Questionnaire on Dealing with Sensitive Primary Species Occurrence Data

Summary of Reponses

Arthur D. Chapman

8 May 2006

Prepared for the Global Biodiversity Information Facility

Contents

Introduction

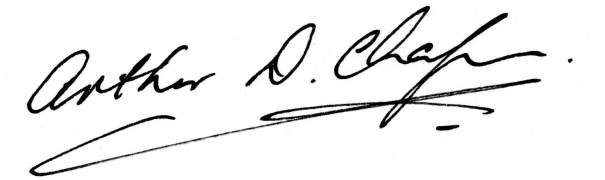
This document summarises responses to the "**Questionnaire on Dealing with Sensitive Primary Species Data**" conducted on behalf of the Global Biodiversity Information Facility between 1 March and 1 April 2006, using Surveymonkey.com.

A certain degree of liberty has been taken to fix spelling errors and certain grammatical errors in the responses to improve readability. The full answers are available via the associated spreadsheet.

In a number of cases, answers to one question will only make real sense when taken in context with the same respondent's answers to related questions. For this reason, to get a full picture, one should examine the full list of responses in context.

Further analysis of the answers, recommendations and guidelines will be forthcoming prior to and following a workshop to be held later in 2006.

Any errors, misinterpretations or omissions that have arisen through the process of summarizing, are the responsibility of the compiler.



Arthur D. Chapman
Toowoomba, Australia

14 April 2006.

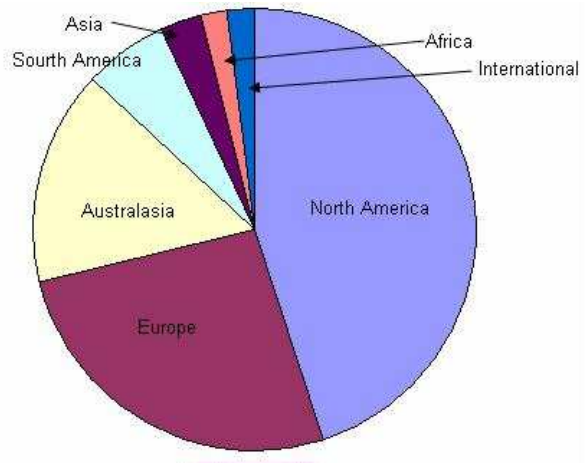
Responses to Questionnaire

A total of 225 People accessed the Questionnaire:

Detailed Responses:	102
Email and base info only:	48
Duplicates:	4
Looked only – no data:	70

Responses were received from 24 countries and two International organizations:

USA	42	Faroe Islands	1
Australia	14	France	1
Spain	8	Iceland	1
Canada	4	India	1
UK	4	Japan	1
Argentina	3	Peru	1
Denmark	3	Philippines	1
Belgium	2	Poland	1
Colombia	2	Sweden	1
Germany	2	Switzerland	1
New Zealand	2	The Netherlands	1
South Africa	2	International	2
Austria	1		



Contact email and some basic information were received from 15 countries (4 additional to those above):

USA	11	Canada	1
Spain	9	Cuba	1
Argentina	6	Germany	1
Denmark	4	India	1
Australia	4	Nepal	1
South Africa	3	Poland	1
Ukraine	2	Portugal	1
Belgium	2		

Question 5: Does your institution house?

	Botanical	Zoological	Other	Respondent Total
Collections of Specimens (e.g. Museum/Herbarium)	**70% (53)**	63% (48)	21% (16)	**76**
Living Collections (e.g. Botanic Gardens/Zoo/Aquarium/Culture Collection)	**79% (19)**	29% (7)	17% (4)	**24**
Observational Records	67% (45)	**82% (55)**	25% (17)	**67**
Survey and Monitoring Data	69% (43)	**81% (50)**	27% (17)	**62**
			Total Respondents	**100**
			Skipped the Question	**2**

Of the 100 responses:

	Specimens	Living Collections	Observations	Survey and Monitoring
Botanical, Zoological and Other	13	1	14	15
Botanical, Zoological	14	3	20	16
Zoological and Other	1	0	2	1
Botanical and Other	0	1	0	0
Botanical Only	26	14	11	11
Zoological Only	20	3	19	18
Other Only (incl. Paleontology)	2	2	1	1
Total	**76**	**24**	**67**	**62**

Questions 6 and 7: How many records or observations does your institution/collection hold? and How many records are you currently making available via the GBIF Portal?

	Question 6 No of Responses	Question 7 No. of Reponses
Number of Specimens	80.0% (76)	79.4 (77)
Number of Living Collections	64.2% (61)	55.7% (54)
Number of Observations	42.1% (40)	40.2% (39)
Number of Survey and Monitoring Records	56.8% (54)	49.5% (48)
TOTAL Respondents	**95**	**97**
Skipped the Question	7	5

A number of Institutions appeared to combine both Observations and Survey and Monitoring Data in their numbers.

	Total Collections (Respondents)	Available to GBIF	Non respondents	Available to GBIF
Specimens	190,242,270	11,614,778	7,675,519	165,200
Living Collections	1,979,603	25,004	10,001	0
Observations	71, 512,061	38,899,302	31,000	0
Survey and Monitoring Records	39,649,718	5,663,818	10,100	0
TOTAL	303,383.652	56,202,902	7,726,620	165,200

In total, the Institutions who responded reported managing around 300 million records, of which GBIF has access to about 56 million (~18%).

Question 8. When making data publicly accessible do you generalize any fields?

Total Respondents	100
Skipped this Question	2

Two thirds of respondents report generalizing data in some way, whereas one-third said they did not.

NB – this summary has required some interpretation and for those that require details, I refer you to the detailed report in the Appendix.

Of the 65 respondents to this question who answered "Yes"; the fields that are generalized or restricted are listed below. Some of these are restricted in different ways for different taxa, and may be permanently (e.g. endangered species) or temporarily (awaiting publication of research results) restricted. I have reported here the fields that the respondents report are restricted, but it should be born in mind that the fields may only be restricted for certain types of sensitive data by some institutions.

Summary of Responses

Field	No of respondents	Comments
Locality	42	Deleted or altered.
Georeferences	42	Deleted or scale decreased (see later questions).
Accuracy	1	
Collectors' / observers' names	16	Restricted for a number of reasons – privacy of living people; could be used to track itinerary, etc.; some countries have privacy legislation. Others note that they NEVER suppress this information.
Determinor's name	4	Usually restricted for privacy reasons.
Dates	8	Could be used to track collections before and after a sensitive species
Taxonomic information (name, rank)	4	
Habitat information	1	
Sex	2	
Character states	1	
Hosts	1	
Traditional Use information	1	
All (except species name and rank and maybe accession number of feature id)	5	
All (i.e. sensitive data not shown at all)	3	
Variable – custodian controlled	1	

Those that stated 'locality' may also have implied 'georeference' and vice versa, so those categories could be increased.

For the detailed responses – see the Appendix.

Question 9. Are you required by legislation to limit access to certain data?

Total Respondents	**87**
Skipped this Question	15

Sixteen (18%) of respondents reported that they were required by legislation to limit access. Seventy Two (81%) responded in the negative. A number of these were not due to legislative instruments per se, but were legal instruments such as data agreements or collecting agreements with data suppliers, landholders, or traditional owners. Several were due to Acts of Parliament that restrict release of information on endangered species, etc., and some were due to privacy acts and the release of personal information. Several responses mentioned that there was a requirement in the United States to make data available in the public domain.

Question 10. What are your main reasons for restricting access to sensitive data for sharing or making available?

Total Respondents	79
Skipped this Question	23

Many of these categories have involved subjective interpretation of the responses, and in a number of cases there is overlap between the different categories.

Category	No. responded	Pertinent comments
Protection of threatened species, economically important species and reduction of the impact on wild populations of sensitive species	37	• less problems with historical data on endangered species, but more concern on recent data • As part of a major conservation agency, we have a duty to protect population's of rare and threatened flora from over-collection or other abuse.
Preclude deliberate sabotage or collection by unscrupulous and commercial collectors, poaching, hunting, disturbance, over exploitation, etc.	24	• many amphibians and reptiles are harvested from the wild to be sold live and as a consequence populations are sensitive to collecting pressure. • Threats of illegal falconry take and disturbance of nesting falcons from the public, particularly with nests located close to roads. • Egg collecting from rare bird nests, rare plant collecting, disturbance, trampling, badger baiting, commercial exploitation of rare species, etc. • Protecting the original sites where the fossils have been found from commercial collectors • Protect populations of ferns, orchids and cycads from plant collectors • There is a history of egg thefts from falcon-nests and although digging up of rare plants has not been a problem until now in Iceland we choose to mask the accurate coordinates of plants on red-lists.

		• We are concentrating on succulent plants incl. cacti, and a majority of our holdings is of CITES-covered taxa, so conservation is the primary reason to withhold data. • Many of Arizona's plant species (esp. cacti) have value on the market, are located on wide open public lands, and illegitimate collecting is easy • Commercial and illegal fishing
Allow for publication of research results	12	• To prevent stealing of unpublished data • Data are a 'resource'. The most immediate way of getting benefit from them (after having spent energy in data-gathering) is publication. This is an effective pressure, since the researchers are evaluated for publishing results in scientific journals, not for making the data available to others. Thus, a 'quarantine' period to analyze the data first, then (second) make them available to others, is by now the pattern that might enhance 'survival'.
Protect rights of landholders	10	
Protection of people's names, privacy considerations, etc.	6	• For general and legislated privacy • To shield people from possible reprisals by animal-rights activists • Observational data can be interpreted to include (possibly illegal) collecting of material, while it usually consists of photographic records only.
Fear the user may make inappropriate use of the data; not knowing purpose to which data will be put; fear of misinterpretation	5	• Uncertainty of the use to which the data may be put. The data, as in most collections-based datasets, is often fragmentary or incompletely representative of the population distributions or relative protection on conservation estate, Occasionally it is incompletely identified, mis-identified or not provided with the most recent applicable name. Without adequate knowledge and usually discussion of the project for which the data is sought, we cannot guarantee the data is 'fit-for-purpose'.

Gain cooperation of landholders	4	
Confidentiality agreements and data agreements	4	
Protect third party data held by institution	4	
Protect the sources of the data and rights of data providers	3	
Protection of IP rights, including need for proper attribution and citation	3	
Reduction of collection pressure and protect sensitive communities	3	
Biosecurity, Quarantine and Trade considerations	3	• Our sensitive data is primarily quarantine interception data, if released this would have market access and biosecurity consequences. • We would be very concerned about making data from quarantine interceptions publicly available because it could create and international trade incident if our collection was to tell the world about the presence of an exotic pest in Australia or a trading partner. These records are tagged. We also have some highly desirable and protected records of jewel beetles and we would not like the exact location of these to be public. Unfortunately these records are not tagged. • Some quarantine records not publicly available.
Control bioprospecting	1	
To maintain competitive advantage	1	• The main reason to limit access to monitoring data (phenotype) is competitive advantage in the development of new food crops.
Assist in conservation and recovery of species	1	
Benefit sharing, need to maintain good relations with countries of origin, etc.	1	
Commercial in Confidence	1	

Question 11. What are your main reasons for granting access to potentially sensitive data?

Total Respondents	**81**
Skipped this Question	21

Many of these categories have involved subjective interpretation of the responses, and in many cases there is overlap between the different categories.

Category	**No. of responded**	**Pertinent comments**
For scientific research and analysis; scientific advancement	30	• Arhus convention • Taxonomic research • Bona fide systematic research • Given to researchers on a case by case basis.
Species and Conservation planning and management and those agencies so involved; conservation assessment	21	• making land managers and agency biologists aware of rare species to improve their chances of protection • If a researcher has need for potentially sensitive data we will provide it, as long as we feel comfortable with the need for the data, the person's willingness and ability to provide proper attribution for the data, and so on. We realize, however, once we release the data to anyone we have lost any real control over its use in the future. • Documented need-to-know for legitimate research, planning, or conservation purposes (each grantee signs a non-disclosure agreement). • The data held in our collection is the most reliable and verifiable source of plant species distributions in the State. We believe that decisions on survey, reservation, protection and conservation, as well as a wide range of scientific and community endeavours, are better made by consulting this significant resource. If a scientist, land manager or community conservation group member can adequately inform us of the purpose to which they will

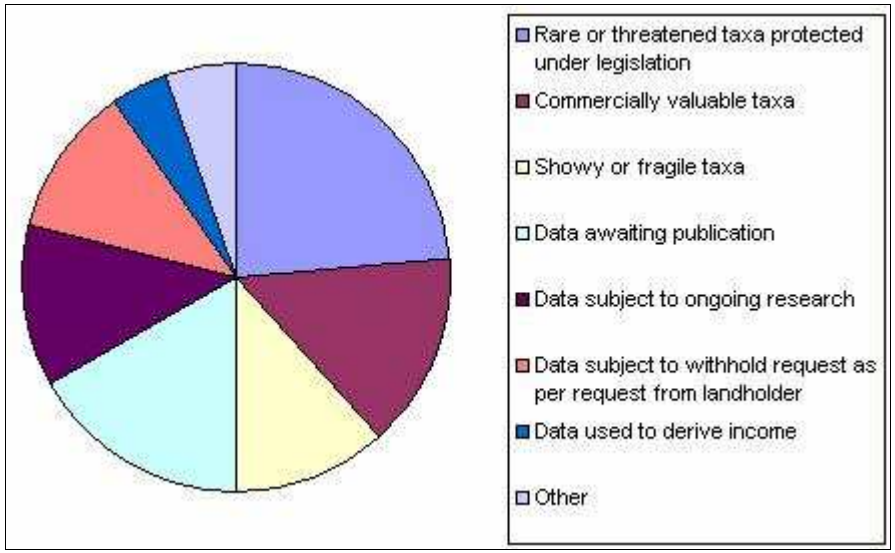
		be using the data, if we feel confident they understand the limitations of the data, and if they commit to only using the data for the intended purpose and not using it again for new projects, or passing it on to a third party, then we attempt to meet their request for access to the data.
Management of the environment and of biological resources; need for continued conservation actions to maintain species and populations; environmental impact studies; biosecurity management	10	• An important concept is that making biodiversity data available should reduce the risk of damage to the environment.
Inquiries from Government agencies and professional organizations, e.g. for policy making and environmental management	8	• Our clients or partners are almost entirely federal and state agencies who need the best information we can provide them, so we don't generalize location data except in the one case already stated.
Species distribution studies, species modelling; vegetation survey and mapping; global scale analysis; monitoring and resurvey	6	
Entire database should be available (free data policy)	6	• We would grant access if such data have already been published in literature • We believe that biodiversity data needs to be freely available to anyone, anywhere, anytime. • Free access to the data increases the utility, usefulness and value of the data. Which increases the value of the institute itself. • Collectors and poachers are usually ahead of scientists, not behind them • It is public data • We are a state agency so and are required by sate law to share all data.
Collaborative projects	4	• By sharing the data, hope that others will also share theirs and have a mutual benefit of it.

Should be available to bona-fide individuals where there is reasonably assurance that data will be put to a non-commercial, serious scientific/scholarly use	3	• If the person has a good reason for needing it, and they are trustworthy
Land Management	2	• Environmentally sensitive information often relates to those species and habitats that are particularly vulnerable to land management activities. It is important that such information is made available to those that control land management activities at a level of detail that is useful. It is important that such information is made available to those that control land management activities at a level of detail that is useful. • To ensure the species are considered during planning, forestry, environmental education etc.
Protection of species – e.g. where lack of disclosure could endanger species	2	• If lack of disclosure could reasonably be expected to result in damage to the species or interfere with the conservation or recovery of the species. • Destruction of occurrences more likely through ignorance than malice. • Those who are conducting developments that may threaten the plants so that mitigation may be effected.
Nothing	2	• Anything you do for protect them will be broken previously by the 'pirate' user. • We do not - people can come here if they want it. They can read the specimens and locate the information. May restrict some science; better than lose some species.
Law enforcement and protection	1	• Law enforcement and protection policies, but access is granted only to scientific or official bodies (not to the general public).
Freedom of Information Act	1	• The agency's legal department issued an opinion that the federal Freedom of Information Act does not provide a shield to protect sensitive zoological observations and location information. The electronic database is not directly accessible to the general public, but requests for data (including the

		location of observations) cannot be withheld.
The work involved in restricting some and not all records	1	
Data contributors	1	• Access is granted only to data contributors, who are vetted for professional qualifications by our Advisory Board.
Data repatriation	1	

Question 12. What categories of data do you regard as sensitive and are liable to restrict information (especially locality information) on?

Total Respondents	**93**
Skipped this Question	9



The two largest categories listed by respondents were ‘Rare or threatened taxa protected under legislation’ and ‘Data awaiting publication’, although all categories (and more) were listed.

The "**Other**" category included:

Data contributed under agreement with another party.
Taxa subject to intense collection pressure.
Traditional Knowledge.
All cycad data, due to intense interest by poachers.
Collectable plants with a cult following - orchids, cacti, cycads.
Taxa facing threat from over-collection.
Data on the use of the material by private breeding companies.
Data points that fall within the Exclusive Economic Zones and within Hotspots or protected areas.
Personal rights, and copyright issues
1) Information with publication restricted by agreements with clients for whom the data were collected. 2) Information where rights are not held in their entirety by us - e.g. images of birds where the photographer has not authorized web publication.
(1) Imperiled or rare according to the NatureServe global conservation status ranking (these do not necessarily have any legal status) (2) restrictions on data use / access passed on by the data provider (source)
Our data providers control access to the data they make available through the NBN Gateway website. The NBN provides best practice advice on how access to information should be managed. There are many reasons why data holders may restrict access to information and the NBN Trust has attempted to catalogue these. The key constraints are listed below… ·Funding and Resource Management ·Authority over data ·Trespass ·Landowner Relations ·Publication interest ·Incomplete or unchecked data ·Concerns over inappropriate use ·Intellectual Property Rights ·Academic Research Rights ·Personalities ·Environmentally Sensitive Data ·Personal Information ·Commercially Sensitive Information ·Acknowledgement and Recognition ·Access Administration Time Constraints ·Cultural Change.
The only criterion for withholding data on sensitive species is if disclosure would result in an adverse effect. We have made an assessment of which species in Wales this would apply to - obviously those species may be protected under legislation, commercially valuable, showy or fragile but that in itself is not the criteria for inclusion on the list. i.e. not all legally protected species are on the list.
Data under systematic revision by museum staff or by specialists. [This could be included under the Research category – AC.]
Quarantine interception; Quarantine records

Question 13. If you have any further comments on any of these categories, please elaborate.

Total Respondents	**27**
Skipped this Question	75

A number of the 27 responses have already been covered under the '**Other**' category under Question 12.

We consider listed species, especially those prone to collections (e.g., butterfly, herps) or often in conflict with potential development to be most sensitive. In the latter case, for example, a potential development site may have a bald eagle nest--if this is "determined" prior to the permitting process, there is concern that the nest (or nest tree) may disappear prior to permit related surveys.
We consider the following to be sensitive occurrences: bat hibernacula, bat day/night roosts in caves/mines, bat maternity roosts; locations of special status (=sensitive) invertebrates; raptor nest locations; federally listed Endangered and Threatened plants. In addition, if we know that a private landowner does not want information released about an occurrence (of any special status species) on his/her land, we regard that as sensitive and do not export it (in other words, we won't generalize the location but will simply exclude the occurrence from our GIS export).
For rare and threatened taxa we follow the IUCN redlist and the U.S. Fish and Wildlife list of Federally Endangered Plant Species. We also follow the advice of specialists in a certain plant families or regions of the world. We would be willing to restrict locally endangered plants if the appropriate lists existed and were well maintained.
Landholders could include both private and public owners.
If a landholder asked us to restrict data, we would comply, but this has not occurred so far.
We have considered the issue of landowners' views. However, we feel that in general it would be unworkable to either consult landowners about release as general practice or withhold data if they object. We feel legally it would be difficult to justify. However, in exceptional circumstances where a landowner objects strongly and it cannot be otherwise resolved, and there is a real threat of environmental harm or breakdown in working relations then we would look at withholding the data. However, when doing new surveys we would ask that surveyors always seek permission and make it clear to the landowner that the data will be released publicly.
Georgia Natural Heritage Program has an exemption to Georgia's freedom of information act. We do not have to give data to anybody unless we deem it important to do so.
In the United States, it would seem the dominant argument is not legislation limiting access to certain data, but rather legislation mandating access to data. However this is the argument posed by some people, and I have not seen legislation that actually states this.
Ours is based on a list of concern as the government agencies in Utah and the US are currently not too concerned about plants. I would welcome

a world list - based on what regional botanists think should be kept general.
Data awaiting publication or subject of ongoing research is withheld at the researcher's request only.
The only "Data awaiting publication" we generally restrict are manuscript names prior to formal Code-compliant published descriptions.
Data awaiting publication or withheld by landholder request is not covered by provincial legislation and must be disclosed in response to requests submitted under the Freedom of Information Act.
In addition to generalizing coordinate data, we protect a small amount of data from any public observation at all for a limited time (usually no more than two years) at the request of data contributors who are actively doing research using these data, and have not published yet.
Re "Ongoing Research": only activated where confidentiality agreements between the State and Bio-assay companies exist.
We do not restrict in the strict sense, in that we don't hide any part of a data record. However, records that are part of ongoing research by museum scientist may be temporarily withheld in their entirely from the public domain at the discretion of the appropriate curator.
All of these are at the discretion of the particular curator.
There is a sliding scale of sensitivity, as new collections are brought into the collections and databased and older ones are analyzed and written up. Only [collections from] the last year or two are considered sensitive.
I clicked on all fields I hear frequently arguments for withholding. Personally, I rarely accept any of the reasons above. I think that these few outliers should never drive the Free Data Access issue.
Some of these categories are more important and more relevant to us than are others, but I believe we would generalize data for any of the reasons stated in question 12. As mentioned earlier, the system will automatically generalize records for rare or threatened taxa; we would need to place a manual flag for the other categories in questions 12.
As mentioned, NGB has tried to avoid sensitive datasets and maintain the complete database freely available. We don't regard the locality data on wild populations of the crop wild relatives as sensitive, but might need to consider this in the future... as these data points are expected to grow with the ongoing activity on EU funding in this field which will produce more data.
Specific cases will vary, but in general all data is made available unless there is a specific reason not to.
Data used to derive income: Images are the main category here, these include rotifers, which due to conflict with CD sold as publication are only served at low resolution over the web, and VIREO where images of birds are provided under a complex set of restrictions and access rights for different categories of users - but where low resolution copies of the images are used to generate interest for purchases of higher resolution copies.

Question 14. What proportion of your collection or observation records do you regard as 'sensitive'?

		Response Percent	Response Total
None		5.6%	5
<1%		25.6%	23
1-10%		**43.3%**	**39**
10-25%		12.2%	11
25-50%		1.1%	1
>50%		7.8%	7
All		4.4%	4
	Total Respondents		**90**
	(skipped this question)		12

The overall percentage for all collections works out at between 2 and 10%, however this is based on very rough statistics and the figure should not be used with any reliance whatsoever. It would appear that many of the higher percentages are with smaller and specialist collections; however there are exceptions to this.

Question 15. Comments

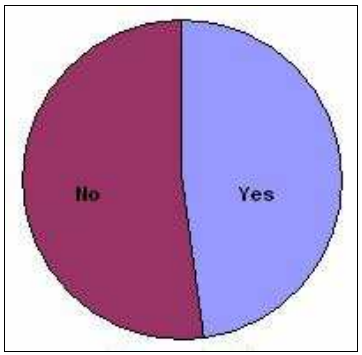
There were 25 Comments, several were just along the lines of '*Never really looked, so this is a guess*' or that the figure is still very tentative. A number of comments have not been included as they only make sense in conjunction with other answers.

Restrictions likely to be applied on taxon and country basis, e.g. Aloes, Orchids, collections from National Parks in Namibia. Also our Seed Bank has a policy of not allowing unrestricted access to collections data; this is usually part of the Access and Benefit Sharing Agreements

under which the Seed Bank works in particular countries.
We treat all rare species data as sensitive for the reasons mentioned above.
Our data providers determine which of their records are sensitive. 1.23% of records are flagged as confidential within the datasets we are given. This figure does not take account of datasets that providers have restricted access to using the online access controls. Of the 150 datasets currently on the Gateway website has no public access. Two of these are trial datasets awaiting removal. Nine are awaiting the appointment of new administrators to set access. Three are new datasets being checked by the provider before being released more widely. A public access position for the remaining two datasets has yet to be agreed with the data provider. 134 datasets are available to the public to access at between 10km and full geographic resolution.
Vast majority receive some basic sensitivity (generalised to 1km with species name attached). A much smaller list (currently three) is so sensitive, they are only displayed at 10km resolution with just "sensitive species" indicated.
For marine biodiversity data rarely the case
Most of the sites were fossils have been collected are unique and need some protection from commercial dealers
A lot of species are protected by law.
Collection: actually very few Other (observation, unpublished records): varies, they might be temporally sensitive.
By 1% I mean the number of records held from the public at the request of data contributors (see item 13), not the number of records with generalized coordinate data (which is all of them).
Most of our data is covered by CITES.
Nature of sensitivity, proportion of records that are sensitive, and treatment varies considerably between collections. Some data sets are considered too sensitive to make publicly available, others are published in their entirety, and others have some degree of data hiding/redaction, with decisions being made entirely on a data_set by data_set basis. Proportion of records that are sensitive varies from no parts of any record, to all parts of all records. 1-10% is a reasonable across all data sets summary.
Restrictions likely to be applied on taxon and country basis, e.g. Aloes, Orchids, collections from National Parks in Namibia. Also our Seed Bank has a policy of not allowing unrestricted access to collections data; this is usually part of the Access and Benefit Sharing Agreements under which the Seed Bank works in particular countries.
As a Public institution paid by the "taxes" of our population, all records are legally publicly available. However for legal reasons species location of "red list" species should not be shown exactly to all Recent data not yet published are generally not made available to all directly.
Essentially all of the precise location data (whether in a text field or georeferenced) is considered sensitive.

Question 16. Does your institution currently use data sharing agreements or data licences to manage access to sensitive data?

Total Respondents	**93**
Skipped this Question	9



It is obvious from this that around half of the institutions responding to the survey use data sharing agreements.

Question 17. Comments

We currently share full information on all records, but will restrict fields when our application is fully implemented.
Special regulations are in place for: 1) loans of herbarium material 2) exchange of silica dried material for DNA analyses 3) exchange of living collections.
We should probably have such an agreement in place. Previous institutional inquiries have run into difficulties over how [to handle] "freedom on information" legislation.
We have a Memorandum of Understanding with the state's natural heritage program. They are the primary custodians of plant location records.
Working with state heritage program.
Confidentiality and non-Reproduction Agreement.
Only use data licenses in some cases. Most users of our data do not sign a data license but receive generalized data instead.

We have two agreements that relate to data exchange through the NBN Gateway website. 1) The NBN Gateway Data Provider Agreement is an agreement between the NBN Trust (as manager of the NBN Gateway) and a data provider wishing to make their biodiversity data available through the NBN Gateway. The agreement gives the NBN Trust permission to hold a copy of the data supplied by the data provider in order to load them onto the Gateway website for the data provider to then control access to using online administration tools. 2) The Gateway Terms and Conditions govern access to and use of all material available through the NBN Gateway. Everyone that visits the site is bound by these conditions. They establish the terms under which the biodiversity data on the NBN Gateway can be accessed and used and meet the needs of most data providers. However, data providers can identify additional terms of access for their own data. Additional terms of access build upon and, where they conflict, supersede the Gateway's terms and conditions.
We had a bioprospecting agreement under Argentinean law.
We do use data sharing agreements with certain federal agencies to spell-out our mutual interest in sharing data. We require all our clients to sign a data use agreement which spells-out our expectations for their use of our data.
+ Data Sensitivity Training (basically awareness training and a prerequisite to be a licensed data user).
Agreement with the networks of data providers (observations).
We have a general license agreement that also covers sensitive data but it does not apply to our internet available data.
I'm not sure what sort of formal mechanisms we have in place for this. For herbarium specimens, the closest things we have is a materials transfer agreement, by which dissemination of physical material is restricted. Some members of our staff work routinely with State and Federal agencies to monitor endangered species. I can check with them about their data sharing protocols.
Probably, 'under the table' agreements and arrangements exist between institutions. If not officially, then this is established as a Culture.
We will use these for institutional data providers to the Avian Knowledge Network. They are currently under development.
These have no legal validation and have more the character of a generally agreed "memorandum of understanding".
These are still in preparation but should be completed before long. We intend to have Restricted Release Licenses to be used when we give sensitive data to trusted partners. If we regularly share data with partners we will be looking to sign Data Exchange Agreements which will apply on an ongoing basis to all the data we exchange, and will include the sensitive data issue and any other pertinent matters.
These agreements are now largely restricted to research data, not general biological collections.
We are setting up sharing agreements in the southeastern US though SERNEC.
We routinely set agreements with various countries where we collect living and preserved materials. This certainly applies post-CBD but it also applies in principle to pre-CBD materials.
At some extent, with specimen collectors who require data occultation until publication.
These are developed case by case basis. No data licensed for non-scientific use. Requests are considered individually for taxonomic researchers.

Some agreements are in place for other institutions' use of our data.
No data licensed for non-scientific use. Requests are considered individually for taxonomic researchers.
Letter of Intents, Code of Ethics and Memo of Understanding.
A material transfer agreement [http://tor.ngb.se/Material/MTA.php] follows all the seeds we distribute. This document is foremost to ensure future free public access to the biological material and data generated from this material. Basically it is not allowed to patent or register property right on the material as received.
We usually grant access to bona-fide researchers, following established current practice of most herbaria, i.e. we need a proven interest concerning the material, usually including a statement from a major professor (in the case of student and Ph.D. work) or institution head.
The wider our data is used the better. We have trouble making money out of it, but if someone else can, more power to them.
Yes, with USDA and USDI agencies, The Nature Conservancy, and people contacted by them.
Our data sharing agreement treats all records as equal.
Yes, but currently ad-hoc and in need of overall policy development. Our main concern is to prevent commercial use without prior consent.
Sort of -- informally now; more formally in the future under PBIN.
There is a draft document about Collection Management plan 2006-2010, which concerns among other things Physical and electronic access to collection "items". For the Zoology department the providing of content to GBIF is clearly mentioned as part of the plan. By providing data to GBIF, our Museum has also accepted the data sharing and data use agreement of GBIF, although the Museum is not providing the data directly but via the National GBIF node.
We have data sharing agreements with our member programs (Natural Heritage Programs and Conservation Data Centres). We have data license agreements with partners who receive the precise location data. Access is granted to sensitive (precise location) data under a Data License Agreement on a case by case basis for projects that support our organization's conservation mission. The Data License Agreement includes specifics of what the data set consists of, who may have access to the precise data, the products, data ownership / copyright, citations, and legal warranties. Data License Agreements are for a limited duration (typically 1 year), and the recipient is encouraged to obtain an updated dataset.
Data sharing with OZCAM & other NT Government agencies

Questions 18 and 19. Would you be prepared to make examples of these agreements or licenses available to GBIF?

Virtually all who answered '**Yes**' to Question 16, were prepared to provide copies to GBIF. Comments included:

Yes - We would need administrative and other organization approval before making it available.
Yes - Copies of all NBN agreements and model licenses are available from the NBN website... http://www.nbn.org.uk/information/info.asp?Level1ID=1&Level2ID=10&Level3ID=7
Yes - http://fishandgame.idaho.gov/cms/tech/CDC/cdc_data_use_agreement.pdf
Yes - We can only provide a German version…
Yes - [http://tor.ngb.se/Material/MTA.php]
Yes - On the condition of continued protection of sensitive information.
Yes - Our data sharing agreement is available via our DiGIR provider metadata. The same will be true for all of the institutions with DiGIR providers participating in the taxon-based vertebrate networks (MaNIS, HerpNet, ORNIS, FishNet II).
Yes - Pending their drafting (they exist in rough draft right now).
Yes - For the moment I have only a draft, which I have seen available in French and Dutch for comments on our intranet. With agreement of our head of department, I guess that there will be no problem to share these guidelines and ideas.
Yes - We would be happy to provide a sample of both our Data Sharing and Data License Agreements.
Yes - Contact Dr Rodney Turner or Dr Ian Naumann, APPD, Plant Health Australia
No - Since they also falls within non-disclosure agreements.

Question 20. If you currently restrict access to sensitive data, would you be prepared to make the data available via data sharing agreements or data licenses, etc. (with suitable password protection; a single download for a specific purpose or some other method) to bona fide users for, or in:

	Government Agencies	Universities and Research Organizations	Non Government Organizations	Commercial Consultants	Public	Other	None	Respondent Total
Conservation Planning or Research?	**90% (63)**	**90% (63)**	66% (46)	43% (30)	29% (20)	16% (11)	4% (3)	**70**
Taxonomic Research?	84% (57)	**91% (62)**	59% (40)	37% (25)	31% (21)	12% (8)	4% (3)	**68**
Genetic and/or Molecular Studies?	82% (50)	**90% (55)**	59% (36)	39% (24)	31% (19)	13% (8)	5% (3)	**61**
Species Distribution Modelling?	84% (58)	**90% (62)**	59% (41)	41% (28)	28% (19)	13% (9)	4% (3)	**69**
Other	**80% (16)**	70% (14)	65% (13)	50% (10)	45% (9)	30% (6)	10% (2)	**20**
							Total Respondents	**74**
							(skipped this question)	28

It would appear that most institutions (over 80%) are prepared to make all categories of data available to Government Agencies and Universities and Research Organizations; around 60% to non Government Organizations and 25-50% to Commercial Consultants and the Public with the use of suitable protection methods such as Password access, or single downloads. Only 4% (except for data under the '**Other**' category – see Question 21) responded to the effect that they would supply no data of this nature under these circumstances.

Question 21. If you selected 'Other' please list categories of data you would make available and to whom.

There were 25 respondents to Question 21, although only 20 of these also ticked the '**Other**' category in Question 20.

1) Private companies other than commercial consultants. For example, we provide our statewide data coverage to Idaho Power Company, a company which includes a wildlife division which utilizes our statewide data (the same data we provide to federal and state agencies; no difference). 2) In the past we have provided data to Native American Tribes and hope to do so again through cooperative agreements. The data was clipped geographically for each Tribe (i.e., for each Reservation).
TNC, Audubon Society, and similar organizations who do not do research but are interested in conservation
We typically allow data access under a data license agreement on a case by case basis for particular projects, rather than a blanket use type / user group.
NGB have user authentication and limit access to some data. In principle all data is publicly available and we have on occasion provided a full database dump on request. The main reason behind user access limit is to provide useful, relevant data to different user groups. For example not to overflow outside users with conservation management details. The person detail data is more limited.
This advice comes from our experience of working with a very broad range of data providers who are beginning to make their data holding available through the NBN Gateway. We currently have over 66 public, private and voluntary bodies sharing 160 datasets containing 20,173,556 records. Our data providers determine which of their records are sensitive and control access to them through the NBN Gateway website. A significant number of data providers are only willing to make detailed or sensitive data available to individuals and organisations with whom they have established a formal exchange agreement. In our experience such agreements are established between the data provider and data user directly. The NBN Trust offers advice on drafting agreements but is not involved in actual negotiations. This has been an important lesson for us. The Gateway is a data exchange tool that helps a data provider supply data to a user, but it does not always remove the need for a relationship between the provider and user. For example, 1) a data provider may want to establish some level of support from the user before supplying data. This support may be financial (e.g. either a one off payment or an annual contribution based on a service level agreement) 2) when dealing with environmentally sensitive information a provider may wish to establish a strict non-disclosure agreement with the data user before supplying the data.
The Secretaria de Agricultura, Ganaderia, Pesca y Alimentacion is preparing with the research institutions a list of which species are important to use in commercial interest.

Eligibility of recipients is currently at the discretion of regional biologists outside of our agency (CDC) but within same government Ministry.
Basically if you have a good reason for needing the data, and you are a bona fide user (are respectable), you can have our data.
In principle we can envisage giving data to any of the above categories of user and use, provided that agreements are in place to prevent commercial exploitation without prior consent. We would wish to examine specific users within the categories in order to assess risk of agreements being breached where sensitive data is concerned. In some cases of data we are bound by agreements which prevent us from giving access.
Taxonomic concept and specimen data. To: Economic & ethnobotany research. Horticultural and agricultural research.
Eligibility of recipients is currently at the discretion of regional biologists outside of our agency (CDC) but within same government Ministry.
Research on breeding systems, pollination, dispersal, biogeography, plant morphology, floristic studies.
I would want a request to this institution in any category - not something that happens without my knowledge.
To my knowledge, we would restrict data access only for unscrupulous or illegal activities, or if law or other binding agreements require it. We have not yet defined how we would establish "legitimate use", but it likely includes more than just the categories listed; and hence "Other".
I haven't ticked taxonomic or genetic or molecular studies because I don’t imagine much of a data that we hold would be relevant to these areas. However, the point is that we be happy to provide sensitive data to others for most purposes providing that:- a) They are a bone fide user and can demonstrate that. b) They want to use it for a relevant and legitimate purpose. c) They can meet the requirement to keep the data confidential.
Economic spp. (to any of the categories given suitable arrangements).

Question 22. If you 'ticked' any of the categories in Question 20, how should users be identified and approved?

Total Respondents	**67**
Skipped this Question	35

Comments: (emphasis is mine)

We think there should be different categories of protection. Some data can be made available to the wide public, while other data should be more restricted. Use of very sensitive data should only be approved after signing of a **written agreement**.
Each case would have to be evaluated separately. As a rule, we allow generalized data sets, or restricted data subsets more to be sent out more freely and with less scrutiny than less generalized data sets covering larger areas.
Users make **requests by email** and these are evaluated on a case by case basis. The number of requests we receive is low, so this is a reasonable solution for us.
By a **written request** to our Director of Science and Public Programs
On a case by case basis. Each request would need to be reviewed for scientific need and validity. The **requester** would need to be **certified and validated**. If granted, the access to the data would need to be **controlled by a tight authorization and verification method**.
Direct contact and on a case by case basis. Tell us what you want and if it's "reasonable" we'll give you our data.
Direct contact with the managers of the Avian Knowledge Network.
Government Agencies should be self-evident. Basically, I would want Universities/Research organizations to be entered into a **reciprocity agreement** with us. If they will sensitive share data with us, we will share with them.
1. Detailed explanation of project; 2. certification from three independent referees; 3. **signed agreement** with zero future cooperation clause if violated. Much the same as Qld NPWS expects of applicants for permits to collect.
Users should be identified by name and provide details of their - parent organisation or agency - project aim, purpose, audience - projected outcomes – i.e. reports, publications - preparedness to abide by **data licence agreements**
On a **case-by case basis** currently. Unsure what the proper policy would be.
In general, we provide specific information to government agencies, academia, and NGO conservation organizations and may not require a data license for those entities depending on the circumstances. Commercial consultants usually only get generalized data.
Any federal agency is a potential user. Any state agency is a potential user. Any well-known conservation organization is a potential user.

Consultants are screened pretty heavily. Any potential user that will not sign a **data use agreement** is not approved.
Users needing access to the data for conservation purposes that belong to organisations large enough and permanent enough to be able to be held legally accountable for any misuse of the information. (e.g. not small lakeholder associations directly, consultants)
Ministerial decision centralized mechanism at the institution level
Registered log in with passwords.
Depending on the level of "sensitiveness" identification and approval can be achieved: First of all, users must be identified and approved for access, via sponsorship of their institution (indicated in data agreements or similar official document). Then, access to data could be made by a combination of **username+password+ipaddress** (pool of addresses), for a predetermined period of time. It can also use some type of graphical password (see: http://clam.rutgers.edu/~birget/grPssw/) Finally, in case of highly sensitive data/information a more restricted protocol should be put in place and the use of biometrics cannot be ruled out.
A user should have a **personal username and password**. The users are assigned to groups and given access or presentation template based on the groups where they belong.
Personal user login so we can track and control usage. Discuss with investigator/custodian of the data the access rights for others.
It is the data provider that has to be convinced of the identity of the data user. On the NBN Gateway there are several ways that a provider can identify a user. All registered users have to provide a valid email address before their account is created by the Gateway. This is checked by sending a **launch account link to the email address** they supply. Our data providers have access to online administration tools that allow them to find and see details that we hold on registered users. **A key method of identification is email addresses**. This is not valid in all cases, but for most large organisations all staff have a standard email address. This can be used to check that Paul Smith is actually a member of English Nature for example. We also have organisation accounts on the Gateway and individual users can be made members of these organisation accounts by an administrator appointed from that organisation. For example, If Paul Smith is a member of the English Nature organisation account on the Gateway then this helps confirm that he is a member of the organisation English Nature.
Through **electronic means.**
We would aim to hold a **list of approved trusted partners**. Most of our local partners are well known to us as we work with them on a regular basis. We have not yet established a system for vetting non-local applicants.
In the **Data Sharing Agreements** with our member programs, the member programs specify the level of access for different user groups (gov't, academic, ngo, public, etc) that can be provided without a data license agreement. For all data that are more precise than the specified level, we discuss with the member programs if they'd like to participate in the project, and if so then we create a **data license agreement** with the end data recipient. We are in the process of developing web services that include an online system for managing data access permissions and allowing users actual online access. We'd be happy to talk more with you about our web services work.

The user should submit a formal **on-line request** which details the planned use of the data, and which should include a pledge to acknowledge the source of data. This request should go to the data provider for formal approval.
Making a **written inquiry** about who (researcher), institution, material, objectives of research project.
IP address of requesters should be shown to data providers GBIF should find a way to let data providers include a table of **approved userids**, and data portals should send **userids/IPs** on data requests sent to providers; the software on data providers should have a way of checking the users table before sending the data response.
At the moment, we essentially restrict access to the collection, borrowing of specimens, etc and access to our database to Govt Departments (incl. museums) and to Universities, CSIRO etc in Australia and internationally. There are CITES restrictions in sending specimens overseas on loan.
They should be able to provide, or make accessible, similar or comparable data. In other instances, by assuming certain (reasonable) conditions for use or. In case that any 'gain' is to be obtained from output (funds, curricular including publication, patents), it should benefit the institution or researchers (feed-back mechanisms are admissible, e.g. facilitating data to institutions that cooperated or funded in the past).
Users from such organizations can be asked to provide proof of their affiliations with (a) Government agencies, (b) universities / research organizations, and (c) NGO. Further they can be asked to **route their request through the head of the agencies/institutions** and need to let us know how they would use the data.
We usually grant access to bona-fide researchers, following established current practice of most herbaria, i.e. we need a proven interest concerning the material, usually including a statement from a major professor (in the case of student and Ph.D. work) or institution head.
Proof of affiliation and an accompanying proposal.
Kew would need to give permission based on information such as who they were, who they were working with, and what the intended uses and results were.
Tough call -- we're still sorting that out.
For all categories of Users, the scientists wish in general to know who has accessed their data and for what purpose. Some go further and would like to be asked personally to allow access to "their" data or not. There are some doubts among many researchers for the category "Commercial". However for the forthcoming EU FP7 projects we are warmly recommended to collaborate and find synergies with industrial partners. For these "users" **acceptance of Access and use of data** may be controlled more than for the others.
Public Users means PUBLIC. There are only very limited and special types of data not to be released. Anyways, such decisions are democratic

team decisions one cannot judge alone on.
Currently **on an ad hoc basis**

Question 23. If there are categories you DID NOT 'tick' in question 20, please give reasons 'why not?'.

Total Respondents	**41**
Skipped this Question	61

Reasons:

Valuable books cannot be checked out, they stay in the library. For the same reasons, public and other users should not be granted even limited uses of data unless they visit the source.
I do not feel they need access to the data for the specified purposes.
'Usage data' are restricted to internal use only.
Obviously, because these data are sensitive by definition
Basically, I am not willing to agree to hand over the ability to do any vetting of a request.
We would like to consider each application individually.
Essentially we deal with kindred collections in Govt Departmenmts (incl. museums), Universities, CSIRO etc and have been reluctant, except in rare circumstances, to deal with NGOs and Commercial Consultants. However, we are currently considering placing the bulk of our database on-line for general access, but restrictions on borrowing specimens will still apply.
We can't share data the contributors don't want to share, period, because not sharing them for a limited amount of time (up to two years) is a condition for them contributing the data. This is long-held and widely known organizational policy. I can't just change this policy because I feel like it!
If data are really sensitive, I do not see any reason why anybody other than what I ticked should have access to them. However, on a case-by-case basis others (i.e. NGOs or consultants) could eventually be granted some access under a suitable agreement.
These have to be handled on a case by case basis.
We would prefer to deal with the unchecked categories on a case by case basis based on their requirements and proposed use of the data. Such users could be registered thru the same system but manually.
We might release data to commercial consultants, depending on the nature of the project. All requests are evaluated on a case by case basis.
Non-government institutions are evaluated according to our answer 17. Any commercially oriented consultations are governed by the application of the Rio Convention, and are evaluated individually.
Unpublished data can cause taxonomic chaos; commercial consultants have no ethical track records; public, media information has already

proven to endanger species.
Accountability with the general public and consultants. If consultants are hired by a larger sponsoring organisation, then access to detailed information can be acquired through that relationship as long as the contact individual at the larger organisation is the person held accountable for the use of the data.
Not sure who 'other' would be. Anyone could conceivably gain access to our data for a one-time use, if the use was considered valuable enough and the user agreed to the terms of our data sharing policy.
It would be optimal if non-government organizations and commercial consultants would need to give more detailed information on their project and specific reasons why they need the sensitive information.
We would provide generalized information to the public, but releasing specific location information would be decided on a case by case basis.
Commercial consultants: concern over using the data for commercial gain/profit. Commercial consultants and the public: need to provide interpretation of the data; danger of users drawing incorrect conclusions from the data.
Difficult to explain, many reasons. Some are due to the 'local' (national) circumstances. There is some stress that institutions and researchers should be able to gather the funds they need for research. Under these circumstances, any data derived from research must be set a price, and an income should be expected from data sharing. This is a logical consequence of how the things are working, I do not mean that this is the way I'd like it.
Commercial and general public access would not likely promote conservation.
In general, commercial consultants and general public are not considered bona fide users of sensitive data.
We do not permit our data to be used for commercial purposes.
I would only restrict data that would be given to the general public, mostly because I feel that any licenses, etc. would be impossible to enforce.
Generally we would not provide sensitive data to the public, however, if an individual made a special case - e.g. they were doing some private research or a private development, then I don’t see why they could not be given the data under license.
These general categories have no formal names/background and would be difficult to identify and qualify users.
Requests from the general public are too difficult to verify.
Because I do not think that those combinations of categories normally happen or because it will be difficult to track future handling of the data.
The "merely curious" may intentionally or unintentionally distribute the information. That is a danger in any case, but efforts should be made to limit the possibilities.
"Other" uses would need to be specifically identified and evaluated before we agree to make data available for such uses. "Public" is excluded because access to all our data is on a written, documented need-to-know basis.
Although the data should be available to everybody based on the mission of GBIF, the right of providers should also be protected. In case the

provided data are used for commercial purpose, the value received should be compensated.
We do not currently restrict access to sensitive data, because our general collection data is not yet available on the web. Only our type specimen data is available through the web.
Because the data was acquired with public monies it is by definition publicly available to anyone that requests a copy unless federal laws prohibit it. Federal legislation has not been established that would limit the availability, precision or accuracy of the data, nor to narrow the definition of "PUBLIC".
In all instances, there are no such restrictions or agreements.
No current restrictions.
Not applicable in Palaeontology.

Question 24. Do you currently restrict the textual locality information for 'sensitive' taxa in any way?

Total Respondents	**89**
Skipped this Question	13

Response	No.	Response	No.
Do not make field(s) available	21	Just provide name of township or county	2
Replace wording identifying specific localities	8	Remove GPS locations	1
Set spatial resolution (grid)	3		

Some specific comments

Replace with alternate wording For example, "This specimen represents an endangered or threatened species. The specific locality has been removed from the on-line record to protect this species from over-collection. These data may be supplied to researchers on request."
Locality information is not displayed on the public Tropicos website. However, locality data is currently being provided to GBIF, which we may now be revisiting.
Those interested in further information should contact us or visit the herbarium, and have a reason prepared for wanting information.
Do not make field available - and I hope we also cut out the lat/long data. Most of our lat/long data are so sufficiently inexact that, given we are a mountain region they are not useful for precision work.
Replace 'Locality' statement with 'Nearest named place'. Refer to HISPID3 (Conn, 1996) for definitions.

Questions 25, 26 and 27. Do you currently generalize the georeferencing information (latitude and longitude) before sharing? If you selected 'Yes' in question 25, how do you currently do this?

	Yes	Generalize (explain below)	Randomize (explain below)	Respondent Total
a. Remove altogether	**94% (16)**	18% (3)	0% (0)	**17**
b. Round to 1 minute	**86% (6)**	57% (4)	29% (2)	**7**
c. Round to 10 minutes	**83% (5)**	67% (4)	0% (0)	**6**
d. Round to 30 minutes	**60% (3)**	**60% (3)**	0% (0)	**5**
e. Round to 1 degree	25% (1)	**75% (3)**	0% (0)	**4**
f. Move to nearest named place	**67% (2)**	**67% (2)**	0% (0)	**3**
g. Report by geographic region	**65% (11)**	53% (9)	6% (1)	**17**
h. Report by bioregion	**86% (6)**	29% (2)	0% (0)	**7**
i. Report by standard grid	50% (7)	**71% (10)**	14% (2)	**14**
j. Report by map sheet (explain scale, etc. below)	**88% (7)**	62% (5)	12% (1)	**8**
k. Combination of >1 above (note which, below)	**100% (4)**	50% (2)	25% (1)	**4**
l. Some other method (explain below)	**89% (8)**	67% (6)	44% (4)	**9**
			Total Respondents	**46**
			(skipped this question)	56

Forty Two answered '**No**' to Question 25 with 15 skipping both questions.

Summary:

Response	No.	Response	No.
Report by a geographic region or bioregion	25	Remove altogether	17
Report by standard grid or map sheet	22	Move to nearest named place	3
Round down (to minutes, 10 minutes, 30 minutes, degree, etc.)	17	Some other method (see comments)	8

- A number (more than the 4 that identified doing so) of respondents used more than one method – usually this depended on the level of sensitivity of the taxa (e.g. rounded to 1 minute for some species, 10 minutes for others).
- Very few examples of using randomization were identified (11); while many more examples of generalization (53) were cited.

Please explain in detail how any of these are done

Do not report lat/long, report only county.
When reporting georeferenced coordinates, the coordinates are rounded to the nearest minute with no seconds (00) reported. Depending on the request, an alternative method is to report occurrences by 7.5' USGS topographic quadrangle (or quarter-quad). When provided data as a GIS file, the coordinates are randomly shifted up to 200 m and buffered to create the polygon file which is provided.
UTM rounded to the inclusive 10x10 km square.
Our data providers are able to use online administration tools to set the spatial resolution at which different users can view records held within a dataset. The control affects the detail at which the locality associated with each record within a dataset can be seen and mapped. Resolution can be set at 10 km^2, 2 km^2, 1 km^2 or full resolution. We use the British and Irish national grids as the basis for geographic locations within our records. We divide the Grid up into 10km, 2km, 1km and 100m grids and plot the records within these. This is used to control the geographic resolution at which records are mapped on the Gateway Grid map and Interactive map. The national gird locations within the textual records are reduced accordingly, depending upon the level of access that the user has (e.g. TL02 for a 10km grid reference, TL028300 for a 100m grid reference). Data providers can also flag individual records as confidential within the database that they supply to us. They can then use the online controls to set whether or not a user is able to see these records. This is an on or off control, you either get to see them at their full detail or you don't. This is not a flexible control and we do advise providers to submit sensitive records within a single dataset and then apply the full range of access controls to it.
Generalizing: taking the center (centroid) of a grid square, usually 10 x 10 km squares (UTM grid).
Initially we would provide only a species list with rank information. Location information is very rarely provided.

For the majority of species, a location is generalised to the 1km squares containing it. i.e. in a GIS, the layer of exact boundaries is intersected with the 1km squares coverage to develop tables of the squares that intersect with a particular occurrence. The species name is still associated with the square(s), but it may be at any location within any particular square. For the most sensitive species (three currently), the same process is followed, only using a ten kilometer square grid, and only "sensitive species" is indicated for the name.
We haven't standardised how we generalise. If the location is expressed as a grid reference we would simply change it from say a 6-figure reference to a 2-figure one. eg. SH346190 to SH31 On maps with point data we would require that shaded 1km or 10km (etc) squares were used instead. We haven't specified how this should be done. Most staff here use a standard OS grid. Alternatively, it would be acceptable to shade in a parish or district boundary to indicate vague location - anything really so long as it is not possible to pinpoint the actual location of the site.
(1) Removing seconds/minutes/part past decimal point. (2) Reporting country, state/province, and county only, omitting any other details.
Although we currently remove the entire lat. and long., we would consider any method of generalization or randomization that was agreed upon by the community.
For GBIF data we generalize all georeferences by removing seconds, convert to decimal degrees, and round to 2 decimal places.
We generalize to the nearest 0.1 decimal degree, which equals 6 minutes.
Generalized by including only map name, county name, and/or ecoregion name, and not supplying data in georeferenced format.
This depends on the species. If it is a wetland orchid, e.g., and there in only one wetland in an area, no specific information is given. If the habitat is widespread, a county may be given or large area (Cliffs in the Grand Canyon) (Arizona has many large counties). UTM or Lon/Lat or Township - Range information are not given.
Take a random location, but always use the same random location
We use the three methods listed (geocode removal, generalising to nearest 10 minutes), refer only to IBRA bioregion) depending on the information product requested. Generalising geocodes - by removing the seconds field altogether and the last digit of the minutes field. Randomising - by referring the IBRA region(s) containing the points.
Randomize location with x,y offset with a min and max distance from center
Generalise by rounding minutes to nearest 10; will soon change to rounding to nearest 0.2 decimal degrees (about 12 minutes).

Question 28. What do you see as the main advantages and disadvantages of the method(s) you are using?

Total Respondents	**35**
Skipped this Question	67

Of course, many of these responses require to be tied to the earlier answers (i.e. methodology), but also are valuable stand-alone. To see the interrelationships, refer to the full spreadsheet of responses.

Advantages	Disadvantages
Very general and coarse scale; have not investigated and dedicated much time to really focus on the issue and implement a more elaborate policy.	
It may be not enough as a "generalization" in all cases, but offers a good balance between protection and still making the data available for some purposes such as occurrence within a region, or even coarse distribution studies.	
The main advantage is that the Gateway can hold and share detailed sensitive biodiversity data that would otherwise not be available or known about. The online controls mean that our data providers retain control over access to their data. This gives them more confidence to supply data through the Gateway and negates the need to negotiate detailed access positions ahead of submission. However, we do ask data providers to identify a public level of access and document any access constraints ahead of submitting new datasets.	The main disadvantage is the resource burden of online dataset administration for data providers. Improvements to the form and function of the tools has helped. The administrative resource burden is less for datasets that are made more available to the public. At some level this has worked as an incentive for providers to increase the availability of their data. Data providers are also becoming more prepared to take on the role of administration as the Gateway website becomes more efficient, relevant and useful to them.
Very safe right now, no problem with trust, also very easy to maintain.	But limits our ability to develop mapping applications and our ability to provide wider access to data.
	It could be time consuming and may involve having to redraw maps etc.

Gives some level of protection to the exact species locations, even though these sometimes may be figured out if knowledge of the habitat is known. For most species, this currently isn't a great concern, but if there were no protection, and future experts decided that there was a need for sensitivity, the genie would be proverbially out of the bottle at that point - the exact data would be in the public domain. For the very sensitive species, 10km is a good compromise between planner needs for detailed data (or at least a flag that sensitive data is there) and the need to protect exact locations. Previously, these species were so sensitive to exploitation that it was decided not to map them at all, that the species was better protected by not even allowing planners to have access to the information than it was to risk that the information would get out to collectors.	Makes security more fiddly to set up, incorporating two different levels of security for the data. Issues with portraying records within small natural areas since associating the name with a small area is higher accuracy than would normally be given. Not globally followed, concerns about standardisation for this species across different provincial and national jurisdictions. leads to questions like "Manitoba gives me the data in this format; why is yours so much different and more protective?"
We provide enough info for users to see that we have specimens of the taxa they are interested in. If people want the data they will ask us.	Generalizing locality info is not always effective, especially when you are talking about political units of very small areas (e.g. some island nations). I also think it is worth restating that allowing access to collector names, numbers and collection dates gives determined poachers an avenue to track down localities based on biographical history of the collector.
Generalized data are adequate for most research purposes. Most researchers who need more precise data are data contributors anyway (250 people can get into the password-protected area). Legitimate researchers who really need the exact data can always write to the relevant data contributors, whose names are given right on the relevant web pages.	We could probably get away with giving coordinates a little more precise than the nearest degree.
The main advantage is that it is very clear that the coordinates have been removed. Generalizing or randomizing could lead to confusion for legitimate researchers if the coordinates are not	The main disadvantage I see is that the records can not be easily used for very general distribution maps (eg. county level distributions).

clearly marked as randomized or generalized. It's also easy.	
Simple algorithm applied to all records. Provides about 1.5 miles of error which is a compromise between usable for statistics and unusable for specimen location.	
Bona-fide requests for which we grant access to our data must be treated individually, which results in substantial work-load, depending on the number of specimens involved.	Access to full data is not available instantly but involves a sometimes considerable time delay, which might be cumbersome to some workers.
Only see advantages: generalization creates/retains "true" data, whereas randomization creates deliberately false data. Generalization can be implemented (or not) on a case-by-case basis, depending on the intended use(s) of the data. Generalization is easily implemented by simply omitting data fields containing more precise data, and supplying data in tabular instead of georeferenced format.	
Those interested in only habitat information may obtain it and Presence may be verified by those who need only that information. In the meantime, the plants are protected.	
Brainless. This is an advantage. I guess we could generalize by rounding to 10 minutes or something. I shall be interested in what most people are doing.	
It works. We need to use several rules	
Simple, and provides information that is still useful at medium scales, without giving away the exact location of populations.	
Allows continued use of point data, but introduces error into modeling routines	

Question 29. Do you treat all categories of Sensitive Data the same, or do you have different levels of generalization?

		Response Percent	Response Total
The same		51.2%	22
Have different levels		48.8%	21
	Total Respondents		43
	(skipped this question)		59

About half of respondents that generalize treat all categories the same, while half have different levels of generalization for different categories of sensitivity.

Questions 30 and 31. If you selected 'No' in Question 25, are you likely to generalize the georeference information in the future if a standard method for doing so is recommended?

Total Respondents	**42**
Skipped this Question	60

Comments

We would consider it, but there are likely losses as well as gains for doing so.
Possibly, but this would be considered for sensitive species only. All other records will not be generalised.
If and when we actually have any data on species listed as threatened or endangered.
None of the data are currently available on line, but we shall protect the data if made available.
Very few of our records have georeference.
Only those familiar with the species can judge what information is too revealing.
Provided VERY sound checks and balances are in place to ensure integrity.
We generalise for 'sensitive' species but not for others. I see no reason for us to generalise for other spp.
Not sure. We are obligated to "share" data in the form that we use it and maintain it. It becomes a legal matter--if a requestor pushed, I think we would still have to share the true data, not something we create specifically for general distribution. Again, I am not sure due to the state laws mandating

that we "share" our data with the public.

Questions 32 and 33. Do you see an advantage in all institutions using a standard (i.e. the same) method of generalization if one was recommended?

Total Respondents	**78**
Skipped this Question	24

Comments

Yes - That would certainly be an extremely useful policy.
Yes - Could allow coarse studies, although I assume that for finer research access to the actual data would be granted (?).
Yes it would consistency would certainly be an advantage when using a interface like GBIF to find and access biodiversity information resources from around the globe. A standard approach has been useful for the NBN Gateway. It allows data providers and users to build familiarity with the system and access constraints more quickly. However, it is not possible to please all of the people all of the time. Data providers have asked us to develop additional access controls based on geography, but presently we have decided that expanding our controls is not a priority and would also slow performance of the site.
Yes, too many for this survey. Working with NatureServe on observational data standard currently.
Yes - Can then aggregate data from multiple institutions and analyze and query more efficiently.
Yes unless there was a way of removing the generalisation that could be exploited by "hackers".
Yes - Standardization is a laudable goal in all bioinformatics endeavors. I am sure you can develop some very elegant algorithms, but I

think they would have to be highly parameterized to adequately protect all species in all areas. And remember please that many of us are still working on getting our collections georeferenced to start with!
Yes - It would facilitate actual use of the data (assessing fitness for use) at least some level of precision.
Yes - Guidance would be helpful. Whether organisations wanted to, or could adopt the guidance is a different matter.
Yes - Then it will easier to use data for making comparisons.
Yes - Improved credibility of studies based upon GBIF data.
Yes - Information being withheld by one institution would be withheld by all, but I can see a lot of time and effort being expended at developing the standards. Be sure the moderate restriction that would result is worth the investment.
Yes - A standard method would be very helpful, especially for smaller institutions without the time to evaluate/develop their own standards. But institutions should also be allowed to deviate from the standards, especially to allow greater protection of sensitive species data when desired.
Yes - To share data requires a common standard - that is obvious so generalization methods need to treated exactly the same. After all, you will be using computer programs to filter and display data and therefore the rules must be explicit, accurate and implementable.
Yes - It would allow users some confidence for using the data at particular scales
Maybe - Two levels of dealing with sensitive data would be required by us. Quarantine records must be completely hidden Rare or threatened taxa protected under legislation and Commercially valuable taxa would need to have localities generalized
Maybe - could save a lot of time and (software) development costs. I don't immediately see other reasons, but maybe they're there.
No - If the standard becomes known, then commercial dealers will more easily deduce the exact location we shall try to protect.
No - Whether to generalize should be up to individual institutions, and so should the level of generalization. For example, 0.1 degree precision might get you within the home range of an individual large mammal, which is dangerous, but nowhere close to the only tiny patch inhabited by a really rare plant, which is fine.
No - Only those familiar with the species can judge what information is too revealing.
No - Globalisation in database structures is more and more becoming absurd; the costs far outweigh the benefits
Just don't get too doctrinaire about it. Collections are providing a heck of a lot for free with no return. People and organizations (including NatureServe) can do grunt work, including pleases and thank yous, if they want data. (Bragging rights plus $2 will buy a cup of coffee).

Question 34. GBIF now recommends either ABCD or Darwin Core as data schemas. Are there additional attributes that you would recommend to facilitate the sharing of generalized data?

Total Respondents	**44**
Skipped this Question	58

Responses:

I don't know enough to answer this question. Use of such schemes would have to be explored with user groups and also would require substantial' re-jigging' of our database and accessory programmes for mapping and such like.
For DC: some generalisation data.
Neither schema represents the details in our (observational) data very completely, so more work is needed on schema development or alternative approaches. However, I am unaware of any alternatives that are practical at the moment.
We thought it was odd that the new DC schema was missing a field for original locality description.
• DataRequester IPadress • DataRequester UserID • Information about type of randomization/generalization made by data provider.
Right now, Darwin Core version 1.2 doesn't include paleontologically relevant (and important) fields at all. I know there's a paleontological extension in the works, but I've looked at it and it's super-basic, only including time terms and stratigraphic unit terms. There are many, many other relevant fields, such as geographic and stratigraphic scale of resolution, paleoenvironment, section name and level, and any number of taphonomy and paleontology-specific collection method fields. Basically, I get the feeling that the people who wrote this extension haven't even heard of the Paleobiology Database, the largest database effort in the whole discipline, so they're completely unfamiliar with our fields. Nobody involved with Darwin Core has asked for our help yet.
GeoreferenceIntroducedError - Yes/No Can't see how anything universal for amount of introduced error could be implemented. There's a tension in revealing how accurate or inaccurate a data point is to those who might want to misuse the georeference location. But, for statistical analysis an ambiguously accurate data point is problematic.
Generally, database structures should reflect which fields are "public" and which not. I would, from our own experience, that there should perhaps be 2 fields for detailed locality data, one "private" with the complete data, and one "public" with generalized data. Beware of the possibility to reconstruct detailed habitat data through (1) combination of generalized locality and altitude, and (2)

itineraries of the collector obtained from combining database resources!
Seems like a minimal set of attributes would be: 1. A boolean flag to indicate whether location data are to be considered restricted for a given record. 2. Restricting Institution/Source: reference to a proximal contact, from or through which permission may ultimately be obtained for the disclosure of full location data.
We have had thoughts. Rather than how accurate are our georef data, we record how were they obtained. We could make an estaimate of how accurate this is but what we do not know is the accuracy of the label info. We put all our ecology info into notes I think.
I have no experience at all, don't even know what abcd or Darwin Core exactly mean.
I am not familiar enough with either ABCD or Darwin Core to judge.
Within Darwin core, a standard degradation indicator so that all implementors know the nature of the generalization and can act or program data interfaces to the data in a like minded manner.
No. Darwin Core 2 is fine. A more generalised consultation BEFORE they were formed would have been useful.
The new TDWG observation standard and the TCS are both important and not part of the above.
It is possible to deliver generalised data now with ABCD, but no field allows to specify in a standard way the type of generalisation that has occurred. The element AccuracyStatement is the closest (http://ww3.bgbm.org/abcddocs/AbcdConcept1007). As a free text field we could insert "generalised to nearest 10 minutes", but a categorised field would be required to ensure data of a known accuracy was retrieved.
Generalization method

Question 35 What is the native data format of the geo-referenced data you make (or could make) available?

		Response Percent	Response Total
Point		**77.2%**	**61**
Line		16.5%	13
Polygon		32.9%	26
Grid		20.3%	16
Text only (not georeferenced)		49.4%	39
	Total Respondents		**79**
	(skipped this question)		23

Question 36. Data mining, combining techniques and co-relational analyses may be used to circumvent some generalisation schemes. Can you comment on how to avoid that?

Total Respondents	**39**
Skipped this Question	63

It would appear that a number of respondents mis-understood the purpose of this question.

No - we have no 'generalisation schemes' - what is recorded on input cards is largely what is recorded in the database, though there are some categories that are not entered; however, there is similar general access to the input cards by those who have access to the database. Access is not explicitly regulated other than users signing an agreement relation to input of their data.
Not a clue other than using such a coarse generalisation so as to render the data useless.
Access to data through the NBN Gateway is primarily controlled by spatial resolution. Data providers can use online controls to set the level of access a user has to 10km, 2km 1km or 100m grid squares. These geographical limitations could be circumvented by other information associated with the record, for example a site name. We have built filters into the Gateway access controls that ensure that such associated information is not available to users that do not have full resolution access.
Perhaps the publicly accessible mirror should not contain the ungeneralized data in the first place - i.e., the generalization happens at some stage between a working copy of the data and a publicly accessible copy, not via query of the working copy.
Avoid putting out too many ranks with generalised records. Avoid putting out your data in more than one spatial reference systems that can be compared. Insist on consistency of publication of the data, so publicly available data do not use differing generalisation techniques that permit discovery through co-relational analyses. Avoid randomisation techniques that can be analysed repeatedly to look for clustering of the locations that gives away the true locations.
I have no comments on how to avoid it. Once data became available to open public, we can impose little restriction on how the data are used. To prevent unwanted use, the only way is not to disclose the data.
I don't have any bright ideas on this subject.
Hard to see how that would work with our data, but maybe you know something I don't.
Not without further specific information on the methodology behind such techniques.
Text only, requiring contact with the institution, and presentation of legitimacy seems to me to most reduce the risk.
To prevent intelligent analysis of data to derive location means also omitting data from ALL records that could be used for

correlation/inference: particularly collector, date, elevation, and province/lower political division. This begins to impinge on the usefulness of the data to other biodiversity analyses.
In general, I believe that most fields pertaining to locality and itineraries must be blocked completely. This is why we (with some exceptions) give only collection numbers, but not collector's names.
They will be very refined and expensive analyses of mixed quality data - but bad science frequently trumps good science especially if it comes across as less expensive. Sorry - I am still burning over being told that the trouble with working with taxonomists is that they expect to be funded to go out and collect material - not just use the stuff in the USDA genebank.
Not if the interlocking fields are not available.
I don't understand. Seems to be more of a risk with randomization than generalization.
We have experienced situations where canny users have found records by the same collector with an adjacent collector's number or date, but of a non-conservation taxon, and extrapolated the collecting locality given to apply to a collection of a conservation taxon. The only way to avoid this would be either to generalise all records, or obscure collectors number and date along with geocode. This would leave sheet number as the only unique key reference for each record, but restrict it to a single institution. (Ie. finding all the institutions holding replicates of 'Bloggs 1234' would not be available.) I am also aware of data mining of other Australian herbaria's contributions to GBIF, where that herbarium has not known or flagged the record as a conservation taxon, and so detailed locality statements for a WA endemic conservation taxon have been freely available via the GBIF portal.
Having a standard would help.

Question 37. Do you have any other comments you wish to add?

Total Respondents	**44**
Skipped this Question	62

We clearly need to think through some issues involving our sensitive data, and look forward to seeing the forthcoming recommendations.
There is lot of emphasis on questions of access and data generalisation. Probably these are serious issues for taxonomic groups that have a high public profile. As yet, this does not apply to freshwater fishes in New Zealand, which is the purpose of our database. Initially, this was established to enable the 'harvesting' of data already being collected by agencies meeting their own objectives and obligations, and without the database that we established, such information was in serious jeopardy of disappearing without trace. So, the database stores such information, making it widely available, and over the decades providing an historic record of what was found/observed.
I am eager to see this field as standardised as possible! I would like to avoid reinventing the wheel and would readily adopt/adapt any widely-accepted, sound scheme that GBIF or TDWG may develop to replace my current (quite crude) scheme. Please don't forget to include provisions for data security while in transit to approved users (PGP, RSA, or even Vernam for us security maniacs).
The advice provided comes from our experience of working with a very broad range of data providers who make their data holding available through the NBN Gateway. We currently have over 66 public, private and voluntary bodies sharing 160 datasets containing 20,173,556 records. Our data providers control access to the data they make available through the NBN Gateway website. The NBN provides best practice advice on how access to information should be managed and decided. An important concept is that making biodiversity data available should reduce the risk of damage to the environment. Environmentally sensitive information often relates to those species and habitats that are particularly vulnerable to land management activities. It is important that such information is made available to those that control land management activities at a level of detail that is useful. The NBN does not want to receive generalised data from data providers, but to give them the controls necessary to let them share their data at the level they are comfortable with the public and specific users.
I would emphasize that GBIF should try to improve propagation to data providers of information about WHO requests data. For example, IP address and some kind of userid -which could be previously registered/validated on GBIF portals-, are data that should be included in XML requests that are sent to GBIF data providers from GBIF portals.
We need a single database or web service that compiles the conservation status of organisms across all jurisdictions. In our case, we have this information for Oregon in our database, but not for the other 49 states, nor other countries. For this reason, we excluded all non-

Oregon specimens [<5% of the total] from our GBIF records.
Through user/password authentication, it should be possible for entities to bypass restrictions on location data, if approved, and if each approved entity agrees in writing to implement the same (or more restrictive) policies regarding access to sensitive location data.
We do not currently restrict access to sensitive data, because our general collection data is not on the web. Only our type collection data is available through the web. This issue must be carefully considered before we make our general collection data web-accessible.
Looking at my responses, not one of my most positive days. Seriously - anything proposed - if it is to be adopted must be very simple. And I do want to feel that the person in charge of a collection really does have some control.
The above answers were based on the fungal collection that I am taking care of, not based on all the information in the National Science Museum.
The answer to Qn.36 raises the issue of the need for a 'Consensus Census' whereby at least conservation taxa in all jurisdictions are given a agreed name and some commitment is made by (eg.) GBIF partners to ensure records of these taxa are a) identified with this name
The information provided in this survey applies to data provided via the AVH. MEL also provides data direct to researchers, on request. Requests are assessed on a case by case basis. Sensitive data is often provided in full, subject to standard terms and conditions.
I would object to standardized systems because situations vary so widely. I will offer my "wetlands in Arizona" and an example.
Data I discussed in this survey include only vertebrate fossils..
Thank you for addressing this very important question. For publicly funded institution, there is an increasing pressure to make "everything available", and as far as I followed the discussions involving US herbaria, there might even be laws to this end.
This is a great area for discussion and collaboration. I look forward to seeing the combined results. Chuck Miller.
Thank you!
Thank you for tackling this important issue!

Appendix

Responses to Question 8. When making data publicly accessible do you generalize any fields?

locality limited to county only, no collector names.
specific locality information may be withheld.
For public web site, we do not provide detailed location information, but rather to the county and USGS topoquad level. For professional staff this is made available at the point level.
Georeferences (R-T-S, Latitude/Longitude, UTM, etc.) are scrambled for forest cover records gathered on Privately-owned forest lands (botanical). Locations for these sites are accurate within the context of a County only.
All except feature_id.
many ways including: locality, georeferences, dates, collector's name, taxonomic name, taxonomic level.
georeferences, description of occurrence (if contains georeference), locality.
10-km UTM grid used in lieu of locality.
online administrative tools that our providers can use to control the availability of their own data. These controls can be applied differently to the different individuals and groups that have registered accounts on the Gateway website. Access to data can be controlled in the following ways: i) Set the spatial resolution at which different users can view records held within a dataset. The control affects the detail at which the locality associated with each record within a dataset can be seen and mapped. Resolution can be set at 10 km², 2 km², 1 km² or full resolution. ii) Set whether or not a user is able to download a copy of all or part of a dataset. This will be at the spatial resolution specified for the user above. iii) Set whether or not a user is able to see individuals records flagged as sensitive/confidential. iv) Set whether or not a user is able to see attributes associated with the data in addition to species name, locality, and date. This may include sex, abundance, life stage etc., but excludes recorder and determiner names. v) Set whether or not a user is able to see any recorder and determiner names associated with records. We make a range of tools available on the NBN Gateway that our data providers use to control the availability of their own data resources. These controls are most developed for species records. The controls can be used to set different access levels for the public, specific registered individuals and specific registered organisations. Data providers are able to... a) Limit the resolution of locality for records within a dataset. Resolution can be set at 10km square, 2km square, 1km square and full (actual) resolution. b) Set whether or not a copy of their records can be downloaded from the NBN Gateway website. c) Set whether or not attributes (additional fields of information within records) can be viewed (standard access gives the user the species taxa, the date recorded and the geographic location at the set resolution). d) Whether or not records flagged as confidential or sensitive within the resource can be seen. e) Whether or not the name of original recorders and determiners (if included)can be seen.
locality, dates, collector's name.

Locality.
On the data base that we are including the collection it is all the field cited, and taxon name, one previous determination and who was the specialist that did it.
Georeference.
Taxonomy. We do not release the species name or site description without a data sharing agreement. We do not release any personal names.
Anything describing the exact location of provincially tracked species (1700).
detailed locality information that is more precise than county or watershed.
We would only provide species name and rank information.
Georeferences.
Sometimes we generalize precise location, sometimes we generalize what species the record refers to.
Occasionally we generalize locality, depending on the client and their intended use of the data.
We restrict all fields except taxonomic info and specimen accession number. The idea is just to let researchers know that we have the material, and they can make a special request if they need the data. This applies to State and Federally listed species in the US as well as species on the IUCN red list that are Critically Endangered, Endangered, or Vulnerable.
locality, collector's name, determinator's name.
We don't typically have a problem, but when we do we scale this to the county level (provide the centroid lat long for the county).
character states for taxa.
locality, dates, observers names.
specific nest locations of sensitive species.
For defined sensitive species we blur the grid reference (e.g. only presenting them at 2 figure rather than 6 figure grid ref) or present maps at a smaller scale. Site names may also be altered if they reveal the exact location of a sensitive species.
Locality for some records where the fossils needs to be protected from fossil dealers.
Locality, georeference.
locality, geocodes.
localities, dates.
Lat & Lon limited to 2 digits. locality data is replaced by a restriction message.
each of our records has a flag that can be set to suppress certain data (once set it suppresses that field and everything below it in the hierarchy); these values include region, country, BRU (World Geographical Scheme), Sub-country1, Sub-country2, Sub-country3, geographic area, locality, lat/long, national grid. In addition any taxon whose IUCN ranking is R, V, E or Ex is automatically suppressed at an institution-chosen level. We do not suppress collector name or number under any circumstances. The above practices apply to our herbarium as well as living collections.
locality, georeferences, collector's name.
UNPUBLISHED INTERESTING DATA: only 5 mandatory DarwinCore fields are shown

SENSITIVE FLORA: locality and georeference are omitted.
locality or fine details on location (e.g. with a precision lower than 10 x 10 km) collector's name (dependng on many circumstances such as the degree of 'protection' of the taxon) Date (precise date, e.g. day and month) only to abbreviate, generalisation usually not necessary.
Concerning red-listed species of plants then we generalize the georeferences to the center of the 10 x 10 km grid.
Latitude and longitude minutes, seconds, or values past the decimal point (i.e., everything but an integer value for the degree).
locality, location, dates, collector's name.
For certain types of sensitive data we do not display any locality information lower than county level (including geocoordinates, elevation, precise locality). We also blur the locality information from any images of sensitive specimens we make available online.
We do not provide geo-coordinates. We generalize to district level.
Selectively, on an individual basis, curators may elect to screen specific locality, geographic coordinates, and collector's name.
Georeferences are generalized in GBIF data. Locality data is suppressed on Tropicos public web site.
locality, coordinates, accuracy.
georeference, locality.
For most collections (except those where data has been published), only collection number, national and first-order administrative unit are given, all other data (incl. collectors names) are withhold.
locality, georeference.
locality, georeferences (only for records deemed sensitive).
May suppress collectors name, locality text and georeference data. If mapping may dither co-ordinates or generalise
Many different scales. Some data sets are simply not made available. Some data sets are served with only a very small subset of fields for all records. Some datasets are served with locality data finer than county redacted either for particular records or for all records - redaction of some records is usually based on name lists of threatened and endangered species. Redaction may involve the addition to the publically accessible records of comments in square brackets or may simply be silent removal of data. Latitudes and Longitudes may be truncated (with notations or silently), and georeferencing in other systems (UTM, PLSS) is likely to be silently removed from the data.
Sensitive species data not shown. Message is to contact curator for further information and access is given to legitimate concerns.
Locality georeference.
Sexo,Determinador,Localidad,Fecha,Colector,Foto.
Currently little data is publicly accessible, but we think it likely that we will not make detailed distributional data available for plants likely to be threatened by over-collection, and we will restrict traditional use information.
We only generalize lat/long fields, and only for selected taxa (low percentage).
grids or polygons instead of points, not always because of sensitivity, but because of uncertainties on the exact location.
specific location (latitude / longitude) and/or GIS features directions to the field site owner name survey site name local jurisdiction (locality).

Collector name and locality.
Locality, dates, georeferences, host data, collector's and identifier's names. Full data.
Geocoordinates.
For viewing external to the Department, or for users with an insuffiecient privilege level for viewing our specimen data, we routinely exclude the detailed Locality statement and substitute 'Nearest named place'. If we provide a geocode in these circumstances, then it is 'generalised' to the nearest 10 minutes. (Actually, it is not necessarily the nearest 10 minutes, we simply lop off the last digit of the minutes field prior to delivery).
Locality and georeference.
Threatened and Endangered species.