# GBIF survey 2015: GBIF Data Fitness For Use in Modelling Ecological Niches and Geographic Distributions

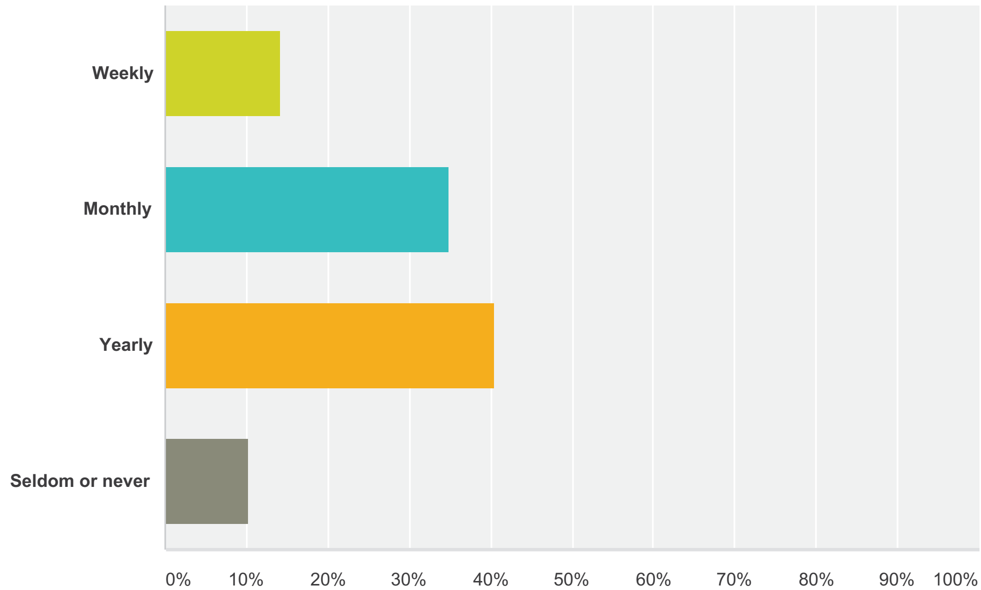
Ensuring availability of relevant data is essential for distribution modelling, a scientific approach increasingly supporting decision-making.

The task group on fitness of GBIF-mediated data for use in distribution modelling invited researchers and users of biodiversity data to provide their opinions through the following survey.

The survey aimed to capture the best available experiences on biodiversity data and particularly GBIF-mediated data, to document limitations in existing GBIF services and in accessing other important sources of data, and to suggest improvements in the functionality of GBIF.org for the needs of distribution modelling community.

## Q4 How often do you, personally or with immediate colleagues, use GBIF.org to access data, on average?

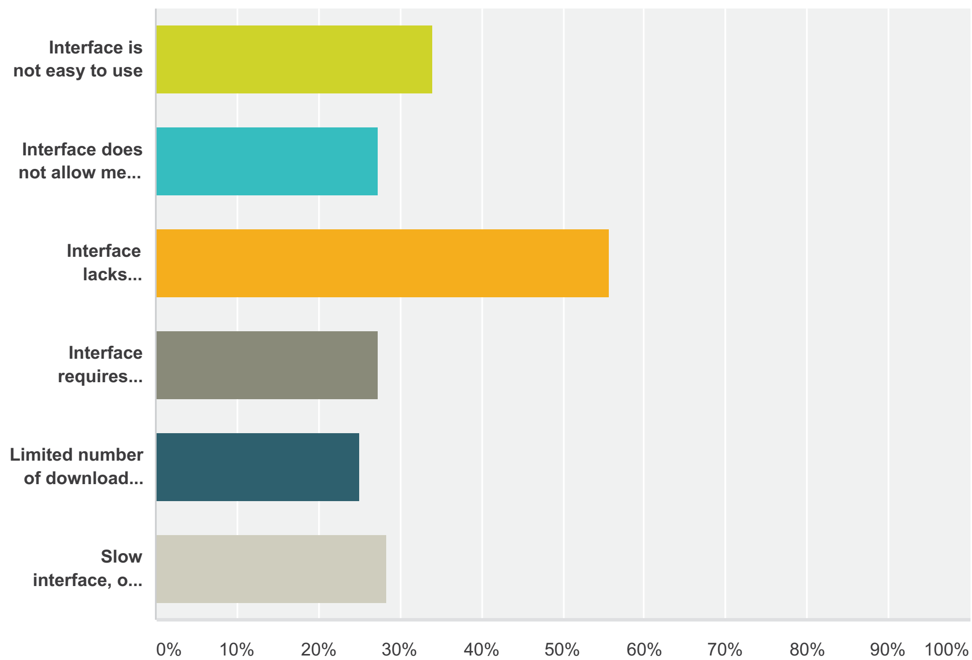
Answered: 126 Skipped: 9




| Answer Choices | Responses | |
|---|---|---|
| Weekly | **14.29%** | 18 |
| Monthly | **34.92%** | 44 |
| Yearly | **40.48%** | 51 |
| Seldom or never | **10.32%** | 13 |
| **Total** | | **126** |

## Q5 Please select which of the following you consider the three main problems with the GBIF.org interface (as opposed to the data)?

Answered: 88 Skipped: 47



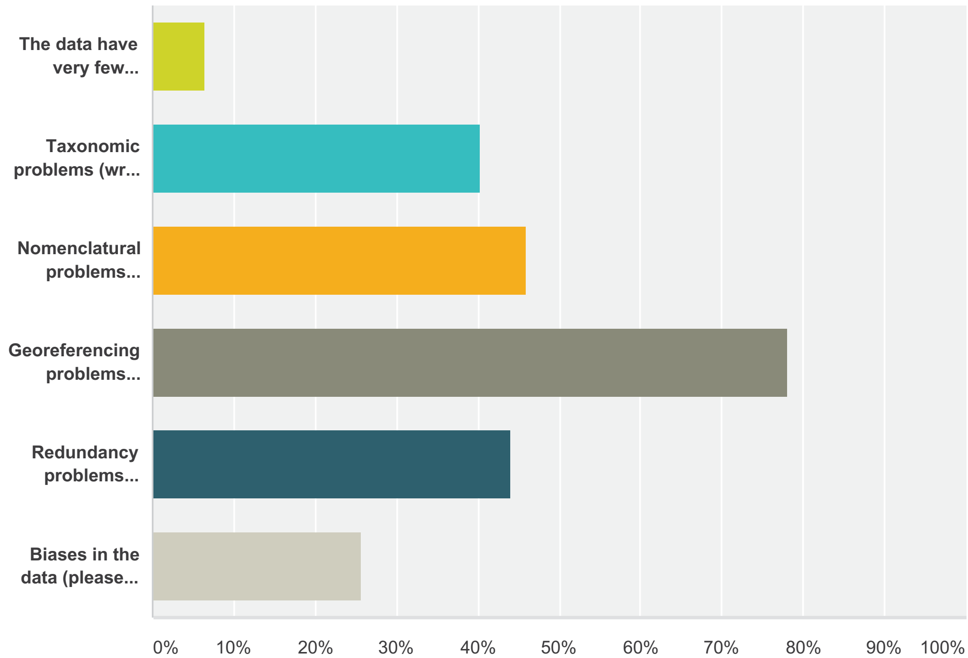

| Answer Choices | Responses | |
|---|---|---|
| Interface is not easy to use | **34.09%** | 30 |
| Interface does not allow me to download all the fields I want | **27.27%** | 24 |
| Interface lacks easy-to-use exploratory tools | **55.68%** | 49 |
| Interface requires registering | **27.27%** | 24 |
| Limited number of download formats | **25.00%** | 22 |
| Slow interface, or low throughput interface | **28.41%** | 25 |
| **Total Respondents: 88** | | |

| # | Other (or comments) | Date |
|---|---|---|
| 1 | I think the new GBIF is very poorly designed, and the "old" GBIF (data.gbif.org) was preferable, though still flawed. | 10/13/2015 5:11 PM |
| 2 | No friendly output of filtered results of a search query, ir order to allow further analysis | 9/16/2015 12:14 AM |
| 3 | batch download based on species list is missing | 9/15/2015 7:15 AM |
| 4 | Often a confusing and clumsy way of feature and data presentation; we want get at the data the most; not more not less. | 9/9/2015 1:13 AM |
| 5 | Poor explanation of fields and (even more importantly) of metadata | 9/8/2015 8:08 PM |
| 6 | Even in simple download formats such as .csv the data is not arranged clearly which makes sorting out difficult | 9/8/2015 10:23 AM |
| 7 | download CSV file was not formatted as CSV | 9/7/2015 1:11 AM |
| 8 | I usually use the API but not the website | 9/4/2015 11:42 PM |
| 9 | JSON output is not the easiest way to handle data; a SQL object would be better | 9/4/2015 9:09 PM |

| | | |
|---|---|---|
| 10 | I have no problems with the intrerface. | 9/4/2015 6:26 PM |
| 11 | interface does not allow downloads for multiple species at the same time | 9/4/2015 5:56 PM |
| 12 | Limted number of records to download at the time | 9/4/2015 5:31 PM |
| 13 | Limited amount of records to download directly from the GBIF.org interface | 9/4/2015 5:09 PM |
| 14 | Data quality hard to verify | 9/4/2015 4:33 PM |
| 15 | Prefer API access via for instance spocc --> https://cran.r-project.org/web/packages/spocc/index.html | 9/4/2015 1:44 PM |
| 16 | Can struggle to find how to download the data, not just view it and it does not tell you that you must be logged in to download data. | 9/4/2015 12:17 PM |
| 17 | I find the current format of defaulting to all species first is a little annoying, preferred the older interface even though was less stylish! | 9/4/2015 12:00 PM |
| 18 | I could be wrong, but I don't believe there is a (easy) way to download data given a list of species. | 8/5/2015 7:20 PM |
| 19 | typically is my students or staff who interact with the site; I work with them to identify questions and methods that use/exploit GBIF | 7/29/2015 6:05 PM |
| 20 | Limitation in the number of records you can download at once. Lost of cellid field in rest services in recent update. | 7/13/2015 4:18 PM |
| 21 | Interface seems to lack visualization tools for record sets | 7/13/2015 2:53 PM |
| 22 | I think the interface is easy! But the download format for CSV has changed. This is confusing. It says it is CSV but tt is really isnt | 7/12/2015 6:17 AM |
| 23 | I actually think it is pretty good! Sure, I'd like it to be quicker... but not a big deal. | 7/9/2015 11:53 PM |
| 24 | Interface does not allow me to choice which fields to download | 7/8/2015 4:46 PM |
| 25 | the new interface is good | 7/8/2015 10:59 AM |
| 26 | I use the api (via R) to download data | 7/7/2015 10:41 PM |

## Q6 Please select which of the following you consider the three main problems the data itself have, from a practical point of view?

Answered: 109 Skipped: 26



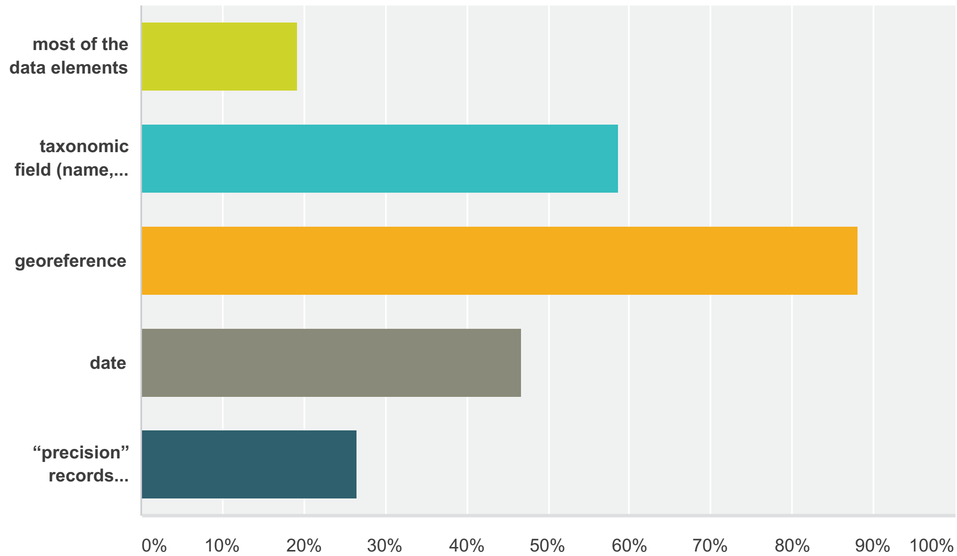

| Answer Choices | Responses | |
|---|---|---|
| The data have very few problems | **6.42%** | 7 |
| Taxonomic problems (wrong identification) | **40.37%** | 44 |
| Nomenclatural problems (synonyms, obsolete taxonomy, associations to multiple higher categories…) | **45.87%** | 50 |
| Georeferencing problems (faulty/imprecise/coarse/non-existent...) | **77.98%** | 85 |
| Redundancy problems (multiple records with same basic info/record batches…) | **44.04%** | 48 |
| Biases in the data (please specify below) | **25.69%** | 28 |
| **Total Respondents: 109** | | |

| # | Other (or comments) | Date |
|---|---|---|
| 1 | Fisheries related data sometimes reported on land | 10/22/2015 12:38 PM |
| 2 | No quality control, e.g. Records/specimens from zoos lead to US records for tigers | 10/19/2015 3:40 PM |
| 3 | It was difficult to prioritise these. The biases in the data seem to be one of the biggest challenges I face now. Data from many ecosystem monitoring programs outside e.g. the EU or North America are not integrated into GBIF. | 10/18/2015 9:52 AM |
| 4 | Incomplete sampling or incomplete consolidation of existing data | 10/13/2015 5:11 PM |
| 5 | risk that people confuse observation and specimen data | 10/2/2015 3:17 AM |
| 6 | No way to fix mistakes so things I improve in my download never get into the main GBIF data. Also the data are heavily weighted toward birds. | 9/23/2015 11:28 PM |
| 7 | Spatial coverage gaps, missing information in some fields, | 9/16/2015 12:14 AM |
| 8 | Data usually suffres from strong geographical biases with large gaps within a species range | 9/15/2015 4:00 PM |
| 9 | geographic gaps, taxonomic incompleteness | 9/11/2015 8:42 PM |

| | | |
|---|---|---|
| 10 | Depends on the application. One can say the data have 60% error, but it's NOT a GBIF problem and never should be. Provider's fault, and state of western science and how interpreted and used ! | 9/9/2015 1:13 AM |
| 11 | Biases in all possible dimensions | 9/8/2015 8:08 PM |
| 12 | Epistemic problems. Do the data represent established populations or transient. | 9/6/2015 2:03 AM |
| 13 | problem when uploading the data from source. For ex. I couldnt find my own collections until I noticed they were labeled as collected by the data manager. | 9/4/2015 9:15 PM |
| 14 | spec.s ex ornamental plantings | 9/4/2015 8:31 PM |
| 15 | Data is biased to only few collections, which corresponds with lack of data for many World regions. | 9/4/2015 7:46 PM |
| 16 | Outdated data (not completely synced with the original datasets) | 9/4/2015 5:09 PM |
| 17 | resolution/accuracy of the data is not always shown | 9/4/2015 4:50 PM |
| 18 | well known and described geographic biases | 9/4/2015 2:47 PM |
| 19 | Temporal biases! Distinction between real survey occurence data and "atlas" or mapping results | 9/4/2015 1:44 PM |
| 20 | sampling bias is obviously a big problem with all of these data but an experienced researcher should know to carefully examine the data before use | 9/4/2015 12:00 PM |
| 21 | The data often has no metadata regarding the date or publisher. The data is spatially biased (e.g. better monitoring programs in some countries) | 9/4/2015 11:00 AM |
| 22 | Biases in taxonomic and geographic sampling | 8/13/2015 12:57 PM |
| 23 | Very little data from ecosytem monitoring programs make it into GBIF. If these data were more typically included, they could contribute to understanding e.g. biotic interactions. Data from extreme envinroments (e.g. the Arctic at a circumpolar scale-- see CAFF's Circumpolar Biodiversity Monitoring Program, or the SCANNET program ) from the edges of species' ranges are often underrepresented or missing in GBIF, though the data may be available (completely or in part) from other publicly available ecosystem monitoring data portals, or directly from institutions that collect and manage the data. Such data would be wonderful to include as they would help refine the shape of species' response curves for improved prediction, etc. | 7/29/2015 6:55 PM |
| 24 | usually a wish for more points or more complete coverage of a species' range | 7/29/2015 6:05 PM |
| 25 | For plants with several duplicate specimens, different determinations can be often found | 7/29/2015 5:09 PM |
| 26 | The spatial and temporal coverages are patchy. The data has been mobilsed opportunisticly, and not in a planned process. | 7/27/2015 1:31 PM |
| 27 | I think the nomenclature problems are well addressed in gbif. The biggest "problem" is that more and more data are being included that are from citizen scientists and may be misidentified or mislocated. It is more onus on the analyst to screen the data. On the other hand if you leave these adta out you may have a very biased sample based only on vouchered specimens | 7/12/2015 6:17 AM |
| 28 | Spatial biases are the biggest problem, and related to this is simply the incompleteness of the data (for many species we may know there are more data but they are not in the system...).) | 7/9/2015 11:53 PM |
| 29 | we have no idea about the completeness of the data and how the data is representative of species' distributions. The lack of precision (accuracy) associated with coordinates for most records is a the second main problem. | 7/8/2015 10:59 AM |
| 30 | Biases stem from the very nature of the aggregation of data. Ensuring data usability for ecological or modelling purposes may require prior modelling of the representativity of the data. Analysis of distribution patterns in the data itself may be key to understand, and eventually compensate, such representativity biases. | 7/8/2015 1:02 AM |

## Q7 Which data elements available through GBIF.org do you use most? Please choose one or more answers.

Answered: 109 Skipped: 26

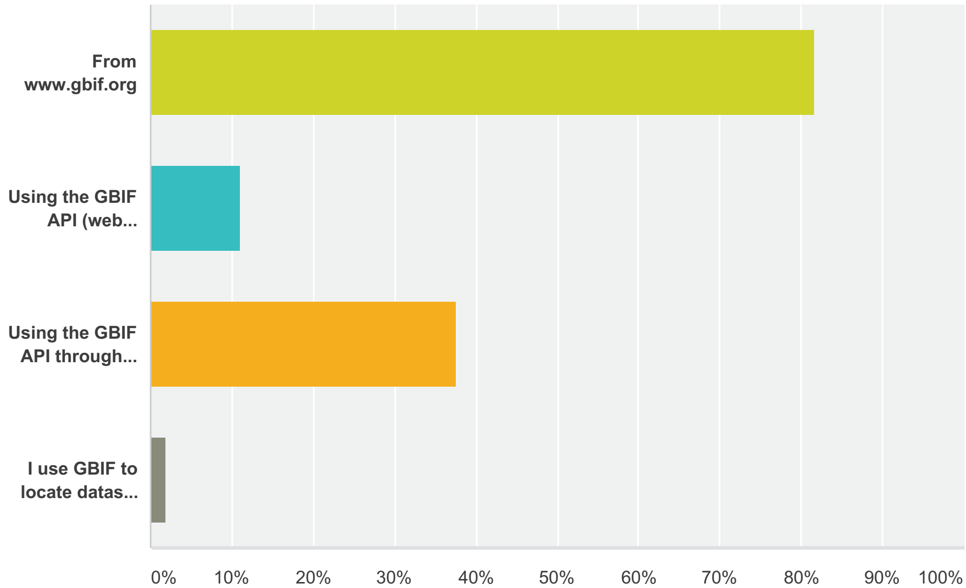



| Answer Choices | Responses | |
|---|---|---|
| most of the data elements | **19.27%** | 21 |
| taxonomic field (name, higher categories) name | **58.72%** | 64 |
| georeference | **88.07%** | 96 |
| date | **46.79%** | 51 |
| “precision” records describing data or name reliability | **26.61%** | 29 |
| **Total Respondents: 109** | | |

| # | Other (or comments) | Date |
|---|---|---|
| 1 | provider | 9/9/2015 1:13 AM |
| 2 | identifier and location | 9/4/2015 9:15 PM |
| 3 | Ecological data like plant host, vegetation, etc. | 9/4/2015 8:13 PM |
| 4 | description data for removing planted specimens | 9/4/2015 12:00 PM |
| 5 | the more info the better! | 7/9/2015 11:53 PM |
| 6 | I would use taxonomy and date if the data was complete and/or reliable | 7/8/2015 10:59 AM |

## Q8 How do you access the records?

Answered: 109 Skipped: 26

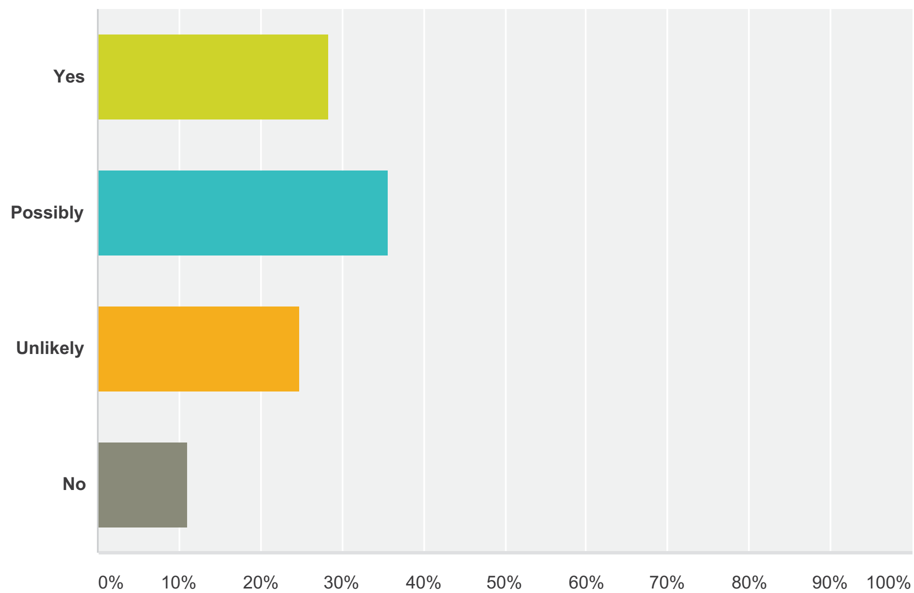



| Answer Choices | Responses | |
|---|---|---|
| From www.gbif.org | **81.65%** | 89 |
| Using the GBIF API (web services) | **11.01%** | 12 |
| Using the GBIF API through some 3rd party library (e.g. the R client) | **37.61%** | 41 |
| I use GBIF to locate datasets of interest and then approach the data provider directly | **1.83%** | 2 |
| **Total Respondents: 109** | | |

| # | Other (or comments) | Date |
|---|---|---|
| 1 | Fishbase | 10/22/2015 12:38 PM |
| 2 | rgbif | 9/4/2015 9:15 PM |
| 3 | I have experienced that for example that not all of the ICES.org dataset, which is an important source of marine data, is included in the GBIF dataset | 7/29/2015 6:55 PM |
| 4 | rgibif | 7/19/2015 10:17 PM |

## Q9 GBIF currently offers record download which can result in large volumes of data. GBIF could offer services that help reduce this volume by projecting records onto a grid size of your choosing, and then return metrics per grid, or a few representative records for each grid cell, which would result in more easily manageable data volumes but at the cost of determining the provenance of the data. Is this something that you would use if implemented?

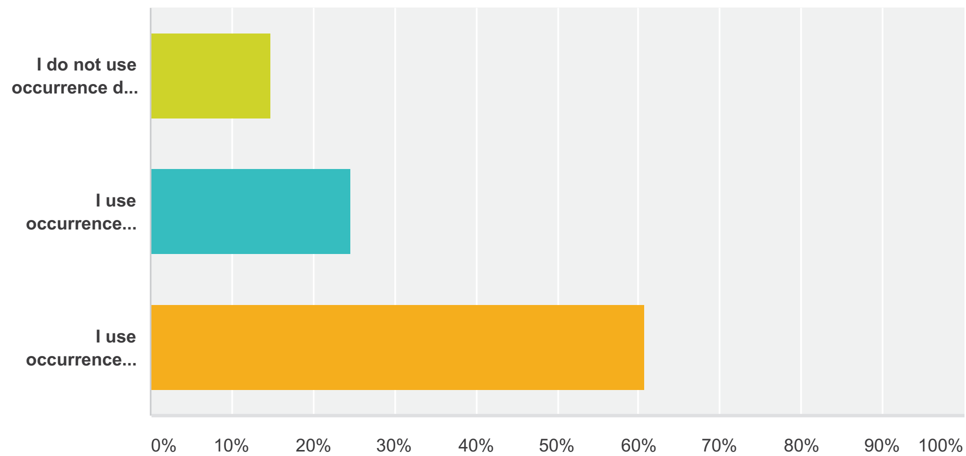
Answered: 109 Skipped: 26




| Answer Choices | Responses | |
|---|---|---|
| Yes | **28.44%** | 31 |
| Possibly | **35.78%** | 39 |
| Unlikely | **24.77%** | 27 |
| No | **11.01%** | 12 |
| **Total** | | **109** |

## Q10 You answered “Seldom or never” in Q4, this is because:

Answered: 61 Skipped: 74




| Answer Choices | Responses | |
|---|---|---|
| I do not use occurrence data in my research | **14.75%** | 9 |
| I use occurrence data, but I find GBIF.org difficult or cumbersome | **24.59%** | 15 |
| I use occurrence data, but I find the data that GBIF serves to be of low/heterogeneous/unknown quality | **60.66%** | 37 |
| **Total** | | **61** |

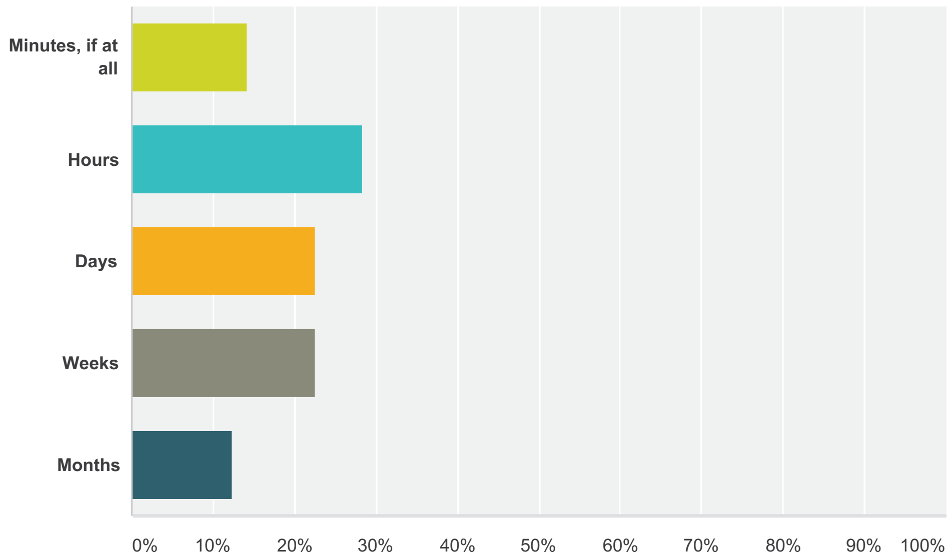
| # | Please elaborate on your problems | Date |
|---|---|---|
| 1 | I only recently discovered this as a source, I plan on exploring it further in the future | 11/2/2015 8:54 PM |
| 2 | For invertebrates the data is generally extremely scarce and of low quality, including numerous errors in taxonomy, georeferencing, etc. | 10/22/2015 3:51 PM |
| 3 | I use sometimes ("yearly"), but through other platforms (fishbase) as i find GBIF to be not very marine friendly in terms of number of species available | 10/22/2015 12:39 PM |
| 4 | So far in my research, I have been able to access the GBIF datasets I required through alternative portals offering improved exploratory capabilities and quality checks (e.g. VertNet, http://vertnet.org/; Holos, https://holos.berkeley.edu/) | 10/19/2015 3:44 PM |
| 5 | I have little information on GBIF data quality; therefore, I refrain from using it in my primary research. | 10/15/2015 8:23 PM |
| 6 | I didn't answer "Seldom or never" I answered "yearly". These kinds of problems are exactly why I am very unimpressed with GBIF as a service. | 10/13/2015 5:12 PM |
| 7 | I actually use occurrence data weekly. I do not know why it is saying I answered it "seldom or never." I do georeference all data on my own using GeoLocate because I have found that most georeferenced data in GBIF has unknown uncertainty or inaccurate coordinates. | 10/13/2015 2:54 AM |
| 8 | I like to try to focus on small areas (landscapes) and the data available on GBIF is sparse. | 10/12/2015 6:11 PM |
| 9 | Maybe these problems are more frequent in the Mediterranean context | 10/9/2015 6:36 PM |
| 10 | I only use GBIF for specific projects and I only work on a few. I download what I want and then it takes a while to analyze it. So I only use it about once a month. | 9/23/2015 11:29 PM |
| 11 | I answered "Yearly", not "Seldom or never". Anyway, often the data provided by GBIF is scarce or incomplete at the regional extent, which is often what I need for my research. Therefore, I usually contact directly with specific data providers. | 9/15/2015 5:00 PM |
| 12 | I answered "monthly". I frequently use occurrence data and GBIF is one of my main sources | 9/15/2015 4:01 PM |
| 13 | I use occurrence data but have never heard of GBIF until today. | 9/14/2015 6:59 PM |

| | | |
|---|---|---|
| 14 | I use mainly my field data | 9/9/2015 5:06 PM |
| 15 | true, but not a serious problem for us because we work with rapid assessments of what is known and there. | 9/9/2015 1:14 AM |
| 16 | I did not select this option in Q4. This is a technical error. | 9/8/2015 10:24 AM |
| 17 | I find the ALA interface and data delivery easier to use so only use GBIF for non-Australian records | 9/7/2015 1:13 AM |
| 18 | I did not answer in his manner. I use the data when I need it (between monthly and yearly. | 9/6/2015 2:04 AM |
| 19 | I did choose "monthly" - is there a problem with the survey? | 9/4/2015 11:42 PM |
| 20 | I answered "Weekly" | 9/4/2015 9:10 PM |
| 21 | Sorry, but I don't remember what Q4 was. | 9/4/2015 8:33 PM |
| 22 | GBIF data is so riddled with inaccuracies that I want to inspect it myself. I don't trust a computer to do it for me. | 9/4/2015 6:27 PM |
| 23 | I answered annually | 9/4/2015 5:59 PM |
| 24 | I answered "weekly" in Q4 | 9/4/2015 5:10 PM |
| 25 | I prefer to access all the occurrences and then select them locally | 9/4/2015 4:51 PM |
| 26 | For many species of mammals, the group I work on, the only way to have good taxonomic identification is by having an expert looking at the voucher specimens and/or molecular data. Based on GBIF information, I could not guarantee this process has been done, particularly using appropriate and updated criteria (regarding taxonomic knowledge in the recent literature and methods used for the identification shown on GBIF). | 9/4/2015 4:37 PM |
| 27 | I do not answered "Seldom or never" in Q4!! | 9/4/2015 4:01 PM |
| 28 | When I use rgbif package 'occ_search', to data downloaded .. they are not same that ones downloaded from gbif web. | 9/4/2015 3:33 PM |
| 29 | My main topic is in epidemiology, so, use GBIF data only irregularly | 9/4/2015 2:48 PM |
| 30 | If the original georeferences of the data points are wrong, then, the smaller grid-sized data sets would be even more biased/incomplete/wrong | 9/4/2015 2:35 PM |
| 31 | I used GBIF data to locate specimens in museums with georeferenced data, but most of these "records" lack a field about who identified specimen (the curatorial ID) and it is impossible to known if a record is precise in its taxonomic identification. | 9/4/2015 2:04 PM |
| 32 | Prefer to use the API rather than a clickable web interface. Mainly because it provides better repetition, reproducibility and speed | 9/4/2015 1:45 PM |
| 33 | i actually answered 'monthly' | 9/4/2015 12:18 PM |
| 34 | I am currently doing a phd so only have used for my study species or occasionally other side projects | 9/4/2015 12:01 PM |
| 35 | I answered "weekly" | 9/4/2015 11:00 AM |
| 36 | I answered yearly. | 9/4/2015 10:33 AM |
| 37 | I actually clicked monthly (and that is what it shows). I use these data both in research and in teaching on a regular basis. | 9/4/2015 10:26 AM |
| 38 | The reason is that I want to have full control of what I am seeing, including the original records; so just getting the output would not be as reliable/adequate for my research. | 8/13/2015 12:58 PM |
| 39 | I actually answered "yearly" - it is probably every few months. I often use data that are not available from GBIF. | 8/5/2015 7:21 PM |
| 40 | None of the above; it is that I am not the one accessing the data that my research group uses (students or staff are) | 7/29/2015 6:06 PM |
| 41 | I generally use eBird data which needs filter restrictions that are more easily set via requesting straight from eBird | 7/29/2015 5:29 PM |
| 42 | I selected "monthly" in Q4. Looks like the survey made an error here (Q10). | 7/24/2015 8:32 PM |
| 43 | I did not answer "Seldom or Never"!!!!!! | 7/21/2015 11:47 PM |
| 44 | Taxonomic and/or georeferencing problems are not uncommon | 7/13/2015 5:43 PM |
| 45 | I did not answer seldom or never, I answered weekly. | 7/13/2015 4:18 PM |
| 46 | I did not answer seldom or never in Q4. My answer was "Weekly" | 7/13/2015 3:58 PM |
| 47 | I always used speciesLink instead, but now most data from speciesLink is also available from GBIF. So maybe I'll start using GBIF more frequently, although GBIF's login wall for data download really bothers me. | 7/13/2015 3:00 PM |
| 48 | i did not answer seldom or never. I answered monthly. your survey is broken | 7/12/2015 6:17 AM |
| 49 | I didn't answer 'seldom or never'... I answered 'monthly' (actually, it tends to be somewhere between monthly and yearly, often for teaching classes). | 7/9/2015 11:54 PM |

| | | |
|---|---|---|
| 50 | I did not answer seldom or never, I answered weekly | 7/8/2015 4:46 PM |
| 51 | I did not answer this | 7/8/2015 11:00 AM |
| 52 | That is not my answer to that question! I use GBIF whenever I look for occurrence data. | 7/7/2015 10:42 PM |
| 53 | I didn't answer seldom or never!! | 7/7/2015 5:10 PM |

## Q11 When you download data from GBIF, how much time do you or your students/associates spend cleaning the data?

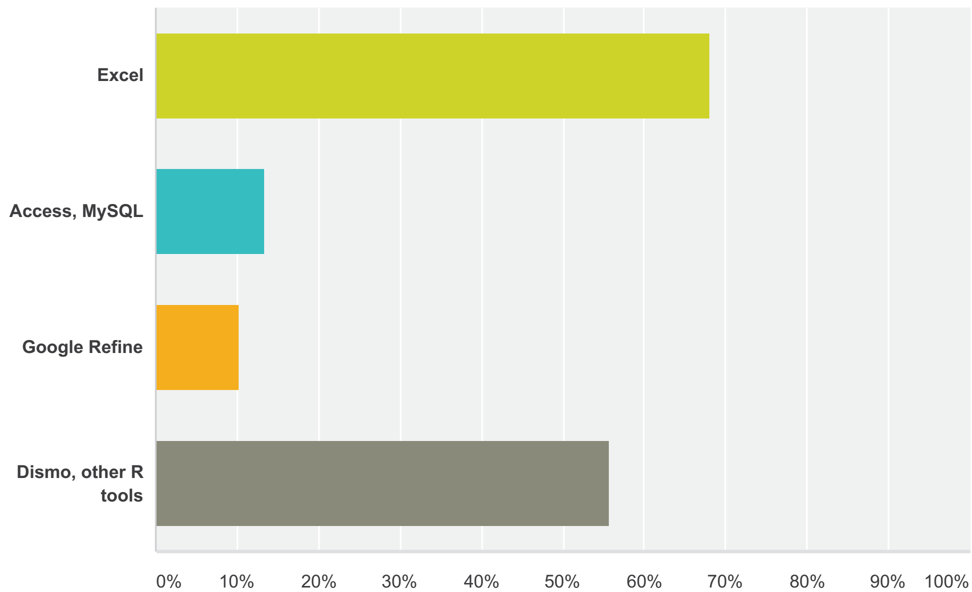
Answered: 106 Skipped: 29




| Answer Choices | Responses | |
|---|---|---|
| Minutes, if at all | 14.15% | 15 |
| Hours | 28.30% | 30 |
| Days | 22.64% | 24 |
| Weeks | 22.64% | 24 |
| Months | 12.26% | 13 |
| **Total** | | **106** |

## Q12 Do you use any software tool to clean the data?

Answered: 97 Skipped: 38



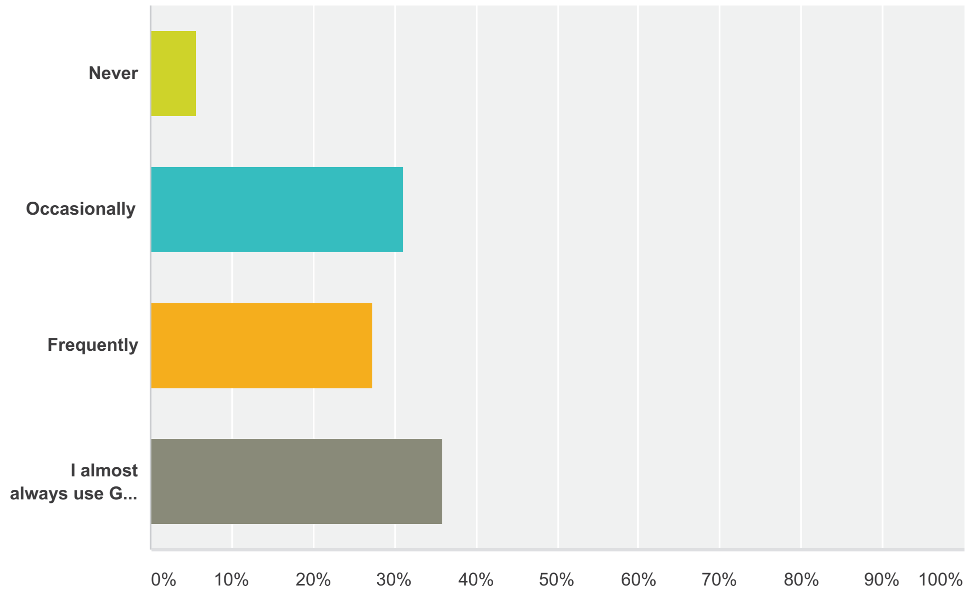

| Answer Choices | Responses | |
|---|---|---|
| Excel | **68.04%** | 66 |
| Access, MySQL | **13.40%** | 13 |
| Google Refine | **10.31%** | 10 |
| Dismo, other R tools | **55.67%** | 54 |
| **Total Respondents: 97** | | |

| # | Ad hoc software (please give us the name, if available) | Date |
|---|---|---|
| 1 | Never tried | 10/22/2015 3:56 PM |
| 2 | ArcGIS | 10/12/2015 7:57 PM |
| 3 | ArcGIS, QGIS | 9/15/2015 4:03 PM |
| 4 | ArcGis | 9/10/2015 5:01 AM |
| 5 | custom-written R scripts for geographic validation and for taxonomic standardization/validation, together with some manual checks | 9/8/2015 8:12 PM |
| 6 | I also use ArcGIS 10.1 as an additional tool | 9/8/2015 7:07 PM |
| 7 | ALA (Sandbox) | 9/7/2015 1:14 AM |
| 8 | CLIMEX, ArcGIS | 9/6/2015 2:07 AM |
| 9 | Arcmap | 9/5/2015 2:30 PM |
| 10 | In house scripts using the ropensci stack. It is done automatically, the time is difficult to estimate. | 9/4/2015 11:44 PM |
| 11 | Always visually in any GIS | 9/4/2015 7:52 PM |
| 12 | Plot points in ArcGIS against a cartographic reference | 9/4/2015 6:11 PM |
| 13 | ArcGIS | 9/4/2015 5:34 PM |
| 14 | Java app to determine if the coordinates are in land, if the country names in the records coincide with the coordinates, and check taxonomy | 9/4/2015 5:13 PM |

| | | |
|---|---|---|
| 15 | Visual examination of voucher information as voucher catalogue numbers and searching on literature where they might be cited; I look for congruencies/incongruencies regarding taxonomy and georreference. I use GoogleEarth and real non-digital maps. I review the literature associated with the involved species and contrast records information in that literature with what GBIF has on its database. I compare the georreference made by papers' authors, what appears on GBIF, and my on georreference. | 9/4/2015 4:48 PM |
| 16 | I am creating a package to cleaning data to Gbif data | 9/4/2015 4:37 PM |
| 17 | QGIS | 9/4/2015 4:05 PM |
| 18 | Taxonomic experience | 9/4/2015 2:48 PM |
| 19 | And python | 9/4/2015 1:50 PM |
| 20 | Google Street view to verify outlying data, arcgis to resample | 9/4/2015 12:03 PM |
| 21 | ArcGIS | 9/4/2015 10:30 AM |
| 22 | www.biovel.eu | 8/13/2015 1:00 PM |
| 23 | BioVeL | 7/27/2015 1:33 PM |
| 24 | GeoLocate | 7/24/2015 8:35 PM |
| 25 | Self created scripts | 7/13/2015 5:45 PM |
| 26 | We check with a R script whether names are currently valid (using catalogue of life) and if coordinates fall inside the municipalities they are suppossed to. | 7/13/2015 4:22 PM |
| 27 | Python | 7/13/2015 4:11 PM |
| 28 | http://biogeo.inct.florabrasil.net (sorry, no English interface yet) | 7/13/2015 3:37 PM |
| 29 | I'm more meaning that we consult the literature and look up the source and then edit the text or excel file. | 7/9/2015 11:59 PM |
| 30 | Visual FoxPro | 7/8/2015 1:05 AM |
| 31 | ModestR | 7/4/2015 6:40 PM |

## Q13 Do you use georeferenced GBIF data for species distribution models or niche models?

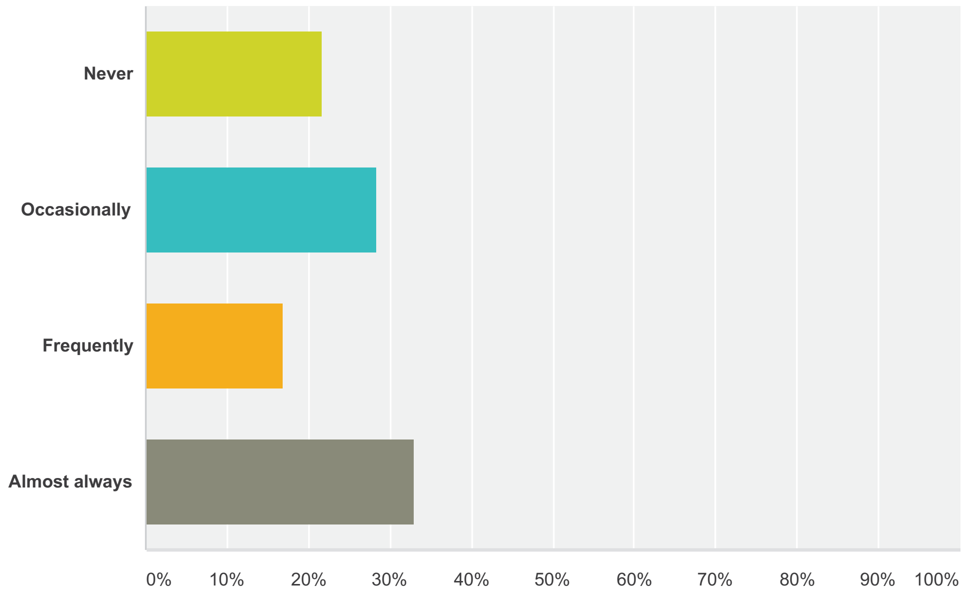
Answered: 106    Skipped: 29




| Answer Choices | Responses | |
|---|---|---|
| Never | **5.66%** | 6 |
| Occasionally | **31.13%** | 33 |
| Frequently | **27.36%** | 29 |
| I almost always use GBIF data for this purpose | **35.85%** | 38 |
| **Total** | | **106** |

## Q14 Do you apply any 'data thinning' procedure (i.e. resampling with some rules, to limit spatial autocorrelation/bias/etc.) to the downloaded GBIF data before building distribution/niche models?

Answered: 106 Skipped: 29



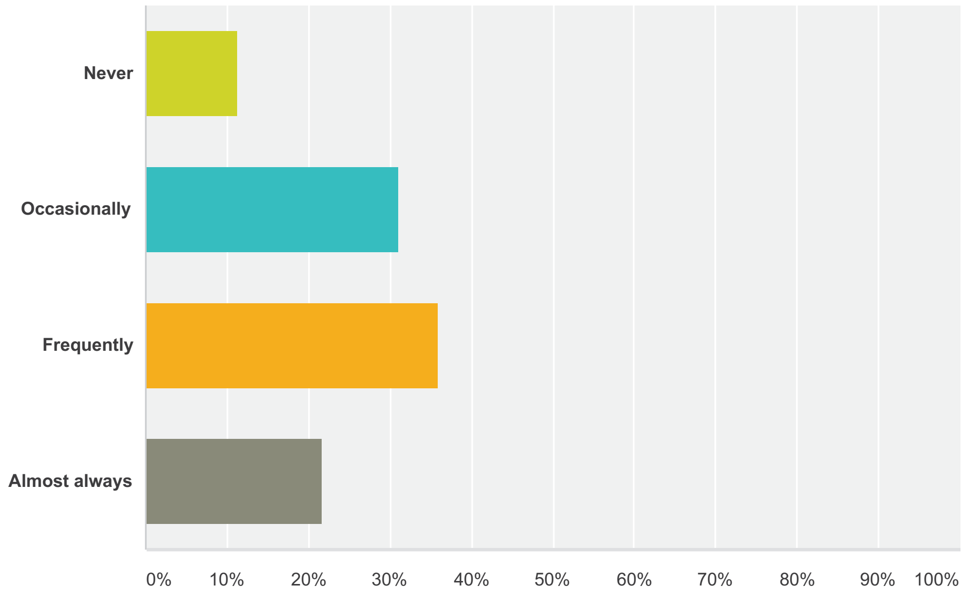

| Answer Choices | Responses | |
|---|---|---|
| Never | **21.70%** | 23 |
| Occasionally | **28.30%** | 30 |
| Frequently | **16.98%** | 18 |
| Almost always | **33.02%** | 35 |
| **Total** | | **106** |

| # | Can you link to the data cleaning algorithm you use when filtering GBIF data? | Date |
|---|---|---|
| 1 | Never used | 10/22/2015 3:56 PM |
| 2 | I clean the records in terms of known range depth for the species, excluding the points that fall outside of this range. | 10/22/2015 12:42 PM |
| 3 | We use custom tools on an adhoc basis for this depending on the project under consideration. | 10/18/2015 9:55 AM |
| 4 | I use the spThin package in R. | 10/13/2015 2:56 AM |
| 5 | I used mant tools for data thinning, accoring with the specific aims of the work. | 9/15/2015 5:04 PM |
| 6 | http://sdmtoolbox.org/ I use "spatial rarefation" within this toolbox. | 9/15/2015 4:03 PM |
| 7 | But I work more on the data than doing any modeling myself. But if I did SDMs, I probably would use such procedures. | 9/8/2015 8:12 PM |
| 8 | First I develop a grid of the desired scale depending on the species being modelled or spatial coverage. Then only extract one point to each grid before niche modelling. This is done in ArcMap 9.3-10.2. I don't use any algorithm for this. It would help if this option is implemented in niche modelling algorithms especially in the dismo package or biomod2 package in R. | 9/8/2015 10:35 AM |
| 9 | No | 9/6/2015 2:07 AM |

| | | |
|---|---|---|
| 10 | R function written by myself | 9/5/2015 1:09 AM |
| 11 | Mainly some R scripts, which I think are already available in some packages. | 9/4/2015 7:52 PM |
| 12 | Ecospat package disaggregation function | 9/4/2015 6:30 PM |
| 13 | After mixing GBIF data with other sources, I've used SDMtoolbox Brown, J.L. (2014): SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeography, and species distribution model analyses. Methods in Ecology and Evolution. | 9/4/2015 6:11 PM |
| 14 | I use algorithm implemented in Bioensemble software | 9/4/2015 5:16 PM |
| 15 | https://cran.r-project.org/web/packages/spThin/index.html | 9/4/2015 4:48 PM |
| 16 | avoidance of duplicate presence records in individual pixels, but most of the work is on the generation of pseudo-absences rules. | 9/4/2015 2:50 PM |
| 17 | Only one species occurrence per grid cell in ArcMap | 9/4/2015 2:48 PM |
| 18 | In some cases I have used a minimum distance criteria (e.g., 5-10 km) to data thinning, mostly when I wish to project niche models to geography. Most recent, I've been using EcoSpat package to do this with ecospat.occ.desaggregation function. | 9/4/2015 2:09 PM |
| 19 | Sub-sampling | 9/4/2015 1:50 PM |
| 20 | Just example in his to a grid of the same resolution as the models | 9/4/2015 12:03 PM |
| 21 | I use ESRI ArcGIS; SDMToolbox | 7/24/2015 8:35 PM |
| 22 | https://cran.r-project.org/web/packages/spThin/index.html and my own scripts | 7/21/2015 11:49 PM |
| 23 | I at least try to ensure that only one point per pixel is used. | 7/13/2015 3:37 PM |
| 24 | No specific algorithm... we'd just do it ourselves in a GIS (ArcMap). | 7/9/2015 11:59 PM |
| 25 | I develop ad-hoc procedures, mostly related to taxonomical depth, according to the problem at hand. | 7/8/2015 1:05 AM |

## Q15 Do you consider the sampling effort of your whole target group when building a model for a given species in that group?

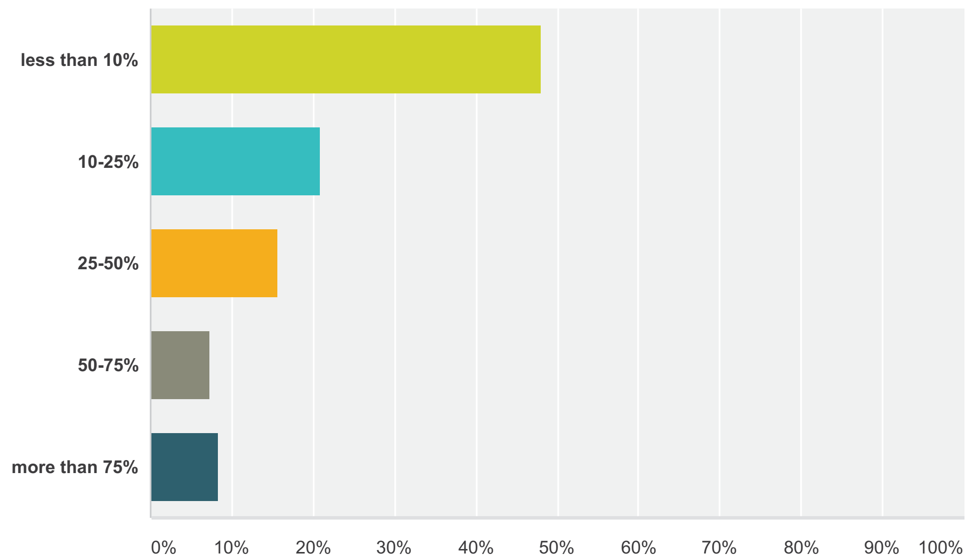
Answered: 106 Skipped: 29




| Answer Choices | Responses | |
|---|---|---|
| Never | **11.32%** | 12 |
| Occasionally | **31.13%** | 33 |
| Frequently | **35.85%** | 38 |
| Almost always | **21.70%** | 23 |
| **Total** | | **106** |

## Q16 In your last modelling exercise using GBIF published data, what percentage of SPECIES did you discard as not fit for you use?

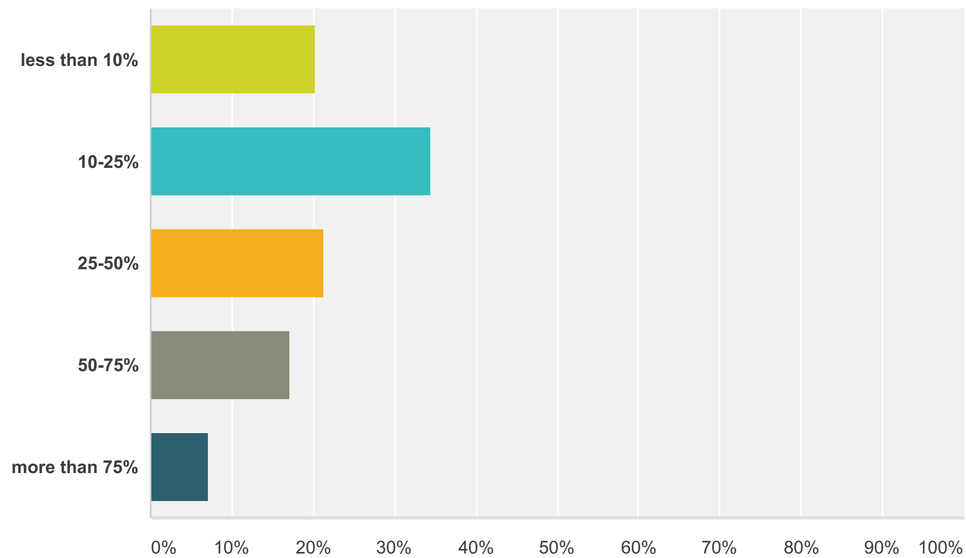
Answered: 96    Skipped: 39




| Answer Choices | Responses | |
|---|---|---|
| less than 10% | **47.92%** | 46 |
| 10-25% | **20.83%** | 20 |
| 25-50% | **15.63%** | 15 |
| 50-75% | **7.29%** | 7 |
| more than 75% | **8.33%** | 8 |
| **Total** | | **96** |

## Q17 In your last modelling exercise using GBIF published data, what percentage of RECORDS did you discard as not fit for you use?

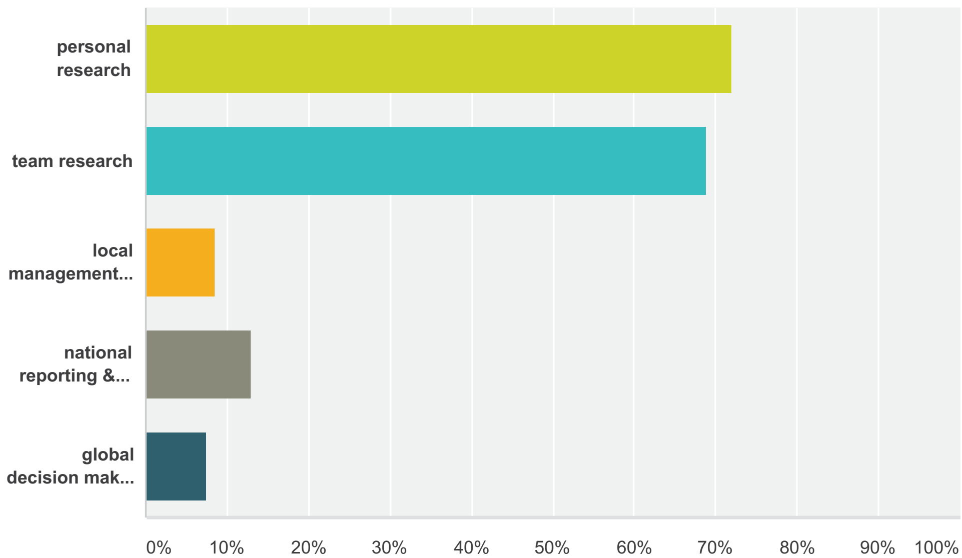
Answered: 99 Skipped: 36




| Answer Choices | Responses | |
|---|---|---|
| less than 10% | **20.20%** | 20 |
| 10-25% | **34.34%** | 34 |
| 25-50% | **21.21%** | 21 |
| 50-75% | **17.17%** | 17 |
| more than 75% | **7.07%** | 7 |
| **Total** | | **99** |

## Q18 For what purpose do you run distribution modelling?

Answered: 93 Skipped: 42

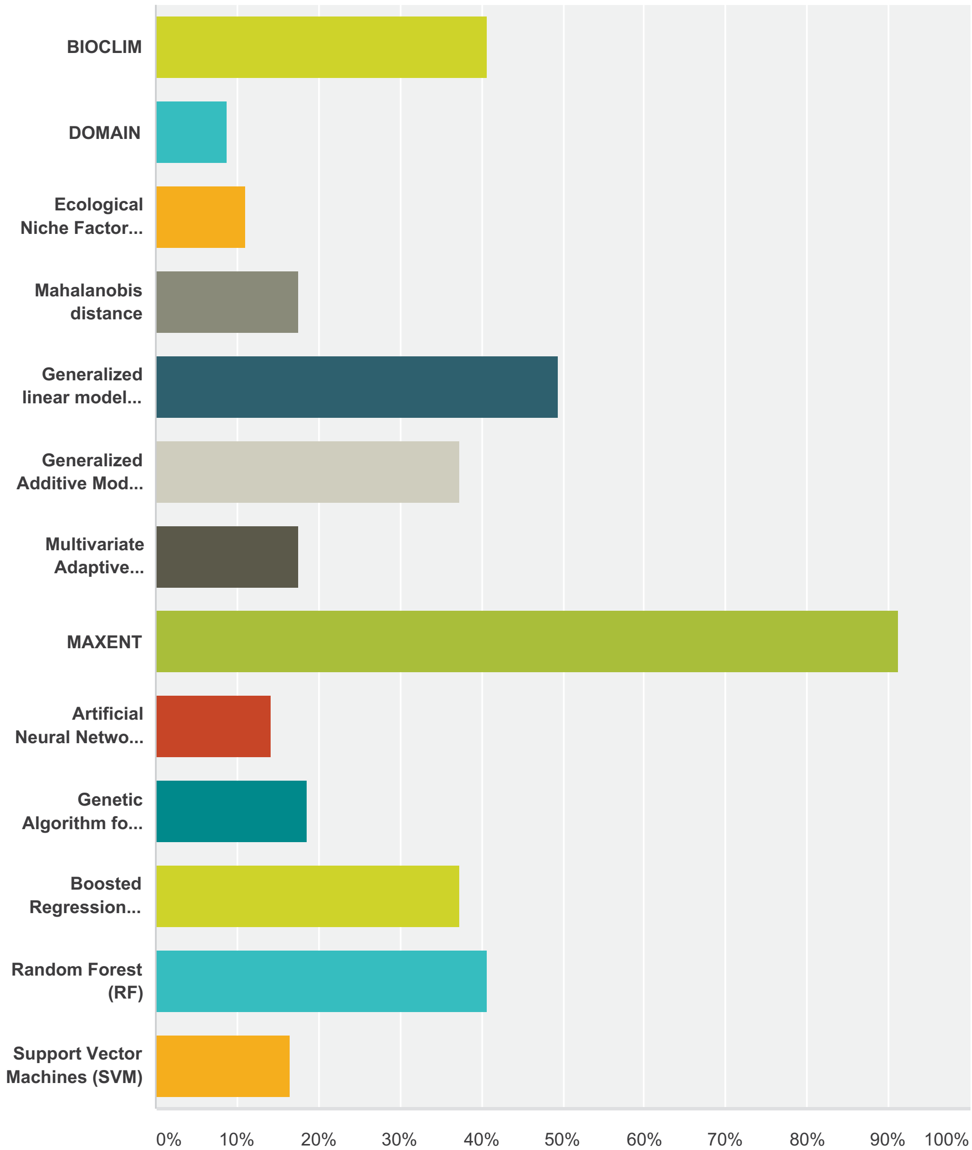



| Answer Choices | Responses | |
|---|---|---|
| personal research | **72.04%** | 67 |
| team research | **68.82%** | 64 |
| local management (land use planning etc.) | **8.60%** | 8 |
| national reporting & policy | **12.90%** | 12 |
| global decision making support | **7.53%** | 7 |
| **Total Respondents: 93** | | |

| # | Other (or comments) | Date |
|---|---|---|
| 1 | exercise on lab | 10/28/2015 7:37 PM |
| 2 | classroom exercises | 10/20/2015 5:04 PM |
| 3 | I am not sure I understand what is meant by team research. I use GBIF for macroecological reserach | 10/18/2015 10:06 AM |
| 4 | class assignment | 10/16/2015 5:00 PM |
| 5 | For the purpose of teaching distribution modeling | 10/13/2015 5:14 PM |
| 6 | discover new populations | 9/9/2015 5:47 PM |
| 7 | Teaching | 9/4/2015 11:46 PM |
| 8 | Workshops, Education | 9/4/2015 1:58 PM |
| 9 | crop wild relative conservation | 9/4/2015 12:22 PM |
| 10 | teaching | 9/4/2015 10:33 AM |
| 11 | Teaching | 7/29/2015 5:32 PM |
| 12 | Clasroom exercises | 7/8/2015 1:20 AM |

## Q19 What algorithms do you use for modelling?

Answered: 91 Skipped: 44



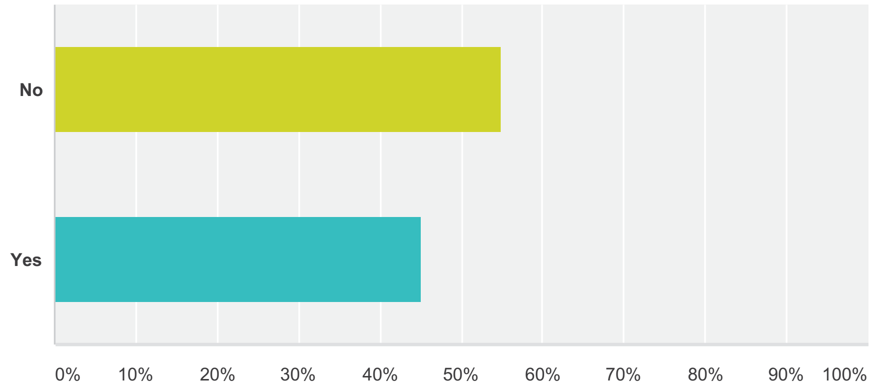

| Answer Choices | Responses | |
|---|---|---|
| BIOCLIM | **40.66%** | 37 |
| DOMAIN | **8.79%** | 8 |
| Ecological Niche Factor Analysis (ENFA) | **10.99%** | 10 |
| Mahalanobis distance | **17.58%** | 16 |
| Generalized linear model (GLM) | **49.45%** | 45 |
| Generalized Additive Model (GAM) | **37.36%** | 34 |
| Multivariate Adaptive Regression Splines (MARS) | **17.58%** | 16 |
| MAXENT | **91.21%** | 83 |

| | | |
|---|---|---|
| Artificial Neural Networks (ANN) | **14.29%** | 13 |
| Genetic Algorithm for Rule Set Production (GARP) | **18.68%** | 17 |
| Boosted Regression Trees (BRT)/Gradient Boosting Machines (GBM) | **37.36%** | 34 |
| Random Forest (RF) | **40.66%** | 37 |
| Support Vector Machines (SVM) | **16.48%** | 15 |
| **Total Respondents: 91** | | |

| # | Other (or comments) | Date |
|---|---|---|
| 1 | EuclidDist, | 10/22/2015 2:41 PM |
| 2 | unknown | 10/20/2015 5:04 PM |
| 3 | We use ensemble forecasting, and the algorithms change from time to time depending on changes in the literature. | 10/18/2015 10:06 AM |
| 4 | I signed the ones I use more often | 9/15/2015 5:13 PM |
| 5 | CLIMEX | 9/10/2015 5:05 AM |
| 6 | CLIMEX | 9/6/2015 2:10 AM |
| 7 | Favourability Function | 9/5/2015 2:04 AM |
| 8 | Euclidian Distance | 9/4/2015 4:59 PM |
| 9 | Bayesian Classification | 9/4/2015 3:40 PM |
| 10 | I have tried many of the above and might use in future but they were not useful for my research question | 9/4/2015 12:07 PM |
| 11 | Generalized Dissimilarity Modeling, Gradient Forest, other "community-level" methods | 8/5/2015 7:29 PM |
| 12 | I've used my own probabilistic models as well (MAPGEOS, 1997, resembling BIOCLIM) for specific purposes such as encounter probability prediction. | 7/8/2015 1:20 AM |

## Q20 Do you use ensemble modelling?

Answered: 93 Skipped: 42



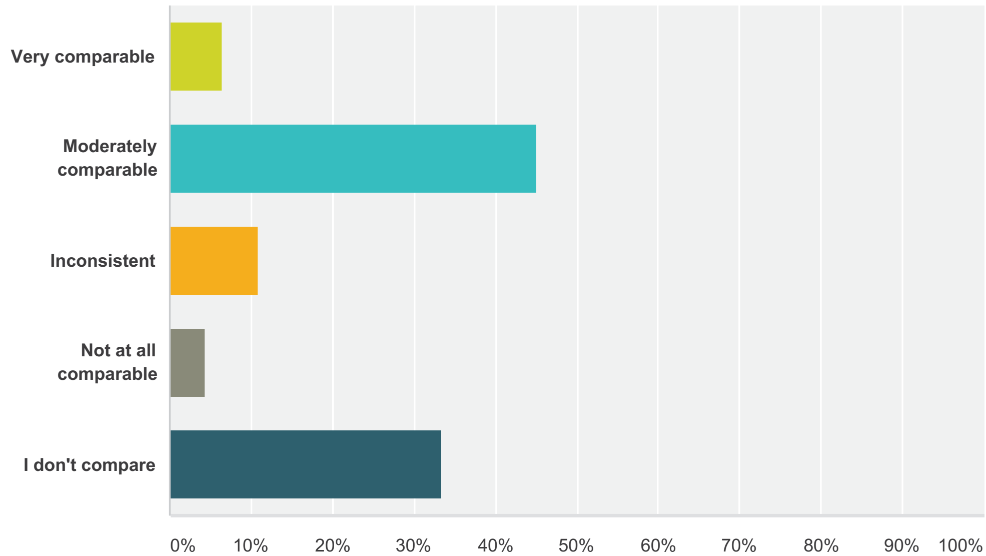

| Answer Choices | Responses | |
|---|---|---|
| No | **54.84%** | 51 |
| Yes | **45.16%** | 42 |
| **Total** | | **93** |

| # | If yes, kindly provide details | Date |
|---|---|---|
| 1 | I use BioEnsembles software (Diniz-Filho et al., 2009) | 10/22/2015 2:41 PM |
| 2 | . | 10/20/2015 5:04 PM |
| 3 | Generally, the cell-wise mean/median projection from about 4/5 of the above algorithms | 10/19/2015 3:55 PM |
| 4 | predictions are a weighted average based on prediction performance | 10/18/2015 10:06 AM |
| 5 | I use BioMod2 | 10/13/2015 3:02 AM |
| 6 | From reclassified models into binary presence/absence, I consider the median of the models for each pixel | 9/15/2015 4:08 PM |
| 7 | various ways and options | 9/9/2015 1:23 AM |
| 8 | I use ensemble models from the top general linear model, general additive model and boosted regression trees. | 9/8/2015 10:54 AM |
| 9 | Via http://www.bccvl.org.au/ | 9/7/2015 1:19 AM |
| 10 | Biomod2 | 9/5/2015 2:32 PM |
| 11 | Average multiple technics (BRT, MARS, etc.) and multiple data sources if uncertain (CCSM4, MIROC5, etc.) | 9/4/2015 10:37 PM |
| 12 | disminuir sesgos | 9/4/2015 7:56 PM |
| 13 | biomod2 | 9/4/2015 6:35 PM |
| 14 | Ensemble mean, median, max, and min | 9/4/2015 6:32 PM |
| 15 | Through the biomod2 platform and I have used the weighted mean ensemble method | 9/4/2015 5:37 PM |
| 16 | For those species with small occurrences available, we use the crossvalidation option in Maxent and calculate the mean of all k to produce a final distribution map. | 9/4/2015 5:19 PM |
| 17 | I use ensemble modelling implemented in R. | 9/4/2015 4:59 PM |
| 18 | using biomod2 | 9/4/2015 4:56 PM |
| 19 | Biomod2 package | 9/4/2015 1:58 PM |
| 20 | Often ensemble of GLM, RF and Maxent models after a cross validation procedure using the R BIOMOD package. | 9/4/2015 11:38 AM |
| 21 | Mainly for forecasting and using software the automatises the procedure | 9/4/2015 10:52 AM |
| 22 | When possible, I prefer to use model ensembles to attempt to quantify algorithm- and/or climate model-based variation. | 8/5/2015 7:29 PM |

| | | |
|---|---|---|
| 23 | predicted by at least one model vs predicted by all models | 7/13/2015 4:26 PM |
| 24 | BIOMOD | 7/10/2015 12:14 AM |
| 25 | mostly as implemented in biomod2 | 7/8/2015 11:13 AM |
| 26 | Those implemented into biomod2, namely weighted mean, committee averaging. | 7/7/2015 5:34 PM |
| 27 | I use different methods of model ensembling... | 7/7/2015 5:24 PM |

## Q21 How comparable are the results from different algorithms when fed the same corpus of data from GBIF?

Answered: 93 Skipped: 42

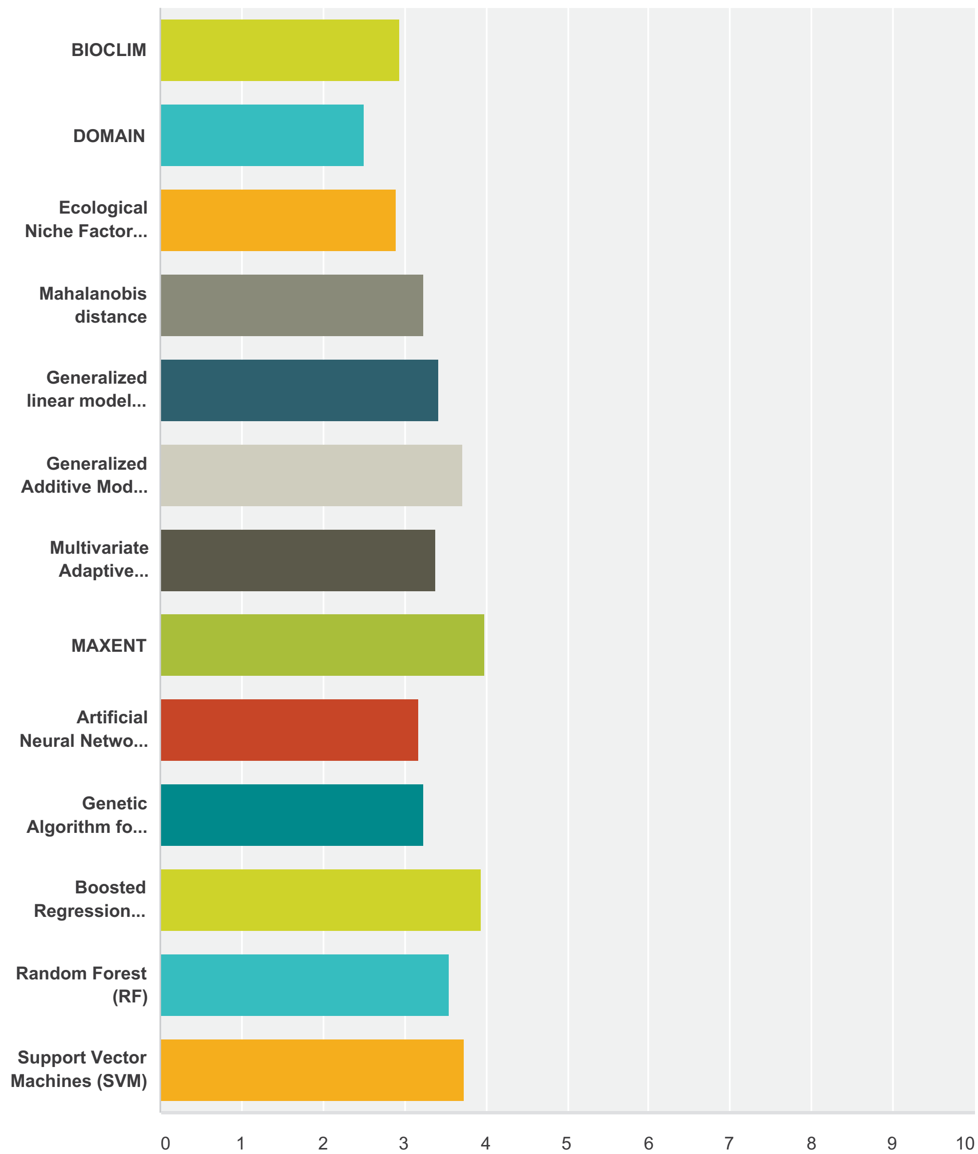



| Answer Choices | Responses | |
|---|---|---|
| Very comparable | **6.45%** | 6 |
| Moderately comparable | **45.16%** | 42 |
| Inconsistent | **10.75%** | 10 |
| Not at all comparable | **4.30%** | 4 |
| I don't compare | **33.33%** | 31 |
| **Total** | | **93** |

| # | Comments or remarks | Date |
|---|---|---|
| 1 | . | 10/20/2015 5:04 PM |
| 2 | Boosted regression trees performs better than general additive models and general linear models. Boosted regression trees spatially predict more sites with high suitability. I have only compared the predicted values among these in my recent unpublished work. | 9/8/2015 10:54 AM |
| 3 | i haven't done it yet... | 9/5/2015 2:32 PM |
| 4 | Dependent on the taxa looked at. | 9/4/2015 11:46 PM |
| 5 | Many papers highlight the different performance of algorithms. Many factors intervene (calibration area, prevalence, transference of model, etc.). | 9/4/2015 8:05 PM |
| 6 | it depends largely on the number of samples available for each species, for brazilian species we have a lack of information (small samples) for many species. | 9/4/2015 4:59 PM |
| 7 | N/A | 9/4/2015 4:53 PM |
| 8 | That depends of the species and the geographical region. In some cases, different algorithms provide similar outputs, but in others are very different. | 9/4/2015 2:17 PM |
| 9 | They are similar for projections same area same time but differ when transferring to different areas and times | 9/4/2015 10:52 AM |
| 10 | a problem well known from litterature | 7/13/2015 4:26 PM |
| 11 | Ive not done this specifically using GBIF data (but with other data they can vary a lot) | 7/10/2015 12:14 AM |

## Q22 Kindly rank modelling algorithms for use on GBIF type data

Answered: 75 Skipped: 60



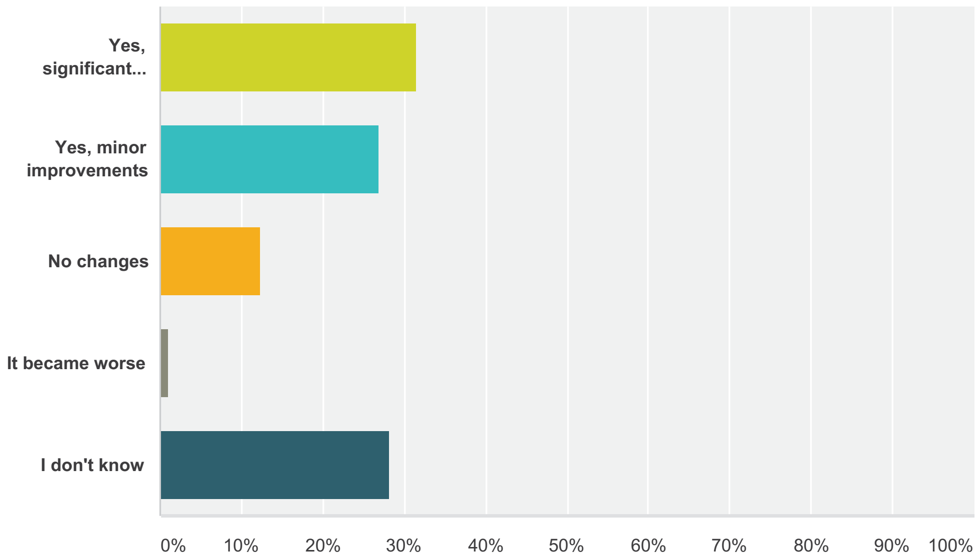

| | Poor | Below average | Satisfactory | Good | Excellent | N/A | Total | Weighted Average |
|---|---|---|---|---|---|---|---|---|
| BIOCLIM | **8.06%** 5 | **16.13%** 10 | **16.13%** 10 | **22.58%** 14 | **3.23%** 2 | **33.87%** 21 | 62 | 2.95 |
| DOMAIN | **9.43%** 5 | **7.55%** 4 | **9.43%** 5 | **5.66%** 3 | **1.89%** 1 | **66.04%** 35 | 53 | 2.50 |
| Ecological Niche Factor Analysis (ENFA) | **1.82%** 1 | **9.09%** 5 | **20.00%** 11 | **5.45%** 3 | **1.82%** 1 | **61.82%** 34 | 55 | 2.90 |
| Mahalanobis distance | **5.45%** 3 | **3.64%** 2 | **12.73%** 7 | **12.73%** 7 | **5.45%** 3 | **60.00%** 33 | 55 | 3.23 |
| Generalized linear model (GLM) | **3.28%** 2 | **9.84%** 6 | **18.03%** 11 | **29.51%** 18 | **8.20%** 5 | **31.15%** 19 | 61 | 3.43 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Generalized Additive Model (GAM) | **0.00%** 0 | **8.47%** 5 | **11.86%** 7 | **27.12%** 16 | **11.86%** 7 | **40.68%** 24 | 59 | 3.71 |
| Multivariate Adaptive Regression Splines (MARS) | **0.00%** 0 | **7.41%** 4 | **12.96%** 7 | **14.81%** 8 | **3.70%** 2 | **61.11%** 33 | 54 | 3.38 |
| MAXENT | **1.43%** 1 | **2.86%** 2 | **17.14%** 12 | **42.86%** 30 | **25.71%** 18 | **10.00%** 7 | 70 | 3.98 |
| Artificial Neural Networks (ANN) | **0.00%** 0 | **12.00%** 6 | **8.00%** 4 | **14.00%** 7 | **2.00%** 1 | **64.00%** 32 | 50 | 3.17 |
| Genetic Algorithm for Rule Set Production (GARP) | **1.85%** 1 | **9.26%** 5 | **12.96%** 7 | **20.37%** 11 | **1.85%** 1 | **53.70%** 29 | 54 | 3.24 |
| Boosted Regression Trees (BRT)/Gradient Boosting Machines (GBM) | **0.00%** 0 | **3.33%** 2 | **15.00%** 9 | **15.00%** 9 | **18.33%** 11 | **48.33%** 29 | 60 | 3.94 |
| Random Forest (RF) | **1.72%** 1 | **10.34%** 6 | **15.52%** 9 | **18.97%** 11 | **13.79%** 8 | **39.66%** 23 | 58 | 3.54 |
| Support Vector Machines (SVM) | **0.00%** 0 | **0.00%** 0 | **9.62%** 5 | **17.31%** 9 | **1.92%** 1 | **71.15%** 37 | 52 | 3.73 |

| # | Something else, what (please specify or comment) | Date |
|---|---|---|
| 1 | Gower Distance (Good) | 10/22/2015 2:41 PM |
| 2 | . | 10/20/2015 5:04 PM |
| 3 | The choice of an algorithm always depends on what the model is going to be used for, and the nature of the data availble. Assuming this questionaire is aimed at prediction, a method such as bioclim will probably work better than a "novel" method, if Bioclim is using predictors that have a stronger causal link to the organism we are trying to study. | 10/18/2015 10:06 AM |
| 4 | Sorry I can't answer this as I have not done enough comparative work | 9/23/2015 11:34 PM |
| 5 | Please note that I only signed the ones that I worked with (i.e. N/A- I never used the algorithm). And that my data sources were from multiple sources (not only from GBIF) | 9/15/2015 5:13 PM |
| 6 | Favourability Function | 9/5/2015 2:04 AM |
| 7 | See previous answer. There is no excellent method for every modeling challenge/problem. | 9/4/2015 8:05 PM |
| 8 | Not sure what's being asked--do you mean accuracy? | 9/4/2015 6:32 PM |
| 9 | I use the algorithms available in DISMO package and a few others I implement Direct in R | 9/4/2015 4:59 PM |
| 10 | Impossible to tell, because the question is unclear. For interpolation (within the spatial domain of the points), machine learning method do well, but not so much when evaluated for their extrapolation capacity. | 9/4/2015 2:53 PM |
| 11 | I think this question is unanswerable. the usefulness of a model depends on the species and research question not the source of input data. | 9/4/2015 12:07 PM |
| 12 | The question is not particularly good. Whether an algorithm performs well or not depends on a number of circumstances chiefly the purpose of the modelling. In the above I followed the simple principle that for presence only data presence only and presence background techniques are better. But this is a theoretical assertion. In practice, depending on the purposes, a presence absence model can do as well. | 9/4/2015 10:52 AM |
| 13 | I have not evaluated them properly; in our last exercise we decided to remove all SDMs since the underlying environmental data for the Amazon was so patchy/unreliable (in our view) | 8/13/2015 1:06 PM |
| 14 | this is a question better answered by my staff or students: jliao@nd.edu; jdzurisin@nd.edu | 7/29/2015 6:13 PM |
| 15 | N/A is for methods I do not use | 7/21/2015 11:54 PM |
| 16 | this question is largely dependent on the purpose of the modeling exersice | 7/13/2015 4:26 PM |
| 17 | Ouch, that's tough - I don't have any insights specific for GBIF data. And it depends what the criteria are for ranking... do you want a complex model that fits well to the GBIf data (but may be overfit)... or a simple, transparent approach... and do you want to predict the niche or the distribution... | 7/10/2015 12:14 AM |
| 18 | the reliability of results depends on the sampling bias of specific taxa in a specific study area, not so much in the modeling algorithms. I guess that BIOCLIM and DOMAIN are more adapted to GBIF data, although they largely overpredict | 7/8/2015 11:13 AM |
| 19 | See the published literature | 7/7/2015 11:33 PM |

## Q23 If you have been using GBIF for some time, have you observed any changes in RECORD AVAILABILITY for the groups of interest to you?

Answered: 89 Skipped: 46

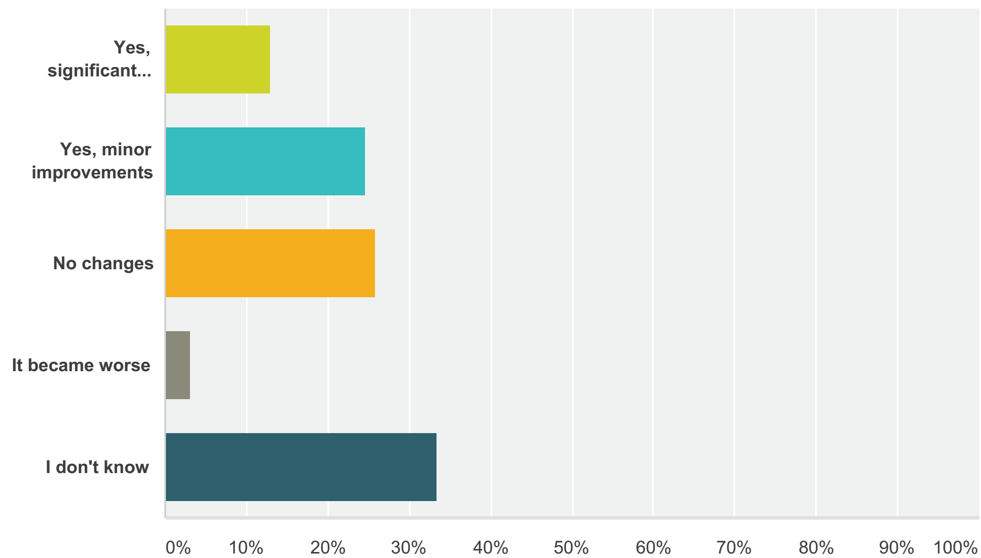



| Answer Choices | Responses | |
|---|---|---|
| Yes, significant improvements | **31.46%** | 28 |
| Yes, minor improvements | **26.97%** | 24 |
| No changes | **12.36%** | 11 |
| It became worse | **1.12%** | 1 |
| I don't know | **28.09%** | 25 |
| **Total** | | **89** |

| # | Comments or remarks | Date |
|---|---|---|
| 1 | . | 10/20/2015 5:04 PM |
| 2 | eBird etc. | 9/9/2015 1:23 AM |
| 3 | The search for different topics is confusing and hidden: Data (species, occurrences etc.). It would be interesting to have a form for more complete search with small explanations of what each search does. A simple form and an advanced search could be more effective. | 9/4/2015 4:59 PM |
| 4 | I rarely use GBIF data. | 9/4/2015 4:53 PM |
| 5 | Mostly from Latin American museums. Now it is possible to find some collections in GBIF, but not for all. | 9/4/2015 2:17 PM |
| 6 | Birds are really well sampled for Colombia | 7/13/2015 4:30 PM |
| 7 | more records for species that already had a lot of records, tropical andes endemic species record availability remains the same | 7/13/2015 4:26 PM |
| 8 | it is noticable how many more data are available now than 5 years ago | 7/10/2015 12:14 AM |

## Q24 if you have been using GBIF for some time, have you observed any changes in RECORD QUALITY for the data of interest to you?

Answered: 93 Skipped: 42




| Answer Choices | Responses | |
|---|---|---|
| Yes, significant improvements | **12.90%** | 12 |
| Yes, minor improvements | **24.73%** | 23 |
| No changes | **25.81%** | 24 |
| It became worse | **3.23%** | 3 |
| I don't know | **33.33%** | 31 |
| **Total** | | **93** |

| # | Comments or remarks | Date |
|---|---|---|
| 1 | . | 10/20/2015 5:04 PM |
| 2 | Vast majority of coordinateprecision field values are blank, and it's inclear what the units should be. | 9/4/2015 6:32 PM |
| 3 | Configure fields and search filter are fantastic, but the access to these windows is a little bit hidden, few people use that. I think it lack in the interface the possibility of submitting a list of species. I do this via an R package, which works very well, but this feature, again, is used by a few people. | 9/4/2015 4:59 PM |
| 4 | I rarely use GBIF data. | 9/4/2015 4:53 PM |
| 5 | Lots of new data is from historical records and has been georeferenced to the centroid of states and/or countries | 8/14/2015 3:45 PM |
| 6 | Mainly concerning taxonomic resolution | 8/13/2015 1:06 PM |
| 7 | Data from eBird is increasinly spatially and taxonomically accurate. | 7/13/2015 4:30 PM |

## Q25 GBIF.org serves taxonomically, geographically and temporarily heterogeneous assemblage of data, reflecting mobilization priorities of data publishers. How do you approach the task of removing data biasing?

Answered: 93 Skipped: 42

| # | Responses | Date |
|---|---|---|
| 1 | ... | 10/28/2015 7:37 PM |
| 2 | spacial filtering, though most of the species data that I focus on is sparse anyway. | 10/24/2015 10:39 PM |
| 3 | Selecting data from similar sources and comparing them with range maps provided by experts | 10/22/2015 2:41 PM |
| 4 | N/A | 10/22/2015 10:12 AM |
| 5 | ? | 10/20/2015 5:04 PM |
| 6 | List length approaches: obtain data at the broad taxon level (e.g. all Odonates) and use number of records per location per time (e.g. year) as a proxy for recorder effort. Occupancy modeling: use spatial/temporal replicates within a larger area/time period as replicates to generate a detection submodel within an occupancy modeling framework | 10/19/2015 3:55 PM |
| 7 | We do many things, but one thing we do is we use temporal matching to link the occurrence data to environmental predictors that reflect the conditions relevant for that time, place and organism under study. | 10/18/2015 10:06 AM |
| 8 | I don't | 10/16/2015 5:00 PM |
| 9 | I don't. | 10/15/2015 9:00 PM |
| 10 | Random subsampling. | 10/15/2015 8:28 PM |
| 11 | . | 10/14/2015 10:46 PM |
| 12 | data research simultaneously through literature review | 10/14/2015 9:42 AM |
| 13 | N/A | 10/13/2015 5:14 PM |
| 14 | Find alternative datasets to fill in the taxonomic group / geographic gaps when possible (i.e. other museum records, observation data, and etc.). | 10/13/2015 3:02 AM |
| 15 | I thin my points. | 10/12/2015 8:12 PM |
| 16 | Since I work on a small scale, I use all the data I can. However, I try to separate by date. | 10/12/2015 6:15 PM |
| 17 | Searching a better environmental and geographical heterogeneity of data | 10/2/2015 6:41 PM |
| 18 | Mostly stick with museum data | 10/2/2015 3:20 AM |
| 19 | Usually try to complete dataset using other sources | 9/29/2015 3:53 AM |
| 20 | Sometimes we limit the number of records that we use so the areas are more comparable | 9/23/2015 11:34 PM |
| 21 | spatial declustering | 9/16/2015 12:32 AM |
| 22 | I would provide a resume measures of data bias | 9/15/2015 5:13 PM |
| 23 | Maybe this should be done after data download according to each study requirements. Data in GBIF could be available in "raw". | 9/15/2015 4:08 PM |
| 24 | Flagging, documenting critical points. Potentially including a metrics/code for sampling effort. | 9/15/2015 7:23 AM |
| 25 | Geographically and temporarily heterogeneous assemblage of data | 9/10/2015 5:05 AM |
| 26 | The most important is the promotion of field studies in less studied regions. Promoting taxonomy would also be welcome (all taxonomic groups). Is very serious the lack of data and the lack of experts. Removing data biasing? Ok in a second step. The knowledge of species distribution patterns is crucial, so congratulations to GBIF to be doing this work. knowledge of biodiversity is very important for several reasons not only to write papers. | 9/9/2015 5:47 PM |
| 27 | data mining and machine learning | 9/9/2015 1:23 AM |
| 28 | dfg | 9/8/2015 8:13 PM |

| | | |
|---|---|---|
| 29 | By making an elaborated cleaning of the data. | 9/8/2015 7:13 PM |
| 30 | NO | 9/8/2015 2:53 PM |
| 31 | Since mine is a collaborative project involving several experts who have 15 years of experience in the region of interest. It will only help if experts sit together and clean all the erroneous records. Further I normally don't restrict my data to GBIF. I also make use of other databases along with regional databases to minimize bias in sampling. | 9/8/2015 10:54 AM |
| 32 | NA | 9/7/2015 11:44 AM |
| 33 | In my case, the more data the better. Bias is far less an issue than spatial and temporal coverage. | 9/7/2015 1:19 AM |
| 34 | This isn't so much of a problem for most of my work. | 9/6/2015 2:10 AM |
| 35 | Checking records with other sources of information. | 9/5/2015 10:27 PM |
| 36 | na | 9/5/2015 2:32 PM |
| 37 | Yes, I agree | 9/5/2015 2:04 AM |
| 38 | With great precaution (and on a case by case basis) | 9/4/2015 11:46 PM |
| 39 | I normally try to have the best knowledge possible of a species distribution before starting compiling records. I normally discard errors in taxonomic identification (e.g., southern localities of Fucus guiry that was previously identified as Fucus vesiculosus) | 9/4/2015 10:37 PM |
| 40 | easier access to related exsicates | 9/4/2015 9:21 PM |
| 41 | Using re-sampling thechiques | 9/4/2015 9:14 PM |
| 42 | Applying spatial filters. | 9/4/2015 8:05 PM |
| 43 | buscar redundancia en los datos, con un análisis estadístico exploratorio | 9/4/2015 7:56 PM |
| 44 | Nothing | 9/4/2015 7:41 PM |
| 45 | N/A | 9/4/2015 7:05 PM |
| 46 | Concerted effort to finance data digitization. Official requests to data repositories such as museums as herbaria to provide data | 9/4/2015 6:35 PM |
| 47 | Geographic or environmental thinning plus bias grid, target background, or inverse p-weighting of records. | 9/4/2015 6:32 PM |
| 48 | Create an optional output that randomly sub-sample the records of each species to the median number of points per species in the region of interest. But this is just a wild guess. Right now, I'm not sure how to handle this problem. | 9/4/2015 6:17 PM |
| 49 | Background manipulation, using bigger grid cells and through new data collection | 9/4/2015 5:37 PM |
| 50 | autocorrelation analysis | 9/4/2015 5:22 PM |
| 51 | I'm starting to use the sample selection bias proposed by Phillips et al., 2009 (Ecological Applications). | 9/4/2015 5:19 PM |
| 52 | I always apply the location filter, and only use the coordinates with no issues. After that I use other spatial filters through QGIS. | 9/4/2015 4:59 PM |
| 53 | I evaluate spatial clustering and combine gbif with other sources | 9/4/2015 4:56 PM |
| 54 | I rarely use GBIF data. | 9/4/2015 4:53 PM |
| 55 | To make sistetematics filters | 9/4/2015 4:51 PM |
| 56 | i don´t aproach | 9/4/2015 4:13 PM |
| 57 | aggregation by grid cell | 9/4/2015 4:10 PM |
| 58 | By including a bias grid to my modelling procedure. | 9/4/2015 3:56 PM |
| 59 | To date I modeled only in South America. I remove records prior to 1950, as the Worldclim database covers the 1950-2000 period. | 9/4/2015 3:40 PM |
| 60 | So far, I just deleted records with some degree of uncertainty. On the other hand, the method proposed by Varela et al. 2014 (Ecography, 37: 1084-1091) to reduce bias seems very interesting. (I have not used it yet) | 9/4/2015 3:40 PM |
| 61 | sub-sampling, and using methods that do not overfit too much. | 9/4/2015 2:53 PM |
| 62 | I add other databases or my own collections from non-digitized repositories (herbaria) | 9/4/2015 2:48 PM |
| 63 | It is necessary to have a good knowledge about taxonomy and distribution (i.e., re-checking primary literature as species descriptions and field guides) of the taxa studied. In some cases, many data are discarded because it is extremely hard to known if a record 100 km apart from others is a good observation or is an error of identification/georeferencing. | 9/4/2015 2:17 PM |

| | | |
|---|---|---|
| 64 | Focussing on specific well covered subsets, supplement with other data (IUCN, Museum, Biodiversity database) if available. | 9/4/2015 1:58 PM |
| 65 | Manually | 9/4/2015 12:27 PM |
| 66 | Difficult to tell if data is biased, therefore do not remove it especially if have few records. | 9/4/2015 12:22 PM |
| 67 | It is a vital first step in modelling and researchers should use any and all tools to do this subject to the question at hand. | 9/4/2015 12:07 PM |
| 68 | Often by selecting a radius around the sampling point from which I eliminate other species presences. | 9/4/2015 11:38 AM |
| 69 | Checking record by record against prior knowledge and removing obvious errors. | 9/4/2015 10:52 AM |
| 70 | In the classroom, we have experimented with several methods | 9/4/2015 10:33 AM |
| 71 | With R scrips we remove duplicate entries (same collection event at different herbaria); remove collections from the same point; jackknife or bootstrap analyses. | 8/14/2015 3:45 PM |
| 72 | I evaluate its impact on the results using different approaches, from automated filters (e.g. removing different thresholds and comparison with thoroughly evaluated data - see http://onlinelibrary.wiley.com/doi/10.1111/geb.12326/abstract). We are also using SpeciesGeoCoder (see http://www.antonelli-lab.net/resources.php) for many analyses with GBIF data. | 8/13/2015 1:06 PM |
| 73 | Spatial / environmental thinning. | 8/5/2015 7:29 PM |
| 74 | We usually augment GBIF data with data from other sources | 7/29/2015 6:13 PM |
| 75 | Taxa-specific communities of experts | 7/29/2015 5:32 PM |
| 76 | Normalising data by working with relative number of records out of all records. Cf. papers on scaling and "abundance from occurrence". | 7/27/2015 1:37 PM |
| 77 | Spatial filtering (SDMToolbox); matching records temporally to environmental data | 7/24/2015 8:38 PM |
| 78 | (1) I always explore the data firs, get acquainted with it. (2) There are envelope methods less sensitive to biases. (3) When there are obvious geographic biases on large datasets, I thin the data by spatial lag. | 7/21/2015 11:54 PM |
| 79 | Spatial thinning; or bias correction via target-group data | 7/19/2015 10:22 PM |
| 80 | Apply spatial and environmental filters | 7/13/2015 5:48 PM |
| 81 | Most of the time we don't. Ocassionally, we use target background sampling. | 7/13/2015 4:30 PM |
| 82 | Searching for additional sources of data that are not part of GBIF | 7/13/2015 4:26 PM |
| 83 | Still not handling bias. | 7/13/2015 3:42 PM |
| 84 | Depaneding on the objective, I remove observations that are not well enough georeferenced, that are older and may not reflect current distributions, and that are from citizen scientists and therefore may have taxonomic, georeferencing or spatial bias problems, I do not trust eBird. | 7/12/2015 6:23 AM |
| 85 | Target group background selection is neat. Or use a simple envelope method (e.g., BIOCLIM) and acknowledge that all the model is telling you is those locations with similar conditions to where you have data... that is useful information in itself if properly interpreted. | 7/10/2015 12:14 AM |
| 86 | correct for overall collection intensity for wider taxonomic group to account for spatial biases in collection intensity - no correction for taxonomic biases. | 7/8/2015 4:51 PM |
| 87 | I compare GBIF data to other distribution sources (e.g. atlases) and thin the data to match the resolution of the associated environmental data | 7/8/2015 11:13 AM |
| 88 | Almost always by pattern analysis. However, often what I study is the data biasing itself. | 7/8/2015 1:20 AM |
| 89 | sub-sampling as discussed above, but the big problem is the data scarcity in the tropics. It would be interesting if GBIF could help find ways to improvat that. | 7/7/2015 11:33 PM |
| 90 | Resampling in the environmental space if possible. If too biased, use other data sources. | 7/7/2015 5:34 PM |
| 91 | I think we need to know (and the required tools) the appropriate methods to get the best use of data given the existing biases, | 7/7/2015 5:24 PM |
| 92 | target group background | 7/7/2015 5:03 PM |
| 93 | imprecise | 7/4/2015 6:44 PM |

## Q26 Where do you share your results? Provide link to examples if possible

Answered: 93 Skipped: 42

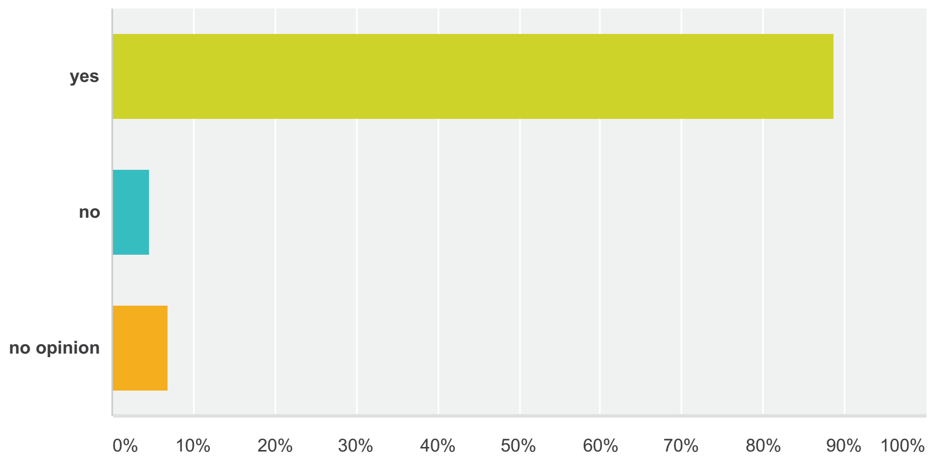
| # | Responses | Date |
|---|---|---|
| 1 | just for my exercise on my class. | 10/28/2015 7:37 PM |
| 2 | . | 10/24/2015 10:39 PM |
| 3 | in peer-review journals | 10/22/2015 2:41 PM |
| 4 | N/A | 10/22/2015 10:12 AM |
| 5 | nowhere | 10/20/2015 5:04 PM |
| 6 | Mostly peer-reviewed journals | 10/19/2015 3:55 PM |
| 7 | https://scholar.google.dk/citations?user=OIefC00AAAAJ&hl=en | 10/18/2015 10:06 AM |
| 8 | I don't | 10/16/2015 5:00 PM |
| 9 | I used the data in a class. | 10/15/2015 9:00 PM |
| 10 | I have only used GBIF in one course for the time being, and in this forum is where I've analyzed, presented, and discussed results. | 10/15/2015 8:28 PM |
| 11 | . | 10/14/2015 10:46 PM |
| 12 | Scientific Articles | 10/14/2015 9:42 AM |
| 13 | N/A | 10/13/2015 5:14 PM |
| 14 | I hope to publish the multiple projects I am part of using this data, but, I am still in the analysis / write up phase. | 10/13/2015 3:02 AM |
| 15 | In press at the moment | 10/12/2015 8:12 PM |
| 16 | Results not yet published/shared | 10/12/2015 6:15 PM |
| 17 | Not apply | 10/2/2015 6:41 PM |
| 18 | peer review publications | 10/2/2015 3:20 AM |
| 19 | Peer reviewed journals | 9/29/2015 3:53 AM |
| 20 | The data have not been made publicly available although the records from my institution are available from our general data base. | 9/23/2015 11:34 PM |
| 21 | peer reviewed papers - http://onlinelibrary.wiley.com/doi/10.1111/mam.12036/abstract | 9/16/2015 12:32 AM |
| 22 | I shere results in scientific appers and in http://atlas.freshwaterbiodiversity.eu/ | 9/15/2015 5:13 PM |
| 23 | Dryad ResearchGate | 9/15/2015 4:08 PM |
| 24 | Currently only at a institutional level. | 9/15/2015 7:23 AM |
| 25 | Distribution of species with interest in agronomy (In current conditions and many climate change scenerios) | 9/10/2015 5:05 AM |
| 26 | Scientific literature, including books http://www.tropicos.org/Person/88229?tab=references | 9/9/2015 5:47 PM |
| 27 | various options and links and outlets. | 9/9/2015 1:23 AM |
| 28 | fgfg | 9/8/2015 8:13 PM |
| 29 | They are not shared yet. | 9/8/2015 7:13 PM |
| 30 | MANUSCRIPT | 9/8/2015 2:53 PM |
| 31 | Only in research papers or publications. The data is also used for ground monitoring of species in reintroduction or translocation programs. | 9/8/2015 10:54 AM |
| 32 | NA | 9/7/2015 11:44 AM |
| 33 | Used for local testing of known species distributions, compared with expert distributions | 9/7/2015 1:19 AM |
| 34 | PLOS one, MEE, Neobiota | 9/6/2015 2:10 AM |
| 35 | http://revistas.ucr.ac.cr/index.php/rbt/article/viewFile/8379/12823 | 9/5/2015 10:27 PM |
| 36 | my PhD | 9/5/2015 2:32 PM |

| | | |
|---|---|---|
| 37 | not yet | 9/5/2015 2:04 AM |
| 38 | Papers, figshare | 9/4/2015 11:46 PM |
| 39 | All results are published in peer reviewed journals. Please follow this link: https://www.researchgate.net/profile/Jorge_Assis2/contributions | 9/4/2015 10:37 PM |
| 40 | Work is in progress, it will be offered for publication in a scientific journal when ms is ready | 9/4/2015 9:21 PM |
| 41 | We are preparing a paper. Results not available at this moment | 9/4/2015 9:14 PM |
| 42 | Supplementary materials of some publications: http://onlinelibrary.wiley.com/doi/10.1111/evo.12343/abstract http://onlinelibrary.wiley.com/doi/10.1111/jbi.12302/abstract | 9/4/2015 8:05 PM |
| 43 | http://www.uacj.mx/SC/Paginas/SELPER-Mexico-UACJ-2015.aspx | 9/4/2015 7:56 PM |
| 44 | not sharing results yet | 9/4/2015 7:41 PM |
| 45 | N/A | 9/4/2015 7:05 PM |
| 46 | Open access publishing | 9/4/2015 6:35 PM |
| 47 | http://www.earthskysea.org/publications/ | 9/4/2015 6:32 PM |
| 48 | https://www.researchgate.net/profile/Ariel_Rodriguez3/ | 9/4/2015 6:17 PM |
| 49 | Published papers | 9/4/2015 5:37 PM |
| 50 | I share results in published papers, e.g. Collevatti et al. 2012, Mol Ecol; Collevatti et al. 2013, JBI; Collevatti et al. 2013, N&C; Lima et al. 2014, JBI; Collevatti et al. 2014, BMC Evolutionary Biology; Collevatti et al. 2015, Frontiers in Plant Science. | 9/4/2015 5:22 PM |
| 51 | www.cwrdiversity.org/distribution-map/ and recent publications: orcid.org/0000-0003-1827-4782 | 9/4/2015 5:19 PM |
| 52 | http://shiny.jbrj.gov.br/modelagem/ | 9/4/2015 4:59 PM |
| 53 | publications, workshops, conferences | 9/4/2015 4:56 PM |
| 54 | I have not shared yet except for appendices of my papers. :( | 9/4/2015 4:53 PM |
| 55 | I cannot share my results yet, because is my thesis | 9/4/2015 4:51 PM |
| 56 | always | 9/4/2015 4:13 PM |
| 57 | journal publications | 9/4/2015 4:10 PM |
| 58 | NA | 9/4/2015 3:56 PM |
| 59 | http://www.sciencedirect.com/science/article/pii/S1616504712000626 http://www.sciencedirect.com/science/article/pii/S1616504715000701 http://link.springer.com/article/10.1007/s10113-014-0604-1#page-1 http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S0327-93832010000200008 | 9/4/2015 3:40 PM |
| 60 | .. | 9/4/2015 3:40 PM |
| 61 | through publications, but none so far using GBIF data sets | 9/4/2015 2:53 PM |
| 62 | Peer-reviewed journals | 9/4/2015 2:48 PM |
| 63 | I'm still in process to share my results. | 9/4/2015 2:17 PM |
| 64 | With workshop participants and students. In scientific articles (currently in review) | 9/4/2015 1:58 PM |
| 65 | Not yet share | 9/4/2015 12:27 PM |
| 66 | n/a | 9/4/2015 12:22 PM |
| 67 | Thesis, hopefully published papers in future. | 9/4/2015 12:07 PM |
| 68 | Only in the published papers | 9/4/2015 11:38 AM |
| 69 | Mainstream literature. | 9/4/2015 10:52 AM |
| 70 | We are working on building a pollen database that includes the geographic distribution of the pollen types. The distribution models are built at the genus level, which is the resolution at which we can identify the pollen. | 9/4/2015 10:33 AM |
| 71 | Not shared yet, but at least four publications recently submitted | 8/14/2015 3:45 PM |
| 72 | We used Zenodo (e.g. http://dx.doi.org/10.5281/zenodo.18425) and twitted about it. | 8/13/2015 1:06 PM |
| 73 | Publications, though now that I think of it I don't recall ever publishing a paper that used GBIF data. | 8/5/2015 7:29 PM |

| | | |
|---|---|---|
| 74 | published journal articles IMPORTANTLY, we also have implemented GBIF in a species niche modeling platform called SPACES (Spatial Portal for Analysis of Climate Effects on Species; https://adapt.nd.edu/groups/cem/wiki/Tools#SPACES). SPACES allows users to import occurrence data from GBIF, select from a set of climatic data and climate projections, and choose niche modeling algorithms for comparison or ensemble calculations. It maps results that are exportable to GIS. I would be interested and willing for GBIF to incorporate SPACES into its modeling platform or work collaboratively to enhance modeling connections between GBIF and SPACES. The next question in this survey asks if GBIF should have a modeling platform. I think we have already built something that GBIF could use/implement. | 7/29/2015 6:13 PM |
| 75 | Publications? | 7/29/2015 5:32 PM |
| 76 | Papers and BioVeL reports. | 7/27/2015 1:37 PM |
| 77 | Publications | 7/24/2015 8:38 PM |
| 78 | We are planning a section of my laboratory Web Page dedicated to share "cleansed data" | 7/21/2015 11:54 PM |
| 79 | Journal publications. | 7/19/2015 10:22 PM |
| 80 | Peer-review publications | 7/13/2015 5:48 PM |
| 81 | biomodelos.humboldt.org.co | 7/13/2015 4:30 PM |
| 82 | http://www.esajournals.org/doi/abs/10.1890/ES13-00049.1 http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0119891 http://onlinelibrary.wiley.com/doi/10.1111/gcb.13027/abstract | 7/13/2015 4:26 PM |
| 83 | Literature and http://biogeo.inct.florabrasil.net | 7/13/2015 3:42 PM |
| 84 | in peer reviewed papers | 7/12/2015 6:23 AM |
| 85 | In papers, ideally with supplementary material with the raw data or GIS layers. I don't have examples using GBIF data (but see 'Data and Software' tab at https://www.ucl.ac.uk/cber/pearson for examples of other data/results sharing). | 7/10/2015 12:14 AM |
| 86 | publications - Global Change Biology, Plos One, Journal of Biogeography Conference talks - ESA, ATBC | 7/8/2015 4:51 PM |
| 87 | In peered reviewed publications? I guess I do not understand the question. | 7/8/2015 11:13 AM |
| 88 | Regular papers and TDWG communications. E.g.: DOI:10.1007/s10531-014-0718-2, DOI:10.3897/zookeys.403.7149, DOI:10.17161/bi.v8i2.4124, DOI:10.17161/bi.v8i2.4125, DOI:10.1371/journal.pone.0055144 | 7/8/2015 1:20 AM |
| 89 | journal articles | 7/7/2015 11:33 PM |
| 90 | Do not understand the question. | 7/7/2015 5:34 PM |
| 91 | I try to put the results of my studies on my website. Examples: http://r-gis.net/?q=positional_uncertainty http://r-gis.net/?q=positional_uncertainty2 | 7/7/2015 5:24 PM |
| 92 | appendices | 7/7/2015 5:03 PM |
| 93 | www.biogeografia.org | 7/4/2015 6:44 PM |

## Q27 Would you find an option in GBIF to automatically provide (and map) the sampling effort/data completeness for a taxonomic group useful?

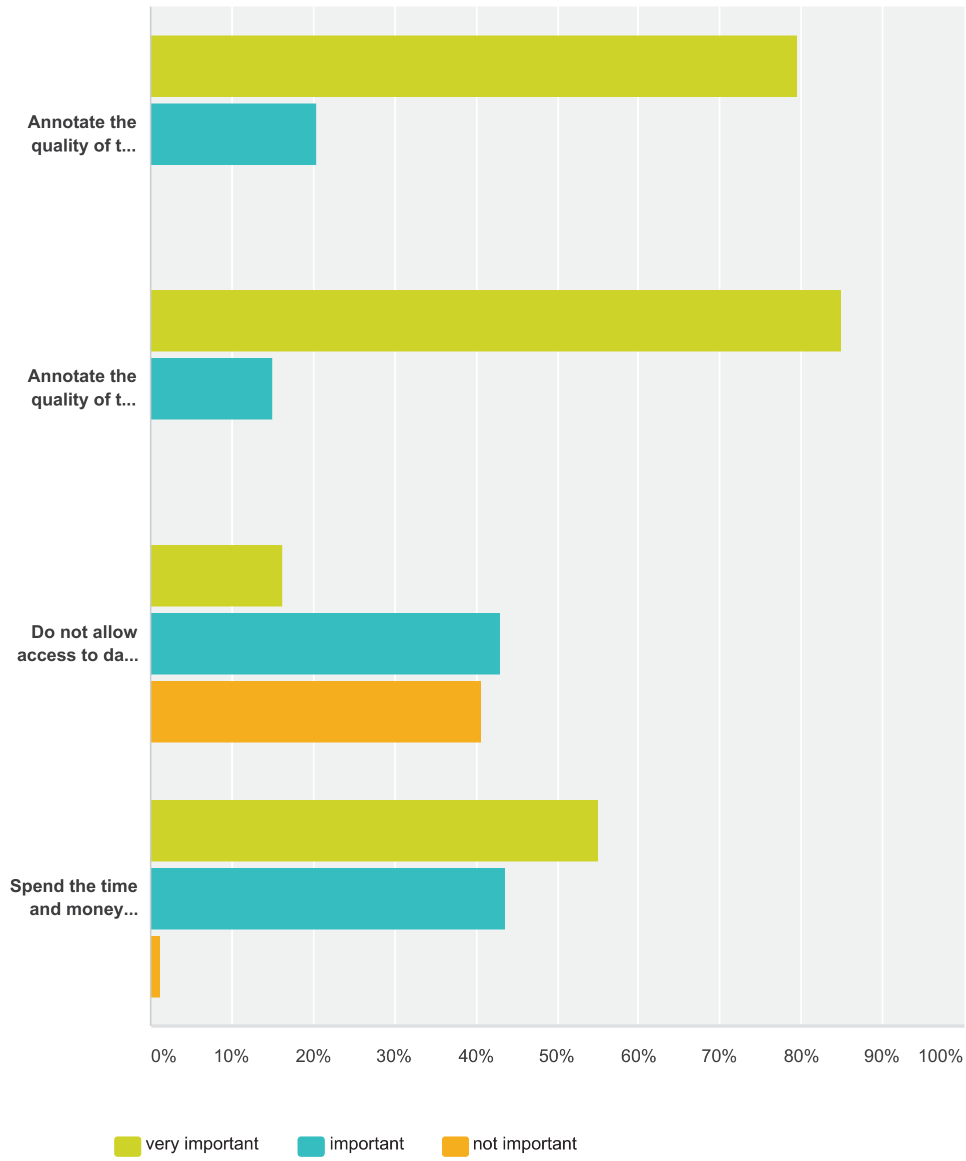
Answered: 88 Skipped: 47




| Answer Choices | Responses | |
|---|---|---|
| yes | **88.64%** | 78 |
| no | **4.55%** | 4 |
| no opinion | **6.82%** | 6 |
| **Total** | | **88** |

## Q28 Which of these suggestions would you consider important for GBIF to transmit to the original DATA PROVIDERS (museums, herbaria, networks of observers)?

Answered: 88 Skipped: 47



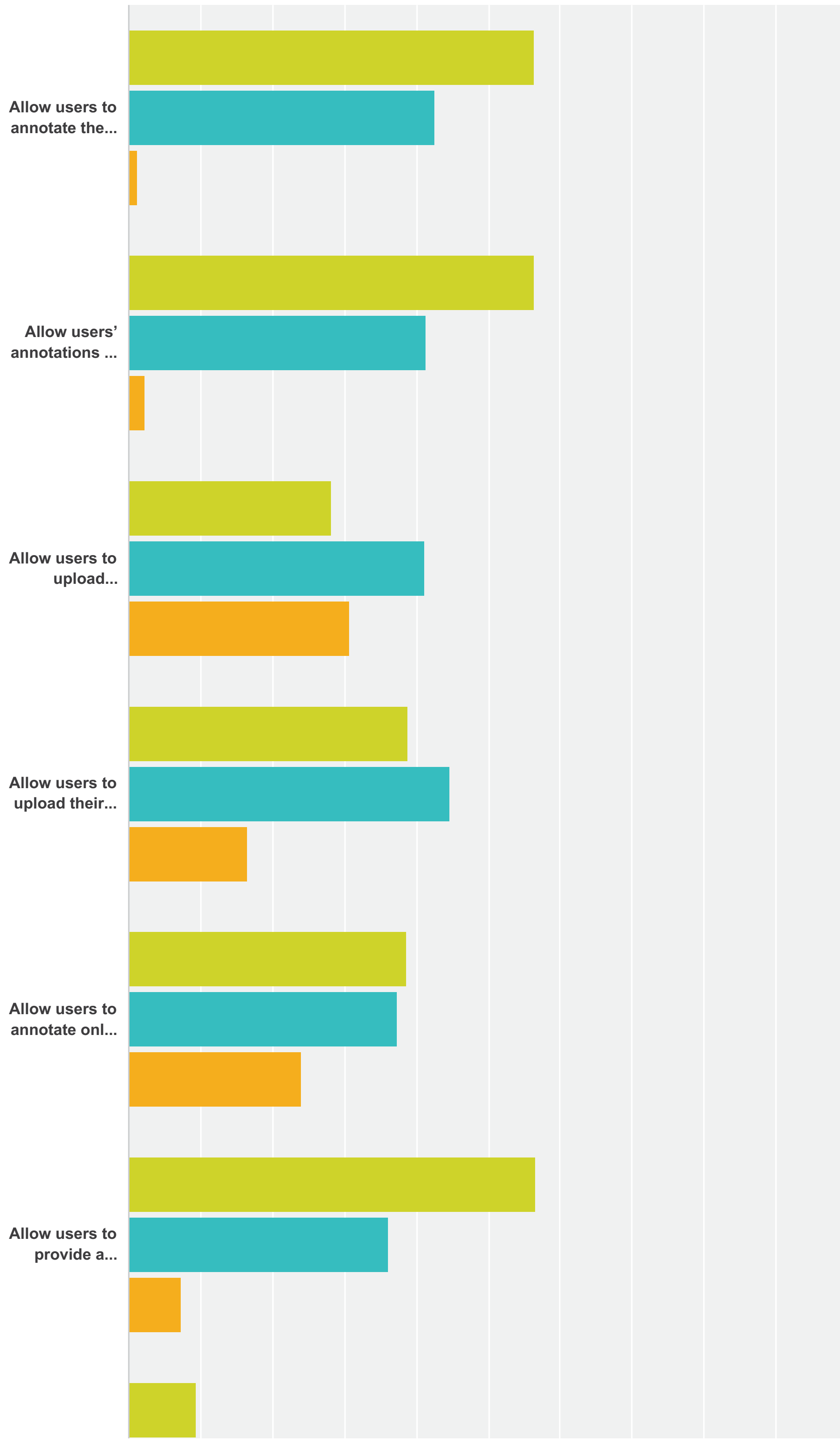

| | very important | important | not important | Total |
|---|---|---|---|---|
| Annotate the quality of the data (identification) | **79.55%** 70 | **20.45%** 18 | **0.00%** 0 | 88 |
| Annotate the quality of the data (georeference) | **84.88%** 73 | **15.12%** 13 | **0.00%** 0 | 86 |
| Do not allow access to data regarded as “low quality” by your experts | **16.28%** 14 | **43.02%** 37 | **40.70%** 35 | 86 |
| Spend the time and money required to correct/update data (taxonomically /geographically) as per observations provided by users | **55.17%** 48 | **43.68%** 38 | **1.15%** 1 | 87 |

# GBIF Data Fitness For Use in Modelling Ecological Niches and Geographic Distributions

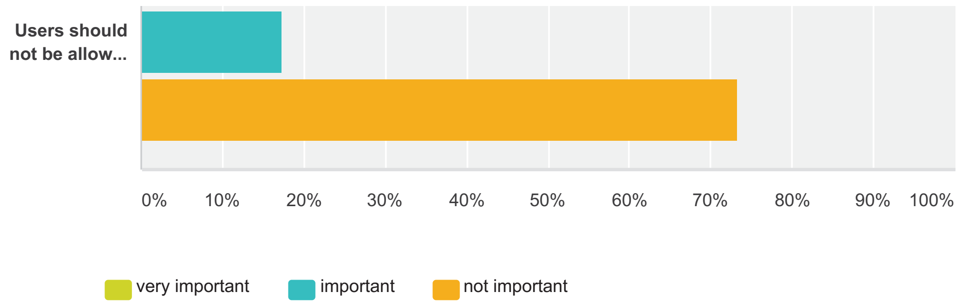
| # | Other (or comments) | Date |
|---|---|---|
| 1 | . | 10/20/2015 5:06 PM |
| 2 | VertNet has made considerable headway towards many of these issues, e.g. bad data can be flagged by users directly to data providers | 10/19/2015 4:05 PM |
| 3 | I think it would be better for biodiversity reserach if there would be a way for data providers to receive credit for providing data that are actively used by the scientific community. The best data collectors are not always those who publish in the best journals, while those publishing in the highest profile journals may merely be using the data that were collected by these valuable data providers. Scientific advancement needs both. If GBIF could quantify downloads/usage of data on behalf of each data provider (if it does not already do so already), this woudl be a useful way to show the value of the data that can be used by data providers to account for their valuable contributions, and a metric that they could show to funding organisations, institutional leaders assessing theri job performance, etc. | 10/18/2015 10:22 AM |
| 4 | All data should be accessible, however a measure of data quality is critical. Researchers should be able to make their own restrictions on data quality. | 10/13/2015 5:20 PM |
| 5 | Please improvise on data download formats and visualization of spatial data online. | 9/8/2015 10:57 AM |
| 6 | Of general importance: What proportion of data is still not georeferenced or digitalized in those collections? | 9/4/2015 8:27 PM |
| 7 | Provide extra field with the contact data of a taxonomic authority on the species. | 9/4/2015 6:27 PM |
| 8 | The third option would only make sense if the data returned to the original data providers. | 9/4/2015 5:06 PM |
| 9 | Consistency in use of field names and their data | 9/4/2015 2:49 PM |
| 10 | Rigorous grouping system in different categories for sources (Citizen science projects, Museum data, peer reviewed studies) and show a data quality index. | 9/4/2015 2:05 PM |
| 11 | full description of collection location and habitat where possible, this is so useful when it is provided | 9/4/2015 12:15 PM |
| 12 | Regarding the last point. Obviously it is important to correct the errors. However, several errors are identified daily by users and yet, they are not reported back to GBIF hence not leading to corrections in the databases. So, rather than spending time and money correcting directly the data it would be useful to create a system of feec backs between users and the GBIF (and data providers) allowing corrections of errors in data as they are identified. | 9/4/2015 11:07 AM |
| 13 | I think the key is to allow crowd-sourcing cleaning and annotation of data in very easy ways | 8/13/2015 1:13 PM |
| 14 | definition of low quality data depends on the question | 7/13/2015 4:38 PM |

## Q29 Which of these suggestions would you consider important to transmit to the GBIF TECHNICAL PERSONNEL to improve the operation of GBIF.org and the interface?

Answered: 87 Skipped: 48



# GBIF Data Fitness For Use in Modelling Ecological Niches and Geographic Distributions

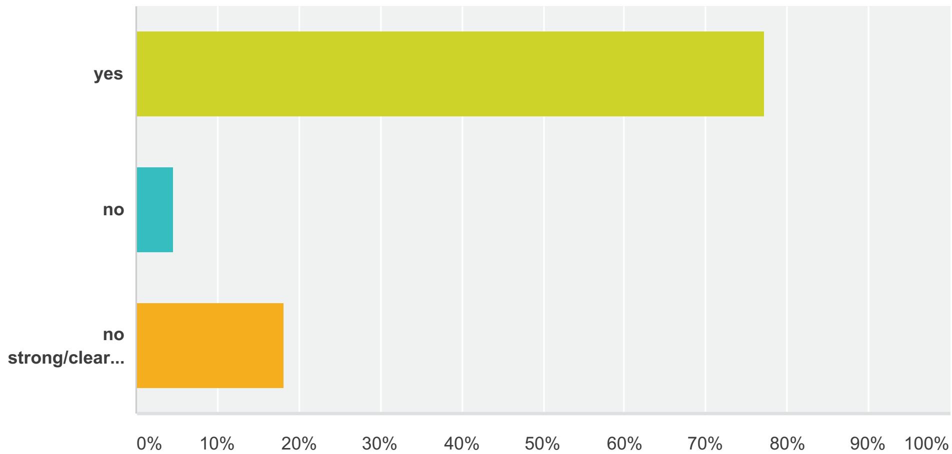



| | very important | important | not important | Total |
|---|---|---|---|---|
| Allow users to annotate the data (pinpoint problems, ask questions) | **56.32%** 49 | **42.53%** 37 | **1.15%** 1 | 87 |
| Allow users’ annotations to be transmitted automatically to the DATA PROVIDERS. | **56.32%** 49 | **41.38%** 36 | **2.30%** 2 | 87 |
| Allow users to upload “improved” versions of a given download. | **28.24%** 24 | **41.18%** 35 | **30.59%** 26 | 85 |
| Allow users to upload their own data | **38.82%** 33 | **44.71%** 38 | **16.47%** 14 | 85 |
| Allow users to annotate only certain parts of the data (taxonomy, geography) | **38.55%** 32 | **37.35%** 31 | **24.10%** 20 | 83 |
| Allow users to provide a quality, or fit-for-use flag for data | **56.63%** 47 | **36.14%** 30 | **7.23%** 6 | 83 |
| Users should not be allowed to comment on data | **9.33%** 7 | **17.33%** 13 | **73.33%** 55 | 75 |

| # | Other (or comments) | Date |
|---|---|---|
| 1 | . | 10/20/2015 5:06 PM |
| 2 | different permission based on users experience/publications | 9/15/2015 7:31 AM |
| 3 | keep GBIF as it is please, do NOT Doctor around much!! | 9/9/2015 1:32 AM |
| 4 | To a large degree, allowing users to provide feedback to data from GBIF would allow these data sets improve their quality much faster. | 9/4/2015 8:27 PM |
| 5 | Cool idea! | 9/4/2015 6:39 PM |
| 6 | The original data should be maintained, but updated versions will be welcomed. A description of all the spoted "bugs" should be included too. | 9/4/2015 6:27 PM |
| 7 | Here, I think that users should be able to tell GBIF when they find errors, but direct annotations should not be allowed. This is, the only one allowed to annotate is the GBIF crew | 9/4/2015 5:08 PM |
| 8 | I think that users could change the data provided this does not change the original information. In other words, it would be the user's choice to use the original data or secondary data. | 9/4/2015 5:06 PM |
| 9 | Edits by users should be monitored or check for errors as well by a third party | 9/4/2015 2:49 PM |
| 10 | this could be useful as long as users cannot modify original records | 9/4/2015 12:15 PM |
| 11 | I think that users don't own the data, so they should have limited capability of modifying it. This should be the responsibility of the providers. However, users should be able to pass their comments to the providers | 8/14/2015 3:50 PM |
| 12 | I think these options would be nice, but not sure I would trust them from any "user" - depends on who these users are. | 8/5/2015 7:49 PM |
| 13 | annotate/comment the data can be important and useful but is not necessarily easy to implement | 7/13/2015 4:38 PM |
| 14 | Consider how to link with initiatives such as iNaturalist, which is collecting data from 'citizen science' type initiatives... this particularly relates to the oint about allowing users to upload their own data | 7/10/2015 12:19 AM |

## Q30 Do you think the field would be well served to have a single online repository; archive for point occurrence data published in peer-review journals (similar to Genbank for genetic sequences)?

Answered: 88 Skipped: 47



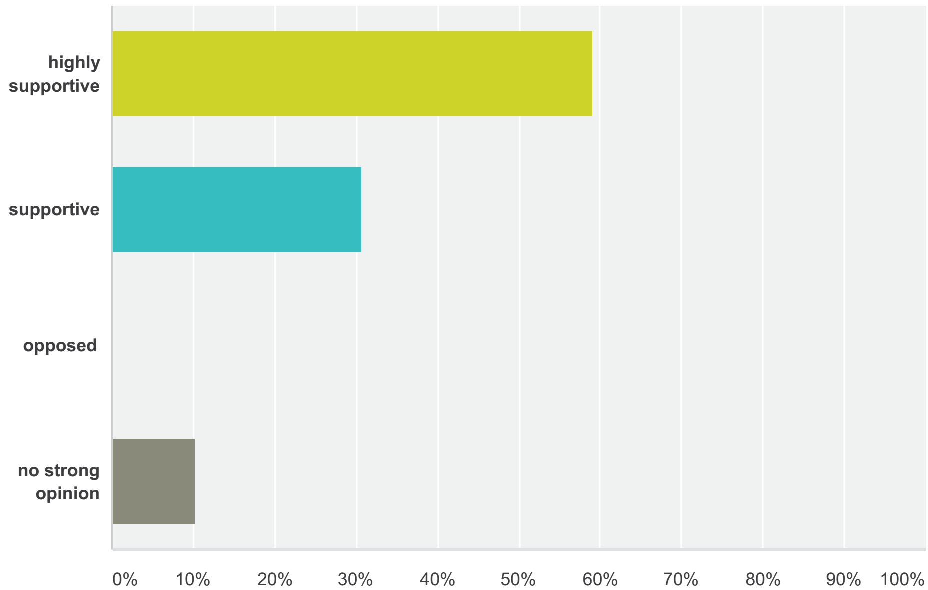

| Answer Choices | Responses | |
|---|---|---|
| yes | **77.27%** | 68 |
| no | **4.55%** | 4 |
| no strong/clear opinion | **18.18%** | 16 |
| **Total** | | **88** |

| # | Comments or remarks | Date |
|---|---|---|
| 1 | For herpetological species in the Tropics, I currently have to aggregate data from 3+ occurrance datasets. Having one, comprehensive suppository that is managed and quality-checked (as Genbank is) would be exptremely useful. | 10/24/2015 10:48 PM |
| 2 | . | 10/20/2015 5:06 PM |
| 3 | It seems to me that it is fruitless to attempt the proliferation of related by somewhat distinct efforts to serve the same data. More localized efforts (e.g. Atlas of Living Australia) enable developing much more functionality than would be possible at the global level. However, what would be useful is for all these parallel efforts to be kept track of in a centralized repository. | 10/19/2015 4:05 PM |
| 4 | That would be amazing. | 10/12/2015 8:20 PM |
| 5 | It is a very important point | 9/9/2015 6:02 PM |
| 6 | But 8t has to be done right. It can't be the output of a small group if individuals. | 9/4/2015 11:48 PM |
| 7 | Absolutely! | 9/4/2015 6:39 PM |
| 8 | Definitively, yes! One that is well-supported with funds to hire experts in taxonomy and georreference, for example. | 9/4/2015 4:59 PM |
| 9 | Although Genbank is useful, many sequences lacks metadata, e.g., specimen identification, georeferenced data, and localization of collected specimens | 9/4/2015 2:25 PM |
| 10 | Improvements needed not separatism | 9/4/2015 2:05 PM |
| 11 | however I think many researchers already simply check gbif without digging for more, personally I found about 1/4 of my records through other sources | 9/4/2015 12:15 PM |
| 12 | These files should be clearly annotated as to how the authors filtered their data | 8/14/2015 3:50 PM |

| | | |
|---|---|---|
| 13 | Yes journals should strongly encourage data deposition of locality data; but scientists should also consider working directly in web-based platforms, rather than waiting until all work is done and THEN submit the data. The Swedish database system for such records is a very good example: https://www.artportalen.se/ (although I'm not sure this is already fully implemented, it will be the 'Artportalen' 2) | 8/13/2015 1:13 PM |
| 14 | the biodiversity community should have already pushed towards this | 7/13/2015 4:38 PM |
| 15 | very much so! | 7/8/2015 11:33 AM |
| 16 | There so many purposes even for occurrence data that I am not entirely convinced that only one repository might be the best option. | 7/7/2015 5:38 PM |

## Q31 Please provide your thoughts on GBIF being a repository/archive for point occurrence data published in peer-review journals.

Answered: 88 Skipped: 47




| Answer Choices | Responses | |
|---|---|---|
| highly supportive | **59.09%** | 52 |
| supportive | **30.68%** | 27 |
| opposed | **0.00%** | 0 |
| no strong opinion | **10.23%** | 9 |
| **Total** | | **88** |

| # | Comments or remarks | Date |
|---|---|---|
| 1 | my experience with GBIF has thus far not been amazing in terms of accurate occurrence records. Better quality-control options would increase favorability of the service in my mind. | 10/24/2015 10:48 PM |
| 2 | . | 10/20/2015 5:06 PM |
| 3 | I'd want to know what points were from PRed lit and the citations | 10/12/2015 8:20 PM |
| 4 | What would describe certain quality for any modeling purpose in a repository like this? If this is restrictive for publication may or may not be advisable--consider discussing. | 9/4/2015 8:27 PM |
| 5 | I think that it may be easier to associate with another database that is already receiving submissions from journals (such as dryad), since authors won't stick to having to submitted to again another database. | 9/4/2015 5:08 PM |
| 6 | What is the sustainability of such effort? Cost? | 7/21/2015 11:58 PM |
| 7 | As long as data quality is annotated | 7/13/2015 5:51 PM |
| 8 | Current Darwin core is somehow limiting | 7/13/2015 4:38 PM |
| 9 | this is absolutely needed! The only alternative is Dryad and it's a mess! | 7/8/2015 11:33 AM |

## Q32 Please make (at least three) free text suggestions to USERS of GBIF.org

Answered: 88 Skipped: 47

| # | Responses | Date |
| --- | --- | --- |
| 1 | explore, use, feedback | 10/28/2015 7:42 PM |
| 2 | 1. Check for georeferencing information 2. Engage in quality control practices for all data 3. Search for alternate scientific names, if the species is frequently misspelled or has recently been changed. | 10/24/2015 10:48 PM |
| 3 | I don´t understand the question | 10/22/2015 2:46 PM |
| 4 | I am not extensively familiar with GBIF. | 10/20/2015 5:06 PM |
| 5 | - Always perform data quality checks, especially in terms of georeferencing errors - Make use of as much data as possible, not only for the focal species but for all closely-related/associated species - Perform sensitivity analyses nearly each step of the data filtering process; what choices most affect the output of models? | 10/19/2015 4:05 PM |
| 6 | I would suggest that GBIF provide on its download page suggestions for quick quality assurance of the data. Suggestions might include: 1.Check for lat/long reversal errors 2. Check for taxonomic errors 3. Be aware of all biases in the data that could undermine your study. | 10/18/2015 10:22 AM |
| 7 | No user account required Allow for download of georeferenced data only Provide easier to interpret metadata Make file sizes smaller/more manageable The site is 'down' a lot, so that could be handled better | 10/16/2015 5:03 PM |
| 8 | Great product. Very useful. What is a free text suggestion? | 10/15/2015 9:02 PM |
| 9 | 1) Applied Spatial Data Analysis with R 2) Dynamic Occupancy Models in R-package, unmarked 3) spacetime: Spatio-Temporal Data in R | 10/15/2015 8:33 PM |
| 10 | . | 10/14/2015 10:51 PM |
| 11 | I think that a such important database will need further improvement on the number of species records. | 10/14/2015 9:50 AM |
| 12 | 1). GBIF's website needs to be easier to use/access. It's currently poor software 2). GBIF's Integrated Publishing Toolkit is quite frankly terrible! It's difficult to set up and maintain, full of bugs, and very poorly documented. This software has been my worst experience with GBIF and is a strong disincentive for having data included on GBIF. 3). Ditch the login for access to data. | 10/13/2015 5:20 PM |
| 13 | Please, do not just download GBIF data and throw it into a model. | 10/13/2015 3:05 AM |
| 14 | Map everything first, check the dates | 10/12/2015 8:20 PM |
| 15 | I don't understand this question | 10/12/2015 6:17 PM |
| 16 | I dont have suggestions | 10/2/2015 7:30 PM |
| 17 | no thanks | 10/2/2015 3:22 AM |
| 18 | Some journals like the new Pensoft Biodiversity journal already bank occurrence data. So maybe it would be better to partner with them. One thing that bugs me is that data that are not specimen based are allowed. If you can't check the ID then the record is suspect. Even photographs and molecular data would be better than nothing. When I was looking at the GBIF records for the Asteraceae (Compositae) of Colombia we had a good species list and downloaded 1.2 million records. We ended up using 34,000. Some of these discarded records were species that had wider ranges than just Colombia-Ecuador-Venezuela (our reference area) but many of them were records of Colombian taxa that were weedy elsewhere or weeds in Colombia. There should be a way to identify indigenous plants from the area. | 9/23/2015 11:43 PM |
| 19 | Report discovered data issues corrections to the publisher of data, and make them clear in the methods section of your reports or publications; Always attribute data usage by correctly citing its use; Support open by publishing your datasets through GBIF | 9/16/2015 12:46 AM |
| 20 | 1 Always consider sampling effort and completeness before modelling, and include that information in the description of your methods and approach when publishing your results 2 Whe possible, give feedback to data providers, regarding data quality every time you identify or suspect a possible error, presenting detailed arguments for that 3 Be part of the data sharing community by also sharing your own data. | 9/15/2015 5:25 PM |
| 21 | - expand data collection to institutions of under sampled regions. - provide online filters in GBIF of coordinate precision before data download. - develop metrics of data reliability in terms of taxonomy and coordinate precision. | 9/15/2015 4:12 PM |
| 22 | assess spatial and temporal biases or likelihood of biases prior to use check for duplicates as in the same records submitted by different institutions develop long/lat data verification methods (plotting, point in polygone, accuracy threshold,...) use and verify species key for download in rgbif | 9/15/2015 7:31 AM |

| | | |
|---|---|---|
| 23 | Validate the points of georeferences | 9/10/2015 5:06 AM |
| 24 | I make suggestions directly to GBIF. Make an global analysis of the current crisis in taxonomy Without taxonomists does not exist GBIF with updated data Disclose in a serious way the Taxonomic impediment Work in scientific score for the data providers. | 9/9/2015 6:02 PM |
| 25 | -focus on transparent and repeatable science -promote more data publication and open access and open source -focus on conservation & sustainability the most, reduce the EU-centric and commercial approach to science please | 9/9/2015 1:32 AM |
| 26 | fgfg | 9/8/2015 8:14 PM |
| 27 | Doble check the information they download. Use an additional software to corroborate the georreferenced ocurrences. When looking for a specific species, make sure to check all the synonyms available to avoid exclusion of some data. | 9/8/2015 7:21 PM |
| 28 | EASY RAPID COMPLETE | 9/8/2015 2:54 PM |
| 29 | Source of data Data visualization Erroneous records | 9/8/2015 10:57 AM |
| 30 | - Taxonomic and geographical tools to comment and correct data. - Increase interaction between users. - Macro projects asking for new data. | 9/7/2015 11:48 AM |
| 31 | 1. A careful examination of all fields in downloaded records is mandatory. Summaries in Excel work well. 2. Probability of correct name match is foundational 3. Generating spatial and temporal views provide the best context 4. Examination of any assertions | 9/7/2015 1:27 AM |
| 32 | Be EXTREMELY careful of this data when you use it for modelling. Even when it is clean, the relationship to covariate stations can be very misleading. Beware of country level records, precisely geocoded to the county centroid. | 9/6/2015 2:16 AM |
| 33 | 1. Always clean the data. 2. Checking that the geographic coordinates agree with the localities of the records. 3. Using clean records of GBIF and other records obtained directly from museums and scientific papers. | 9/5/2015 10:46 PM |
| 34 | Check the origin of the data | 9/5/2015 2:34 PM |
| 35 | - cite the data and accession date - interact with gbif through API because this is reproducible - use gbif as a teaching resource | 9/4/2015 11:48 PM |
| 36 | To users making SDM or ENM using GBIF data I suggest: 1) Consider having a deep knowledge of target species distribution. 2) For marine species, revise the number of decimal degrees in the coordinate reference. Maybe records are offshore just because of coarse resolution. 3) For marine species, use (when available) the depth field to validate the accuracy of the record. This can sometimes resolve issues concerning coarse resolutions. | 9/4/2015 11:18 PM |
| 37 | -better interaction techinical support x data providers to improve data quality more efficiently -better documentation of taxonomy concept -better way to geographycally clip the occurance data desired -an easy way to combine georreference with location as labeled. This is important to avoid excluding registers because of lat x long in swaped field or N/S W/E swaped. | 9/4/2015 9:30 PM |
| 38 | -Learn R/Perl/JSON -Do not trust GBIF until the data has been cleaned/annotated -Re-check your data BEFORE submitting it | 9/4/2015 9:18 PM |
| 39 | - If used for modeling, check for spatial/taxonomic/environmental consistency according to known distributions or regionalizations (in each of these spaces). - Apply spatial filters to even out your sample and "homogeneously represent environments." - Don't include dubious data. | 9/4/2015 8:27 PM |
| 40 | don't know | 9/4/2015 8:00 PM |
| 41 | Quitar las redundancias en los datos datos georeferenciados de alta calidad y exactitud mayor difusión | 9/4/2015 8:00 PM |
| 42 | References, photos, filter to subfamilies | 9/4/2015 7:08 PM |
| 43 | Be smart user. Make use of tools available to clean and verify data. Support using and improving GBIF. | 9/4/2015 6:41 PM |
| 44 | ALWAYS inspect data downloads from GBIF because they often contain no information on georeferencing accuracy. Even when georeferencing precision is reported and all else seems well, inspect the final map of records to determine if locations are accurate. Taxonomy is inconsistent. | 9/4/2015 6:39 PM |
| 45 | -Please verify the nomenclature of records and contact the pertinent taxon authority in case of discrepancies. -In case known locations appear geographically misplaced in GBIF data, reposition these data with you own georreferences. -Please provide feedback when you find nomenclatural /geographic errors in the data (include each record ID). | 9/4/2015 6:27 PM |
| 46 | Add an option to help identify duplicated records Add collection method Include absence data when possible | 9/4/2015 5:41 PM |
| 47 | -check occurrence data before using (coordinates in wrong places) -check data for taxonomical errors -check data for synonymies | 9/4/2015 5:27 PM |
| 48 | N/A | 9/4/2015 5:21 PM |

| | | |
|---|---|---|
| 49 | - check for duplicates - consider "precision" - check for distribution of points (search for biases) and complement GBIF with other sources if possible | 9/4/2015 5:08 PM |
| 50 | 1) Configure fields and search filter are fantastic, but the access to these windows is a little bit hidden, few people use that. 2) I think it lack in the interface the possibility of submitting a list of species. I do this via an R package, which works very well, but this resource, again, is used by a few people. 3) I think users could change the data provided if this does not change the original information. In other words, it would be the user's choice to use the original data or secondary data. | 9/4/2015 5:06 PM |
| 51 | At moment, the most important job is the taxonomic data update | 9/4/2015 5:04 PM |
| 52 | Daniel Paiva Silva , Bruno Vilela , Diogo Borges Provete , Marcial Quiroga | 9/4/2015 4:59 PM |
| 53 | thanks fort teh info | 9/4/2015 4:54 PM |
| 54 | n.a | 9/4/2015 4:12 PM |
| 55 | 1. Always keep the whole database. 2. Identify the records with more uncertainty (coordinates, taxa, etc.) 3. Map the bias of your species. | 9/4/2015 4:04 PM |
| 56 | 1) Revise carefully the georeference of each record, especially if the resolution used to model is fine grained 2) Records located well outside of known distributional range, or located in a different ecoregion of that known for the species being modeled, should be treated with care, as can be a species misidentification 3) Contact with the data provider if a suspect of species misidentification exists | 9/4/2015 3:57 PM |
| 57 | Be carefull with the data (ckeck for taxonomic and geographic mistakes), correct bias and inform to the local GBIF nodes of the detected mistakes. | 9/4/2015 3:48 PM |
| 58 | - never model blindly, always look at the data and cure them; - be careful with pseudo-absence assumptions; | 9/4/2015 2:55 PM |
| 59 | Always clean your data before using it, i.e. do not trust original data Incorporate other sources of data into your study, i.e. do not use ONLY GBIF data Consult with taxonomy specialists to check both species determinations and occurrences | 9/4/2015 2:49 PM |
| 60 | As many of my concerns with GBIF are on the taxonomic issues, I would suggest that GBIF would request to data providers information about accuracy in taxonomic identification (e.g., how were identified specimens, using primary literature, photographies, field guides). Knowing this, I think would be possible to suggest for each users what records might be potentially problematic and not discarding many of them before mapping. | 9/4/2015 2:25 PM |
| 61 | - Treat the data with caution! - Quality is almost always better than quantity - Fill / correct all the metadata ;) - Biodiversity data is nice | 9/4/2015 2:05 PM |
| 62 | Double check synonyms for species prior to downloading data. | 9/4/2015 12:24 PM |
| 63 | Always dig into the data thproughly before using it, always resample and address spatial bias where possible, Google Street view is so useful for checking the accuracy of records combined with location descriptions. | 9/4/2015 12:15 PM |
| 64 | Always check the quality of the geo-referenced material, many of the presence records may be wrongly geo-referenced. For species distribution models always pay attention to the date the data was collected Pay attention to the sampling bias and sampling effort from samples extracted from GBIF | 9/4/2015 11:44 AM |
| 65 | Check errors prior to using the data Report errors back so that other users benefit from the time you spent data cleaning Be careful to adapt your analytical approach to the characteristics of the data | 9/4/2015 11:07 AM |
| 66 | sampling bias annotations accuracy of georeference coordinates notation of the occurrence (i.e. was it collected from a botanical garden or from an actual ecological survey) | 9/4/2015 10:35 AM |
| 67 | Curate your own data. Curate your own data. Curate your own data. | 8/14/2015 3:50 PM |
| 68 | - assess your data, but don't be afraid of using it (all knowledge is provisional) - make your data freely accessible and citable - cite your data providers and the tools you use | 8/13/2015 1:13 PM |
| 69 | (1) Recognize that the data have significant issues & should be used carefully and evaluated critically. (2) Using GBIF data with models (e.g., ENMs, SDMs) compounds data issues. (3) Consider building models using only those species with at least ~10 points per predictor varaible. | 8/5/2015 7:49 PM |
| 70 | See comment in Question 26. | 7/29/2015 6:13 PM |
| 71 | Question unclear. | 7/29/2015 5:34 PM |
| 72 | 1 2 3. | 7/27/2015 1:41 PM |
| 73 | No suggestions at this time. | 7/25/2015 1:03 AM |
| 74 | 1) Always explore your data visually (plot maps, plot data in environmental space, per species, etc.) 2) Whenever possible, consult with an expert on the group/region of the world 3) Always document the steps taken to clean the data, save the intermediate files, and publish the steps as Supplementary Materials | 7/21/2015 11:58 PM |

| | | |
|---|---|---|
| 75 | Beware of taxonomic inaccuracies. Beware of hugely erroneous (and easily catch-able) errors in georeferences. Keep in mind that the lack of a record in a particular place (or perceived changes over time) probably reflects sampling effort at that place/time, just as likely as it does the real absence of the species. | 7/19/2015 10:25 PM |
| 76 | Provide data quality fields. Mark dubious records Allow users interaction through annotation | 7/13/2015 5:51 PM |
| 77 | The level of data quality required depends on each research question. Be aware that taxonomy is by far the most complicated issue on GBIF data. One very interesting avenue of research worth to explore is how to cross reference GBIF data with other information to detect potential taxonomic, temporal and spatial errors. | 7/13/2015 4:38 PM |
| 78 | 1) Always check the data for taxonomic and geographic quality 2) Acknowledge bias 3) be conscientious about the data publishing process and all the parties involved in getting you free data and properly acknowledge them all. | 7/13/2015 4:36 PM |
| 79 | Suggestions to GBIF users (?). Sorry, I didn't understand what you're asking here. | 7/13/2015 3:49 PM |
| 80 | 1 2 3 | 7/12/2015 6:25 AM |
| 81 | . | 7/10/2015 12:19 AM |
| 82 | - - - | 7/8/2015 4:53 PM |
| 83 | I have the feeling that GBIF is a valuable source of occurrence information but that modern data mining techniques and user annotations should be used to evaluate and strengthen the quality of the data. If google can do it, gbif can do it! My feeling is also that a lot of effort has been done on the taxonomic side of gbif (I guess for historical reasons), but that much is still to do on the georeferencing side. A lot of existing data is not in GBIF for political or monetary reasons. Some resources of GBIF should be devoted to convince reluctant countries and institutions to join the movement. Using the argument that GBIF could become the official repository/archive for point occurrence data when publishing in peer-review journals, with clear and accessible references to the data providers, would help. A program of data digitization of herbarium data in some countries (e.g. eastern european countries and ex-USSR countries which have a solid floristic tradition) would be relatively cheap and efficient. | 7/8/2015 11:33 AM |
| 84 | 1. Understand that GBIF serves data that have been made available by owners as owners see fit--not all data will be there and often it'll be PART of what the owner holds. 2. Understand that data served is NOT a random sample--it is a selection of the universe of data. That said, if that selection is what interests you (your chosen universe) then you may actually sample that selection. 3. If you discover a pattern, bias, or systematic error that may affect the representativeness of the data, let it be known by others. 4. If your cleanup procedures satisfy you, try to protocolize them and let them be known. | 7/8/2015 1:27 AM |
| 85 | - Analyse occurrence distributions along environmental variables to detect outliers. - Analyse occurrence distributions along lat/long to detect outliers - Always visualize the occurrence record from GBIF and compare them to know extent of occurrence (even if coarse) for data consistency. | 7/7/2015 5:38 PM |
| 86 | NA | 7/7/2015 5:25 PM |
| 87 | ? | 7/7/2015 5:04 PM |
| 88 | provide maps of survey efforts provide maps of survey zones by experts even if occurrence data are not available | 7/4/2015 6:47 PM |

# ANNEX A: ACKNOWLEDGEMENT

GBIF thanks the following researchers and users of biodiversity data for providing valuable input.

Achille Hounkpevi
Anna Carter
Arthur Chapman
Angela Cuervo
Haidy Rojas
Brian Rolek
Adam Smith
Andrea Thomer
Fabián Viteri
Jesus Aguirre Gutierrez
Mariano Altamiranda
Diego Alvarado-Serrano
Robert Anderson
Andre Andrade
Luciano Anjos
Alexandre Antonelli
Miguel Araújo
Jorge Ari Noriega
Arturo H. Ariño
Jorge Assis
Lee Belbin
Joana Boavida-Portugal
Olivier Broennimann
Kevin Burgio
Camilo Calderon
Vanderlei Canhos
Carola Canon
Pedro Cardoso
Monica Carlsen
Nora Castaneda
Rosane Collevatti
Anna Csergo
Jose Don De Alban
Renato De Giovanni
Andrew Dennhardt
Yoannis Dguez
Maria Elena
Anahi Espindola
Sergio Estay
Natalia Estrada Carmona
Kathryn Everson
Kenneth Feeley
Miguel Fernandez
Marinez Ferreira de Siqueira
Rui Figueira
Ana Filipe
Matt Fitzpatrick
Janet Franklin
Vicki Funk
César Garcia
Carol Garzon-Lopez
Marius Gilbert
Gustavo Gonzalez
Duane Griffin
Moisès Guardiola
Adriana Guerrero
José Guerrero
Eliécer E. Gutiérrez
Jessica Hellmann
Robert Hijmans
Ned Horning
Falk Huettmann
Raul Jimenez Rosenberg
Erica Johnson
Martin Jung
Riddhika Kalle
Roland Kays
Fabrice Kentatchime
Wanja Kinuthia
Karen Kiser
Darren Kriticos
Joona Lehtomäki
Andres Lira-Noriega
Jorge Lobo
Glenn Lowe
Maria Crisitna Mangano
Crystal McMichael
Glenda Mendieta-Leiva
Cory Merow
Carsten Meyer
Fabian Michelangeli
James Mickley
Daniel Miranda
Daniel Miranda-Esquivel
David A Moo Llanes
Martha Munoz
Ryan Nagelkirk
Babak Naimi
Elkin Noguera
Julia Nordlund

Jennifer Pannell
Monica Papes
Andrea Paz
Richard Pearson
Jade Phillips
Timothée Poisot
Erika M.D Porporato
Magdalena Precoma de la Mora
Luis Quijada
Chanel Rampartab
Giovanni Rapacciuolo
Hugo Rebelo
Maria Cristina Rios-Blanco
Danielle Rivera
Philipp Robeck
Ariel Rodriguez
Leidy Viviana Romero Alarcon
Hannu Saarenmaa
David Sánchez-Fernández
Hartanto Sanjaya
Babu Santhanakrishnan
Stephanie Shaffer
Jorge Sigala
Thomaz Sinani
Jorge Soberon
Hugo Suárez
Allison Sussman
Wilfried Thuiller
Morgan Tingley
Jen Tinsman
Ricardo Torres
Christian Vazquez
Greta C. Vega
Julián Velasco
Jorge Velasquez
Jaris Veneros
Dan Warren
Jennifer Weaver
Darice Westphal
Marten Winter
Mary S. Wisz
Veronica Zamora
Gabriela Zuquim