

THE IMPACT OF THE BRAZILIAN VIRTUAL HERBARIUM IN E-SCIENCE

RESEARCH THEME: MOTIVATION AND OUTCOMES DUE TO OPEN AND COLLABORATIVE SCIENCE THROUGH E-INFRASTRUCTURES

The Brazilian Virtual Herbarium (BVH) is one of the country's National Institutes of Science and Technology, a program of the Ministry of Science, Technology, and Innovation. It integrates 160 datasets from more than 100 herbaria of Brazil and 11 from abroad. More than 4.8 million records associated to more than 685 thousand images of vouchers, live material and pollen are freely and openly available to all interested.

The project developed two important mechanisms to allow users to collaborate:

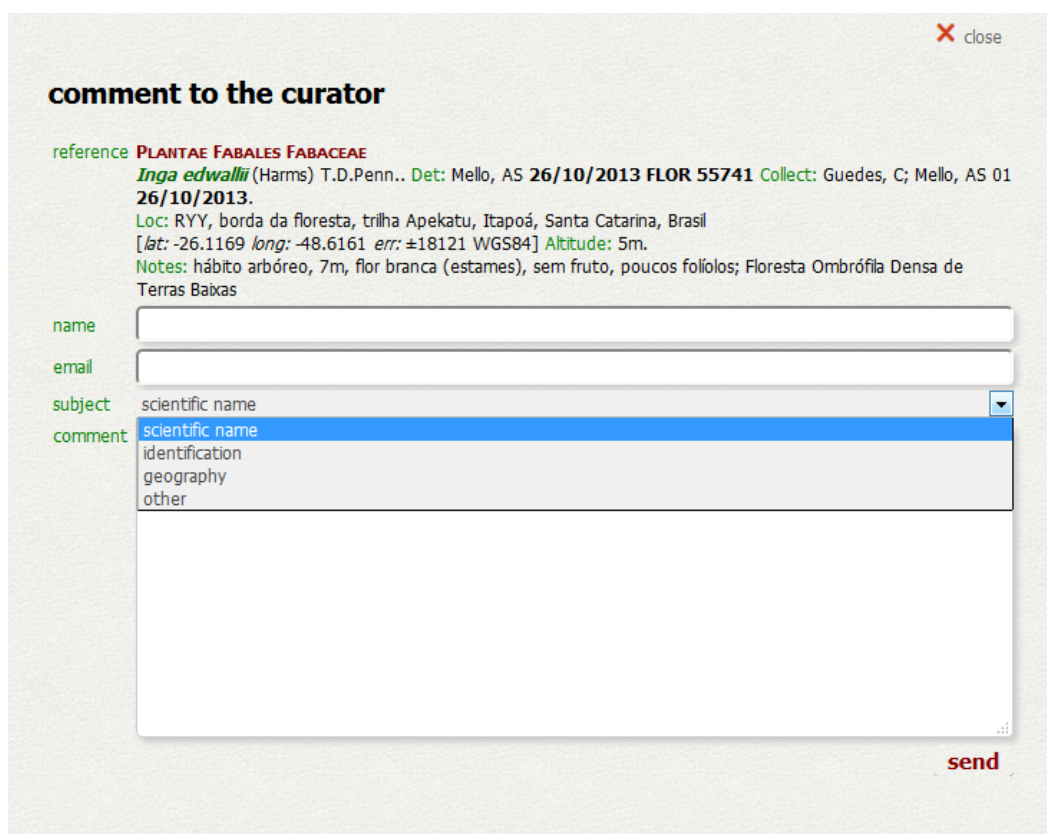
- An annotation system; and,
- A workflow (BioGeo) to enable users to produce and publish species ecological niche models.

Questions include:

- What motivates users to send their comments (*Annotation System*)?
- What motivates researchers to produce and publish their models through *BioGeo*?
- Are there mechanisms that could be used to increase participation?

Annotation System

Within the Brazilian Virtual Herbarium project, feedback mechanisms were developed to allow users to send their comments about specific data records. When clicking on the “new comment” icon, a window pops up for users to provide their input (figure 1).



comment to the curator

reference **PLANTAE FABALES FABACEAE**
Inga edwallii (Harms) T.D.Penn.. Det: Mello, AS 26/10/2013 FLOR 55741 Collect: Guedes, C; Mello, AS 01 26/10/2013.
 Loc: RYY, borda da floresta, trilha Apekatu, Itapoá, Santa Catarina, Brasil
 [lat: -26.1169 long: -48.6161 err: ±18121 WGS84] Altitude: 5m.
 Notes: hábito arbóreo, 7m, flor branca (estames), sem fruto, poucos folíolos; Floresta Ombrófila Densa de Terras Baixas

name

email

subject

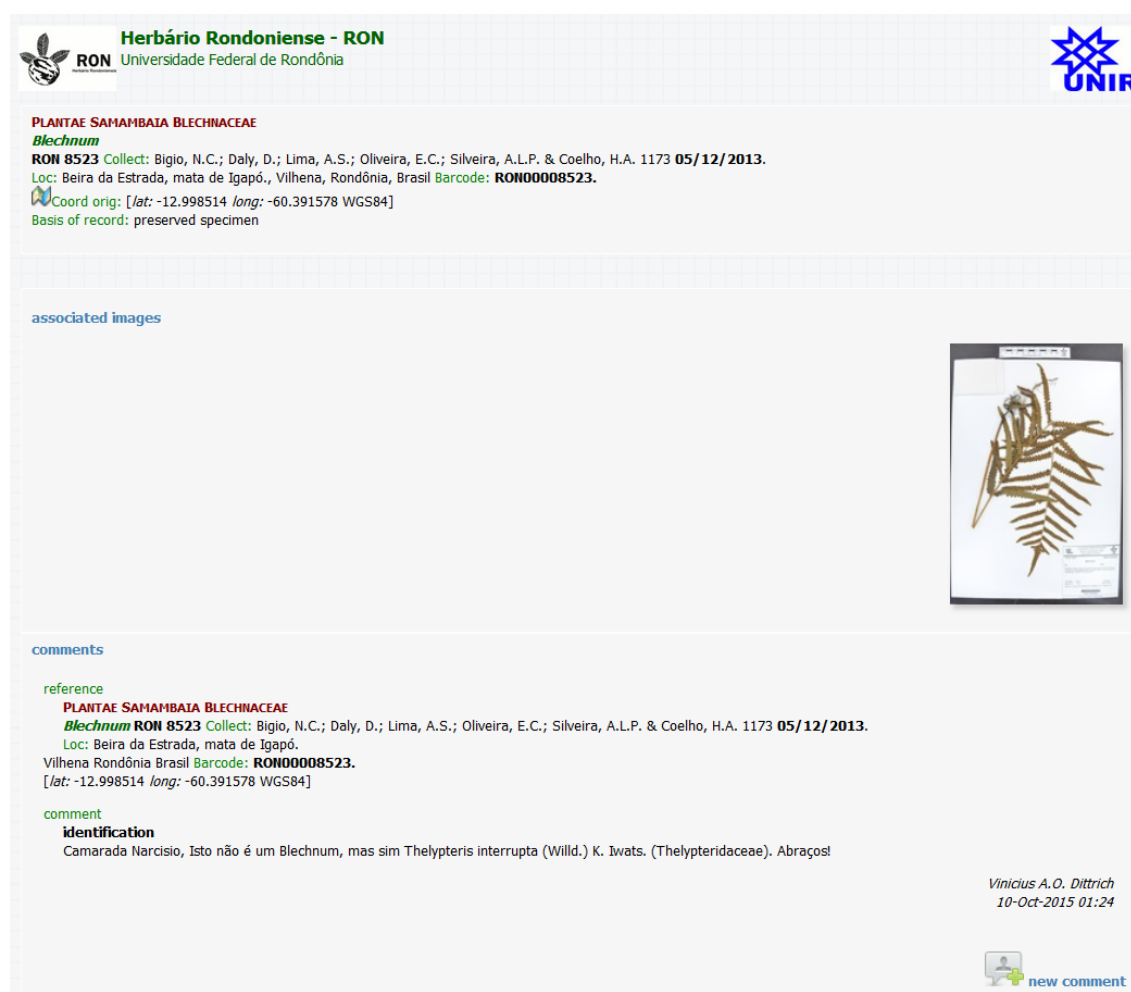
comment
 identification
 geography
 other

send

Figure 1. Popup window to enable users to send comments to curators

As can be seen on figure 1, the system presents the full data record and users must fill out the form with their name and email, select the subject – scientific name, identification, geography, and other – and write their comment. When clicking on **send** the comment is sent to the email indicated by the user for

confirmation. Once confirmed, the curator receives the email that is also archived in a database. Figure 2 shows a record with an associated comment.



Herbário Rondoniense - RON
Universidade Federal de Rondônia

PLANTAE SAMAMBAIA BLECHNACEAE
Blechnum

RON 8523 Collect: Bigio, N.C.; Daly, D.; Lima, A.S.; Oliveira, E.C.; Silveira, A.L.P. & Coelho, H.A. 1173 **05/12/2013**.
Loc: Beira da Estrada, mata de Igapó., Vilhena, Rondônia, Brasil Barcode: **RON00008523**.
Coord orig: [lat: -12.998514 long: -60.391578 WGS84]
Basis of record: preserved specimen

associated images

comments

reference
PLANTAE SAMAMBAIA BLECHNACEAE
Blechnum **RON 8523** Collect: Bigio, N.C.; Daly, D.; Lima, A.S.; Oliveira, E.C.; Silveira, A.L.P. & Coelho, H.A. 1173 **05/12/2013**.
Loc: Beira da Estrada, mata de Igapó.
Vilhena Rondônia Brasil Barcode: **RON00008523**.
[lat: -12.998514 long: -60.391578 WGS84]

comment
identification
Camarada Narcisio, Isto não é um *Blechnum*, mas sim *Thelypteris interrupta* (Willd.) K. Iwats. (Thelypteridaceae). Abraços!

Vinicius A.O. Dittrich
10-Oct-2015 01:24

new comment

Figure 2. Record RON 8523 with an associated comment.

Figure 2 shows that the herbarium RON has voucher 8523 identified as *Blechnum* and the comment indicates that it is *Thelypteris interrupta* (Willd.) K. Iwats. (Thelypteridaceae). This comment was sent on October 10, 2015, but the last time this database was updated was in June 2015. Even though the record was not altered by the curator, it appears as an on-line annotation associated to the specific record. Users can therefore benefit from a specialist comment even before the data has been revised and altered. Users can also check this information as there is an image associated to the record.

Methodology

When this survey was prepared, the system had received 622 comments from 141 people. 473 comments referred to the scientific name, 68 to the identification of the material, 59 to the geographic data, and 22 classified as "other". 85.5% of the comments referred to data records of plants, 13.5% to animals and 1% to microorganisms.

An email was posted to all 141 users who sent their comments through the annotation system asking what was the motivation for using this tool. In order to facilitate the analysis, six options were offered:

- (1) Contribute to the improvement of the quality of the data;
- (2) Correct errors in order to enable the use of the data in their research;
- (3) Correct errors in order to use the data in the BioGeo workflow;

- (4) Check the determination and/or geographic information to use this information in the List of Species of the Brazilian Flora;
- (5) Check the determination and/or geographic information to use this information in the red list assessment (CNCFlora);
- (6) Others. In this option, users were asked to specify what other reasons they had.

People could choose more than one option. We also asked whether the herbarium accepted their comments and corrected possible errors, asking them to choose one of the four options below:

- All records were corrected
- Most records were corrected
- Some records were corrected
- No record was corrected

Results

Of the 141 emails sent, we received 20 answers, around 14% of the total.

- 85% indicated that their motivation was to contribute to the improvement of the quality of the data
- 50% to correct errors in order to enable the use of the data in their research
- 5% to correct errors in order to use the data in the BioGeo workflow
- 5% to check the determination and/or geographic information to evaluate the species' distribution and include this information in the List of Species of the Brazilian Flora

No one indicated the use of the tool to use the data in the red list assessment and no other motivation was included.

As to whether, to their knowledge, the collections are benefiting from their comments to correct possible errors, only 16 of the 20 specialists answered this question.

- 15% indicated that all records were corrected
- 15% indicated that most records were corrected
- 15% indicated that some records were corrected
- 15% indicated that the records were not corrected
- 20% indicated that they do not know whether the data was corrected

Comments

The most important driver for participation is to contribute to improve data quality and to enable the use of the data in research. It is probable that the 20% that did not answer the second block probably did not check to see whether the records were modified. However, we can conclude that 60% did not only contribute with their comments but also checked to see if the data was changed.

BioGeo (Biogeography of the Flora and Fungi of Brazil)¹

BioGeo is a system developed to expand the knowledge about biogeography of plants and fungi of Brazil, using modeling techniques of potential distribution and counting with an active participation of specialists. A diagram of the workflow is presented in figure 3.

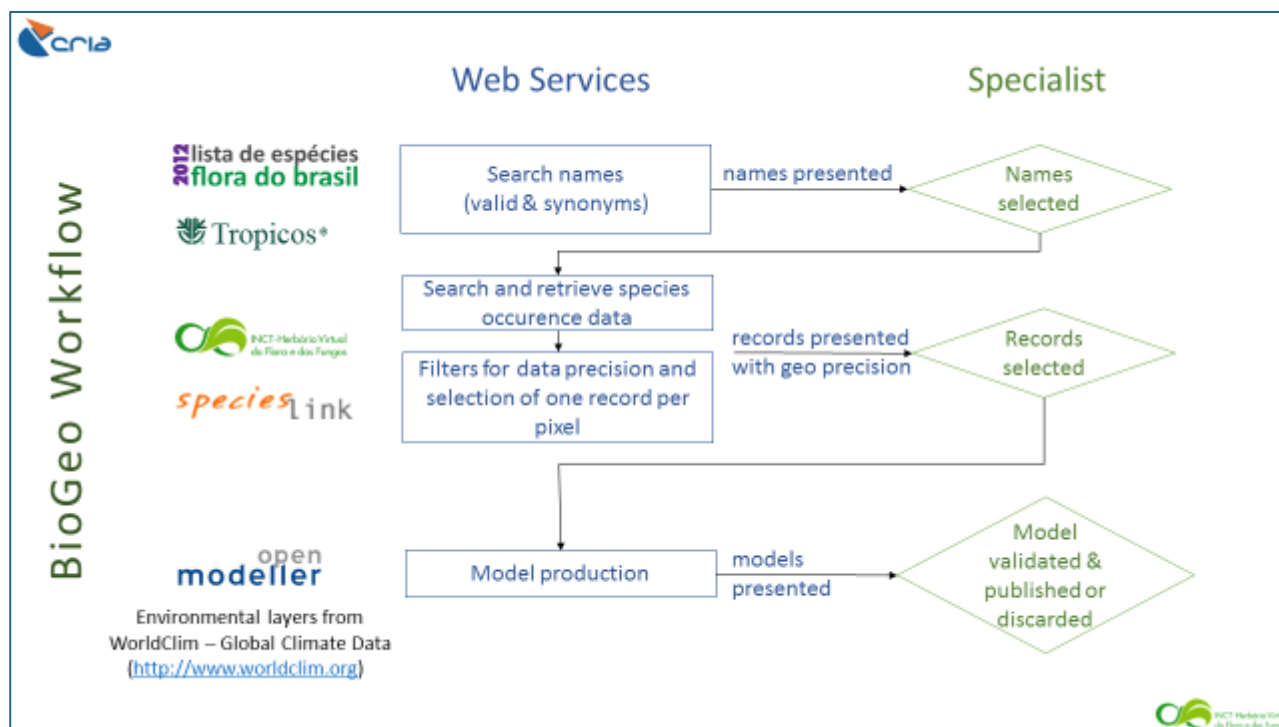


Figure 3. BioGeo Workflow

When a specialist registers in the system, he/she indicates the species or genera that he/she wants to model. The process begins when a specific species is selected. Through the workflow, the system using web services for the List of Species of the Brazilian Flora (provided by CRIA) and Tropicos (provided by Missouri Botanical Garden) presents a list of names (valid names and synonyms) to the specialists who selects those to be included for searching. The name as searched through the speciesLink web services and the results go through a filter for data precision (lat/long) that selects one record per pixel. Records selected by the system and all other records found are presented to the specialist who defines which records will be used in modeling. Depending on the number of point data, 1 to 5 different algorithms are used and models are produced using the openModeller web services and WorldClim data. The resulting models, together with a consensus model, are presented to the specialist who then decides whether it should be published or discarded (Figure 4).

¹ See <http://biogeo.inct.florabrasil.net> (in Portuguese only)

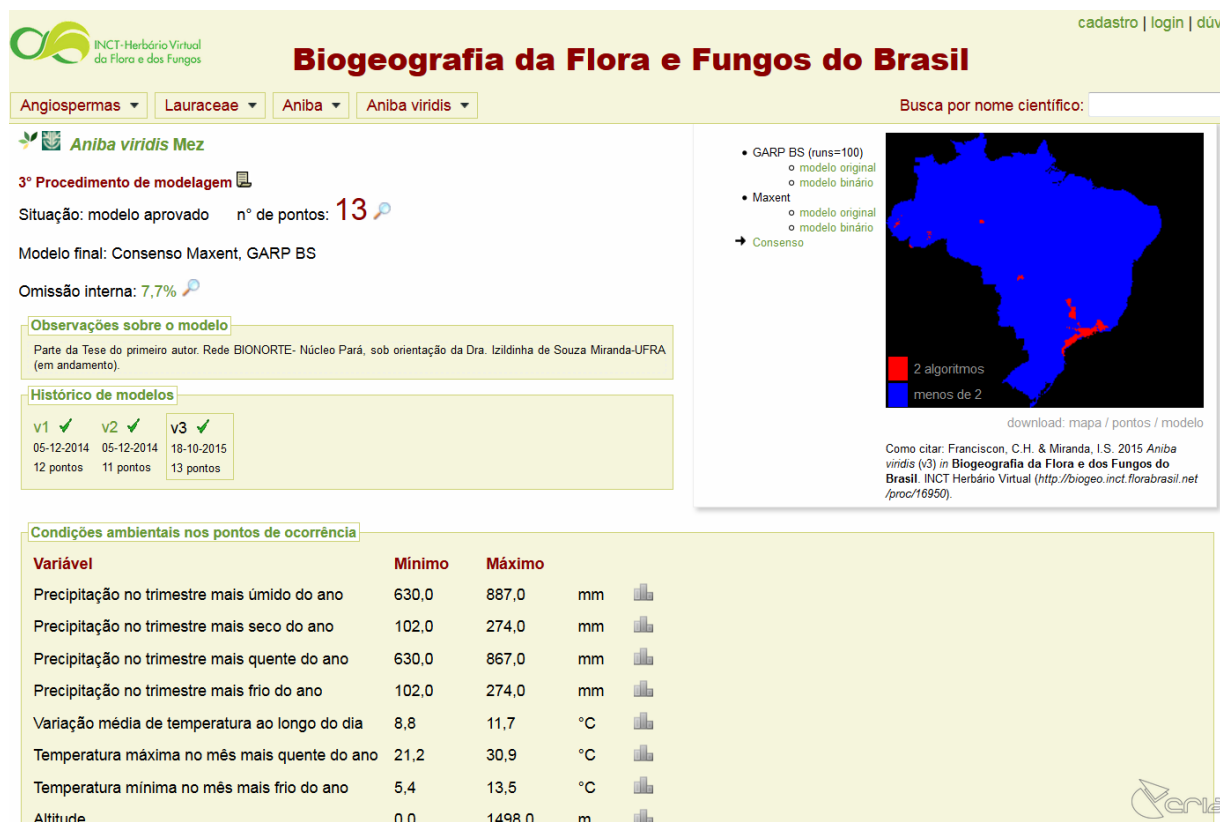


Figure 4. Example of a distribution model published on-line

Methodology

When the survey was sent out, there were 177 specialists registered in the system, meaning all were authorized to produce distribution models. Nevertheless, only 43 had actually published models on-line.

An email was sent to all 177 specialists who were asked to fill out the following information:

- (1) Institution
- (2) Academic level
- (3) Field of interest
- (4) If registered and did not publish any model, what was the impediment
- (5) Use of the model published through BioGeo
 - a. Planning new collecting efforts
 - b. Public policies
 - c. Articles
 - d. Others (please specify)
- (6) Weaknesses of BioGeo
- (7) Strengths of BioGeo
- (8) What would you like to see in BioGeo in the future

Results

We received 17 answers from 10 PhDs, 4 students doing their PhD, 2 Masters and 1 Bachelor (biology).

Nine (53%) did not publish their models for diverse reasons:

- They are still compiling the data
- They just carried out some tests

- They do not like the idea of having another specialist authorizing their participation – NOTE: this is not correct and an email was sent to this specialist to clarify this point.
- Problems with understanding the system
- The resulting model did not answer the hypothesis raised about the species
- Lack of time
- I am not a specialist

As to the use of the models:

- Two (12%) have not used any model available
- Four (24%) have used the models to plan new collecting efforts
- Three (18%) for public policies
- Seven (41%) to publish articles
- Two (12%) indicated using the models for their research (this was under “others”)

Weaknesses

- The system does not include data from other South American countries (restricted to Brazil)
- The fact that non specialists can generate models
- Many specialists in biogeography are not participating. Suggests a better communication strategy to make BioGeo known.
- Data of the models can only be exported in xml. It would be interesting to have other formats such as ascgrid and float.
- The models use a fixed set of environmental layers. It is not possible to select layers, algorithms, and other parameters.
- Not intuitive
- Insufficient data, this is not a problem of the system but it affects the quality of the models generated
- Not being able to project future scenarios with climate change

Strengths

- Easy to use (5 – 29%)
- Available distribution models
- The system that generates the maps is very good
- The system is fast and the graphic output of the models is good
- Extremely easy to use, principally for decision making, policy or research
- Many species have models
- Data sharing without restrictions
- Standardization, enabling the comparison of models
- Treatment of the occurrence points to generate the models
- Reduction of processing capacity of desktops to generate models
- Utility in planning new collecting efforts
- Visualizing the models, data used, liberty to select the data (validate or not and justify)
- Facility to manipulate and interpret
- Potential for diverse uses, both academic and for public policies

Future Requirements

- Inclusion of data from other South American countries
- Historical data about areas where specimens were collected
- Specialists that participate in the Flora of Brazil initiative should be invited to participate

- A link to data use, restrictions, models and results to give proper credits and stimulate new contributions
- A feedback mechanism for public policies
- Distance training in GEOstatistics
- Possibility of selecting geographic areas (such as states, regions, phytogeographic limits, among others) and bioclimatic layers.
- Where specialists are listed, include the species and families they are modeling
- Enable projection for scenarios of climate change

Comments

Although the number of answers was small (less than 10%) there are important contributions that can help guide future activities.

The number of people that registered compared to the number of specialists that are actually producing models indicate that there may have been a misinterpretation as to the usage of the system. It seems that people may have thought it necessary to register in order to access the models. This seems plausible when some of the answers received by those that have not published any model includes “I am not a specialist”.

One of the motivation in participating is obviously to use the model that was produced. Our main focus as to data users is the scientific community. Therefore, it is interesting to recognize the usage of the model to plan new collecting efforts, for research, to publish articles, and for public policies. These were all aims of this tool.

As to weaknesses, as the focus of the system was to help plan new surveys in Brazil, the geographic scope was Brazil and with current environmental conditions. It would be interesting to further develop the system increasing the geographic scope to South America – more data and possibly a better resulting model – and to build scenarios considering climate change.

A weakness mentioned referred to the necessity of a better communication strategy, as important specialists are not involved. This is true. The system was not publicized as it was under development and now, as the project ended, it is just being maintained. If we are able to obtain new grants, a communication strategy must be in place.

As to strengths, although one user said the system is not intuitive, five said it is easy to use. Given the fact that the system shares distribution models for 3.562 species without having provided any training courses, one can assume that it is intuitive for a knowledgeable person. Table 1 shows the number of species with models per taxonomic group.

Table 1. Species distribution models produced through BioGeo

| Taxonomic Group | No. of Species in the List of Brazil | Species with Distribution Models | % | No. of Approved Models | No. of discarded models | Models awaiting approval |
|----------------------|--------------------------------------|----------------------------------|-------|------------------------|-------------------------|--------------------------|
| Algae | 4.747 | | 0 | | | |
| Angiosperms | 32.831 | 3,471 | 10.6% | 4,046 | 126 | 147 |
| Bryophytes | 1.524 | 5 | 0.3% | 5 | 3 | 12 |
| Fungi | 5.712 | 10 | 0.2% | 10 | 6 | 1 |
| Gymnosperms | 30 | 4 | 13.3% | 4 | | |
| Ferns and Lycophytes | 1.253 | 59 | 4.7% | 68 | 5 | 7 |
| Total | 46.097 | 3,549 | 7.7% | 4,133 | 140 | 167 |

Despite the lack of a more substantial dissemination of BioGeo, almost 8% of all fungi and plant species that occur in Brazil have a distribution model published and openly available on BioGeo.

It is interesting to see that one user expressed as strength of the system the fact that data is shared without restrictions.

The answer as to future requirements certainly will help us when writing a new proposal for funds to enable the continuity of this initiative.