Survey on georeferencing of Natural History Collections

December 2019 – February 2020

Arnald Marcer, Quentin Groom, Elspeth Haston, Francesc Uribe

Mobilise Action
Mobilising Data, Policies and Experts in Scientific Collections

Survey on georeferencing of Natural History Collections

SECTION		Nr of questions
1	INSTITUTION	3
2	SPECIMEN HOLDINGS	3
3	COLLECTION CHARACTERISTICS	6
4	GEOREFERENCING EFFORT	3
5	COLLECTIONS, PROTOCOLS, RESOURCES AND TOOLS	10
6	GEOREFERENCING UNCERTAINTY	6
7	PUBLIC DATA REPOSITORIES	3
8	FEEDBACK	5

[Survey questions](#)

Example questions

Mobilise Action
Mobilising Data, Policies and Experts in Scientific Collections

Survey on georeferencing of Natural History Collections

1. INSTITUTION

* 1. Where is your Institution located?

○ Africa   ○ Australia and Oceania   ○ North America (Canada to Panama)

○ Asia   ○ Europe   ○ South America

Specify country

[   ]

* 2. Institution type.

☐ Museum

☐ Botanical Garden

☐ Other (please specify)

email

Survey on georeferencing of Natural History Collections

Dear Natural History Museum / Botanical Garden staff member,

First, apologies for cross-posting if this is the case.

We would like to ask you for a few minutes of your time to answer this online survey regarding georeferencing practices at Natural History Museums and Botanical Gardens. We ask you to answer it in relation to the collection at your institution you may be responsible for, either directly or indirectly. We will also appreciate it if you can circulate this link to other people you may know who may answer it for other collections.

This survey is an activity of the Working Group 2 of the EU MOBILISE Cost Action. You can take a look at what the cost action is about at: https://www.mobilise-action.eu. The MOBILISE Cost Action is under the umbrella of DiSSCo, the Distributed System of Scientific Collections, which is an EU research infrastructure for natural science collections (https://www.dissco.eu).

This survey is being sent to Botanical Gardens and Natural History Museums around the world for research purposes. Its results will help to take a picture of the current state of georeferencing of natural history collections. Results will be presented at a MOBILISE workshop which will take place in Warsaw, Poland, in February 2020. The survey is anonymous but, we will happily send the results to you if you opt for it in the survey.

You can start at the end of this message with the first question.

We very much appreciate and thank you in advance for your time,

Sincerely,

Arnald Marcer [1], Quentin Groom [2], Elspeth Haston [3] and Francesc Uribe [4]
[1] CREAF, Spain
[2] Meise Botanic Garden, Belgium
[3] Royal Botanic Garden Edinburgh, UK
[4] Museu de Ciències Naturals de Barcelona, Spain

Legal note:

Please be informed that your personal data (only name and work email in this survey) are protected under European and Spanish regulations on personal data protection (Regulation (EU) 2016/679, of 27 April 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) and Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales). The CENTRE DE RECERCA ECOLÒGICA I APLICACIONS FORESTALS (CREAF) is liable for data handling.

Your contact data have been collected from two public access databases: Botanic Gardens Conservation International (http://www.bgci.org), GRSciColl: The Global Registry of Scientific Collections (https://www.gbif.org/grscicoll); or provided by CETAF (https://cetaf.org).

The sole aim of handling these data is to send you a single email asking for participation in this survey conducted by CREAF [Survey on georeferencing activities at institutions holding Natural History Collections]. Additionally, if you give your consent and provide your contact data in the corresponding question in the survey, we will use it to send you back the results of the survey when finished.

These data will be kept only for the duration of this particular survey. They will not be handled to third parties except if bound by a compulsory legal obligation.

At any moment, you may ask for access to your personal data, for modification, opposition, cancelation, portability and to be forgotten. For these matters, you only need to send an email to dpo@creaf.uab.cat. You may also claim your rights in front of the Catalan Data Protection Authority.

In answering this survey I agree with CREAF's privacy policy (http://www.creaf.cat/privacy-policy).
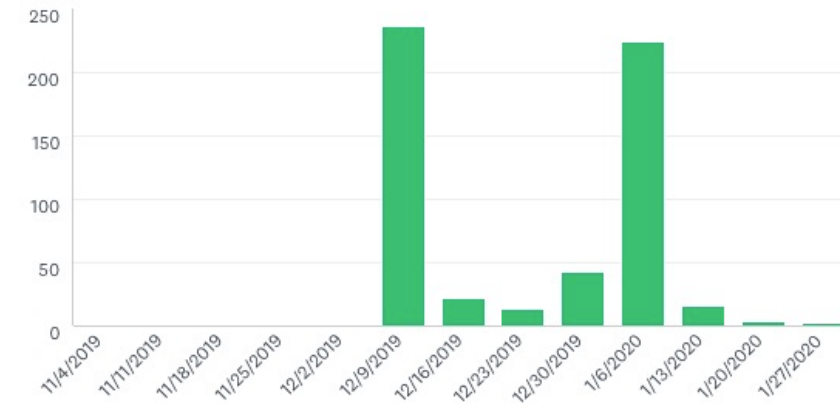
Where is your Institution located?

○ Africa

○ Asia

○ Australia and Oceania

○ Europe

○ North America (Canada to Panama)

○ South America

Survey summary

- Started on December 9th, 2020
- Will end February 29th, 2020
- Used SurveyMonkey
- 39 questions, average time to complete entire survey: 11' 25"
- Direct contacts obtained from
  - CETAF
  - Botanic Gardens Conservation International
  - GRSciColl: The Global Registry of Scientific Collections
- 4270 direct contacts, some mailing lists (e.g. iDigBio members, NHCOLL-L) and some person-to-person ones.
  - 552 answers till February 5th, 2020 (12.9 %)
  - Completion rate: 57%
- Over 200 respondents interested in final report

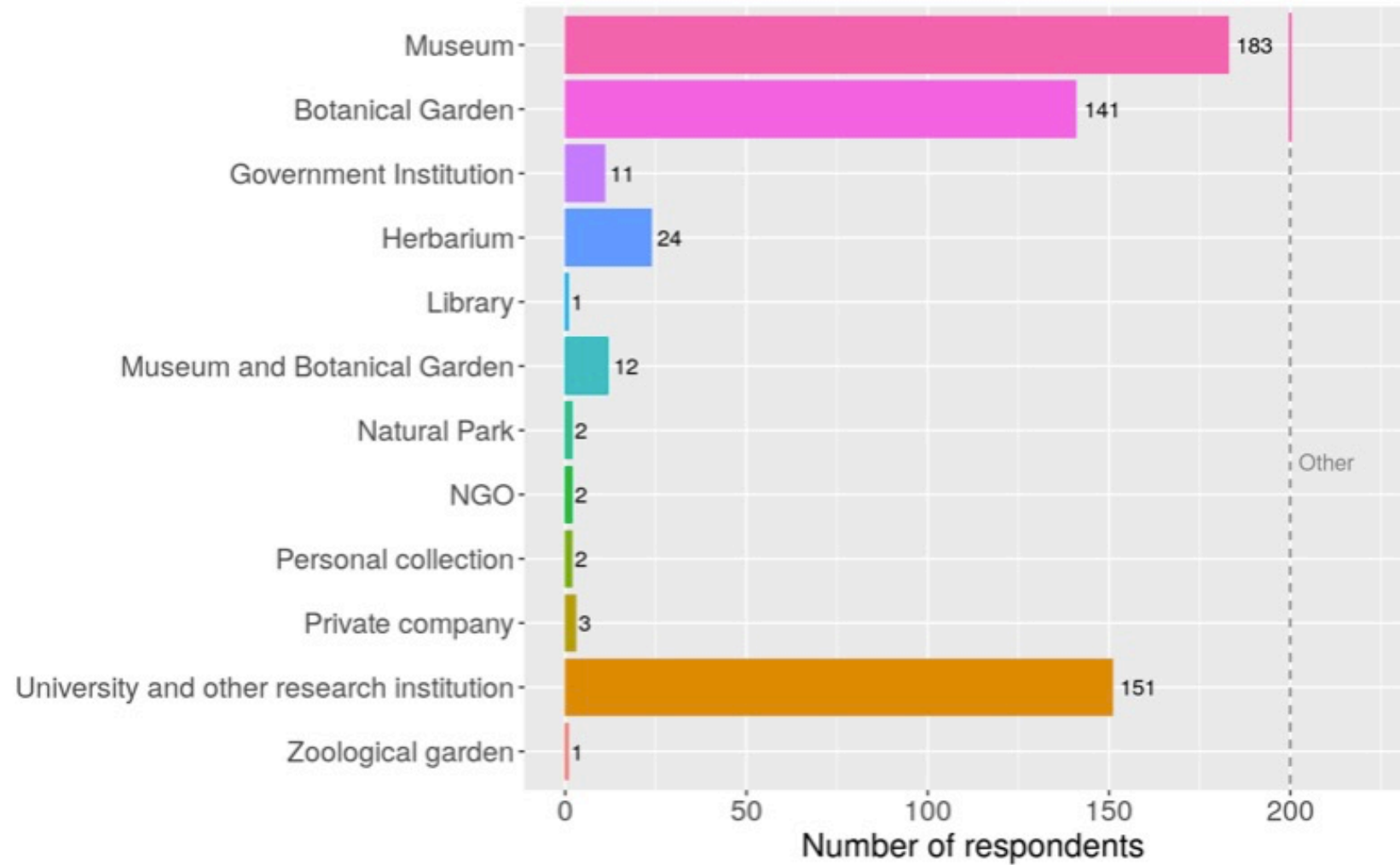# Q1: Where is your institution located?
Specify country
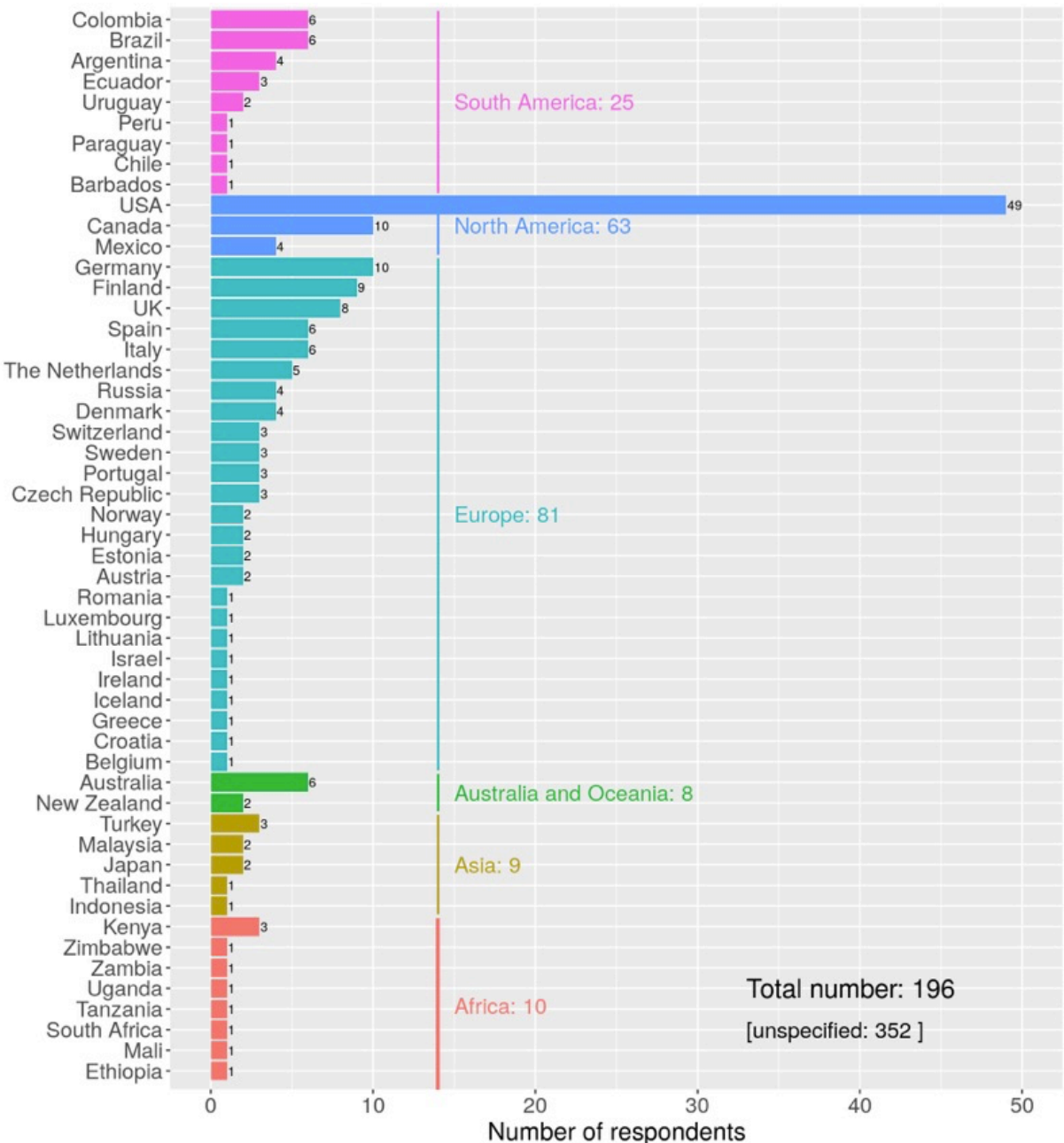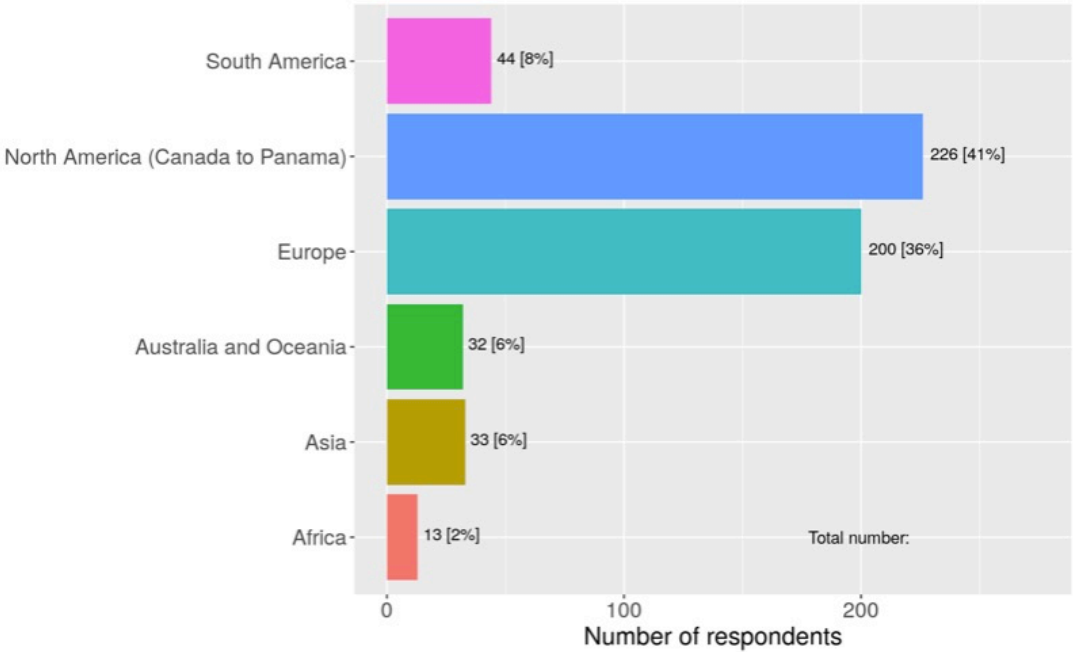
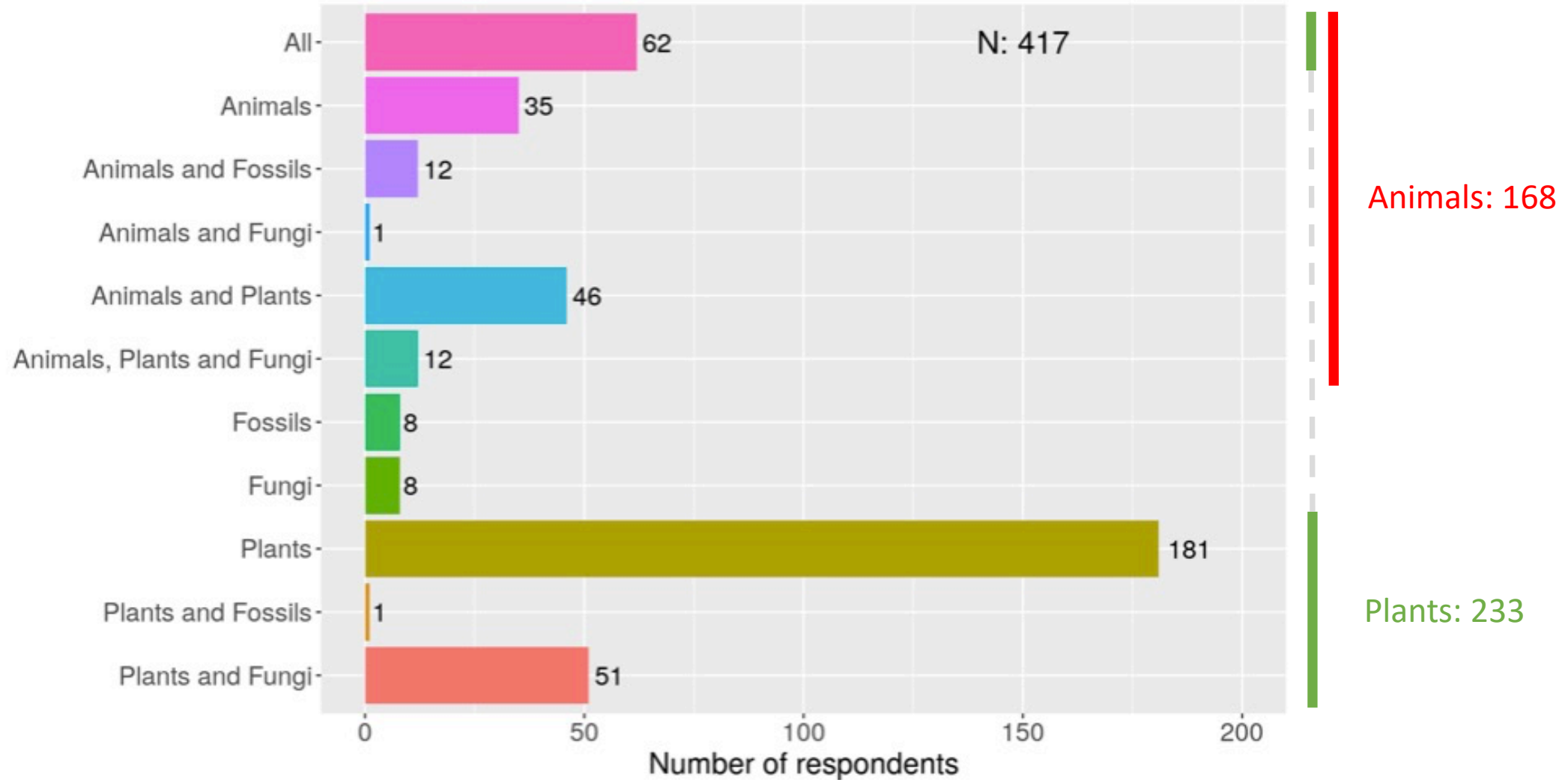## Q2: Institution type

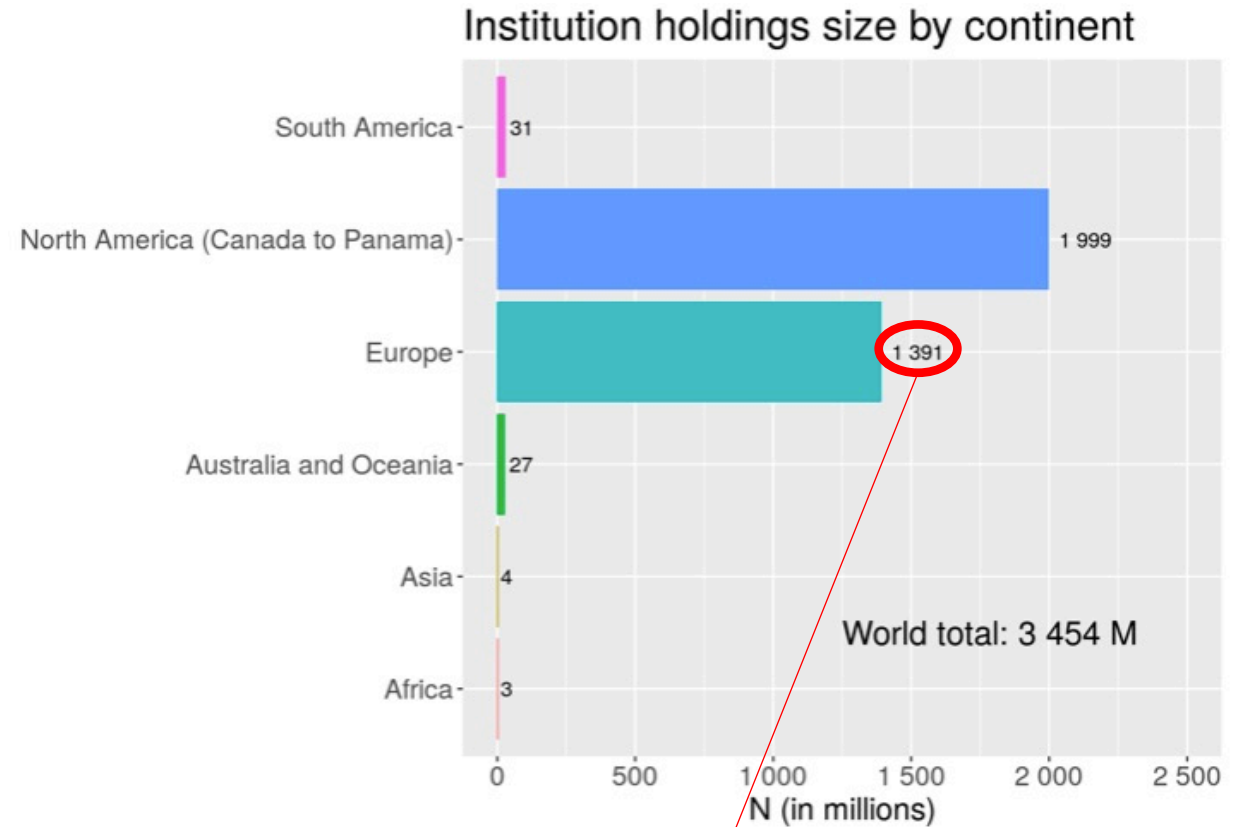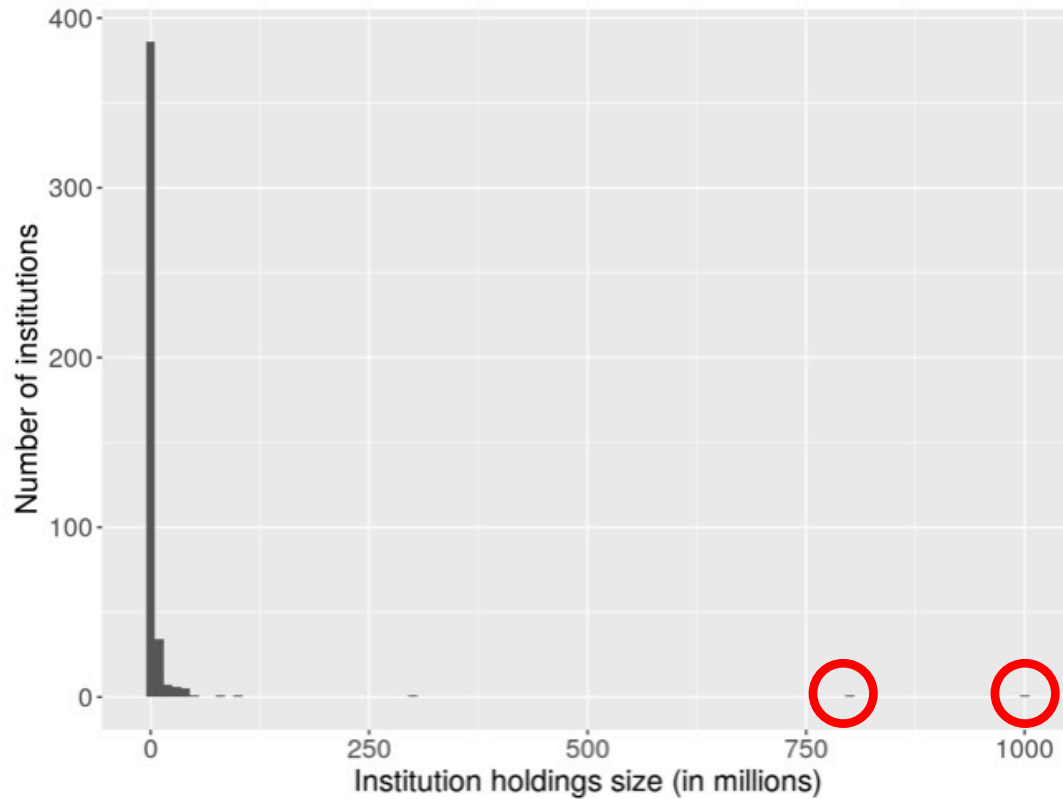**Q4**: What kind of collections does your institution host? (e.g. animals, plants, fungi, etc.)

(please write a whole number with no thousand separators, e.g. 1000000)

Ariño (2010) estimates maximum of 2.1 billion
CETAF's collections in 2018 around 1.5 billion (Smith et al., 2018)

Q5: continued …: For those respondents which did give the institution name (size 1M and above)

Legend: ■ Africa ■ Asia ■ Australia and Oceania ■ Europe ■ North America (Canada to Panama) ■ South America

Institution	Value
Naturalis	42 000 000
Natural History Museum of Los Angeles County	35 000 000
Staatliche Naturwissenschaftliche Sammlungen Bayerns, SNSB IT Center	30 000 000
Florida Museum of Natural History	23 500 000
Harvard Museum of Comparative Zoology	21 000 000
Field Museum of Natural History, Chicago	17 000 000
University of Michigan Museum of Zoology	15 000 000
National Museum Prague	15 000 000
Carnegie Museum of Natural History	13 000 000
KU Biodiversity Institute	11 500 000
University of Illinois Urbana-Champaign	11 000 000
Swedish Museum of Natural History	11 000 000
State Museum of Natural History Stuttgart	11 000 000
Museu de Zoologia, Universidade de São Paulo	11 000 000
Royal Ontario Museum	10 000 000
Natural History Museum, University of Florence	10 000 000
Biodiversity Institute, University of Kansas	10 000 000

Q7: Please enter the collection for which you are submitting your response.
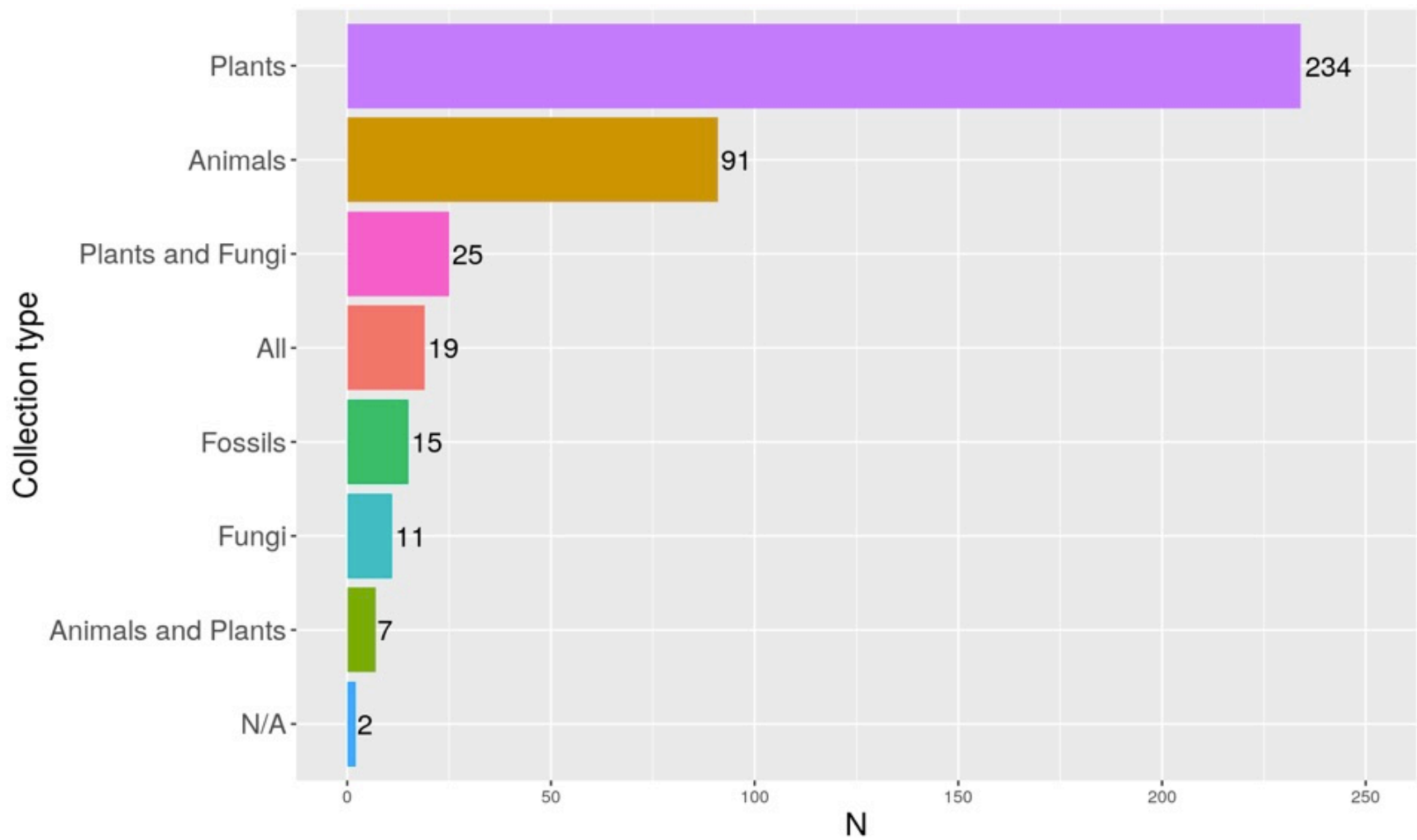
Q8: What is your role within your institution?
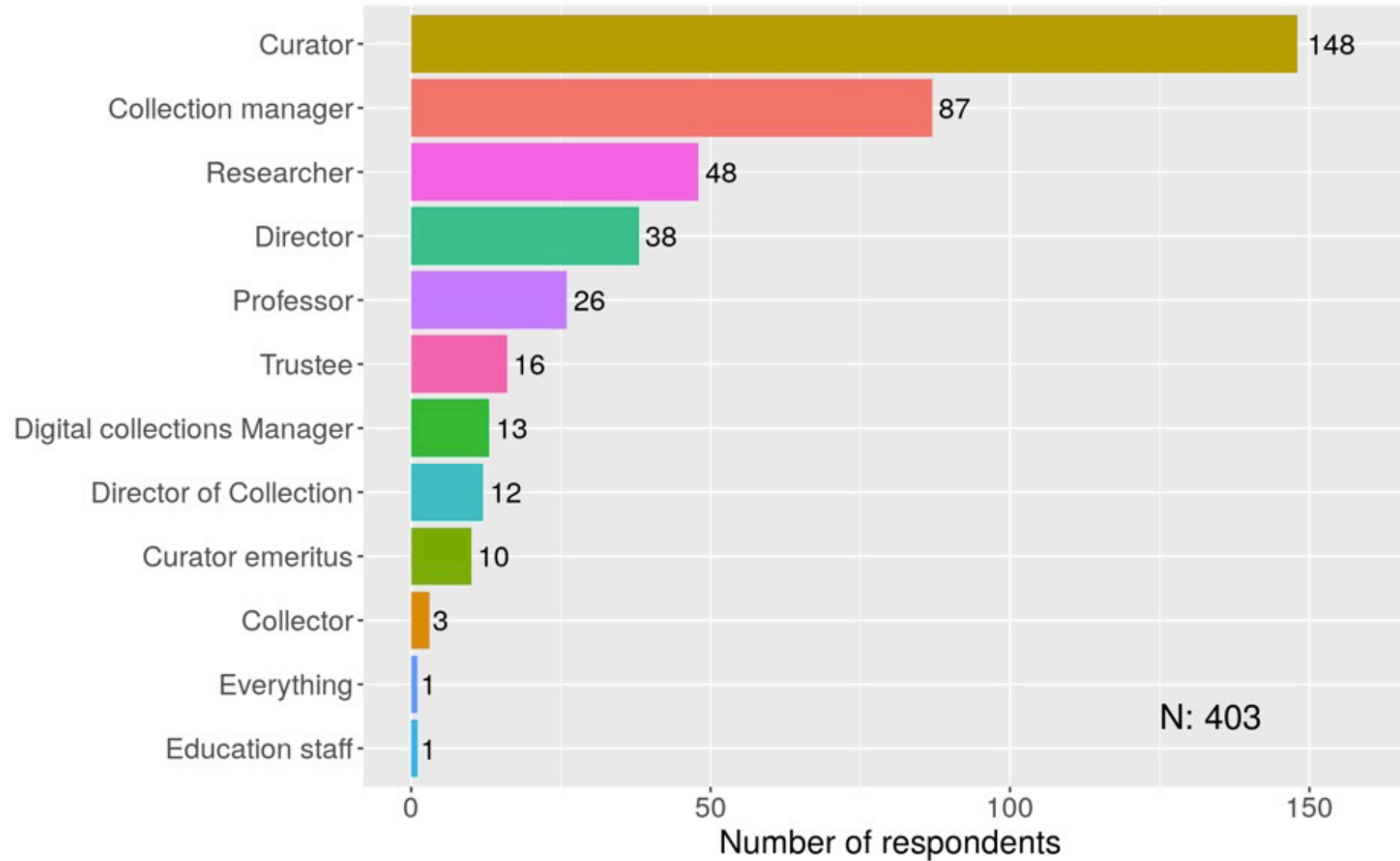
Q11: What is the size of this collection (number of specimens)?

Total number: 1 304 274 220

(Current GBIF holdings for preserved and fossil specimens ~ 178M)

+ two more collections of 80M and 800M
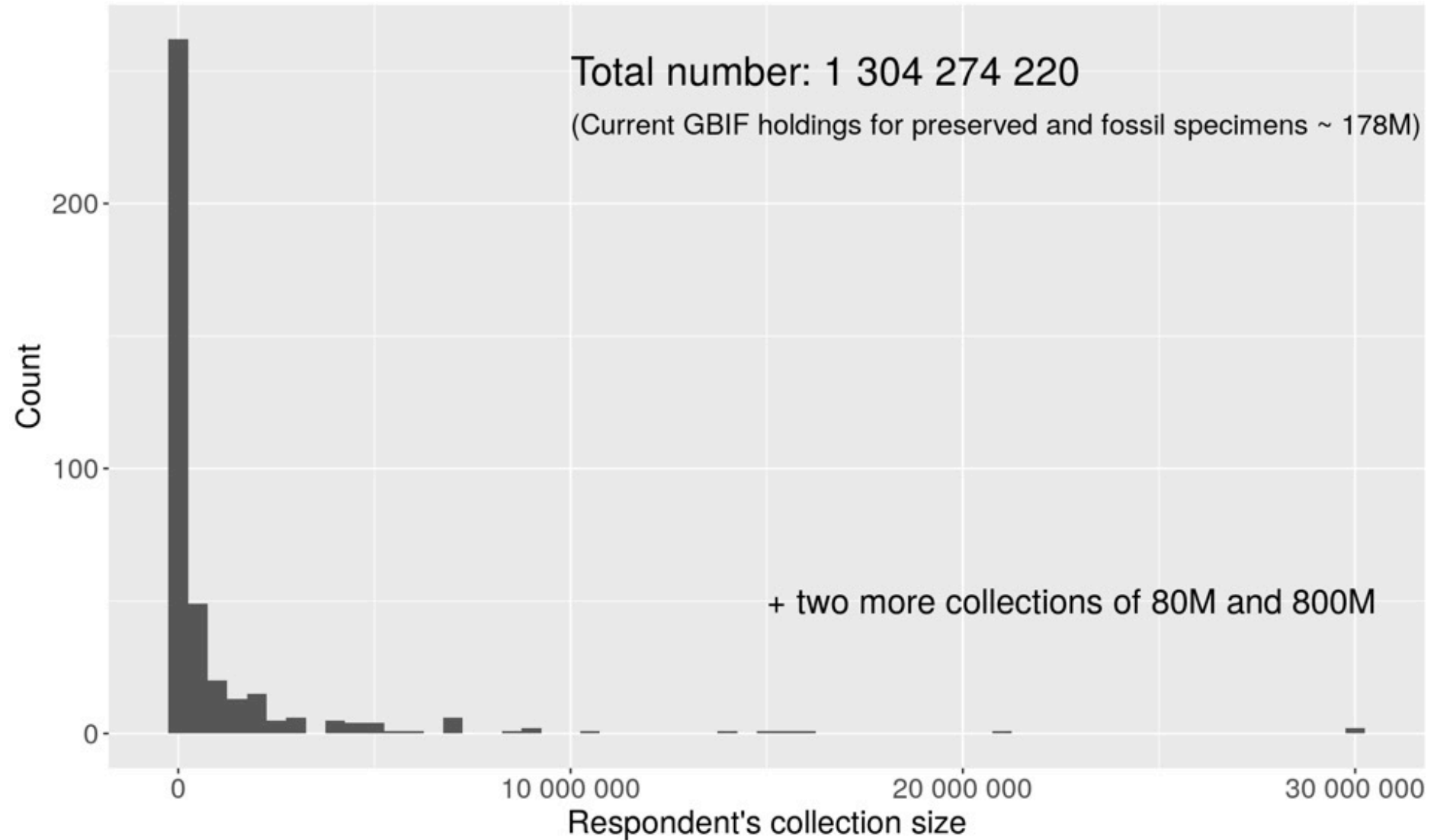
Q12: Is this collection digitized?
i.e. the specimens exist as a digital record, not necessarily including scan and media files
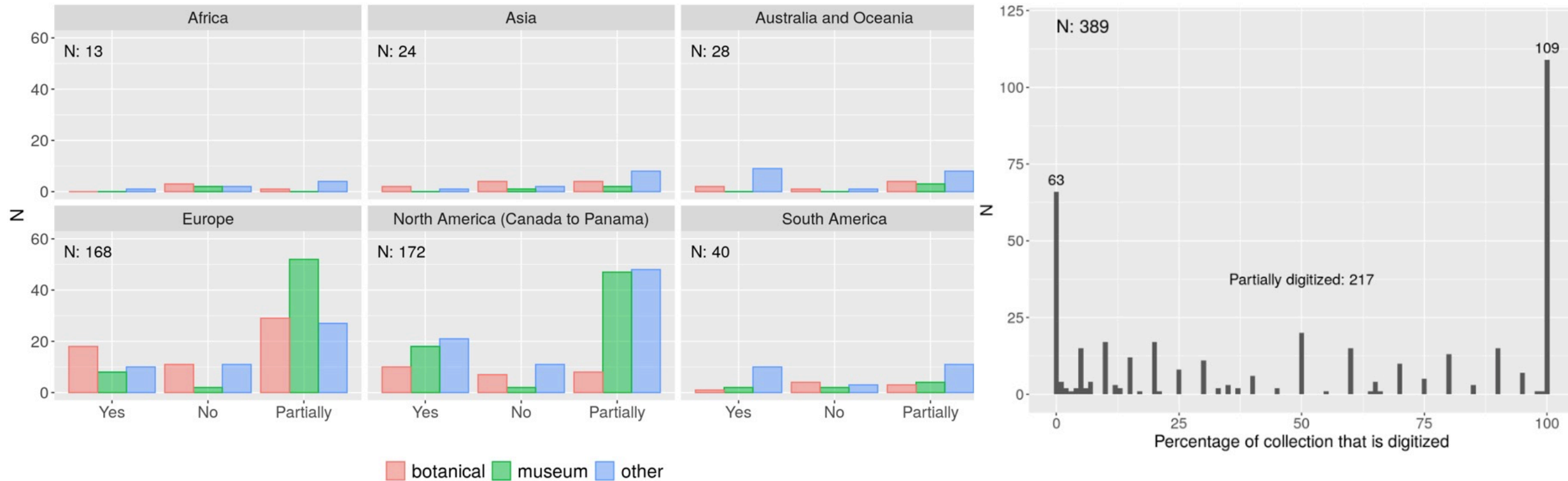
Q13: Approximately, what percentage of them are georeferenced?
In reference to the total collection, not only the digitized part
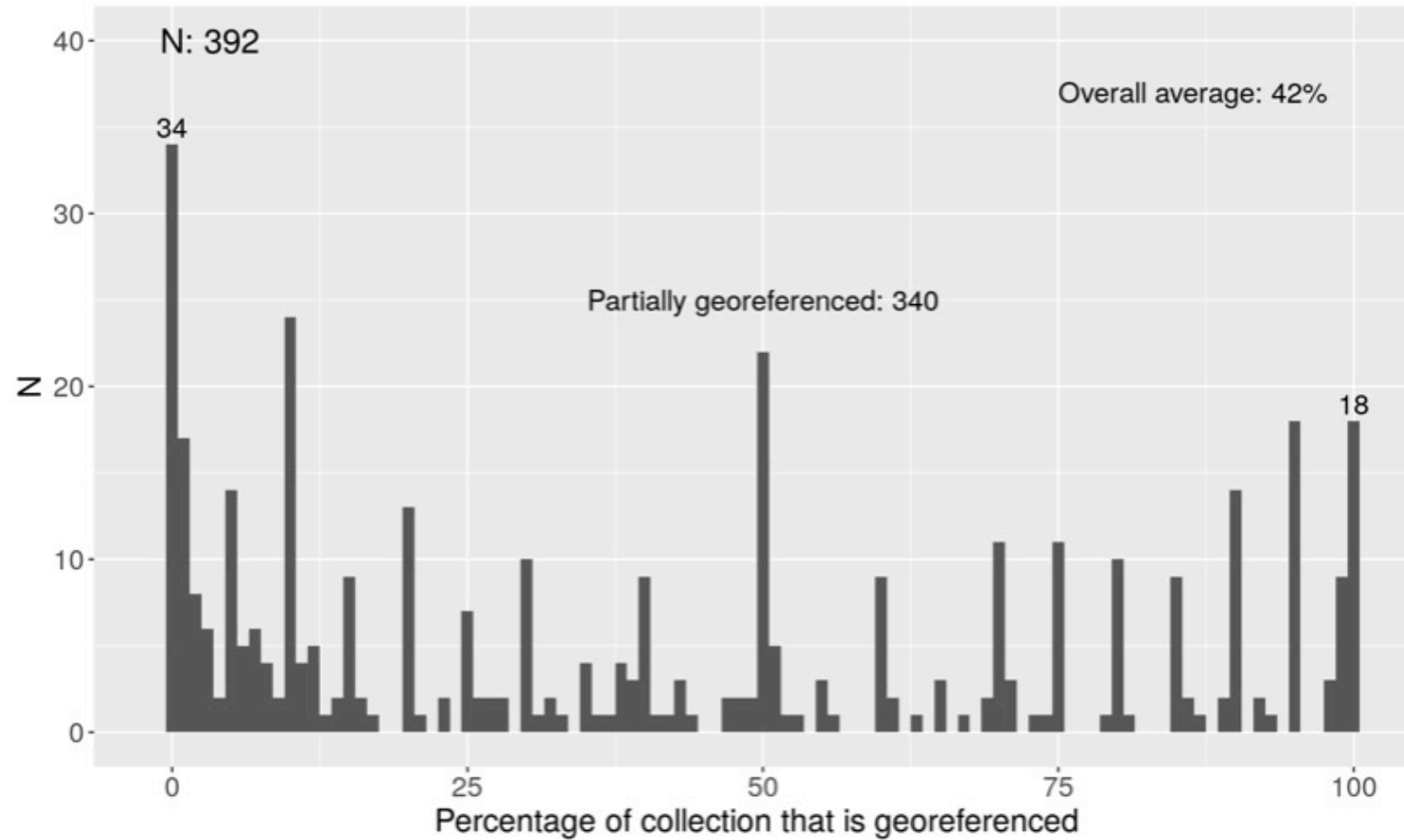
Q16: Which protocol do you follow to estimate coordinates from textual description of sites?
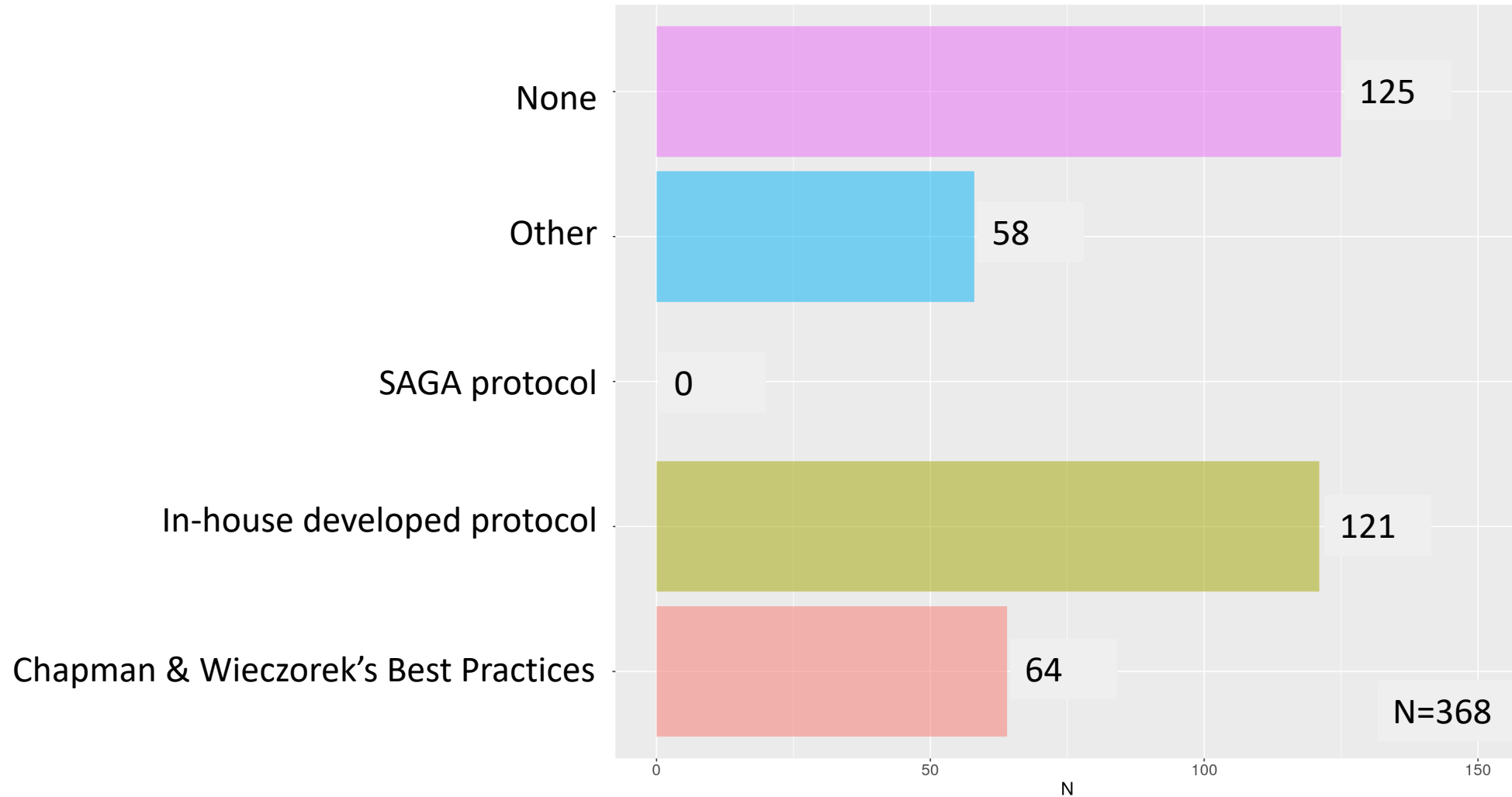
Q20: When you need to estimate coordiantes from textual information, which tools do you use?

Tool	N
Other	101
Marineregions.org	10
Wikidata	26
Geolocate	108
Getty Thesaurus	30
GeoNames	56
Bing maps	11
Google maps/Earth	275
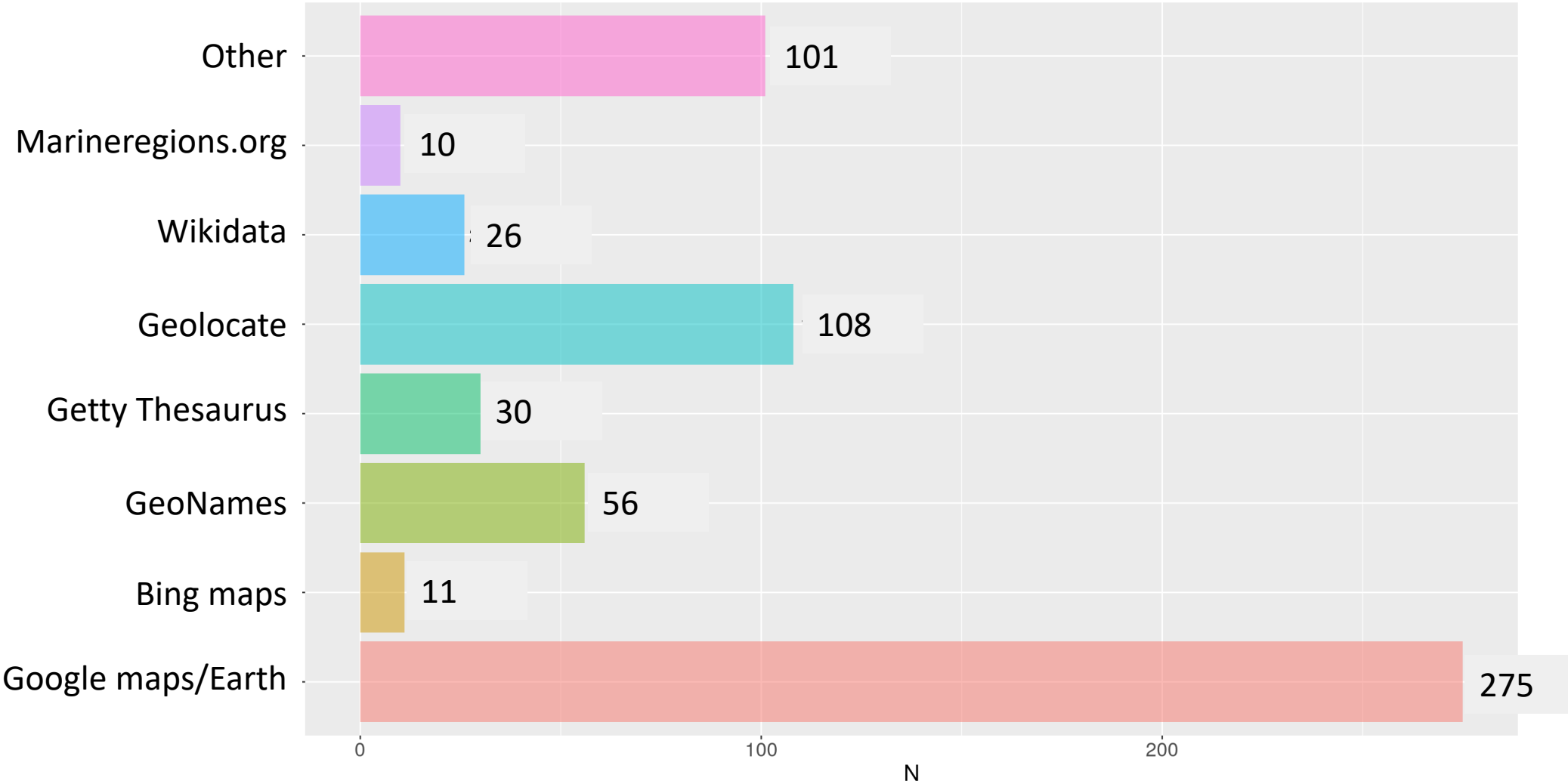
Q23: Which software do you use for managing your georeferenced site names database? If none, please say so. (e.g. custom application, PostgreSQL, Microsoft Access, etc.)

Responses: 324

Did not respond: 16

Legend:
- None
- Custom application
- Data format
- Data Management Software
- Geographic Tool
- Biodiversity Content Management System
- Collection Management Software

Software	N
None	110
MS Access	41
Specify	29
Custom application	21
MS Excel	18
File Maker	12
PostGreSQL	9
Symbiota	8
EMu	8
Arctos	8
QGIS	5
Oracle	5
MySQL	5
IrisBG	5
BRAHMS	5
BG-BASE	5
ArcGIS	4
MS SQL Server	3
Diversity Workbench	2
TMS Collections	1
SEINet	1
PlutoF	1
PastPerfect	1
MUSE	1
MariaDB	1
MapSource	1
MapInfo	1
jabot	1
GRIN-Global	1
Google Earth	1
GeoLocate	1
Fulcrum	1
FoxPro	1
earthcape	1
CALM	1
BIOTA	1
Bespoke	1
Autocad	1
ASDF	1
Ad-Lib	1

Q26: Do you report coordinates uncertainty?
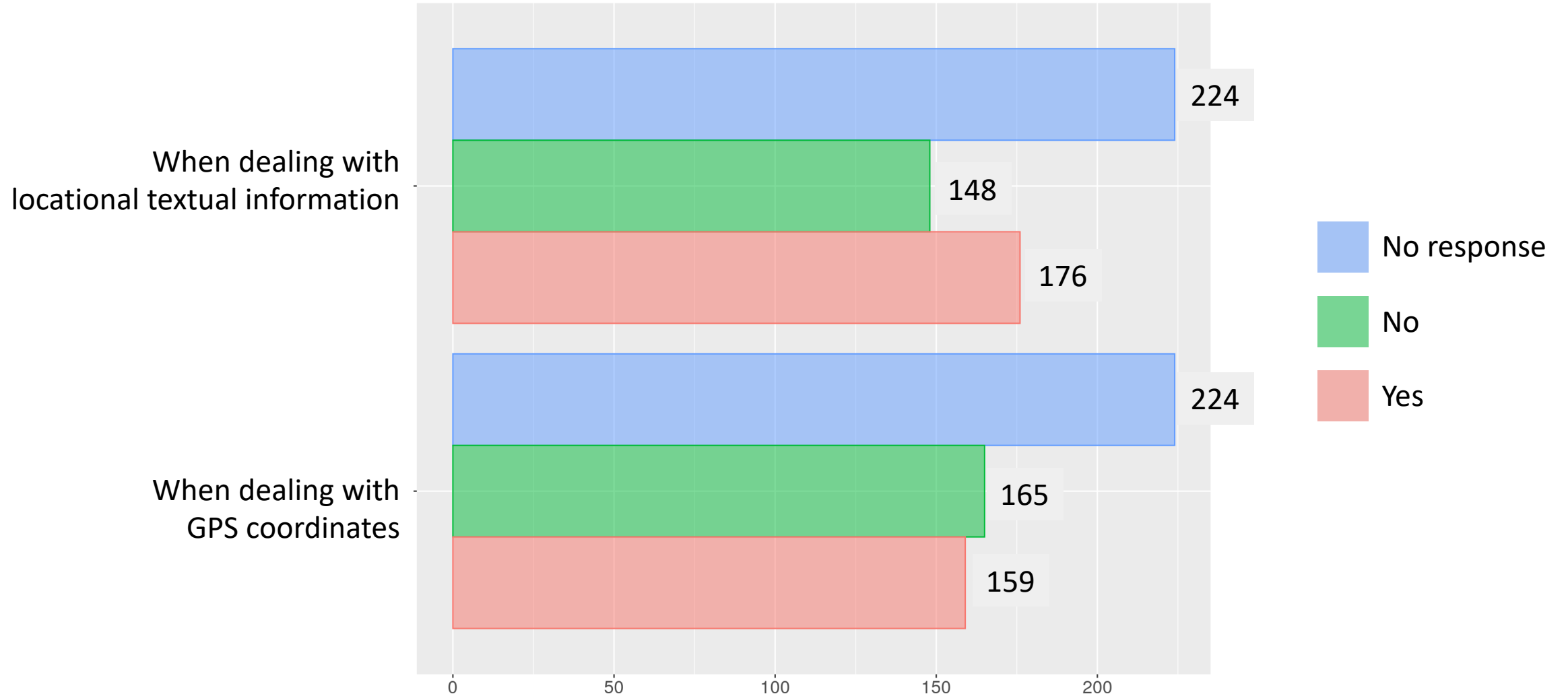
Q27: Do you apply any decimal rounding when publishing
coordinates?
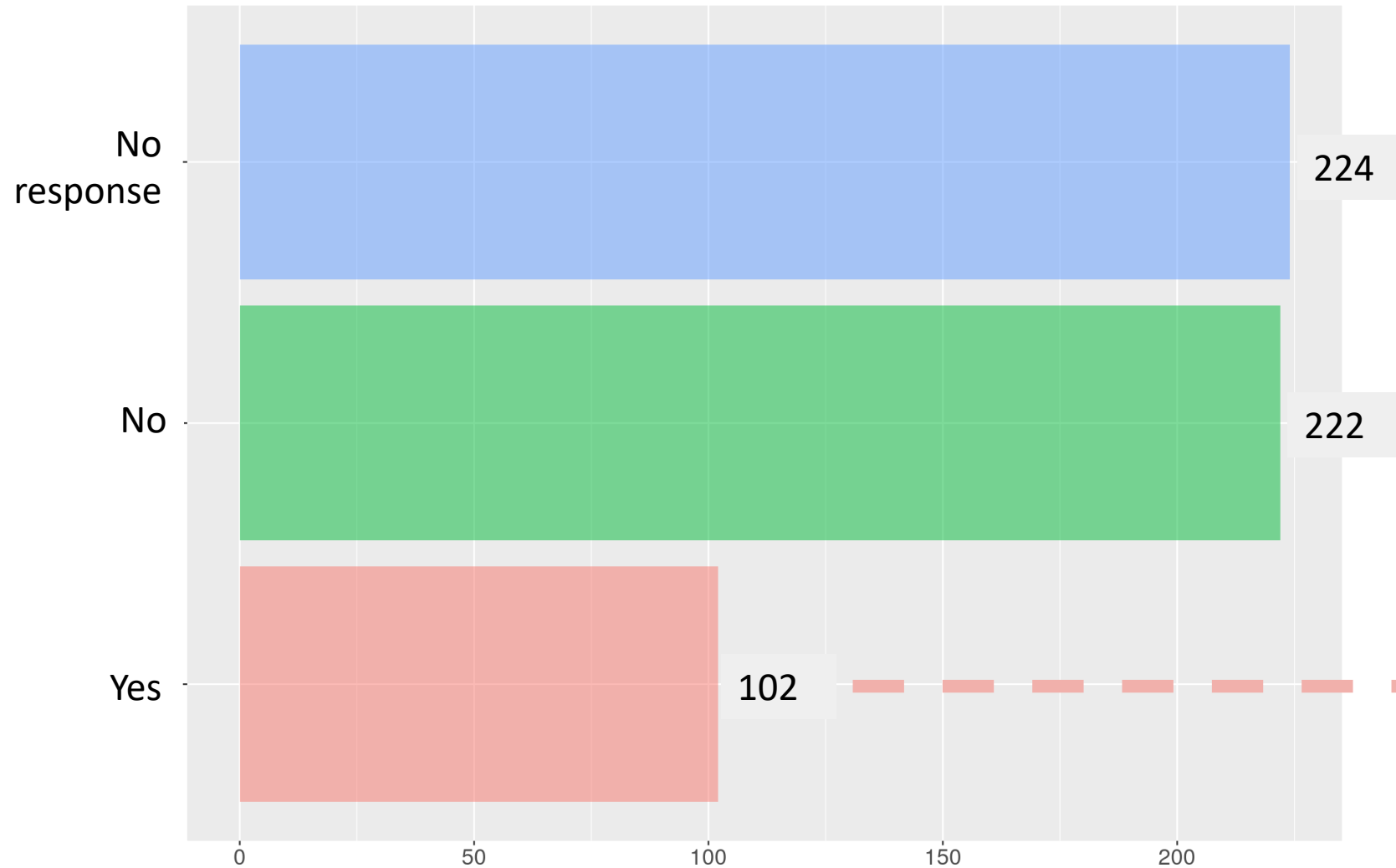
Q28 – Reasons why:

- 8%   data standardization

- 53%   technical reasons

- 39%   sensitive data

Q29: Do you consider it important to aggregate all factors contributing to uncertainty?
(e.g. unknown datum, rounded original units, projection changes, etc.)

Tools used
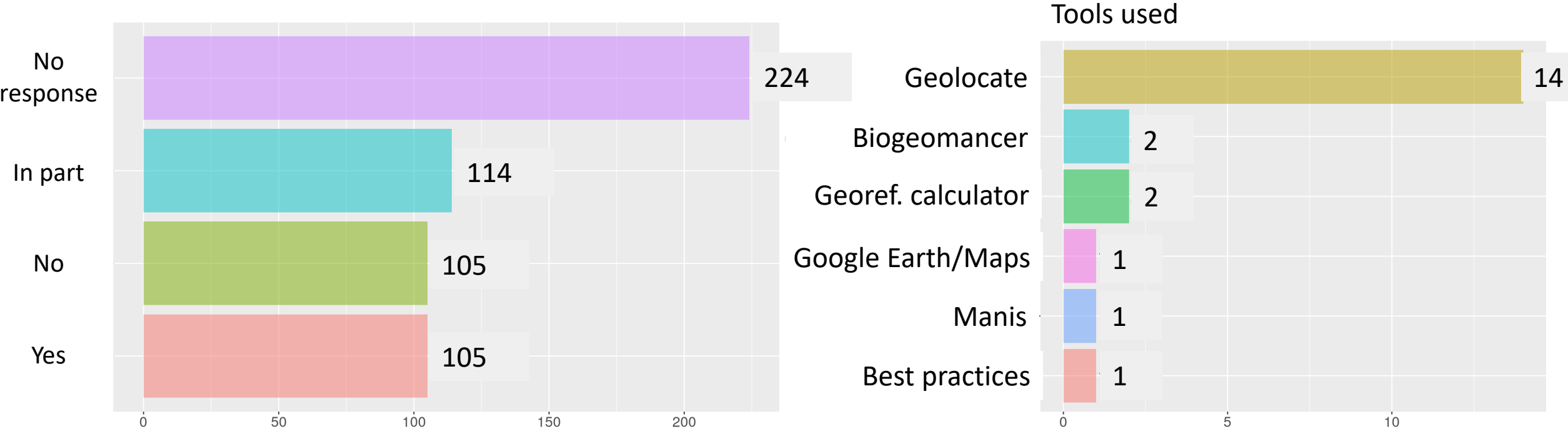
Q30: Do you apply a methodology to detect georeferencing errors after the georeferencing process? If yes, can you provide some details?, if no, please say so.

Q32: Do you use data from public biodiversity data repositories?

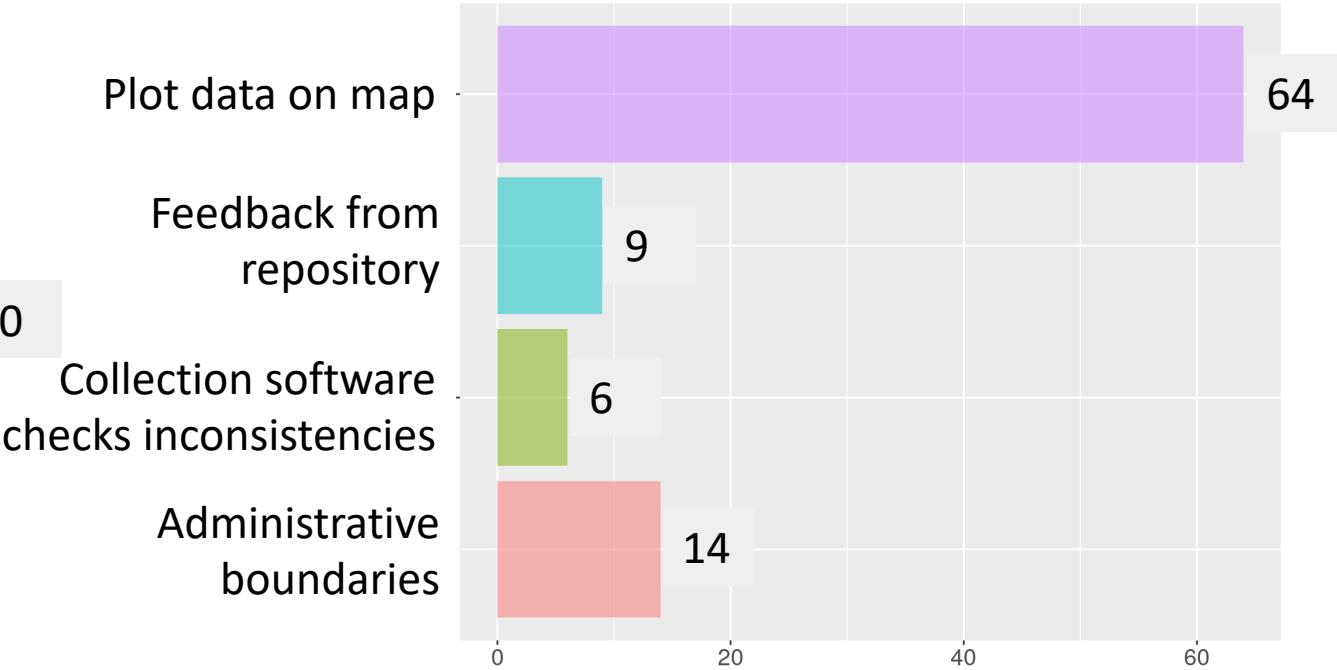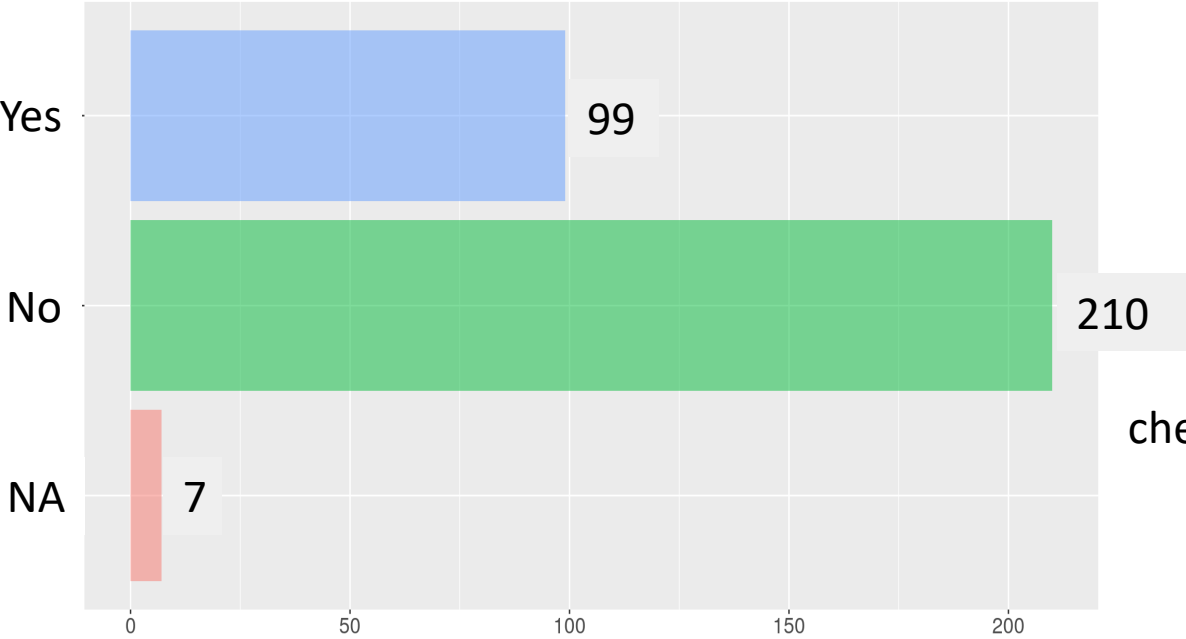| Repository	Count
GBIF - Global Biodiversity Information Facility	193
iDigBio - Integrated Digitized Biocollections	17
Atlas of Living Australia	8
VertNet	5
Symbiota	5
Canadensys	5
SpeciesLink	4
SEINet North American plant network	4
SouthEast Regional Network of Expertise and Collections	3
Mycoportal	3
U.S. Geological Survey	2
Paleobiology Database	2
INaturalist	2
World Flora Online	1
Tropicos	1
Tasmanian Natural Values Atlas	1
Species Map	1
Sistema de Informação sobre a Biodiversidade Brasileira	1
SIB Sistema de Información sobre Biodiversidad de Colombia	1
Red de Herbarios del Noroeste de México	1
Norwegian Biodiversity Information Centre	1
National Center for Biotechnology Information	1
National Biodiversity Network	1
Mycobank	1
Insectis Online	1
Index Fungorum	1
Global Plants	1
GEOLocate	1
Genesys	1
FloraWeb	1
Flora-On	1
eBIRD	1
Discover Life	1
Deutschlandflora	1
CONABIO - Comisión Nacional Biodiversidad, México	1
Biota of North America Program	1
Biological Records Centre	1
BGCI - Botanic Gardens Conservation International	1
Australasian Plant Pest Database	1
AntWeb	1
AlgaeBase	1
Aculeata	1

N: 282

Q34: Has your collection been uploaded to GBIF / Other?
If 'other than GBIF', can you specify?

Left chart (Did your collection get uploaded):
- No: 132
- Yes: 189
- Did not respond: 227

Right chart (Which repository):
- GBIF - Global Biodiversity Information Facility: 191
- iDigBio - Integrated Digitized Biocollections: 29
- VertNet: 13
- Atlas of Living Australia: 12
- Symbiota: 7
- SouthEast Regional Network of Expertise and Collections: 5
- BGCI - Botanic Gardens Conservation International: 5
- SpeciesLink: 4
- SEINet North American plant network: 4
- OBIS - Ocean Biogeographic Information System: 3
- Canadensys: 3
- Mycoportal: 2
- BISON - Biodiversity Information Serving Our Nation, EUA: 2
- WFCC - World Federation for Culture Collections: 1
- Virtual Herbaria: 1
- USGS National Digital Catalog: 1
- Tasmanian Natural Values Atlas: 1
- Tanzania Biodiversity Information Facility: 1
- Swedish Species Information Centre: 1
- Species Map: 1
- SIBBr - Sistema de Informação sobre a Biodiversidade Brasileira: 1
- Plant Search: 1
- PLADIAS - Database of the Czech flora and vegetation: 1
- National Biodiversity Data Centre, Ireland: 1
- InvertEBase: 1
- INaturalist: 1
- HerpNET: 1
- GGBN - Global Genome Biodiversity Nerwork: 1
- GFBio - German Federation for Biological Data: 1
- Genesys: 1
- GBDB - GeoBiodiversity Database: 1
- Flora-On: 1
- FishNet: 1
- Euro+Med PlantBase: 1
- Discover Life: 1
- Deutschlandflora: 1
- CCH2 . Consortium of California Herbaria Portal: 1
- BioGeoSilesia: 1
- AntWeb: 1
- Aculeata: 1

N: 282

Q31: Would you say your georeferenced site names may also be georeferenced by other institutions?

Q34: If you wish, feel free to provide us with any comments or suggestions you may have regarding georeferencing of digital collections.

It would be useful a database of ancient toponims of some countries that have language difficulties, like ancient languages, other alphabets or countries or provinces that have changed the administrative delimitation.

Our collection is small but is representative from this state region. Actually exist few personal related on it. We need equipment and people who support that work collection. Our herbarium is known as CIMI. Thanks

… strictly speaking we only enter locality data as they appear on the label, i.e. without georeferencing. This latter is only done in the context of specific research projects or revisions

Understanding the importance of the issues raised, we plan to develop our work in this area

We simply don't have the resources to address georeferencing as it is not a sufficient priority. We aim to address this by sending data to our national GBIF node who have the software routines to process textual geographic data and prepare georeferences from this. While this data may be useful nationally for Ireland, locality information for other countries is generally too vague (e.g. 'India') to be useful internationally

Into the future, I envision that there will be several georeferences for a single record since some georeferences are fairly rough (or even automated), while others are done with much attention to detail. Creating a venue for storing these different data and tracking their reliability would be helpful.

Q34: If you wish, feel free to provide us with any comments or suggestions you may have regarding georeferencing of digital collections.

I do not trust GBIF and other data repositories based on my analysis of fitness for use. However I di feel that these repositories could be hugely important for advancing digitization and "crowd sourcing" error discover.

Lacking a budget for digitization other than what I've managed to "string together" with consortium based grants (like the Consortium of Mycological Herbaria/Fungaria that is contributing data to MyCoPortal) and lacking available personal time and personnel, I struggle with balancing time spent doing georeferencing with simply entering the available label data and linking images (letting users do the georeferencing based on the available label data).

Some emphasis should be spent on historically avid collector locations. Using Hymenoptera as an example. Just having geo-referenced locations by day,mon,year for either Bohart would greatly ease geo-referencing for smaller institutions.

Thank you for including us in this survey - adding georeferences to the large portion of our collection that doesn't have this data is on our list of things to do, and I have noted the protocols listed in an earlier question.

Georeferencing is almost "pie in the sky" for smaller institutions.   Curating a medium-sized collection of vascular plants that had lain dormant (or essentially so) for twenty years has been a Herculean task over 7 years, and is not even half completed. For example, previously the family classification followed Cronquist (more or less) alphabetically; now it is APGIV organized phylogentically. The generic classifications were dreadfully out of date. Virtually all folders were acidic and needed replacing. Nothing was data based. Nothing was imaged. Again, georeferencing is important, but in my list of priorities, it is nowhere near the top

Q34: If you wish, feel free to provide us with any comments or suggestions you may have regarding georeferencing of digital collections.

Links to online maps and gazetteers for all countries would be really helpful- it can take a lot of time to find these. Another idea would be to have people willing to help interpret localities in foreign languages. Or a way to post localities that can't be found so others might be able to help. I'd also like to see a standard about coordinate rounding- I've seen this error flagged for some of our specimens on GBIF but the coordinates in question mostly have 5-6 decimal points, which doesn't seem like they are really rounded.

We have a system of automatic georeferencing ISTRA. Currently, 45% of geodata are performed by it. ISTRA mainly uses collector-date matches for automatic georeferencing.

Regional natural history museums lack the resources required to follow any kind of strict protocol on geolocating. There is often one member of staff caring for 100,000s of specimens. Interpreting archaic location names, while important, are not at the top of the list of priorities. Anything that makes the process easier is beneficial, but it will always require a deal of time which may not be available.

For now, we have not yet started with systematically georeferencing our digital collections, but we would like to do it in the near future. Therefore, all feedback on the best tools to use is welcome.

Q34: If you wish, feel free to provide us with any comments or suggestions you may have regarding georeferencing of digital collections.

A "one-stop-shop" online reference tool for georeferencing would be great to have. Currently volunteers spend hours searching various references to find obscure location names like small towns, local roads, and small creeks and rivers. If all of these could be available in one place would be revolutionary for time spent on this. If a site like Geolocate allowed the user community to submit place labels for inclusion in the layers that would ultimately build a robust location data source. Perhaps something like what iNaturalist does where a person can submit but then it requires a certain amount of vetting before it can be included for the broader user community might be workable.

In summary

- Survey on NHC holding institutions across the world: 552 respondents

- Accounting for an overall of 3.4B records across institutions

- Responses referring to a total collection of 1.4B, mostly plants and animals, mostly in Europe and North America

- 28% digitized, 56% partially digitized, 16% not digitized

- 42% georeferenced, 50% report uncertainty

- 34% do not follow guidelines, 33% their in-house developed, 17% C&W

- Only 1/3 apply post-georeferencing quality check

- Uncertainty not a widely accepted important issue

- 25% of reported collections not uploaded to repositories (350M)

- User feedback: lack of resources and training, not priority given work load, need for publicly available common georeferenced site names, collector information may help, need to pull efforts across institutions