi	29	87	194		1	12-12-2009 07:5
Silvana C. Ferreira	5	17	88		1	12-12-2009 01:4
Genise Somner	18	11	106		2	11-12-2009 23:4:
Renée H. Fortunato	8	2	14			11-12-2009 22:2
Mara Rejane Ritter	12	3	111		7	11-12-2009 22:0
Tania Regina dos Santos Silva	6		25			11-12-2009 20:5
Luiza Kinoshita	21	33	133		2	11-12-2009 20:0
Rodrigo Duno De Stefano	6	13	65			11-12-2009 19:4
Luisa Senna	11	17	96		1	11-12-2009 19:3
Rafaela Campostrini Forzza	108	13	751		974	11-12-2009 18:5
Maria Leonor D. Rei Souza	55	97	532		4	11-12-2009 18:3
Milton Groppo	22	15	152		9	11-12-2009 18:3
Lúcia Rossi	7	27	65		-	11-12-2009 18:2
Angela M.S.F. Vaz	32	8	121		15	11-12-2009 18:0
Adriana de Mello Gugliotta	39	67	613		40	11-12-2009 17:5
Inês Cordeiro	8	100.00	40		2	11-12-2009 17:4
Pedro Acevedo Rodríguez	33	64	528		26	11-12-2009 17:4
Maria Regina Barbosa	24		117		4	11-12-2009 17:3
Daniela Santos Carneiro-Torres	6		110			11-12-2009 17:2
Elsie Franklin Guimarães	18	13	304		74	11-12-2009 17:2
Roberto Manuel Salas	42	15	464		10	11-12-2009 17:1
Maria das Graças Lapa Wanderley	39	1	327		5	11-12-2009 17:1
	9,5	20	445		24	11-12-2009 17:0









Colanthelia Eremocaulon Filqueirasia





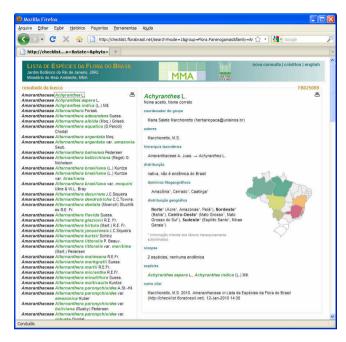






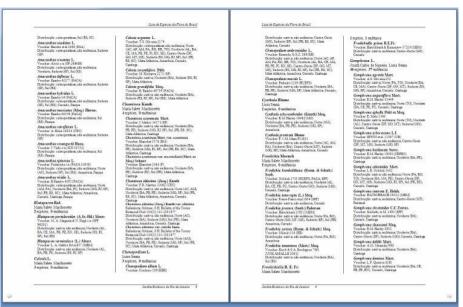




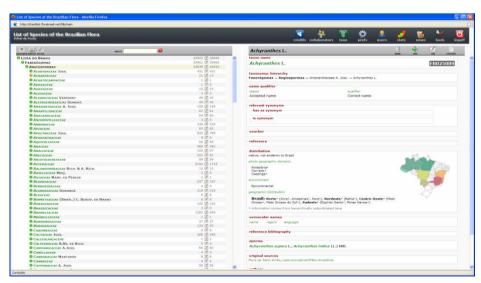


Public online output





Printing output



Working system for continuous update

Data Repatriation



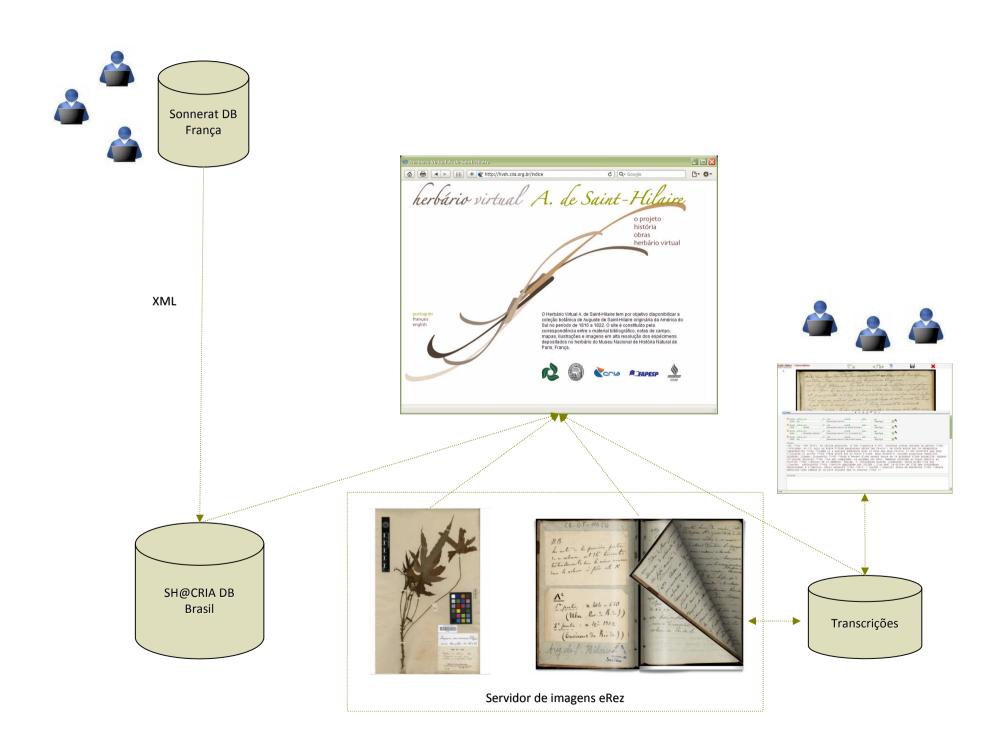


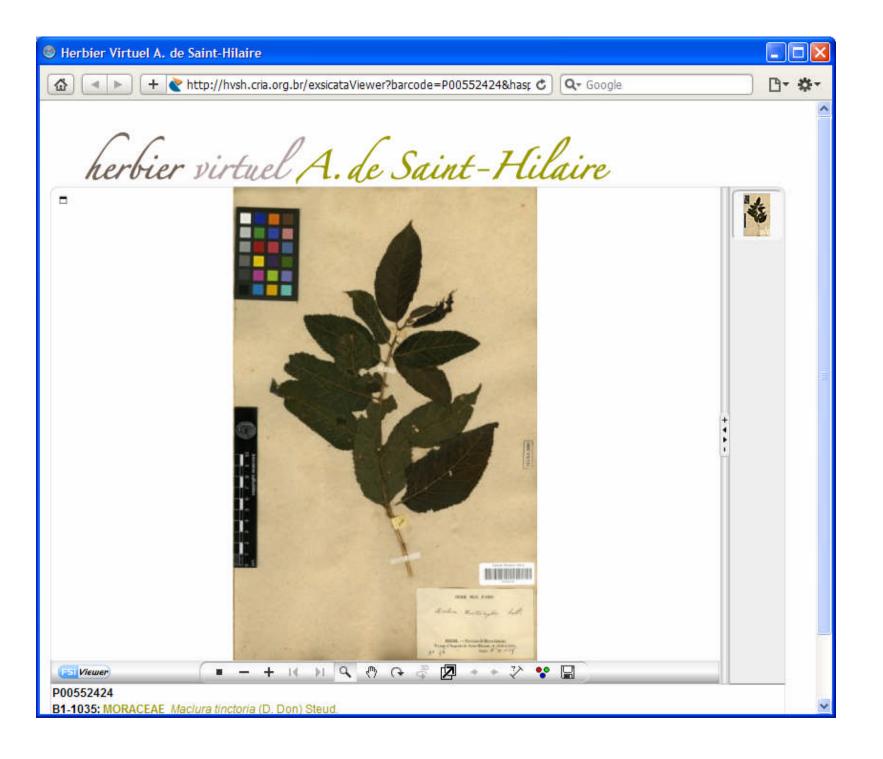
filter

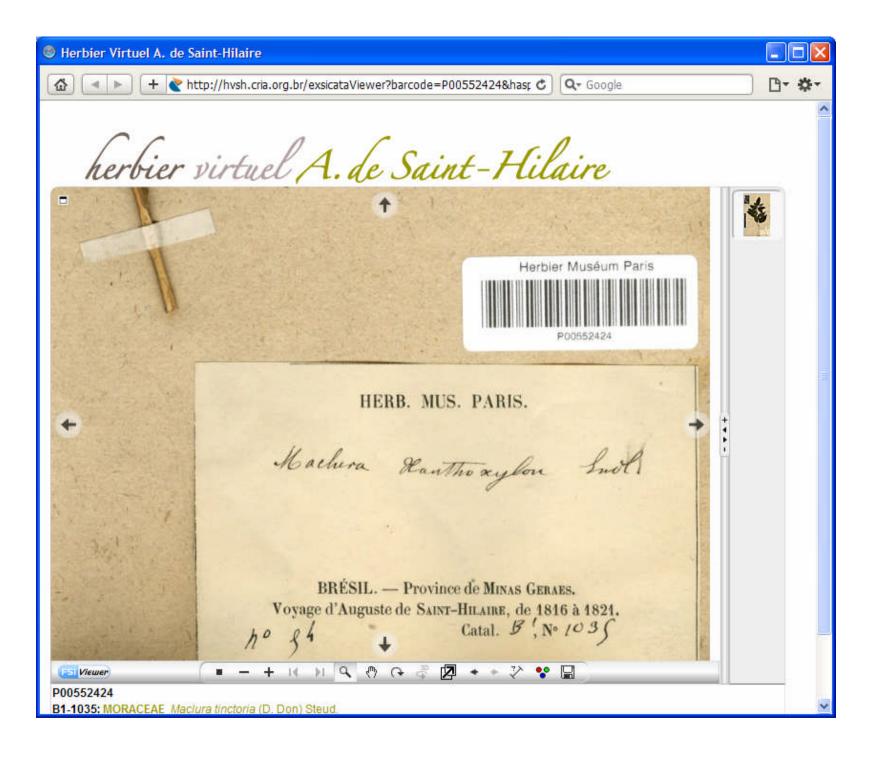
networkManager is a tool used to manage the collections of speciesLink network. In order to obtain more information on each collection, please click on the acronym. It is possible to "filter" the data by typing in keywords such as "plants" or "Ribeirão Preto" in order to view only information about plant collections or about collections locate in the Ribeirão Preto city. See the network indicators.

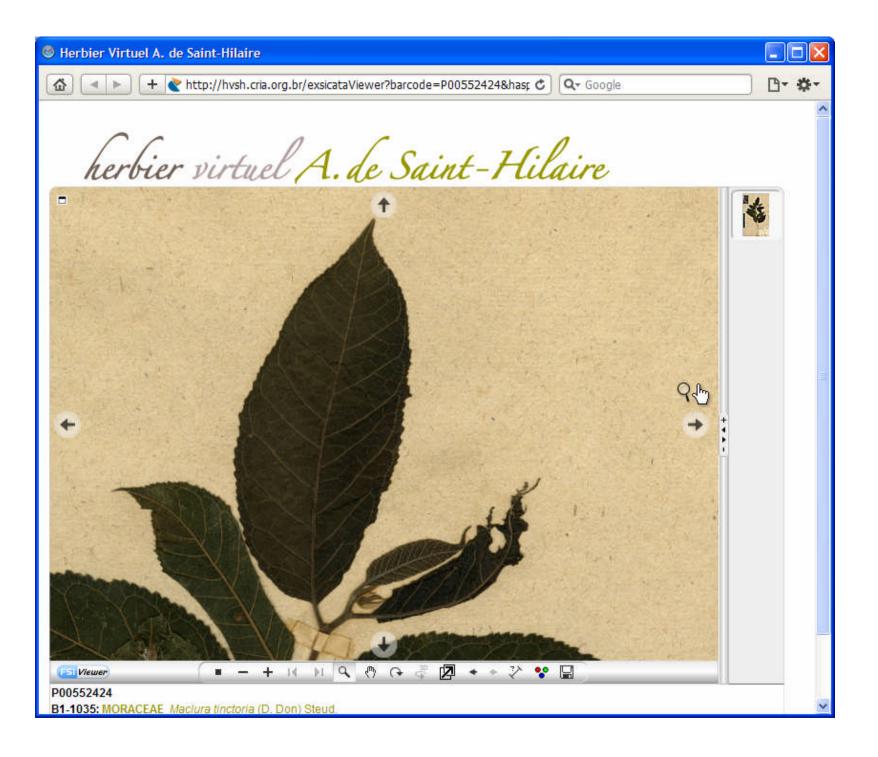
networkManager

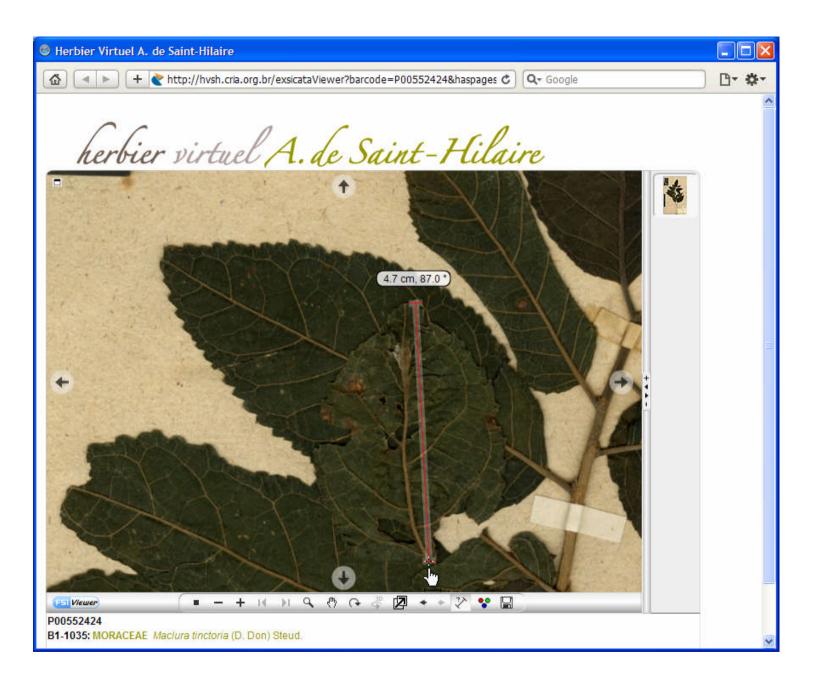
Click here to also see collections that are off-line								int	international			
acronym * •	county	state	network	software	records **	on-line	%	georef ***	%	auto georef****	update	
FPR Colombia	Bogotá	Cundinamarca	International	MS-Excel	47.021	47.021	100%	45.121	96%	0	13/02/2008	
MOBOT BR	St. Louis	Missouri	International	DiGIR provider	177.874	177.874	100%	44.465	25%	0	22/09/2009	
MVZ BR	Berkeley	California	International	DiGIR provider	5.778	5.778	100%	5.132	89%	0	17/04/2008	
NMNH Botany BR	Washington	District of Columbia	International	DiGIR provider	37.662	37.662	100%	2.289	6%	4.532	17/10/2009	
NYBG BR	Bronx	NY	International	MS-Access	241.281	241.281	100%	98.849	41%	83.998	11/09/2009	
ollections					509.616	509.616	100%	195.856	38%	88.530		
Date of last	update: 12/12	/2009 16:36										



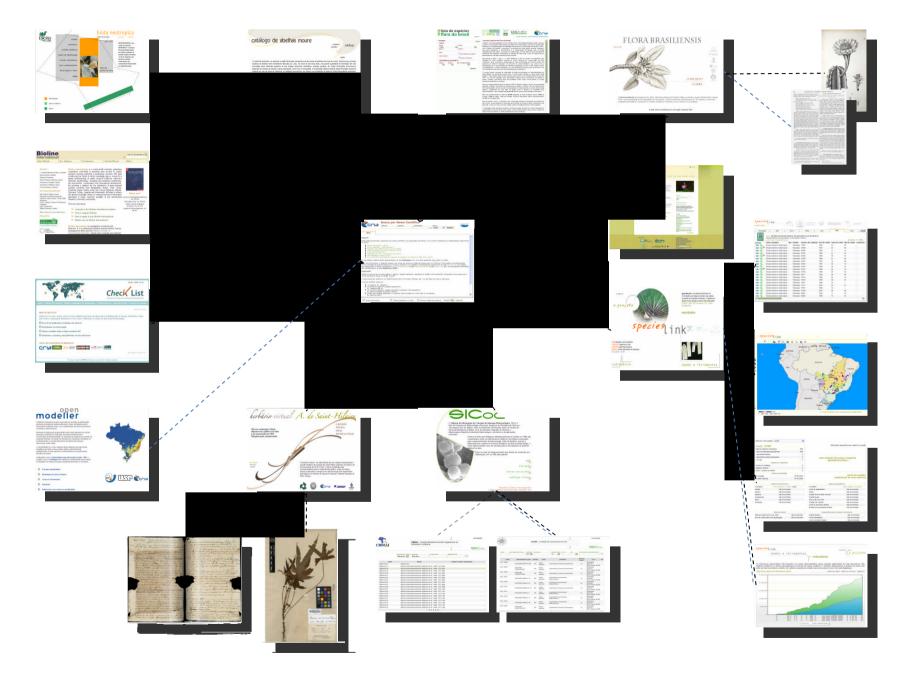








Sistema CRIA de informação



Financiadores e patrocimadores





CNPa









Parceiros internacionais



































THE NEW YORK BOTANICAL GARDEN







Parceiros nacionais

























MUSEU MACIONAL

UFRI













UNICAMP



Embrapa

E:DA

Eingreso Briano de Electrophismento Septendo S.A.













































UFERSA



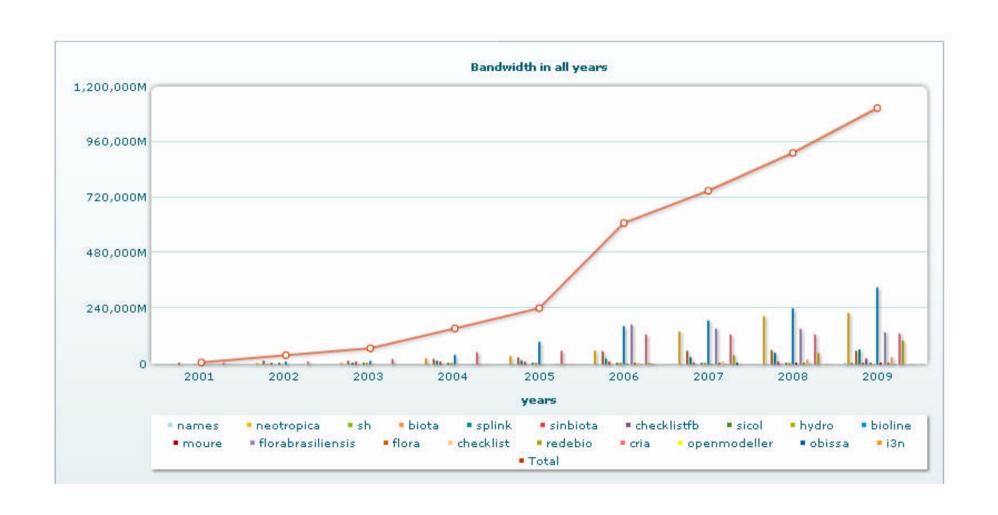




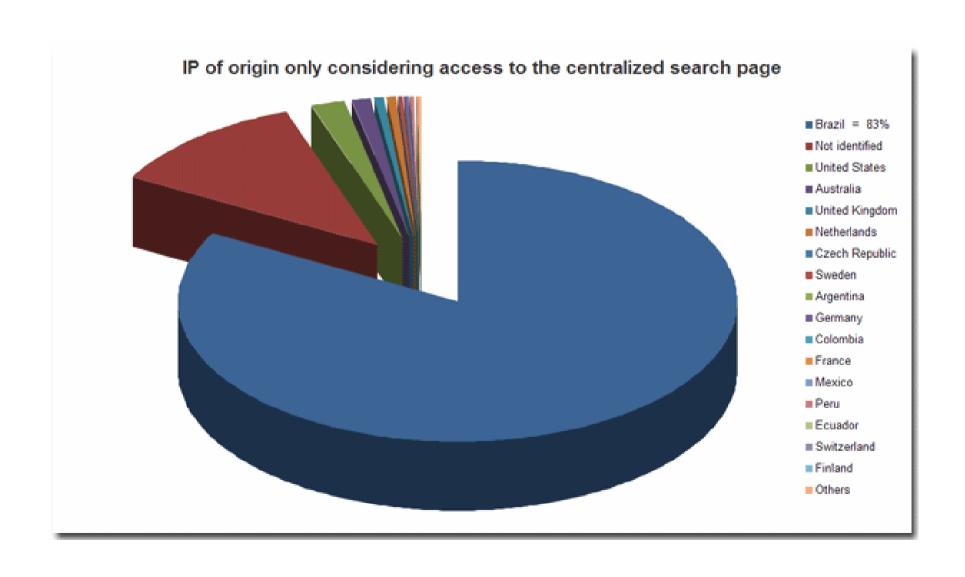
PUCRS



Usage is increasing!



Usage: mainly Brazil



Lessons learned

☐ Adoption of internationally agreed standards and protocols is key
☐ Support unlocking and sharing of data (make it simple and easy !)
☐ Enable data providers to have full control of their data determining what can be openly shared and what is sensitive
☐ Full credit and acknowledgement to the data providers at all levels!
☐ Data providers must see the benefit to participate in the network
☐ Data flagging and data cleaning tools are key to support the identification of data inconsistencies
☐ Stable and long term funding is necessary to ensure development and the persistency of open and free data networks (persistent repositories are critical; funding mechanisms need to be improved!)



Centro de Referência em Informação Ambiental

http://www.cria.org.br

Vanderlei Canhos vcanhos@cria.org.br