

Brazil's Virtual Herbarium

Dora Ann Lange Canhos

dora@cria.org.br

<http://inct.splink.org.br>

Types of data and format

Herbarium data – Darwin Core (data model)

- Collection Acronym & Code (unique identifier)
- Scientific name, who & when identified
- Field Data: where, when, who collected
- Notes: ...
- Images (live plant, dried sample, pollen, ...): barcode; tiff, jpeg, png, ...

Associated data (add value): web services/download

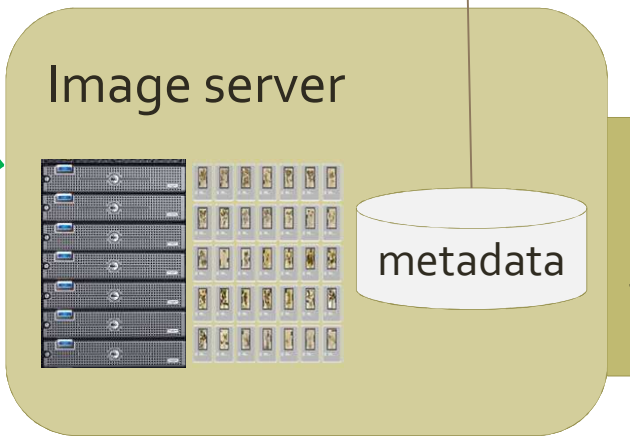
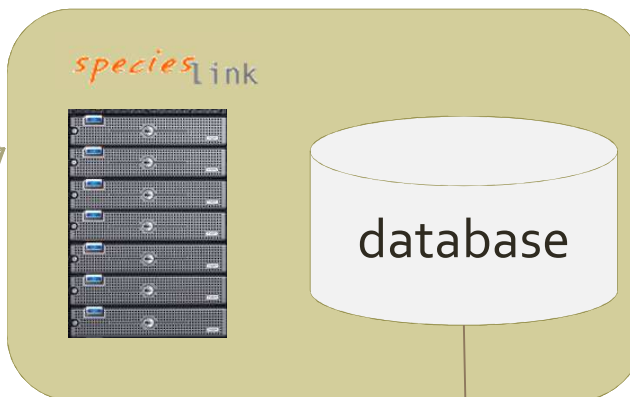
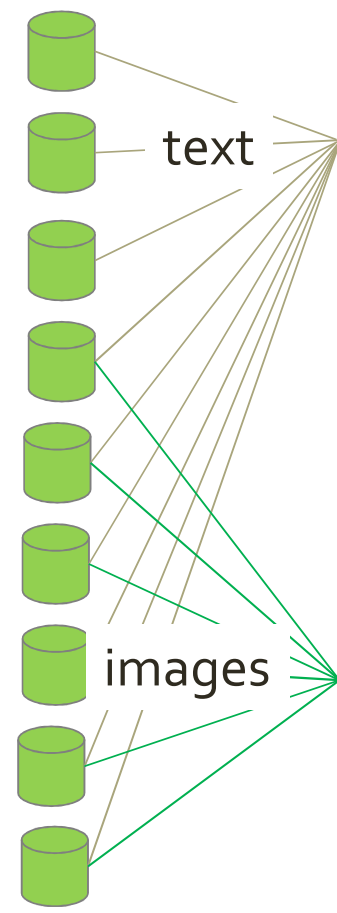
Internal: Flora Brasiliensis, Lacunas, BioGeo, Journals

External:

- Taxonomic lists: Catalogue of Life, Flora 2020 Brazil
- Red List (endangered species)
- Encyclopedia of Life,
- Biodiversity Heritage Library
-



Brazil's Virtual Herbarium



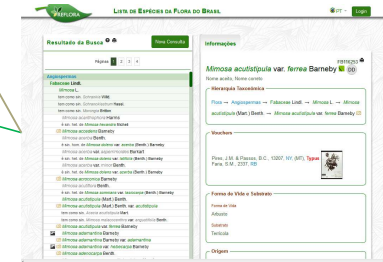
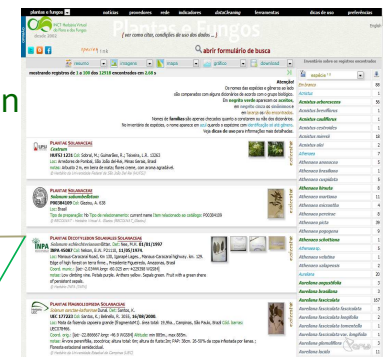
web interface
web services

web services

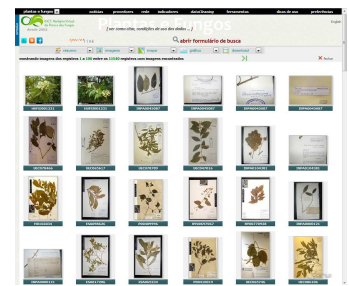
web service

2016 usage: 416 million records, 2.6 million images

190 data sets
5.4 million records
1.3 million images



Search interface



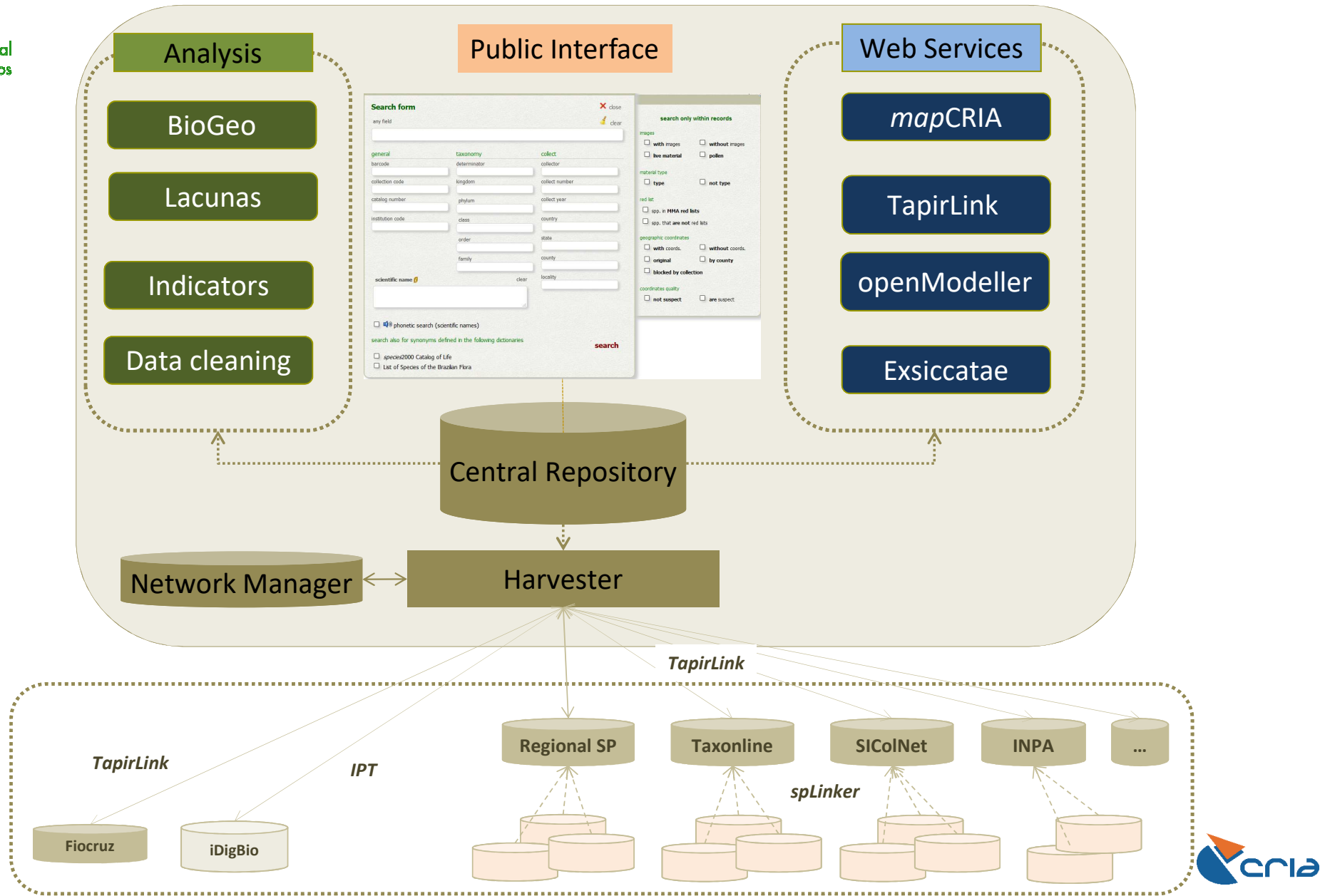
Biogeography: Workflow for Ecological Niche Modeling

4.629 models for 3.899 species (8.4%)

Flora 2020

Indicators, data cleaning, Lacunas, Usage ...

How data is accessed, modified, & contributed



how to cite

*Herbário da Universidade Federal de Minas Gerais (BHCB), Herbário Irina Delanova Gemtchújnicov (BOTU), Coleção de plantas medicinais e aromáticas (CPMA), Herbário da Escola Superior de Agricultura Luiz de Queiroz (ESA), Herbário Padre Balduino Rambo (HPBR), Herbário Rioclarense (HRCB), Herbario da Universidade Estadual de Feira de Santana (HUEFS), Missouri Botanical Garden - Brazilian records (MOBOT_BR), The New York Botanical Garden - Brazilian records (NY), Herbário de São José do Rio Preto (SJR), Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo" - Coleção de Fanerógamas (SP), Herbário Dom Bento José Pickel (SPSF), Herbário da Universidade de Brasília (UB), Herbário da Universidade Estadual de Campinas (UEC) available at **INCT - Herbário Virtual da Flora e dos Fungos** (<http://inct.splink.org.br>) in 2016, Nov 28.*

data use restrictions



The **INCT - Virtual Herbarium of Flora and Fungi** promotes free and open access to data, information, and tools available to any individual or group. All data providers have previously accepted to make their data available in order to promote scientific development and research all over the world, and to stimulate the public use of scientific information.

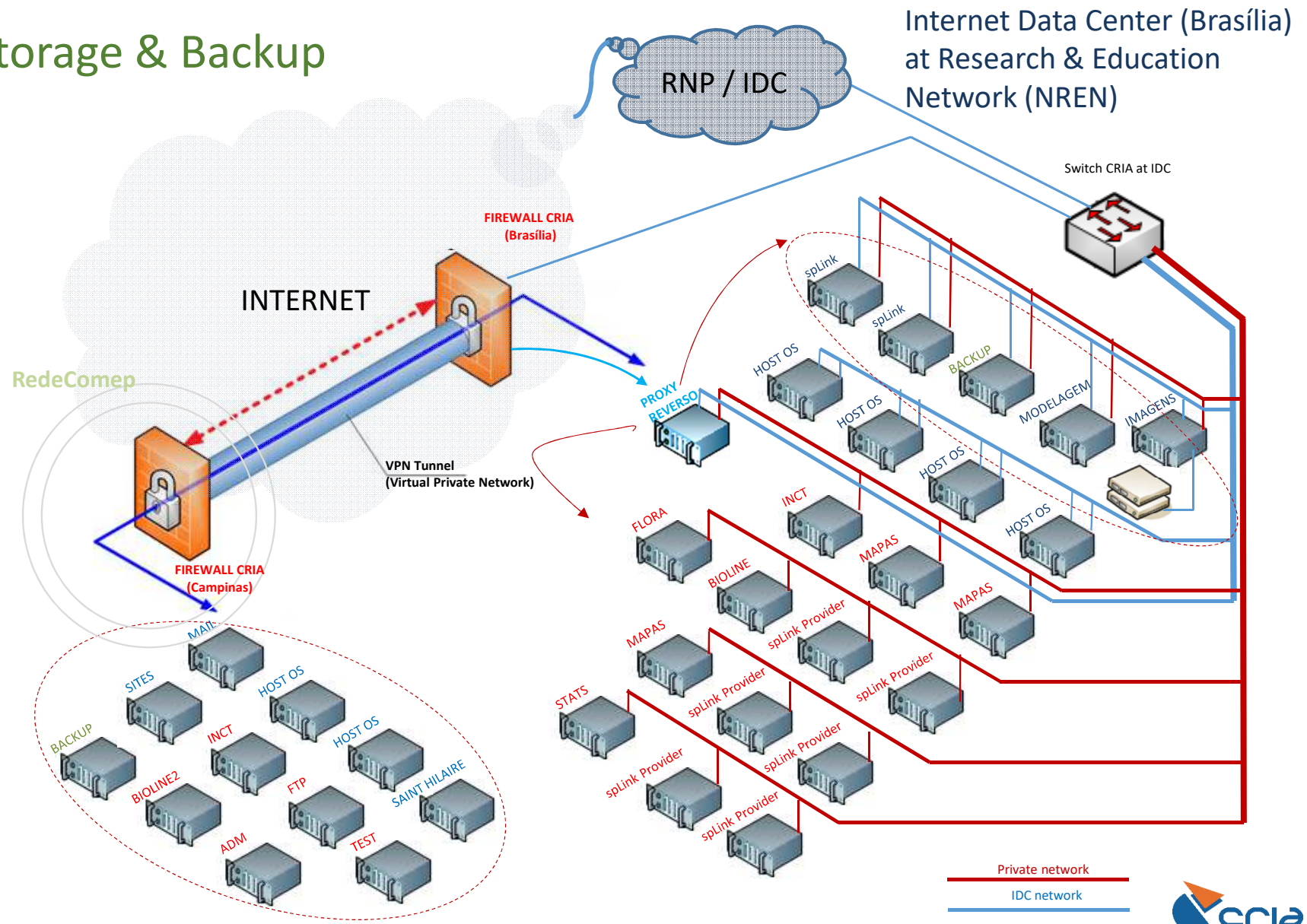
Storage & Backup

Storage

~ 22 TB space
~ 15 TB used

Backup:

- daily on disks
- weekly full backup
- monthly on tape
- Every 6 months external safe deposit



End-user license and Publicising

- All on-line data is shared through a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 (CC BY-NC-AS 3.0) license
 - Free to share and adapt
 - Appropriate credit must be given
 - You may not use the data for commercial purposes
- How to publicize our existence
 - Interacting with data providers and data users (feedback mechanisms)
 - Blog (<http://blog.cria.org.br>)
 - Participating in congresses, seminars, workshops, ...
 - Participating in advisory committees, working groups, ...

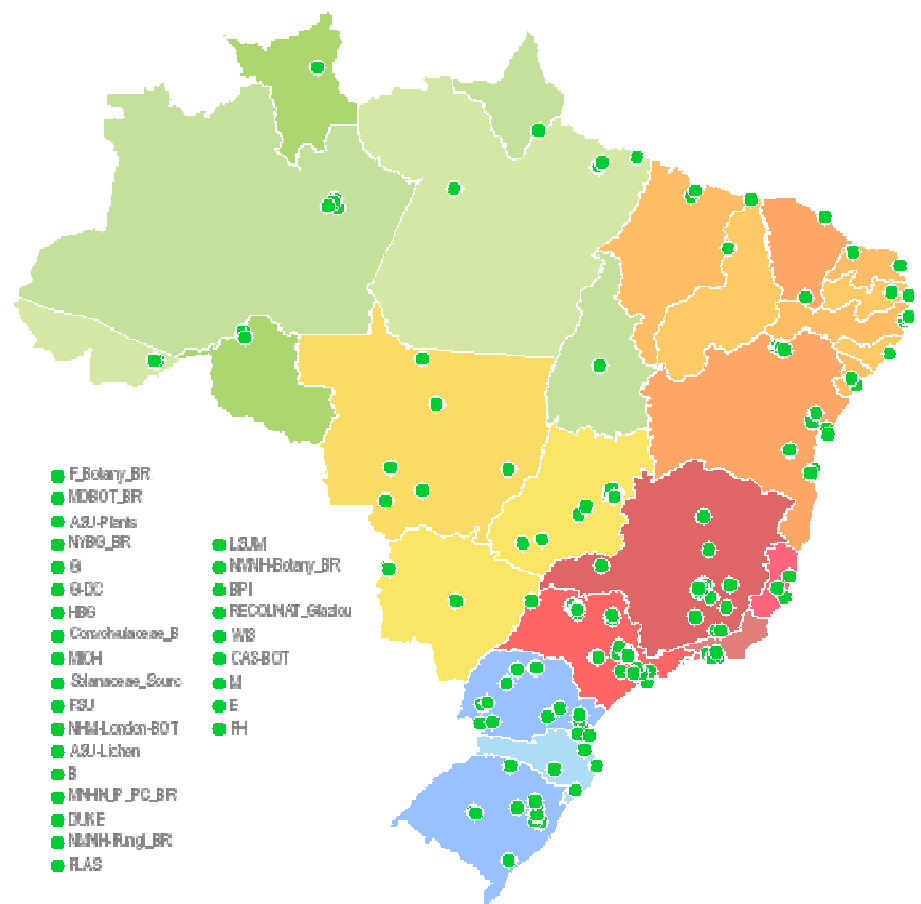
Responsibilities and Resources

- Responsibilities
 - Each data provider is responsible for its data
 - CRIA is responsible for data integrity
- Annual Budget (CRIA is understaffed)
 - ~ US\$ 400 K
- Ethics and Legal Compliance
 - Sensitive data: data providers withhold data they consider sensitive
 - All data sent to the network is openly and freely available on-line
 - A non-binding memorandum of understanding is signed between each data provider and CRIA

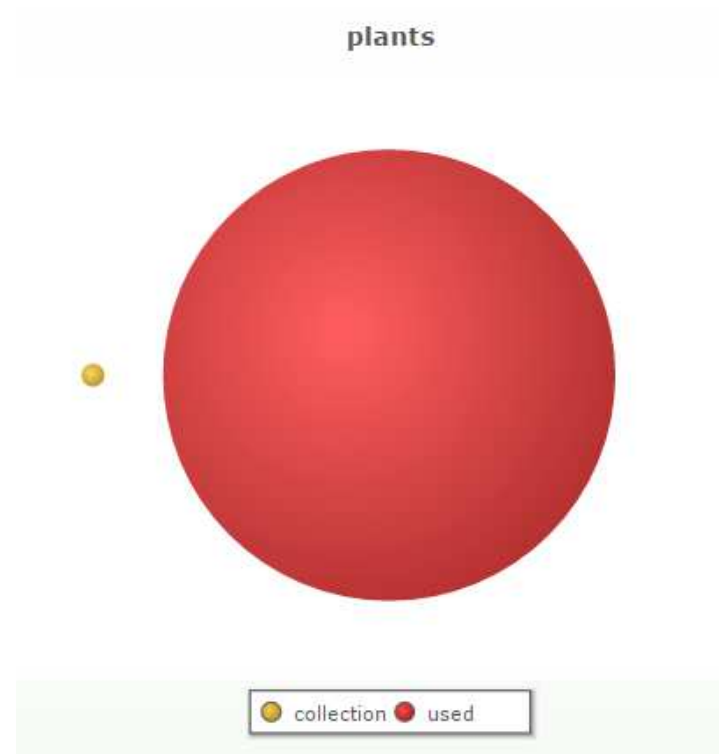
Thoughts about the project

- What was a challenge
 - The data management plan is aimed at a Project that must plan how its data will be organized. Brazil's Virtual Herbarium is an ongoing project, that uses an e-infrastructure that is under development since 2001.
- What went well
 - When I understood that our experience as an e-infrastructure could be of use to the IDRC Data Sharing Pilot, things began to move. So instead of trying to answer how my data is/will be organized, my answers were more about how does the way we work with data help or inhibit data sharing
- What have you learned through the process
 - The importance of a DMP
 - The necessity of e-infrastructures to receive “all” data (different quality levels) and the complexity involved in preserving and sharing this data in a useful and usable way to target users
 - The importance of sharing our experience as to how certain barriers were overcome

Thank you!



~ 120 national herbaria
 21 from abroad



81 times
<http://www.splink.org.br/showUsage>