CRIA's contribution to open access to data from biological collections and biobanks in Brazil

Vanderlei Perez Canhos
vcanhos@cria.org.br
www.cria.org.br
blog.cria.org.br

português

the project

***species*Link** is a distributed information system that integrates primary data from biological collections. The development was funded by FAPESP, GBIF, JRS Foundation, MCTI, CNPq, FINEP, RNP and CRIA.

news

speciesLink

458 collections and sub-collections
8,745,560 online records
4,063,789 georeferenced
489,999 different species names
14 aug 2017 - 03:06

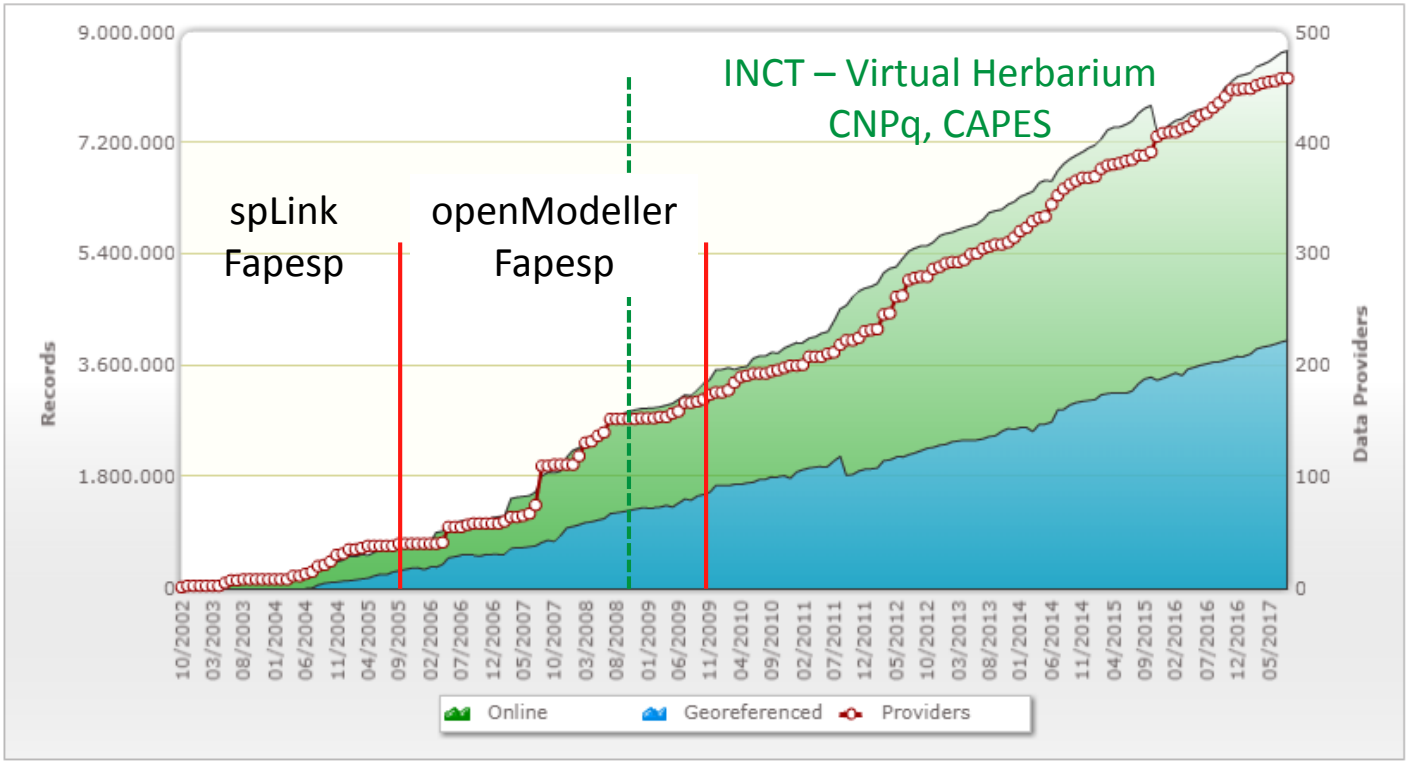
indicators

data & tools

All indicators presented are based on online data only. They are dynamic or daily reports presented as charts or graphs. The indicator reflects only the analysis of data that is available online so therefore may not reflect the reality of each collection, especially those with a low percentage of digitized data. Click here **to view the table data contribution x dependency per state**.

See other options for indicators here

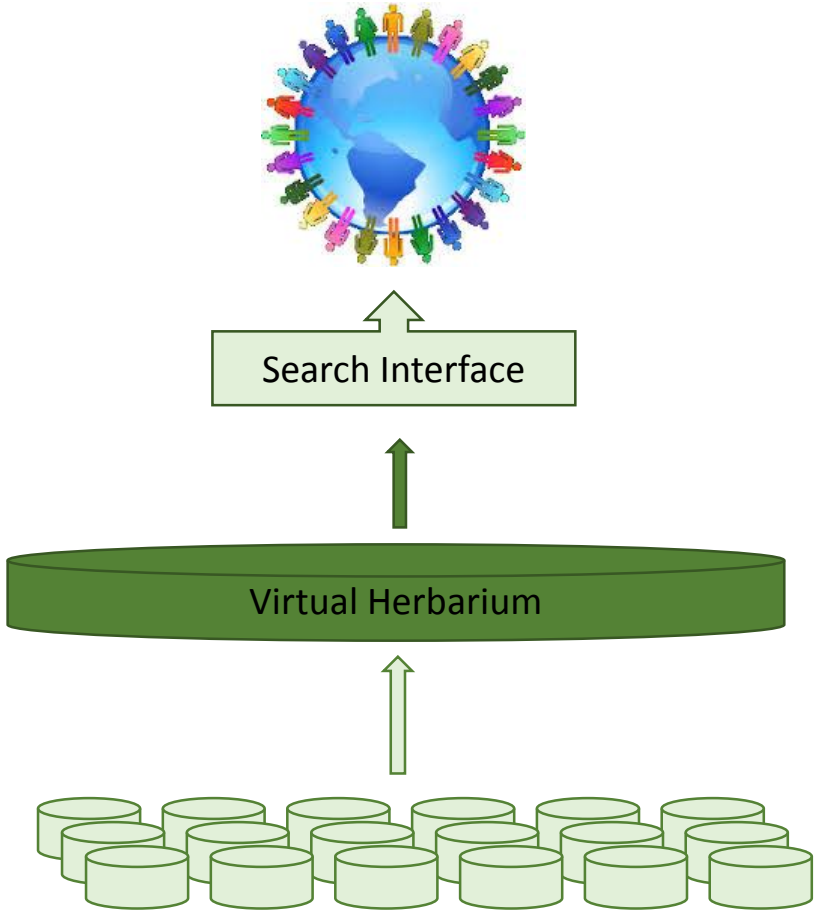
All Networks - All Collections - records

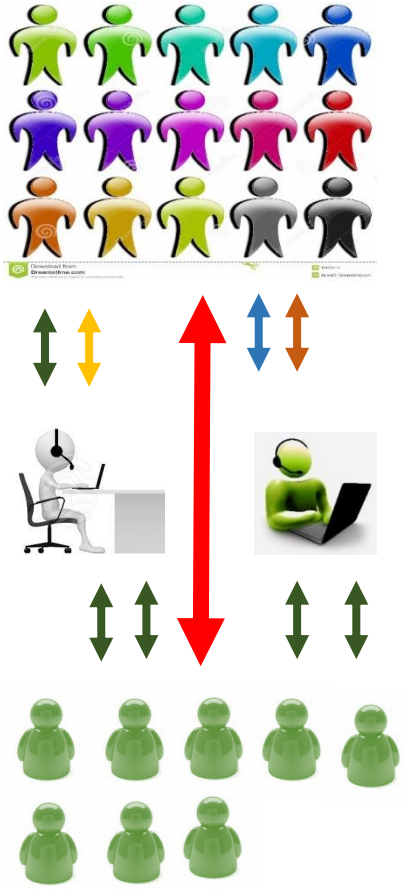


This graph reflects the movement of data (entry and removal) in the network. Monthly averages are presented for both total online records and total georeferenced records. The red line shows the number of data providers per month.

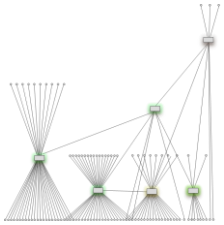
Updated on 14/08/17 05:01

Local e-Infrastructures





Herbaria and other biological collections



- species occurrence data
- Images (vouchers, live plants, pollen)
- maps, graphics, synthesis, catalogues (dynamic)

- data
- images
- maps
- ecological niche modeling

Search Interface

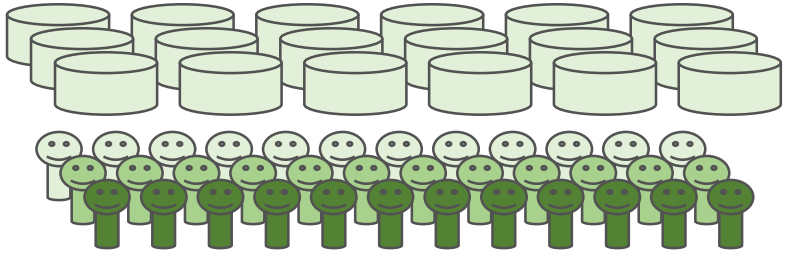
Web Services

INCT Virtual Herbarium of Flora and Fungi

Tools

- species occurrence data
- Images

- data cleaning
- automatic georeferencing
- indicators



- Cybertaxonomy
- Annotation system
- Lacunas
- BioGeo

Data and Tools: open and free access

speciesLink

português | the project

data & tools

search

Query the collections databases using the search engine

Assista ao vídeo com dicas de uso da rede *speciesLink*

YouTube

tools

applications

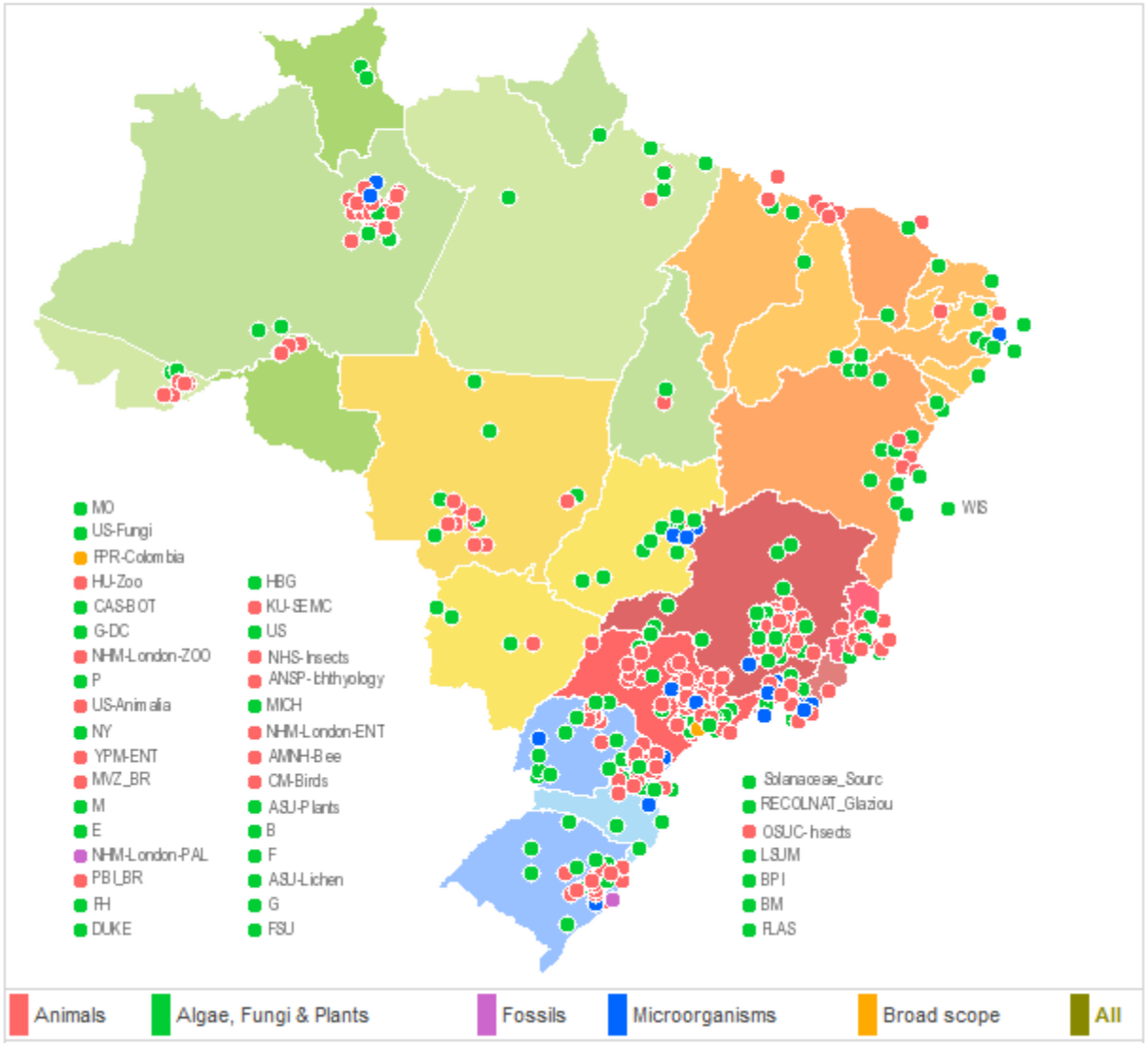
- **dataCleaning**
  This tool was developed to facilitate the detection of possible errors and to help collections in the process of standardizing data. (only portuguese version available)
- **geoLoc**
  Geo reference your data based on locality descriptions.
- **infoXY**
  Get locality information using geographic coordinates.
- **spOutlier**
  Detect outliers in latitude, longitude and altitude using techniques modified from Chapman, 1999.
- **speciesMapper**
  Plot geo referenced points on a map.
- **converter**
  Convert different types of representations of geographic coordinates and datum's.
- **spSearch firefox plugin**
  Optimize the search of species scientific names in the speciesLink network, which can be done directly on your Mozilla Firefox without having to visit the web page.
- **networkManager**
  Manager of the metadata of each participating collection of the speciesLink network.
- **indicators**
  Dynamic or daily reports are submitted in the form of maps and charts.
- **Lacunas**
  Tool that presents the status of the data available through the speciesLink network for each species listed in the List of Species of the Brazilian Flora.
- **Biogeo**
  System that uses ecological niche modelling techniques to improve our understanding of plant biogeography in Brazil. Potential distribution maps can be generated, searched, visualized and downloaded for each species.
- **openModeller**
  Open source spatial distribution modelling tool used to predict species distribution.
- **spLinker**
  Application that enables biological collections to mirror their data in regional servers integrated to the speciesLink network.
- **speciesBase**
  Its system to record specimens information based on Darwin Core.
- **plant collectors in Brazil**
  The aim was to offer a standard list of collectors' names, important to taxonomic studies and herbaria digitization projects. This work ended together with the project, but the results obtained are available online.

cria

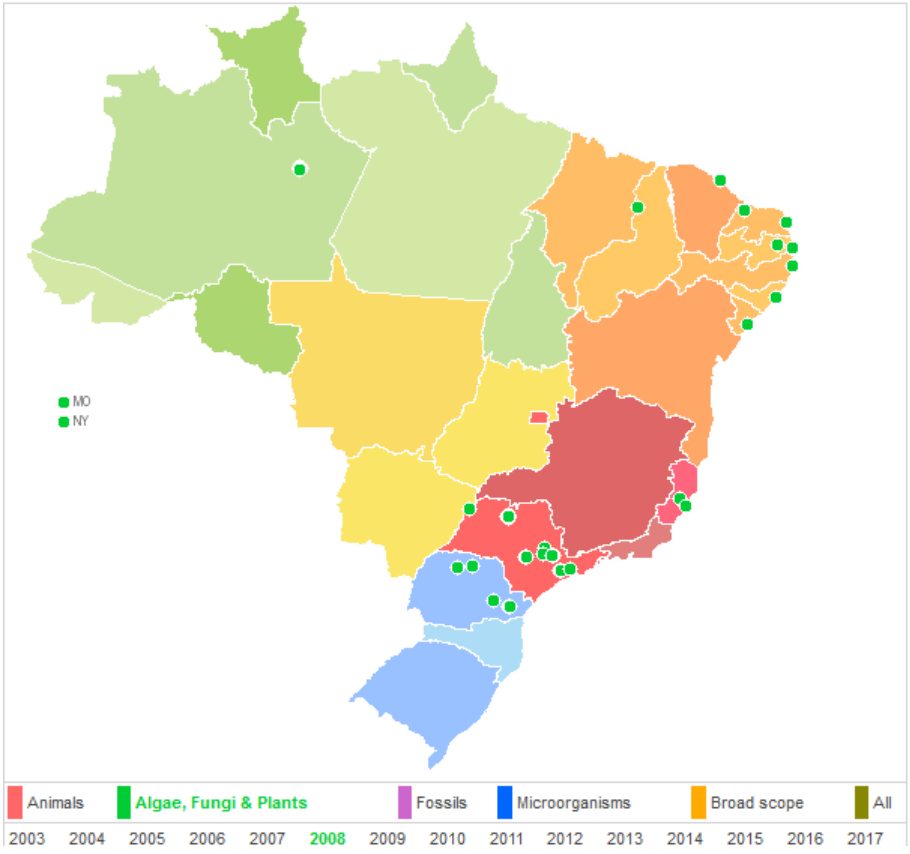
Geographic location of data providers

speciesLink

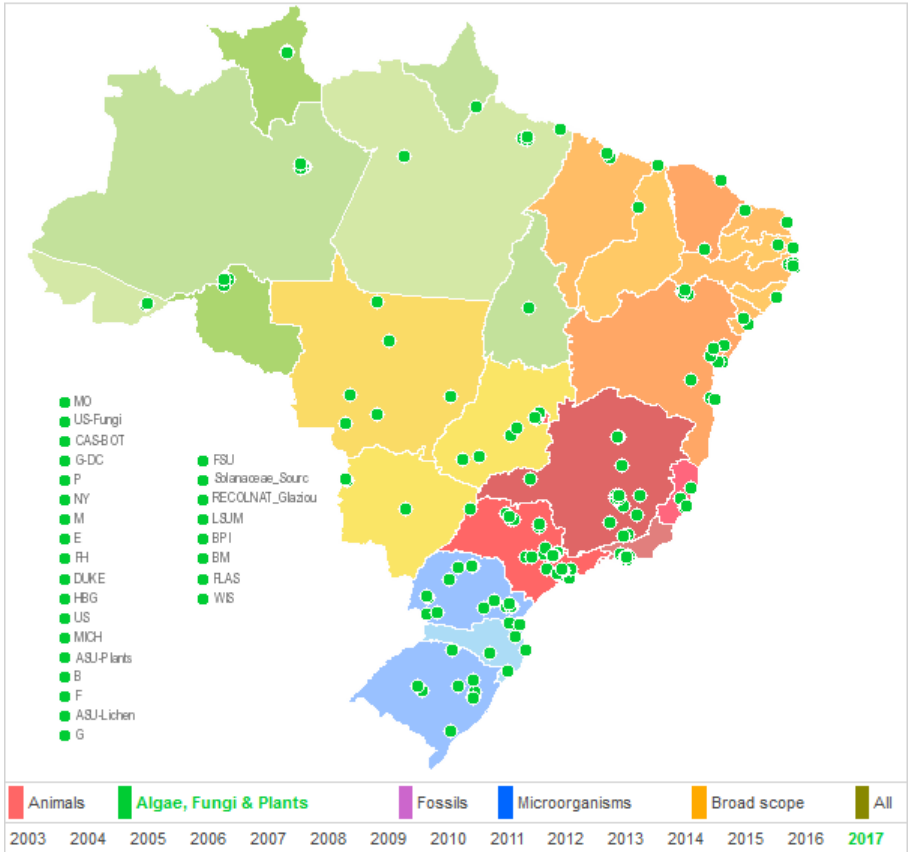
458 datasets in 2017



Virtual Herbarium of Flora and Fungi

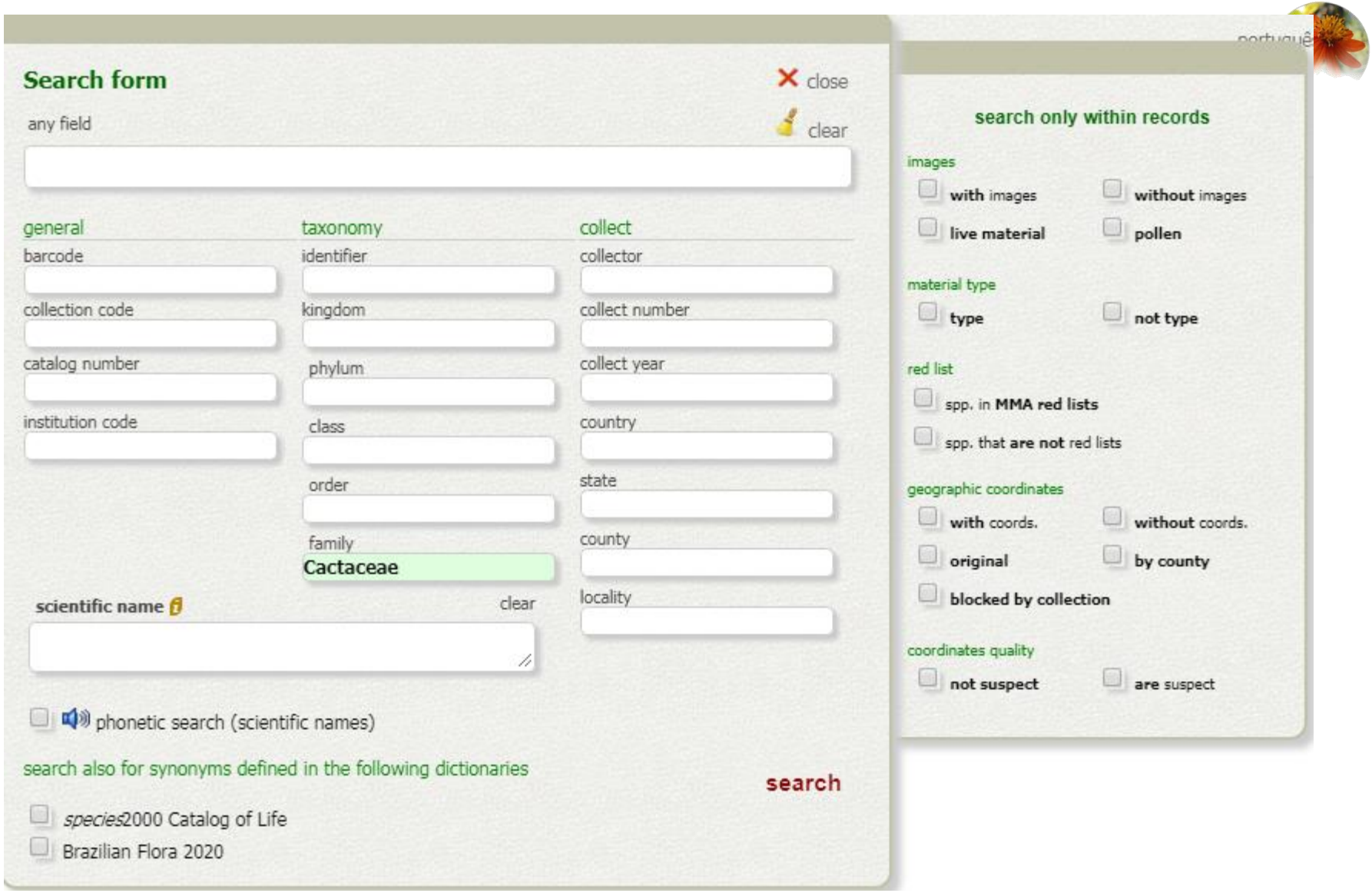


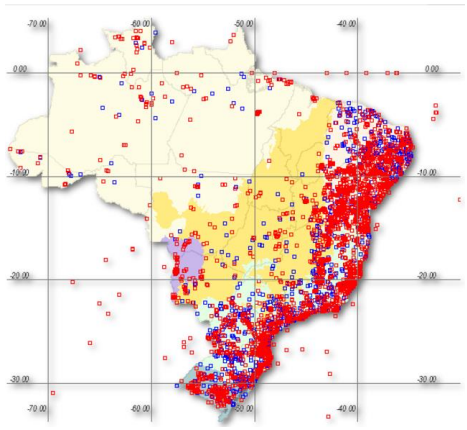
2008: 47 datasets

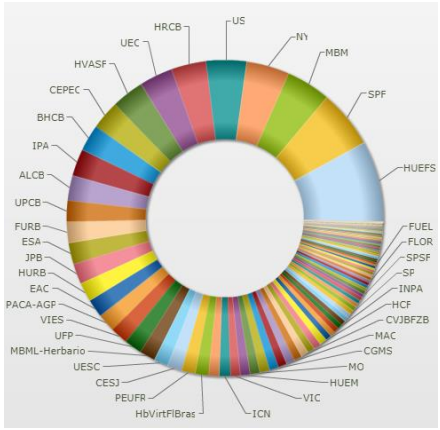


2017: 192 datasets

Search interface: data & tools







spp. names	records
accepted	14,025
synonyms	1,271
ambiguous	13
Genus id only	1,210
not found	909
not identified	564

135
collections and subcollections

17,979
online records

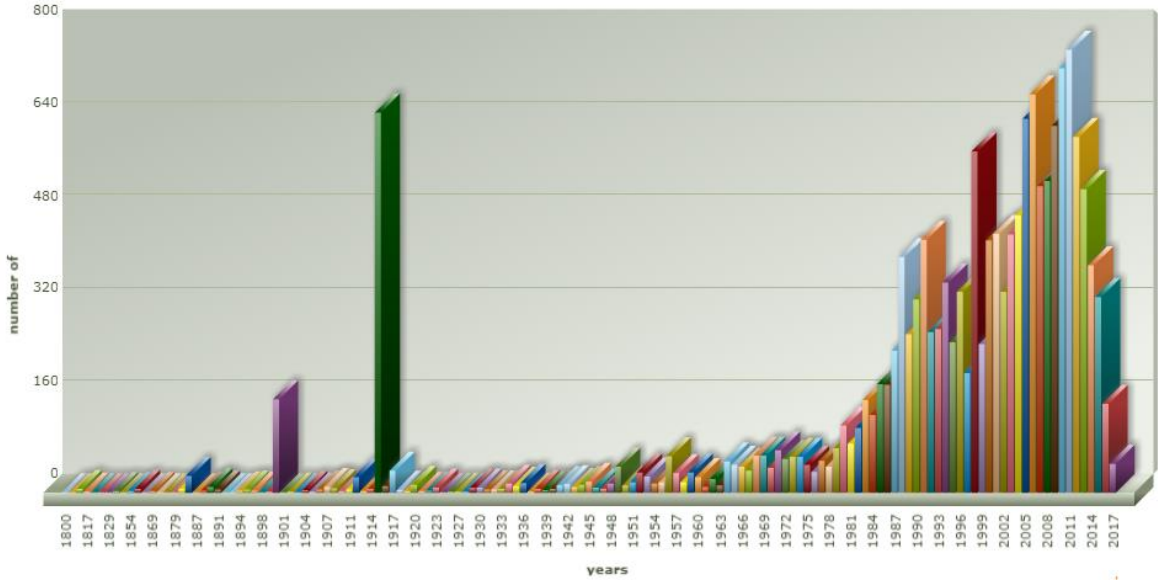
14,387
georeferenced records

5,456
records with images

202
type records

755
records of species in redlists

337
distinct accepted species



PLANTAE ANGIOSPERMAE DICOTYLEDONEAE CARYOPHYLLALES CACTACEAE

Melocactus glaucescens Buining & Brederoo **MMA Flora Red List**

HUEFS 112604 Collect: Lambert, S.M. 4 **22/10/2004**.

Loc: Ca. 05 Km após o distrito de Brejões, estrada para o Lajedo Bordado., Morro do Chapéu, Bahia, Brasil

Coord orig: [*lat:* -11.439722 *long:* -41.085 WGS84] Altitude: 748m.

Notes: Campo rupestre.

Basis of record: preserved specimen

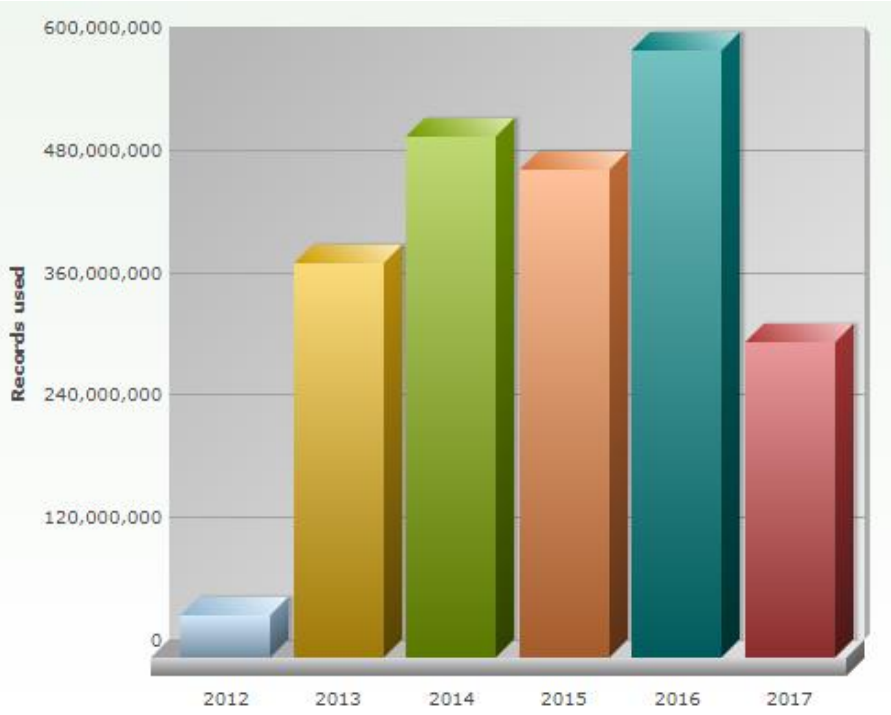
*species*Link usage

2.324.645.784

billion million thousand

total records used since Oct 05, 2012

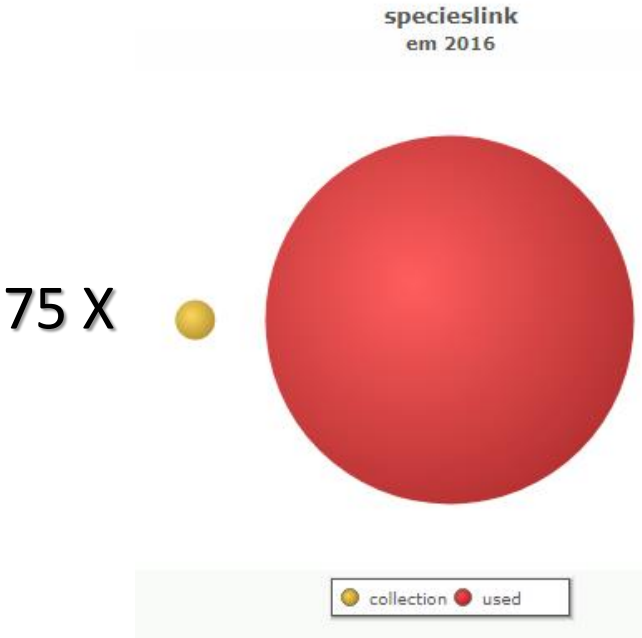
period	records
2012	42.246.860
2013	387.217.394
2014	510.601.392
2015	478.023.079
2016	595.891.817
2017	310.665.149
TOTAL	**2.324.645.691**



2016 > 1.6 million records used per day

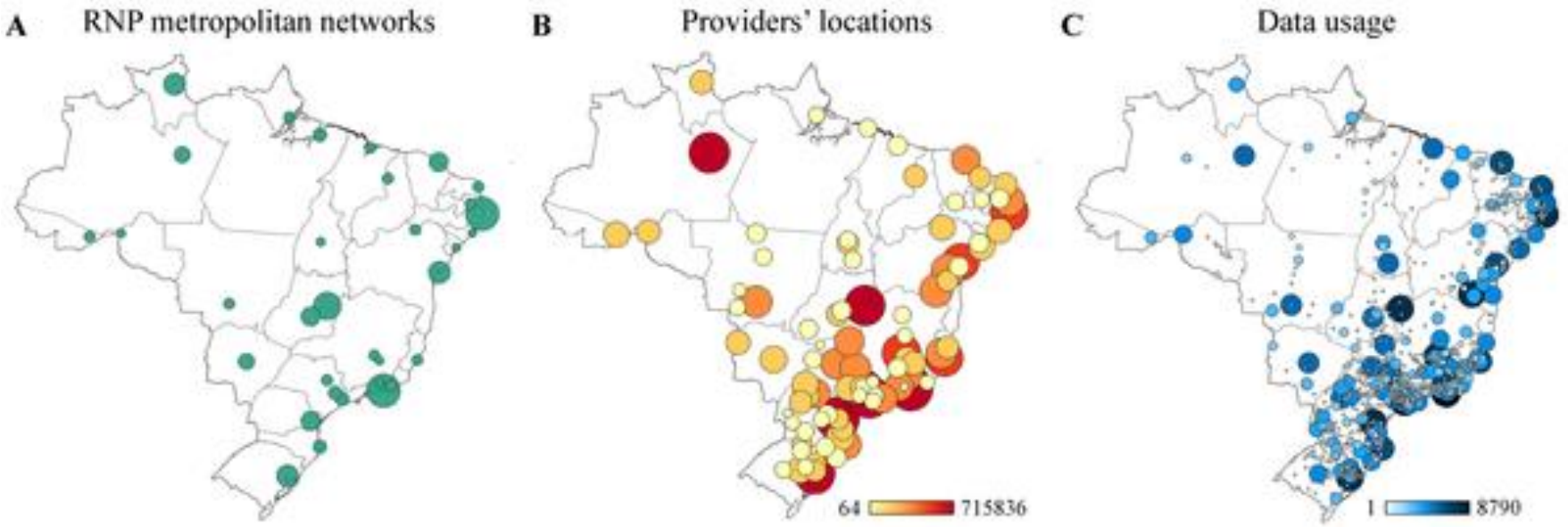
Images served

period	images
2012	49.715
2013	552.847
2014	2.067.836
2015	3.170.330
2016	3.058.288
2017	1.842.076
TOTAL	**10.741.092**



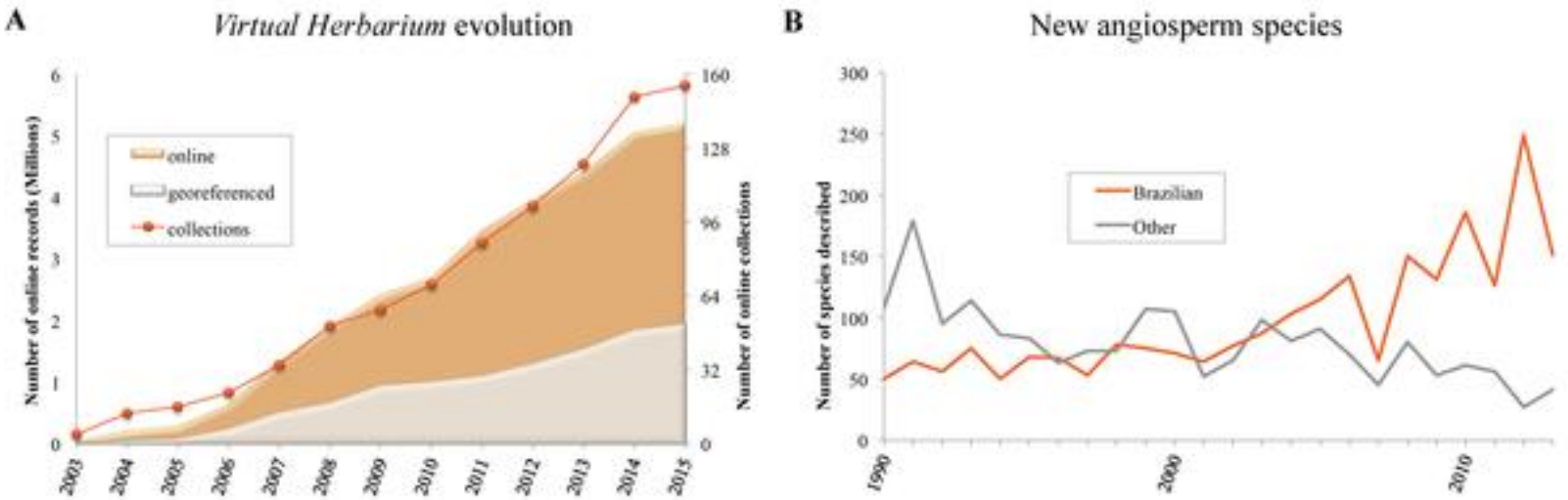
Research infrastructure and biodiversity data usage in Brazil.

1.4 million records/day



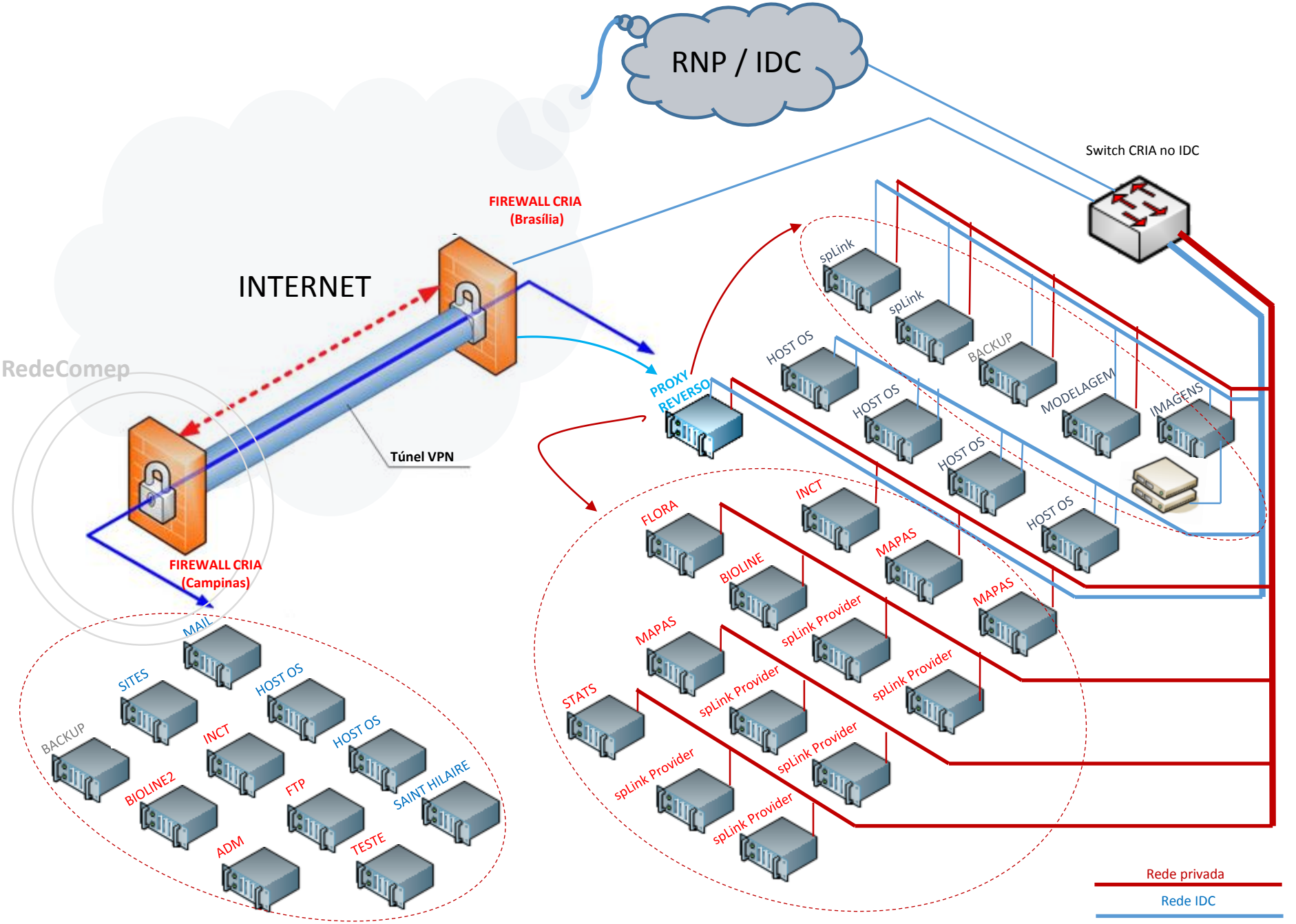
Canhos DAL, Sousa-Baena MS, de Souza S, Maia LC, Stehmann JR, et al. (2015) The Importance of Biodiversity E-infrastructures for Megadiverse Countries. PLoS Biol 13(7): e1002204. doi:10.1371/journal.pbio.1002204
http://journals.plos.org/plosbiology/article?id=info:doi/10.1371/journal.pbio.1002204

The evolution of Brazilian flora primary data available online and species described by Brazilian specialists



Canhos DAL, Sousa-Baena MS, de Souza S, Maia LC, Stehmann JR, et al. (2015) The Importance of Biodiversity E-infrastructures for Megadiverse Countries. PLoS Biol 13(7): e1002204. doi:10.1371/journal.pbio.1002204
http://journals.plos.org/plosbiology/article?id=info:doi/10.1371/journal.pbio.1002204

RNP - CRIA Partnership: Public information systems at RNP'S *Internet Data Center*



Timeline: Microbial Resource Centres Network

OECD

GBRCN
GLOBAL BIOLOGICAL RESOURCE CENTRE NETWORK
DEMONSTRATION PROJECT

OECD	OECD +MCT		Pilot Study
1998-2002	2002-2006	2007	2008-2012
	FINEP	FINEP	CT-CRB e prog. acc. Inmetro

2013

ISO

DarwinCore2

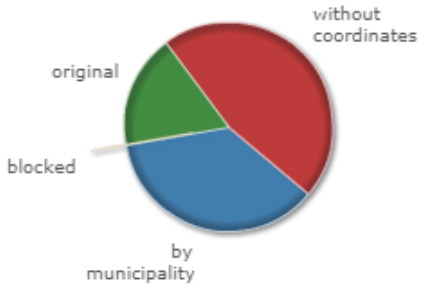
+ microbial

SICol net

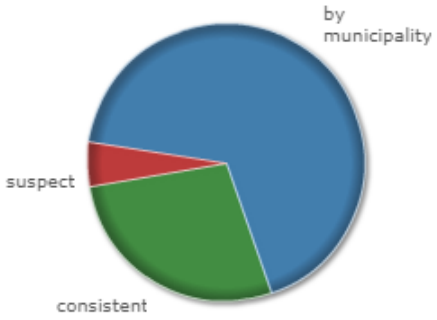
speciesLink

coordinates	records
by municipality	12,061
original	5,843
blocked	0
without coordinates	15,445
suspect	911
consistent	4,932
by municipality	12,061

coordinates origin



coordinates quality



34
collections and subcollections

33,349
online records

17,904
georeferenced records

629
type records

1,581
distinct accepted species

+ More...

The figures represent the **number of records** in the speciesLink network with geographic coordinates. **By municipality** are records with coordinates automatically assigned by a georeferencing tool. **Blocked coordinates** are records that do have coordinates but the data provider considers this information sensitive. **Suspect coordinates** refer to records with original coordinates that do not fall within the municipality informed.

images	quantity
voucher	198
live material	0
pollen	0
subtotal	**198**
labels	0
Flora brasiliensis	0
total	**198**

type of material width images

The figures represent the **number of distinct images** associated to the retrieved records according to what they represent: vouchers, pollen, live material, labels or *Flora brasiliensis* drawings.

spp. names	records
accepted	15,460
synonyms	3,596
Genus id only	4,425
not found	9,533
not identified	335

names quality

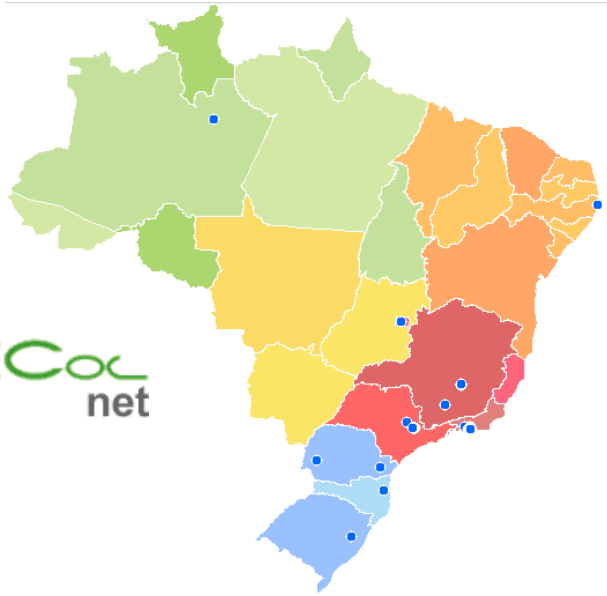
The figures represent the **number of records** for each name status: accepted, synonym, identified to genus level only, not found or not identified (meaning genus and species equal blank). The system only recognizes names with correct spelling based on the available lists. Please check the "tips" link above for more detailed information.

DarwinCore2

+ microbial

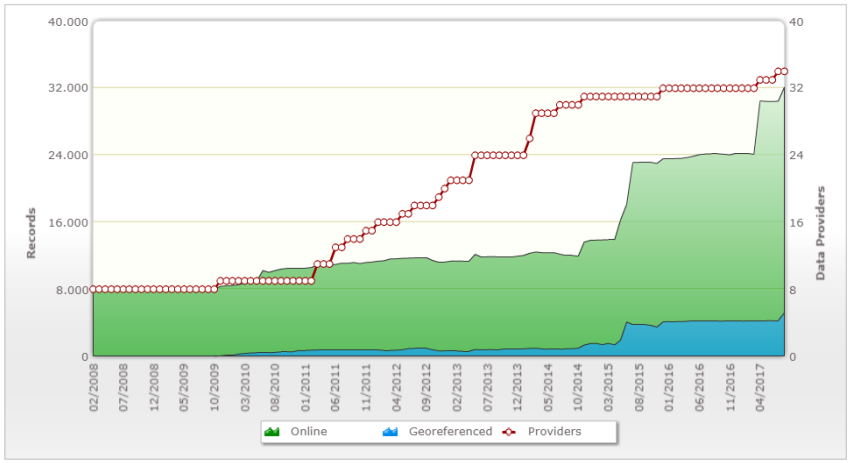
SICol net

speciesLink



SICol net

See other options for indicators here



This graph reflects the movement of data (entry and removal) in the network. Monthly averages are presented for both total online records and total georeferenced records. The red line shows the number of data providers per month.

Updated on 14/08/17 05:39

Source: BGB, BR, CBMAI, CCDCA, CFAF, CFEUnioeste, CG, CICG, CMRP, CRM-UNESP, Fepagro-SEMIA, Fiocruz-CBAM, Fiocruz-CBAS, Fiocruz-CCAMP, Fiocruz-CCBH, Fiocruz-CCFF, Fiocruz-CCGB, Fiocruz-CENT, Fiocruz-CFAM, Fiocruz-CFP, Fiocruz-CLEP, Fiocruz-CLIOC, Fiocruz-CLIST, Fiocruz-CMRVS, Fiocruz-CMT, Fiocruz-COLPROT, Fiocruz-COLTRYP, Fiocruz-CYP, Funed-Virus, IBSBF, LaFMA, UFMG-CM, URM-Micoteca

34 datasets of microbial collections

209 X

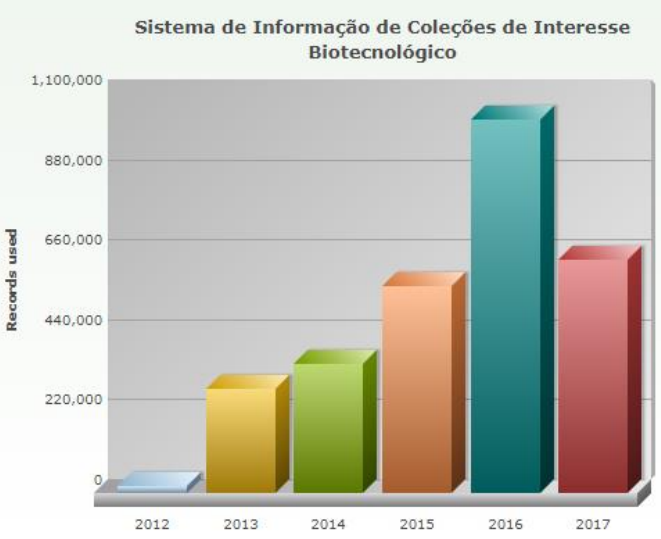
Records used

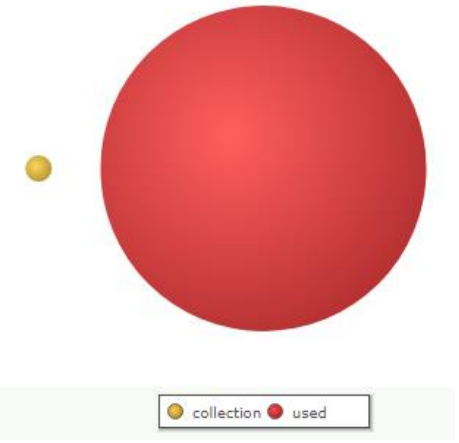
2.909.042

million thousand

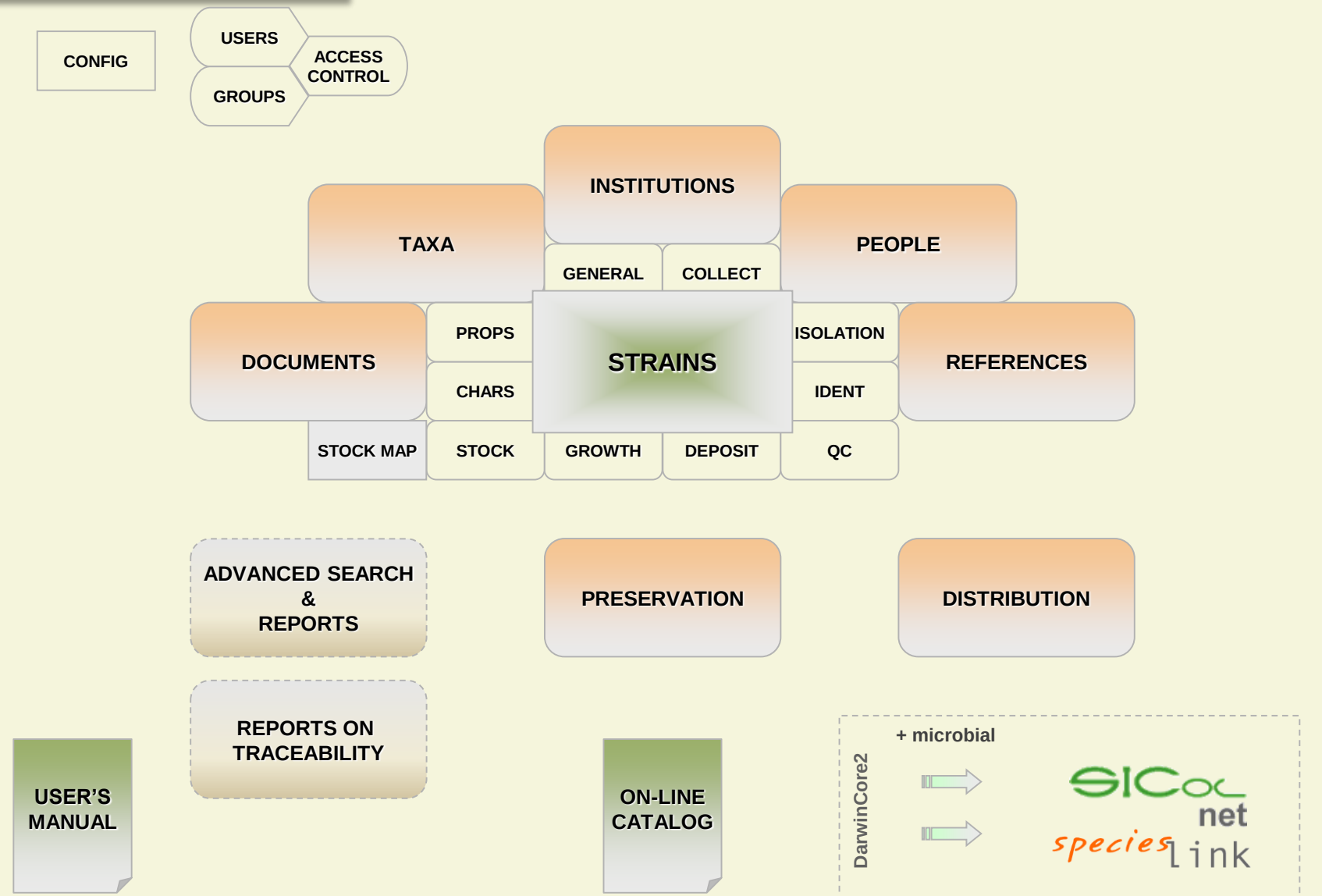
total records used since Oct 05, 2012

period	records
2012	22.091
2013	288.893
2014	356.322
2015	569.602
2016	1.028.965
2017	643.169
TOTAL	2.909.042





µSICol • components



ISO/TC 276 - Biotechnology

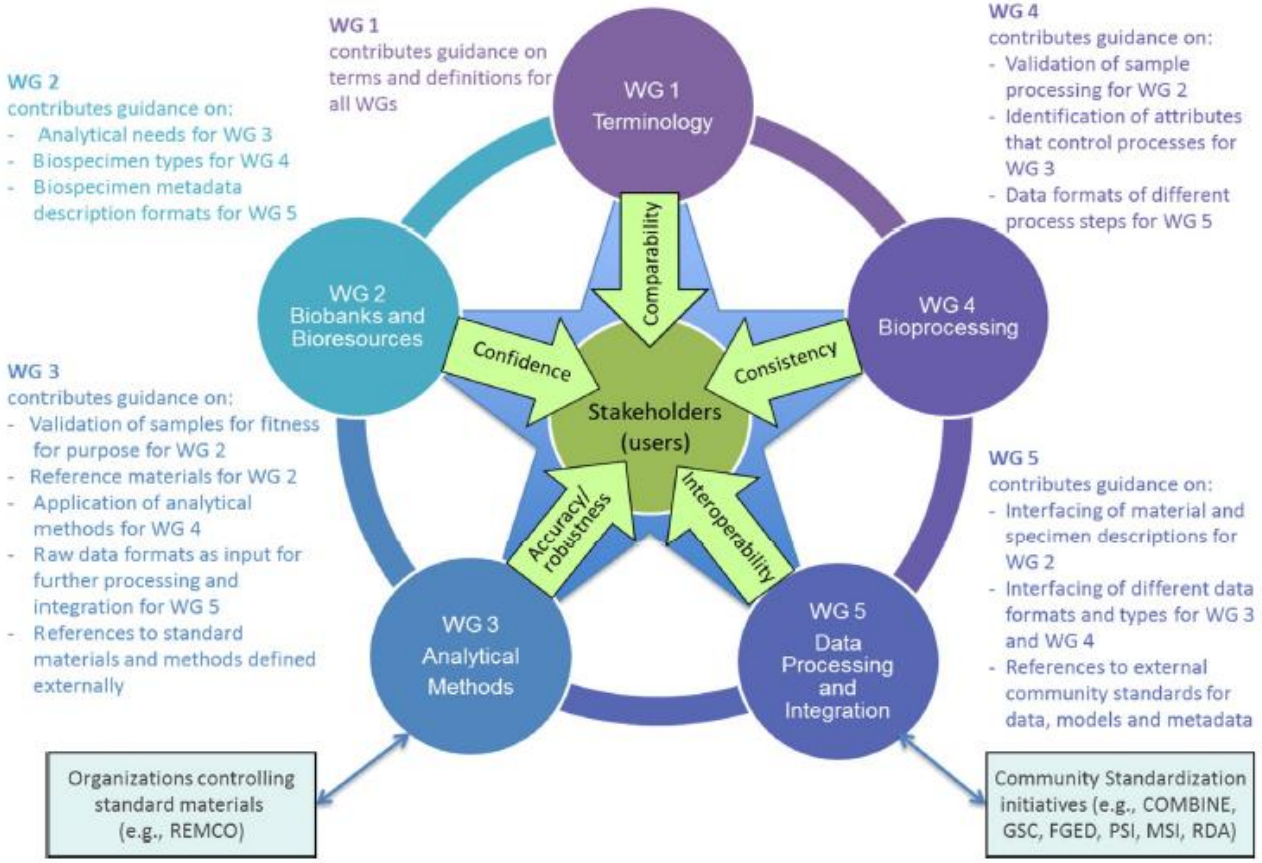
ABNT/CEE 276 - Biotecnologia

ABNT/CEE 276 - Biotecnologia

- Comissão de Estudos Especiais (CEECT) established in August/2016 to mirror ISO/TC 276 activities
- Coordinator: Paulo Holanda (Bioquallis)
- Secretary: Antonio Saraiva (Inmetro)
- Analist: Helena Mendes (ABNT)
- 5 working groups
  - WG1 Terminology
  - WG2 Biobannks and Biological Resources
  - WG3 Analitical Methods
  - WG4 Bioprocessing
  - WG5 Data Processing and Integration (Rapporteur – Vanderlei Canhos)

www.abnt.org.br/cee276

ISO/TC 276 Biotechnology



Thank you!

Vanderlei Perez Canhos
vcanhos@cria.org.br
www.cria.org.br
blog.cria.org.br