DATA CITATION
IN THE ELECTRONIC ENVIRONMENT

A white paper commissioned by GBIF

4 September 2008

Meredith A. Lane

Table of Contents

SECTION 1: Philosophy and Future of Citation 1

Introduction.....1

Nature of attribution 2

Data attribution 2

History of citation 4

Data citation at present.....5

Challenges of data citation in a distributed environment 8

Development of data citation standards.....10

Proposed data citation formats.....11

Development of a data citation infrastructure.....12

RECOMMENDATONS TO GBIF 15

SECTION 2: Survey of current practice 19

REFERENCES 21

APPENDIX 1: Survey

APPENDIX 2: GBIF whitepapers on citation (2005)

SECTION 1: Philosophy and Future of Citation

INTRODUCTION

Scientific publication is central to the scientific enterprise, which itself is fundamental to meeting the challenges that society faces. And, the inclusion of a "Literature Cited" in every scientific publication is a tradition that has had an extraordinary influence on the progress of science (Altman & King 2007), although of course citation is not limited to scientific publication alone.

Citation of previous works enables a writer to place the present work in context, and readers to re-trace the history of the ideas and verify the results reported in the present scientific paper or government report. A citation as included in typical publications today has the following purposes:

- Direct the reader to the source of quotes, ideas, data, etc.
    - Makes them recoverable
    - Enables verification of studies
- Give credit to provider(s) for their
    - Ideas, hypotheses, analyses, results
    - Occasionally for providing the data that were analysed

These characteristics of citations, coupled with the longevity ("persistence") of the print record, have given rise to the metric most commonly used to evaluate not only the impact and importance of a given publication but also the overall consequence of a scientist's work (Belew 2005). It is no wonder that scientists are concerned about citation/attribution, linked as it is to their receiving what has come to be called "career credit."

However, even though the open sharing of raw data is quite possibly the best way to ensure repeatability of experiments and analyses (RIN 2008), the electronic publication of datasets has not achieved the same bibliometric status as has the publication of papers in print media (Brase 2004). Nor, on the whole, have

institutions that have been asked to serve science and the public in this way been adequately funded to do so.

At the same time, there is a greater and greater call from governments, intergovernmental organisations, the public, *and scientists themselves* for more and more data to be made Internet-available, not only by individual and groups of researchers, but by entities such as natural history collections, compilers of bird watching data, or conservation NGOs. The only real, existing incentive for individuals or institutions to make data available is to have their contributions acknowledged, to be given attribution/credit for the intellectual effort and expense laid out in developing databases populated with data that enable the furtherance of others' studies.

NATURE OF ATTRIBUTION

Such attribution, for scholarly endeavor that is reported in print media, is given through the citation process. A traditional citation carries "career credit" weight (Brase 2004, Buneman 2005, CLADDIER 2007), because it implies that the cited object

- Can be unambiguously identified,
- Has verifiable provenance and/or authorship,
- Has been peer reviewed,
- Is retrievable in the exact form used previously, and
- Is persistent (has permanency).

Citations with these characteristics are included in citation indices (e.g. Science Citation Index), which provide the fodder for the bibliometric evaluation of the impact and importance of scholarly works—the measure of "career credit" that directly affects scientists in most institutions.

DATA ATTRIBUTION

There is no intrinsic reason that databases published online cannot be included in abstracting services (Pepe 2008). However, for such publications to be included in

citation indexes, there must be standards (Altman & King 2007) for the format of citations of these resources that provide the same kinds of assurances as those provided by citations of print media.

Further, in order to enable database publication citations to carry as much "career credit" meaning as do print citations, there must be in place an underlying infrastructure that imbues the cited database publication with the characteristics listed above (Costello 2008, Jones et al. 2007, Lowry 2008, Paskin 2005, Pepler & O'Niell 2008, RIN 2008, Wallace 2008, and others). Such an infrastructure is inherently possible in a Web environment, although there are significant technical challenges to be overcome (Brase et al. 2005, Buneman 2005, Pepler & O'Niell 2008, Valle 2008).

The clear need to establish mechanisms whereby the data used in a scientific study that is reported in the literature can be retrieved in the same form that was analysed by the authors has led to at least two national-level projects that are addressing these challenges.

1. The German National Library of Science (TIB) project *Publication and Citation of Scientific Primary Data* of the program *Information Infrastructure of Network-based Scientific Cooperation and Digital Publication*, started by the German Research Foundation in response to a CODATA initiative that in turn is related to the 2004 ICSU report "Data and Information" (see Brase et al. 2005).

   In addressing the issue of data citation, the TIB project is relying heavily on the idea of the registration of datasets, which in the registration process are assigned a Digital Object Identifier (DOI), and then recorded in the TIB library catalogue following the ISO 690-2 standard for the metadata obligatory for citing electronic media, together with the Dublin Core metadata attributes that are standard to library practice. Datasets are categorised by the TIB as *citable* or *core.* Both types receive DOIs, but only the *citable* datasets are included in the library catalogue. The TIB began with geoscience datasets but widened their process to other sciences in 2006.

2. The CLADDIER (Citation, Location, and Deposition in Discipline & Institutional Repositories) Project (http://www.claddier.badc.ac.uk), was funded by the United Kingdom Joint Information Systems Committee (http://www.jisc.ac.uk) for support under the 2005 Call for Projects in Digital Repositories to investigate the issue of linking publications held in institutional repositories to the underlying data held in specialist repositories (see CLADDIER 2007, Jones et al. 2007, Pepler & O'Niell 2008).

   This project involved the investigation of the "themes" of citations (Jones et al. 2007) and the means by which data are made permanently available. Issues identified by the authors of the papers cited here are particularly relevant to GBIF's needs and will be discussed further below.

THE HISTORY OF CITATION

Historically, research was conducted by individual scientists, who generated and then analysed their own data. They reported their results in peer reviewed papers that were published in print (persistent) media. In these papers, they placed their work in the context of other scientific investigations by citing relevant publications by other researchers (see Figure 1). The "citation chain" was simple and stepwise. It was from this simple model that bibliometric evaluation of a scientist's career emerged—the more papers that cited a particular paper, the more "impact" the latter had on the scientific enterprise.

The actual citation of prior publications was also simple (see Figure 1 inset). There were a limited number of types of print media: journal articles, book chapters, books. A fairly standard way of citing those works also developed: Author (generator of the intellectual property), date (when was the intellectual property generated), and title of the work. This was followed by an identification of how (from whom) the work could be retrieved. Because the medium itself was the product of a printing house, citing the publisher became the standard way of indicating the source, followed by a "localisation" within the intellectual property resource of the specific ideas being considered (i.e., page numbers).

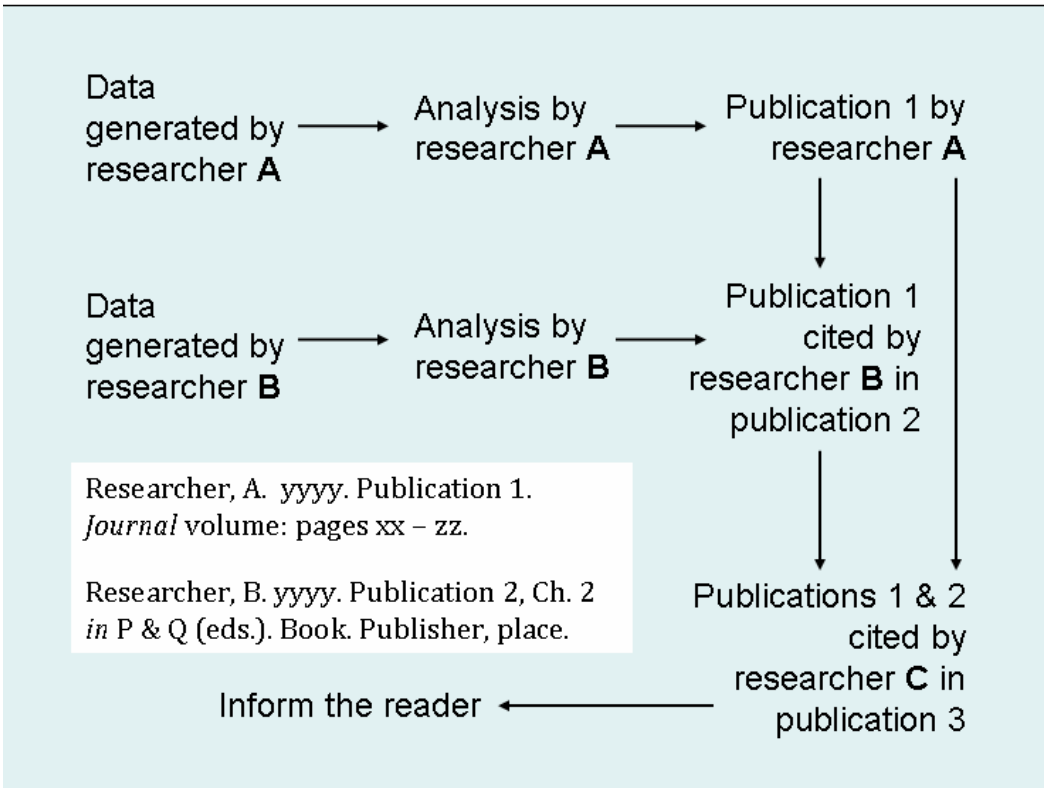


Figure 1. Diagram of a simple citation chain, in which the Literature Cited sections of the publications provides the attribution/credit to previous research, and informs the reader about how to retrieve these resources from a permanent medium. *Inset*: Basic citation formats for typical print publications (these are subject to variation based on the style of different journals, but all contain the same elements). Such formats are familiar and understandable to anyone who has perused or used the print literature any time in recent centuries.

DATA CITATION AT PRESENT

More recently, researchers in some fields (including biodiversity science) have been drawing on data resources provided by others, either in print, online or by special request. The paper that results from the analysis of these data then acknowledges the source of the data, thus providing attribution to an institution or career credit to an individual (see Figure 2). That acknowledgement may or may not be in the Literature Cited; it might instead be found in the text of the Materials and Methods or Acknowledgement sections of the paper, if it exists at all. If it does exist, it is rarely in a form that allows the retrieval of the identical dataset; further, authors may believe

that they have satisfied their duty to give credit by citing a URL, but URLs are notoriously impermanent (Altman & King 2007).

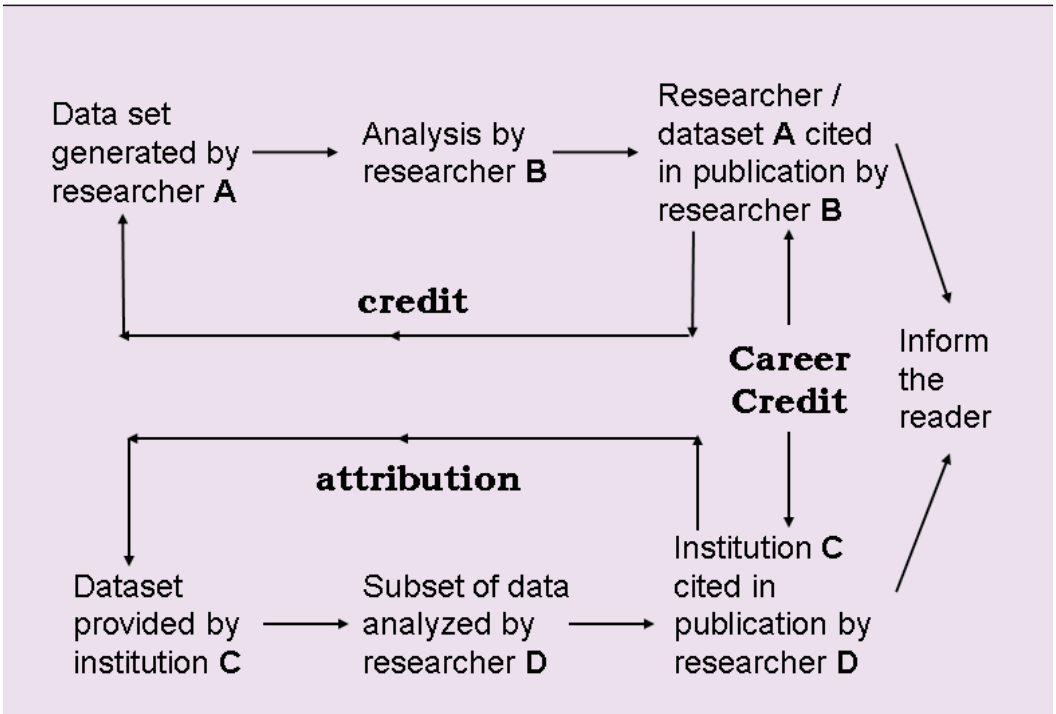


Figure 2. Diagram of a citation chain in which the data analysed and reported upon were not generated by the researcher doing the analysis. Acknowledgement in the resulting publication provides 1) attribution/credit back to the data source (and the publications themselves provide career credit to the researchers) and 2) serves to inform the reader about the source of the data, although the dataset actually used may or may not be retrievable (which means that the current analysis may or may not be verifiable by subsequent workers).

Such "citation" can serve, at least minimally, to inform the reader about how to retrieve the data, and to provide acknowledgement for the scholarly contribution of the data provider(s). What it does less well is to ensure that the cited source is persistent. For example, the person who assembled data in response to a request to his institution may have retired or died without leaving a permanent record in the institution's archive; even had he done so, the ease of retrievability would be minimal. As retrievability decreases, verifiability of the analysis and therefore the scientific value of the work also decreases.

Even the quality of the attribution is reduced. If the author does not provide a copy of the paper to the data provider, that entity does not necessarily know that its support for the scientific enterprise has been utilised, and therefore has less to report to its funding agency about its importance to that enterprise. Of course, staff of the institution can themselves expend time searching the literature for such citations to include in their reports, but this means that fulfilling such requests is a double expenditure (once for filling the request and secondly for hunting for the evidence that this had been done).

These shortcomings are exacerbated when the data source is an online resource, particularly if that resource is dynamically updated and/or growing. Scientific verifiability depends on the ability to exactly replicate the experiment or analysis, which means that the dataset analysed the second (or subsequent) time(s) must be identical to the original. Independent reconstruction of datasets for purposes of analytical replication is not possible (and therefore verifiability is lost), unless

1. provision has been made for a mechanism that can reach back into the past of the database and retrieve only those records, unaltered by updates, that were in a dataset constructed from a subset of that database *at the time of original retrieval*; or

2. the dataset actually analysed is stored in an unchanging state in a trusted repository, thus making it persistent.

Some journals (e.g. *Science*) are providing online storage space for authors to provide "supplemental materials," including raw datasets, for their papers published in that journal. This helps to achieve the goals of persistence and accessibility, but it typically renders the data static (in PDF) rather than electronically retrievable and interoperable, which is needed for re-use of the data, a capacity that is as desirable as repeatability of past analyses.

CHALLENGES OF DATA CITATION IN A DISTRIBUTED ENVIRONMENT

The challenges of citation/attribution that are introduced when a single dynamic database is made available online are both magnified and added upon if one provider makes several databases (or "citable datasets" in the sense of Brase et al. 2005) available. Yet more complications arise when data from multiple providers are made available through one Internet access point by an "aggregator," such as a World Data Center or other repository such as the British Atmospheric Data Centre (BADC), which is participating in the CLADDIER project. Datasets that are obtained by issuing queries to an "aggregator" data portal are *typically* going to require complex citations in order to satisfy the needs of the institutions and individuals who share data.

The Global Biodiversity Information Facility (GBIF) is an "aggregator" portal; indeed, it is an aggregator of aggregator portals (see Figure 3). As such, it has set itself a goal more difficult of achievement than was envisioned when the first Memorandum of Understanding (MOU) was written. The founders of GBIF, cognisant of the needs of data providers, included in the operating principles of the organisation the assurance that they would receive attribution/credit. The GBIF MOU Paragraph 3 reads in part:

> *The purpose of GBIF is to promote, co-ordinate, design and implement the compilation, linking, standardisation, digitisation and global dissemination of the world's biodiversity data, within an appropriate framework for property rights and due attribution.*
>
> and
>
> [GBIF will]*..... make biodiversity data universally available, while fully acknowledging the contribution made by those gathering and furnishing these data.*

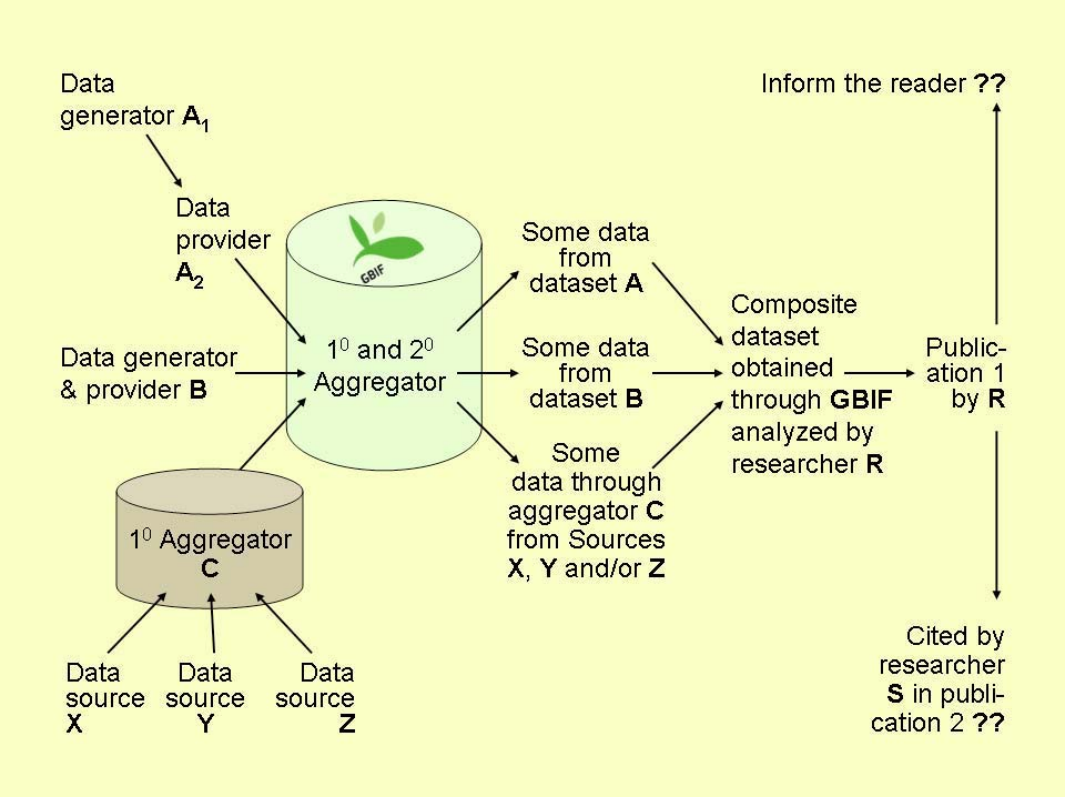


Figure 3. Simplified diagram of possible chains of provenance (the description of the origins of data and the process by which they arrive at a given dataset, see Simmhan et al. 2005) for data that are mediated by GBIF, which can be considered to function as an "aggregator" of data sources. The 1^0 aggregator indicated on the left of the diagram (an example is HerpNET) may of course be drawn upon directly as a portal through which data are obtained by a researcher, although for simplicity's sake that is not shown as part of the diagram. In a situation such as that diagrammed here, the data are made easily accessible to researcher R, but the chain of attribution/credit that should be cited may be difficult to construct, or exceedingly bulky, or both. Even if researcher R manages to do so in a manner satisfactory to both readers and data providers, how should downstream citations concerning R's data, such as S's, be constructed?

GBIF has long recognised the need to develop a standard format that will allow authors to acknowledge contributions by those who share data (data originators, data providers) as well as GBIF itself. However, neither GBIF nor its partners have yet established a standard format that adequately satisfies the weighty requirements for an adequate "data citation" (see below), particularly one that accommodates the multiple steps in the chain of provenance of data as well as the likelihood that datasets constructed by queries of the GBIF Data Portal will include only some

records from each data source rather than entire datasets, as is presumed by both the TIB and the CLADDIER projects.

DEVELOPMENT OF DATA CITATION STANDARDS

GBIF and its partners are not alone. There are no standards for such complex citations yet developed in what is a rapidly evolving electronic world – but that is not to say that such standards are not urgently needed. The universally accepted general format that developed for publications in print media, no matter how it is tweaked, will *not* serve for citations of electronically published data.

The CLADDIER project (CLADDIER 2007, Pepler & O'Niell 2008) identified requirements for the functionality of data citation from the perspectives of both the users (items 1 - 4) and the producers/providers (items 5 - 7) or both (item 8) :

1. Unambiguous
2. Refers to a permanent entity—the repository (= "provider" in Figure 3, above) is as important as the producer (= "generator" in Figure 3, above)
3. Understandable by humans (author and year of publication, or equivalents)
4. Includes reference to the manner in which the data were produced – that is, must account for
    - Granularity (data record, dataset, database)
    - Incrementality (updates/revision, growth)
5. Producer must be traceable
6. Accommodates usage metrics
7. Makes it possible for machines to search for papers that cite the data
8. Renders published datasets equivalent to published papers

A survey of 45 websites (see Appendix 1) that serve scientific data, including GBIF's own (see Appendix 2) was conducted to ascertain the current primary data citation methods and practices of various organisations and disciplines. The results are summarized in Section 2 of this paper, with regard to the functional requirements discussed in the previous paragraph. In short, all of the current practices are clearly derived directly from print citation formats, and fall short of providing the

functionalities described in the previous paragraph and demonstrated as necessary in the discussion of Figure 3.

PROPOSED DATA CITATION FORMATS

Altman and King (2007) have proposed that a data citation should comprise 6 components (3 understandable to humans and 3 to machines):

- Author of dataset
- Date the dataset was first published on the Internet (not the date it was accessed by the researcher who is citing it)
- Dataset title
- Globally unique identifier
- Universal numeric fingerprint (UNF)
- Bridge service (e.g. PURL)

The addition of the machine-readable components begins to move a citation away from the standard format for a print citation, but the usefulness of the formulation of Altman and King would be limited to those datasets that are re-used in their entirety. While that kind of dataset does exist and is used in that manner in some disciplines, the data that are mediated by GBIF are handled at the level of database record rather than dataset (though the result of a query can be considered a dataset that would be analysed as a whole), and thus this formulation is unsatisfactory for GBIF's purposes. As noted by Buneman (2005), citation formulations must take into account other questions than "What form (human- or machine-readable) should citations of data take?":

- How can particular records that are extracted from one database and inserted into another be "localised" within the first for purposes of citation?
- How can the composition of a dynamic database at a certain point in time *in the past* be retrieved?

Another attempt (Jones et al. 2007) addresses some of these concerns. A data citation as envisioned by these authors would comprise:

- Author (including both persons and institutions/organisations)
- Title (clear, human-readable identification of the data resource)
- Edition / Version
- Publisher (the institution / organisation responsible for maintaining the primary copy)
- Date of publication, stated as "year onwards" to indicate dynamic nature of data resources
- Date of revision/update or access
- Availability (URL)

Although Jones and colleagues have also started from a print-citation basis (Patrias 2001), their formulation at least attempts to give attribution to both individuals and institutions, to acknowledge that the producer and the provider may or may not be the same entity, and that datasets are likely to be evolving so that a citation must indicate the time at which a snapshot was taken. It does not, however, provide for "localisation" of parts (records) within a database, analogously to citing particular pages within a book or article. Most importantly, it does not include the globally unique identifier (GUID) that is required to render a citation completely unambiguous and aid in retrievability in the electronic world (Brase 2004, Brase et al. 2005, CLADDIER 2007, IOCCP 2006).

DEVELOPMENT OF DATA CITATION INFRASTRUCTURE

The development of standards and formats for data citation is not something that can be undertaken without *also* adjusting the information infrastructure to

1. assure persistence of the resource (Altman & King 2007, Buneman 2005),
2. facilitate dataset retrieval and interoperability (Jones et al. 2007, Paskin 2005),
3. understand data provenance (the ultimate and successive sources of the data and the transformational processes through which it has passed into the proximate dataset) and attribution tracking (Science Commons 2008, Simmhan et al. 2005, Groth 2007), and

4. make adherence to the standards not only possible but something that authors will in fact do (Lowry 2008).

Thus, it will only be possible for data publication citations to be included in the Science Citation Index and other bibliometric sources and therefore carry the same weight as publication citations with regard to "career credit" for scientists (Lowry 2008, Groth 2007, RIN 2008) and attribution for institutions, if both infrastructure and standards are developed.

This development process will involve:

1. A move away from copyright, even existing Creative Commons licenses, because the IPR protection applied to databases because these carry the problems of copyrighted works into an area where in most cases these do not exist (Science Commons 2008). Science Commons "are launching certifications of the Open Data Commons License (which came out of the Talis efforts, and was drafted by Jordan Hatcher) and the CC0 legal tool. [They] are now developing a certification process for submission of conforming implementations and expect to release that process in 2008."
2. Involvement of learned societies (RIN 2008) in the development of
    a. citation formats
    b. peer review mechanisms for data publications
    c. determining a "least publishable unit" for data publication
3. Clear distinction that there are different types of data and therefore datasets (RIN 2008), and so the citation format and infrastructure must be flexible enough to accommodate them:
    a. Experimental
    b. Modeling and simulation
    c. Observations of time and location
4. Assurance of the persistence (RIN 2008) of data/datasets, through one or the other or both:
    a. Centralised data repository/ies with expertise to curate and migrate the data as technologies change, or

   b. Funding to provide for such expertise at the local level.

5. Technological incorporation, among other things, of
   a. Some method of keeping a record of the ultimate derivation and passage of an item through its various "owners" (attribution tracking, per Science Commons 2008), probably either by annotation (Bhagwat et al. 2005), data provenance tools (Simmhan et al. 2005, Groth 2007), and/or "trackback" (Jones et al. 2007).
   b. A way to capture (snapshot) and preserve the status of a dynamic data system at a specified moment in time (Buneman 2005), and/or to automatically provide/preserve versions (CLADDIER 2007).
   c. Mechanisms for assigning and resolving GUIDs to data objects as citable and persistent identifiers at appropriate levels of granularity (Pepler & O'Niell 2008); these might include DOIs (Brase 2004, Brase et al. 2005, IOCCP 2006, Paskin 2005), Life Science Identifiers (LSIDs), UNFs, URNs, or a combination of these (Altman & King 2007).
6. Requirement by publishers, editors and reviewers that underlying data be published and cited (Costello 2008, Wallace 2008), probably by incorporating a mechanism similar to that used in the molecular community (a GenBank sequence number must be reported in publications), and that there be a "live" link to that data publication cited by subsequent users of those data.

RECOMMENDATIONS TO GBIF

1. **Short term**

   GBIF could explore the use of a more complex citation style than the one that it currently proposes as a part of its Data Use Agreement, but in so doing would not on the whole achieve a result that is any more satisfactory.

   The shift to something that approaches more closely the functionalities that have been ascertained as necessary to appropriate data citation in a distributed database environment is both a medium - and a long -term proposition (see next recommendations).

   For the short term, GBIF should continue to remind users of data that they have an obligation to give credit generously (Wallace 2008) to the providers of data, and to assist them to cite the data within the constraints imposed by the lack of infrastructural developments that enable and facilitate it.

2. **Medium term**

   Within the next 18 months to two years, however, there are steps that GBIF could take to assist users to give proper attribution to data providers. These suggestions arise from conversations held with a number of biodiversity and environmental scientists, especially Drs. Town Peterson, Jorge Soberón, and Robert Buddemeier, and with GBIF's own informatics expert, Tim Robertson.

   1. GBIF already has, in effect, a mechanism for making a "snapshot" of its "database" that is characterised by a point in time. The index database that facilitates searching across the GBIF network is updated approximately monthly. An archive of each update is kept; any search conducted between updates is in effect permanently associated with a particular version of the index database. These archived versions, clearly labeled as such, need to be made web-accessible, with assurance that that accessibility is persistent (GBIF may wish to collaborate with a trusted repository).

2. Currently, users of the GBIF Data Portal need only click once to accept the Data Use Agreement before they are permitted access to the search portal. This is in contrast to many portals and websites from which data are viewable only after registering as a user of the site, but it need not change for the majority of visitors to the GBIF portal. However, GBIF should consider instituting a registration mechanism for those who seek to download datasets for analysis and publication.
    a. This mechanism would provide metadata (including the database version that was queried) to be associated with query strings.
    b. Queries issued against the index following registration should be captured, associated with the registration metadata, labeled with a GUID, and stored in a database of such queries. GBIF will need to assure that this database is accrete-only and persistent (again, collaboration with a trusted repository may be required).
    c. The GUID (and any information necessary to make that GUID resolvable) issued to each query by a user should be automatically provided to that user for use as a citation link in any subsequent presentation, analysis or other use of the dataset obtained.
    d. These GUIDs should also be contained in GBIF's regular reports to its data providers if the particular query called on their resources. This would allow them to monitor at least electronic publications for the GUID, and thus know when their data were being cited/used, without a great expenditure of human time spent searching.
3. When the GUID is resolved, the person should be presented with a human-readable metadata screen that provides the characteristics of the query, the database version that was queried, and other pertinent details such as taxa included, number of records returned, data providers whose databases were tapped, etc.
4. From this screen, should the user decide to do so, it should be possible to run the identical query against the identical version of the GBIF index

addressed by the original query and thus produce a dataset identical to that obtained previously.

This mechanism addresses most of the citation-function concerns raised by CLADDIER (see p. 10, above):

a. The GUID label makes it an unambiguous reference to a unique dataset.
b. It enables repeatability/verifiability by making it possible to recreate a dataset that resulted from a query at some point in the past.
c. It refers to a permanent entity (the archived version of the index database), particularly if GBIF collaborates with a trusted repository, thus giving the citation a weight that approximates that accorded persistent print media.
d. It includes reference (via the metadata screen) to the manner in which the dataset was produced that makes transparent both the granularity and incrementality of the query and the result.
e. The data producers (providers) are traceable.
f. It accommodates usage metrics, because the format of the GUID can be such that
    i. it will be possible to perform machine-searches for papers that cite the data, and
    ii. resolving the GUID will reveal the data sources.

With this mechanism in place, GBIF should then be able to formulate a bibliographic citation format that has both human- and machine-readable components (Altman & King 2007), as well as a set of simple rules for data users to follow to construct the data citation to use in their publications (Lowry 2008). Such a process would both encourage data citation (Costello 2008) and assure due attribution to data providers (as per the GBIF MOU).

3. **Long term and ongoing**

GBIF should engage with its partners, the learned societies, other scientific disciplines and libraries to promote the development of an information infrastructure to facilitate fully functional data citation with the characteristics described herein.

GBIF should take an active role in this process, because the kind of data sharing that it promotes is more distributed than that of dataset repositories, more fine-grained in its data selection than many disciplines, and more complex in its data provider relationships. Because of these things, it has a greater stake in developing a mechanism for attribution and citation than many of the other players.

Because GBIF has already spearheaded the development of some aspects of the information infrastructure that will be needed, it is uniquely positioned to lead the developments of this next aspect of that infrastructure. And, because its own founding document states that it will ensure proper attribution to those who provide data, GBIF is obligated to do so. It cannot build the infrastructure on its own, but it is uniquely suited to be central to the task.

SECTION 2: Survey of Current Data Citation Practice

A survey of 45 websites was undertaken to ascertain the current practices for data citation encouraged by those data providers (see Appendix 1), which were drawn from a spectrum of scientific disciplines, including astronomy, geology, medicine, molecular biology and biodiversity science (both organismal and ecological).

Of the 45 sites, 41 are portals to data or information provided by other source(s), and 4 were evaluated as the direct provider. Looking at these resources in another way, 32 are making *data* available (from single or multiple sources), and 13 are portals to *information*. Among these sites, GBIF is unique in its mechanism of interoperability—a GBIF search taps all appropriate data providers and returns the data to a single screen. No other aggregator site surveyed currently allows this type of searching, but rather requires the user to peruse a dataset at a time.

Various forms of assertion of intellectual property rights using a legal term are indicated by 25 of the sites, although most request attribution for use, and 5 request notification of such use. Among the aggregator sites, 24 request citation of their data providers and 23 request citation of the aggregator site itself (these sets are overlapping but not coincident). Among those that request citation, only 15 provide a format for doing so (AmphibiaWEB, BIRN, CRBIF, FISHBASE, FLYBASE, GBIF, GBIF-ES, HerpNET, LTER, MorphBank, NatureServe, OBIS, REEFBASE, SCAR-MarBIN and TOL).

There is no standard citation format among those provided (among 15 providers, there are 11 different formats suggested), although all suggestions are clearly attempts to adapt the universally standard print citation format to the purposes of citing an electronic resource. Table 1 summarizes the citation formats suggested for "unitary" database sites (the aggregators' individual data sources), and Table 2 those for data or datasets obtained via the aggregator.

Of the formats in current use (including GBIF's own, see Appendix 2), none satisfy the functional requirements for a data citation as discussed in Section 1.

TABLES 1 and 2: Generalized citation formats summarized from those provided by the web sites surveyed (see Appendix 1). In actual citations, table cell labels are replaced by the appropriate information (values), unless the words used here are in italics, in which case those words actually appear. "Kind of publication" is intended to indicate media used (World Wide Web, electronic, etc.).

Table 1. Generalized citation formats: Unitary Database Sites

1	Person(s)	Year	Database name	Kind of publication	URL	Date accessed	
2	Person(s)	Year	Database name	Version	Date accessed	Kind of publication	*Available online:* URL
3	Person(s)	Dataset	Data publisher	URL	Date accessed		
4	Database name	URL	Date accessed	Database owner			
5	Database name	URL	Database owner	Date accessed			
6	Database name	Kind of publication	Year	Data owner	*Data available*: URL	Date accessed	

Table 2. Generalized citation formats: Data or information from Aggregator Sites

1	Person(s) or institution	Database name	Date accessed	URL				
2	Person(s)	Database name	*Retrieved* date	*from* aggregator's URL				
3	Type of data	Data Provider X	Data Provider Y	Data Provider….*n*	*Accessed via* aggregator URL	Date accessed		
4	Person(s)	Year	Database name	Version	Date accessed	Database URL	Aggregator name	Aggregator URL
5	Some aggregators surveyed suggest that rather than trying to cite data sources a bibliographic citation format, users of data obtained from them should describe this in the Materials and Methods section of their publications *and* in the Acknowledgement section, where not only the Aggregator itself but also its funders should be mentioned.							

REFERENCES

Altman, M and GA King. 2007. A proposed standard for the scholarly citation of quantitative data. *D-Lib Magazine* 13(3/4) March/April 2007. Available at http://ssrn.com/abstract=1081955

American Geophysical Union. 2006. Importance of archiving and availability of geophysical data. Statement adopted by Council 29 May 1997 (Reaffirmed May 2001, May 2005, May 2006). Available at http://www.agu.org/sci_soc/policy/positions/datapos.shtml

Anonymous. 2006. DOIs for datasets. *In* the International Ocean Carbon Coordination Project (IOCCP) Progress Report. Available from http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=2081

Belew, R. K. 2005. Scientific impact quantity and quality: Analysis of two sources of bibliographic data. arXiv:cs/0504036v1 [cs.IR].

Bhagwat, D., L. Chiticariu, W.-C. Tan and G. Vijayvargiya. 2005. An annotation management system for relational databases. *VLDB Journal* 14: 373-396. 10.1007/s00778-005-0156-6. http://www.springerlink.com/content/y473517301028ln4/

Brase, J. 2004. Using digital library techniques – Registration of scientific primary data, pp. 488-494 *in* Heery, R. and L. Lyons (eds.). ECDL 2004, LNCS 3232. Springer-Verlag. http://www.springerlink.com/index/1PGLBMJV95TQBY9E.pdf

Brase, J. *et al*. 2005. Standards for the publication of scientific data by World Data Centres and the National Library of Science and Technology in Germany. http://www.kbs.uni-hannover.de/Arbeiten/Publikationen/2005/brase_jcdl2005.pdf

Buneman, P. 2005. What the web has done for scientific data—and what it hasn't. *In* Fan, W, Z Wu and J Yang (eds.). Proceedings: Advances in Web-Age Information Management: 6th International Conference, WAIM 2005, Hangzhou, China. http://www.springerlink.com/index/pu6861u4288506t0.pdf

Buneman, P. 2006. "How to cite curated databases and how to make them citable," *ssdbm*, pp. 195-203, 18th International Conference on Scientific and Statistical Database Management (SSDBM'06). http://doi.ieeecomputersociety.org/10.1109/SSDBM.2006.28

Buneman, P, S Khanna, K Tajima and WC Tan. 2004. Archiving scientific data. *ACM Transactions on Database Systems 27: 2-42*. http://portal.acm.org/citation.cfm?id=564693

CLADDIER (Citation, Location, and Deposition in Discipline and Institutional Repositories) Project Team. 2007. CLADDIER briefing on data publication. http://claddier.badc.ac.uk

Costello, M. 2008. How to motivate scientists to publish data online. PowerPoint presentation at SCOR/IODE Workshop on Data Publishing, June 2008. Available from http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=2253

Emerson, C. 2008. Connecting researchers with data. PowerPoint presentation at SCOR/IODE Workshop on Data Publishing, June 2008. Available from http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=2257

Groth, PT. The Origin of Data: Enabling the determination of provenance in multi-institutional scientific systems through the documentation of processes. Thesis, Faculty of Engineering, Science and Mathematics, University of Southampton. http://eprints.ecs.soton.ac.uk/14649/01/ThesisSubmitted.pdf

IOCCP (International Ocean Carbon Coordination Project). 2006. Progress Report: DOIs for datasets. http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=2081

Jagadish, HV and F Olken. 2004. Database management for life sciences research. *SIGMOD Record* 33: 15-21. http://portal.acm.org/citation.cfm?id=1024697

Jones, CM, KA Bouton, JMN Hey, SE Latham, BN Lawrence, BM Matthews, AJ Miles, S Pepler, and K Portwin. 2007. Data publication outputs of the CLADDIER project. In Proceedings: Ensuring the long-term preservation and value adding to scientific and technical data (PV2007), Munich, Germany, 9 – 11 Oct 2007. http://epubs.cclrc.ac.uk/bitstream/1838/PV2007_Jones_CLADDIER-final.pdf

Lagoze, C, H van de Sompel, M Nelson, and S Warner. 2004. The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0 of 2002-06-14. Document Version 2004/10/12T15:31:00Z http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

Lowry, R. 2008. Citable data publication: Objectives, Status and Issues. PowerPoint presentation at SCOR/IODE Workshop on Data Publishing, June 2008. Available from http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=2130

Paskin, N. 2005. Digital object identifiers for scientific data**.** *Data Science Journal* 4: 12 – 21. http://joi.jlc.jst.go.jp/JST.JSTAGE/dsj/4.12?from=Google

Patrias, K. 2001. National Library of Medicine recommended formats for bibliographic citation. Supplement: Internet citation. U.S. National Library of Medicine. http://www.nlm.nih.gov/pubs/formats/internet.pdf

Pepe, R. 2008. Aquatic sciences and fisheries abstracts (ASFA) database. Document for SCOR/IODE Workshop on Data Publishing, June 2008. Available from http://www.iode.org/index.php?eventID=273&option=com_oe&task=viewEventRecord

Pepler, S and K O'Neil. 2008. Preservation intent and collection identifiers. *In* CLADDIER Project Report II (Identifier Migration Issues for Repositories). http://epubs.cclrc.ac.uk/work-details?w=43640

RIN (Research Information Network). 2008. To Share or not to share: Publication and quality assurance of research data outputs (Executive Summary). Report prepared for SCOR/IODE Workshop on Data Publishing, June 2008. Available from http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=2127

Science Commons. 2008. Database protocol. http://sciencecommons.org/resources/faq/database-protocol

Simmhan, YL, B Plale, and D Gannon. 2005. A survey of data provenance in e-science. *SIGMOD Record* 34(3): 31-36. http://sigmod.acm.org/sigmod/record/issues/0509/p31-special-sw-section-5.pdf

Valle, M. 2008. An introduction to scientific data management. http://personal.cscs.ch/~mvalle/sdm/scientific-data-management.html (accessed 1 Sep 2008).

Wallace, D. 2008. The risk-reward basis for data publication. Document for SCOR/IODE Workshop on Data Publishing, June 2008. Available from http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=2090

INTERVIEWS

A. Townsend Peterson, PhD
Dyche Hall, room 713B
1345 Jayhawk Boulevard
Lawrence, KS 66045-7561
Telephone: 785 864 3926
Email: town@ku.edu

Jorge Soberon, PhD
Dyche Hall
1345 Jayhawk Boulevard
Lawrence, KS 66045-7561
Telephone: 785 864 4540
Email: jsoberon@ku.edu

Robert W. Buddemeier, PhD
KS Geological Survey
414 Moore
Lawrence, KS 66045
Telephone: 785 864 2112
Email: buddrw@ku.edu

Tim Robertson
GBIF Secretariat
Universitetsparken 15
2100 Copenhagen
Telephone: +45 3532 1470
Email: trobertson@gbif.org

PERSONAL COMMUNICATIONS

Robert Penn Guralnick, PhD
265 UCB, Museum-Administration
University of Colorado
Boulder, CO 80309
Email: robert.guralnick@colorado.edu

William Piel, PhD
170 Whitney Ave
Yale University
New Haven, CT 06511-8902
Telephone: 203 436-4957
Email: william.piel@yale.edu

SURVEY OF CITATION PRACTICES OF ONLINE, PUBLIC DATA AND INFORMATION RESOURCES

Forty-five online resources are represented in the following pages by screenshots of the headline (banner) of the home page of their websites, followed by screenshots of citation-related information on pages elsewhere in the sites.

Summary characteristics of the surveyed websites*:

Type of content

Website of data originator	4
Web portal to scientific data	28
Web portal to scientific information	13

Scientific discipline

Astronomy	1
Biodiversity (organismal)	31
Ecology / Environmental	9
Geology / Paleontology	3
Molecular / Genetic / Medical	8

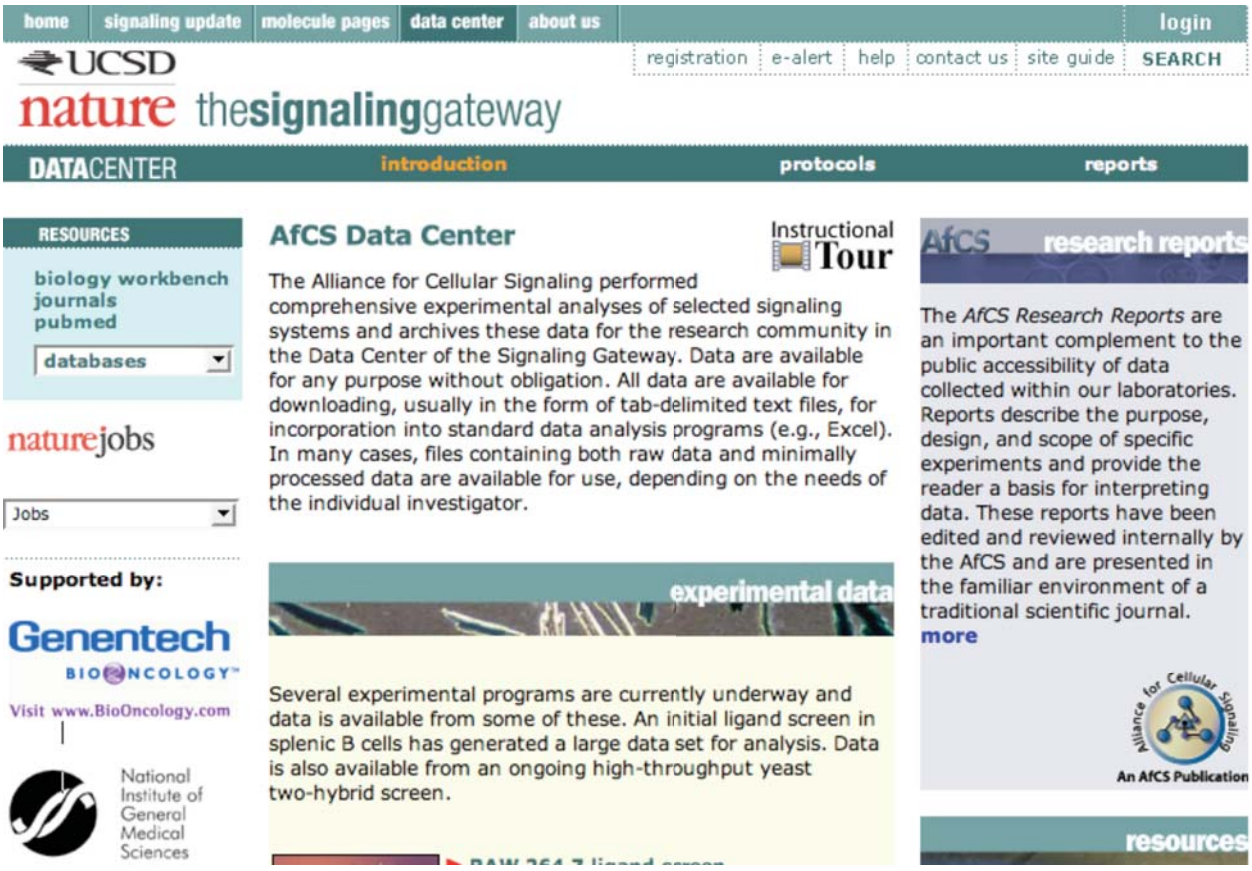
IPR assertion (by actual term used)

None or none stated	20
Public Domain	4
Open access	2
Fair Use	3
Copyright	
Copyright	10
Creative Commons	4
European Union	2

* Numbers in the tables do not necessarily total 45, because some sites may have been scored in more than one category

ALLIANCE FOR CELLULAR SIGNALING DATA CENTER

http://www.signaling-gateway.org/data/Data.html



ALA

http://www.ala.org.au/

Copyright

Copyright on materials available at this site is owned by the participant organisations belonging to the Atlas of Living Australia.

search now

Queensland Government

Copyright Statement

The Queensland Museum retains ownership of all material on this site unless otherwise specifically stated. Permission is granted to copy material owned by the Queensland Museum under the terms of the Copyright Act 1968 (as amended) and in accordance with fair dealing provisions; provided that the content is not altered and the source is acknowledged.

Contact details

If you have any questions about our copyright statement or practices, please contact:
The Copyright Officer
Queensland Museum
PO Box 3300
South Brisbane Q 4101
Australia

Fax: 61 (0)7 3846 1918

Email: copyright@qm.qld.gov.au

Queensland Museum
About Us
Calendar
Education
Foundation
Inquiry Centre
Loans
Media Centre
Publications

Museum Campuses
Brisbane, South Bank
Queensland Museum
South Bank
Brisbane, Woolloongabba
Museum of Lands
Surveying and Mapping
Ipswich
The Workshops
Rail Museum
Toowoomba
Cobb+Co Museum

AMPHIBIAWEB

http://amphibiaweb.org/

A M P H I B I A W E B

AmphibiaWeb is an online system that provides access to information on amphibian declines, conservation, natural history, and taxonomy.

See what's *New* on AmphibiaWeb.

- Search the Database
- Browse by Family
- Country Search
- Amphibian Declines
- Site Map
- About AmphibiaWeb
- Make a Donation

amphibiaweb

2008 - Year of the Frog

Species of the week:
Centrolene ilex

Photo by Twan Leenders

Current number of amphibian species: 6,367 (Aug 25, 2008)

Related projects:

A M P H I B I A W E B

Home
Search the Database
About Amphibians
About AmphibiaWeb
Taxonomy
Links
Use of Data
Site Search
Site Map

Using Data From AmphibiaWeb

Restrictions on Use: Permission to use, copy, or distribute small amounts of information or data delivered from this server is hereby granted under the following conditions:

1. Data available on AmphibiaWeb is intended for personal, academic and other educational purposes and should not be used for any commercial purposes;
2. Data available on AmphibiaWeb must not be used in other web applications without specific permission;
3. AmphibiaWeb must by cited in all copies (see citation format below);
4. Some data may be downloaded to files and altered in format for analytical purposes, however the data should still be referenced using the citation as formatted below;
5. Users may not download multiple images or large amounts of data (including text, maps, sound files, and video) from AmphibiaWeb without permission. It is not in our interest or that of our contributors to have uncontrolled subsets of these holdings available elsewhere on the Internet. We make corrections, add text, add images, etc. on a continual basis, and we want the most current set of images and text to be the only one generally available to all Internet users;
6. Images taken from AmphibiaWeb may only be used in accordance with the usage guidelines provided with the image and the photographer must be clearly credited with the image; for more information click here;

For questions about use, please contact AmphibiaWeb.

AmphibiaWeb updates its pages regularly, but assumes no responsibility for the accuracy, currency, or completeness of content. Please report any errors or inaccuracies.

Citation: AmphibiaWeb: Information on amphibian biology and conservation. [web application]. 2008. Berkeley, California: AmphibiaWeb. Available: http://amphibiaweb.org/. (Accessed: Aug 25, 2008).

AmphibiaWeb's policy on data use.

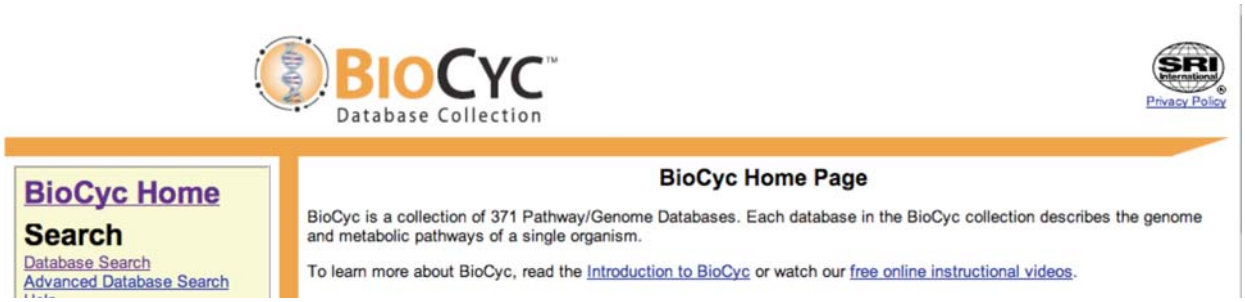
ARKIVE.ORG

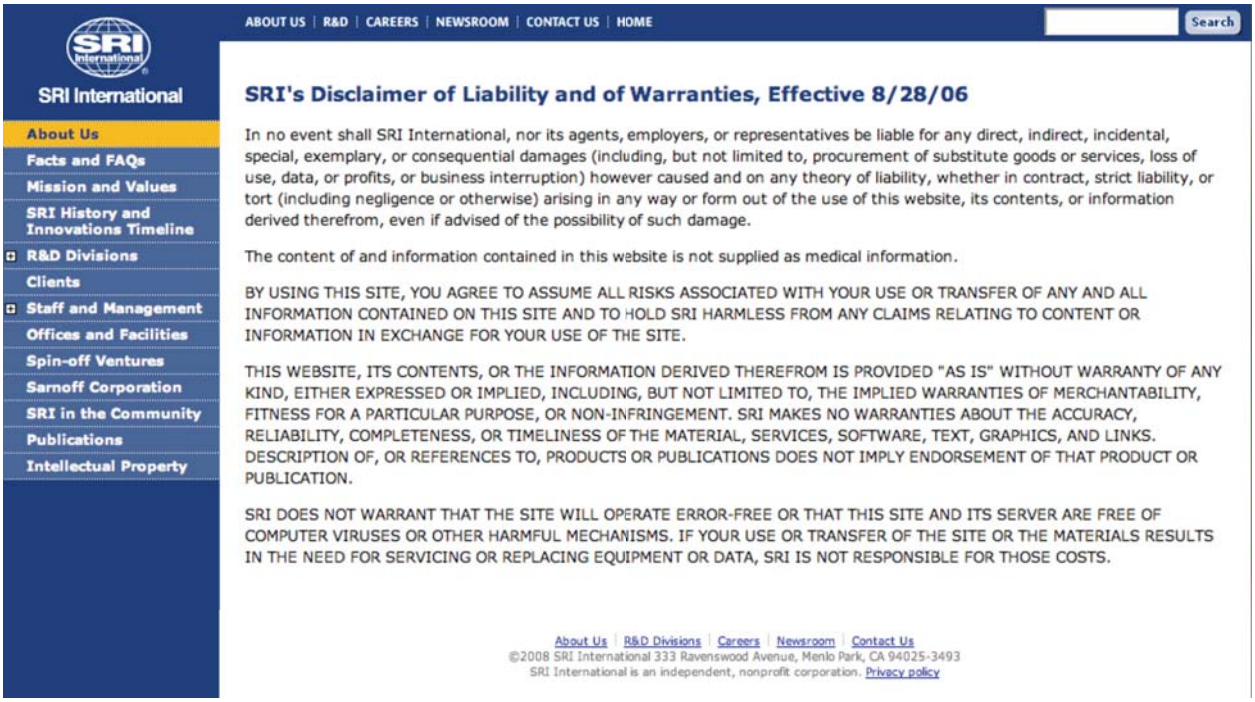
http://www.arkive.org

ARKive Website - 'Terms of use'

BIOCYC DATABASE COLLECTION

http://biocyc.org/





BIOMEDICAL INFORMATICS RESEARCH NETWORK

http://nbirn.net/index.shtm

BIRN Data Repository
Data Use Agreement

This document outlines the provisions of a non-exclusive license for use of data shared through the Biomedical Informatics Research Network (BIRN) Data Repository (BDR). In line with one of the unique aspects of BIRN, the goal of the BDR is to make data originally acquired with investigator-specific interest available to the community for further study. All data are freely available to the public with the provision of an active e-mail address and acceptance of this standard BDR Data Use Agreement. All Users are encouraged to consider sharing resultant data through the BDR to extend and strengthen this collaborative environment.

By your acceptance of these terms, you agree to the following provisions for yourself *and any collaborators with whom you share these data*:

1. **Security:**
   I will make no attempt to determine the identity of any individual whose data are included in the BDR.

2. **Data Guarantee:**
   I will employ these data at my own risk, as the quality of these data cannot be guaranteed.

3. **Publication / Acknowledgement of Data Use:**
   I will acknowledge the use of BDR data when publicly presenting any findings or algorithms that benefited from their use. Such presentations include but are not limited to papers, books, book chapters, conference posters, and talks.

   a) I agree to acknowledge the BIRN Data Repository and funding mechanisms along with the relevant project-specific accession number(s) associated with data downloaded from the BDR. *For example:*

      Methods Section: "These data were downloaded from the BIRN Data Repository, Project Accession Number 2007-BDR-6UHZ1."

      Acknowledgements: "Data used for this study were downloaded from the Biomedical Informatics Research Network (BIRN) Data Repository (http://www.nbirn.net/bdr), supported by grants to the BIRN Coordinating Center (U24-RR019701), Function BIRN (U24-RR021992), Morphometry BIRN (U24-RR021382), and Mouse BIRN (U24-RR021760) Testbeds funded by the National Center for Research Resources at the National Institutes of Health, U.S.A."

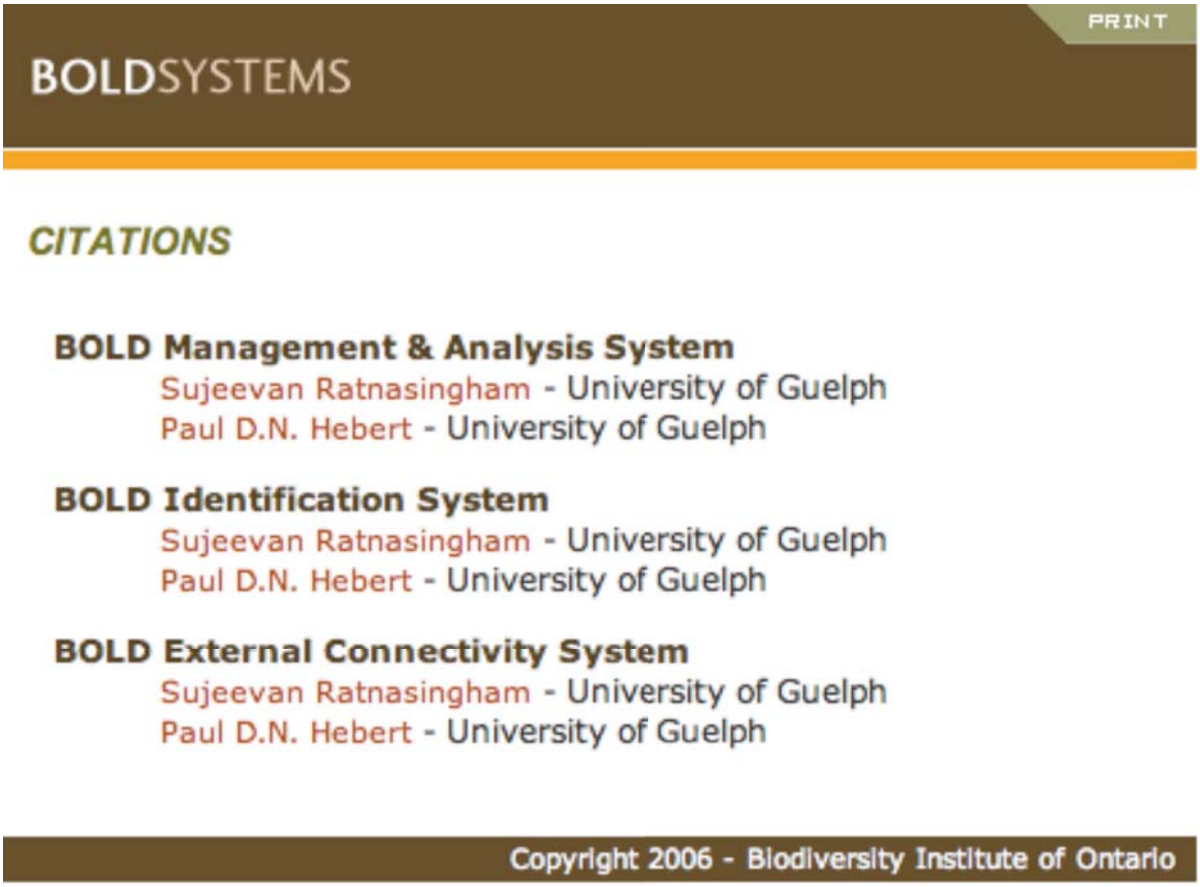
   b) I agree to provide to BIRN a bibliographic citation of the final published presentation or article for inclusion in the BIRN literature archive. I will send this information to **BDR_info@nbirn.net.**

4. **Redistribution of Data:**
   a) Redistribution of original BDR data is permitted so long as the data are redistributed under the same terms and conditions as described in this Data Use Agreement.

   b) Data derived from original BDR data may be distributed under terms and conditions established by the creators of such derived data. Users must comply with the terms and conditions of use set by the creators of these derived data.

BOLD

http://www.barcodinglife.org/views/login.php



CANADIAN BIODIVERSTY INFORMATION FACILITY

http://www.cbif.gc.ca/

Government of Canada Gouvernement du Canada

Canada

Français	Contact us	Help	Search	Canada site
Coming events	About CBIF	Links	Reports and publications	Home

Canadian Biodiversity Information Facility

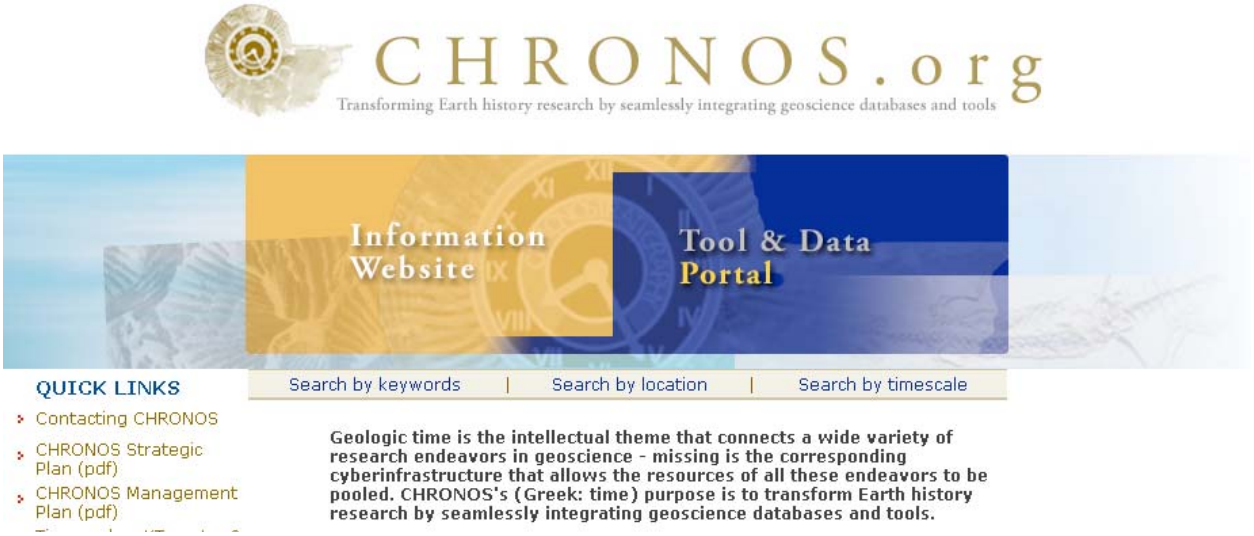
Conditions for use of data

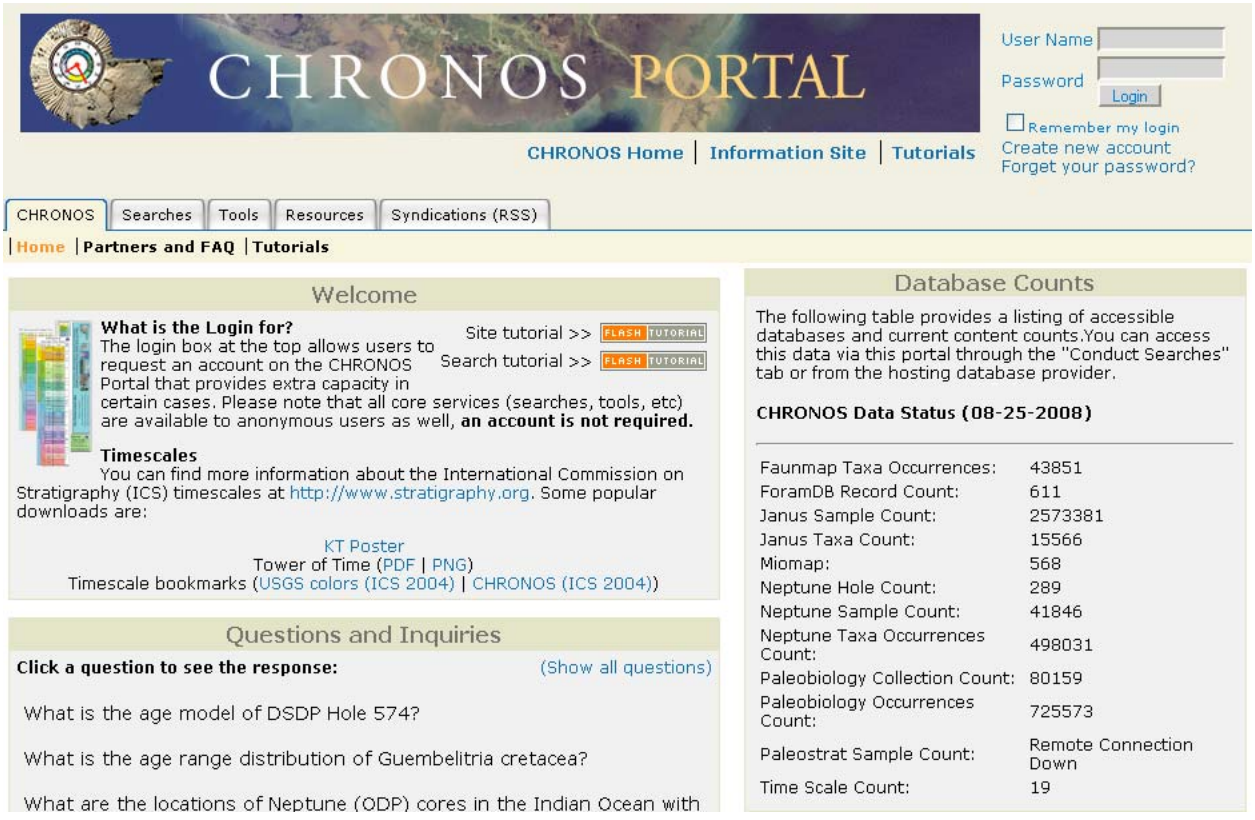
Disclaimers

- Data are provided on an "as is" basis -no responsibility by the network or by the caretakers of the original data is accepted for uses or misuses.
- Data are not to be used on a for-profit basis - use of the data for commercial or for-profit applications are permitted only via written permission from the managers of the data network and from the component data caretakers.
- Data are provided on an end-user basis - that is, data are provided to users, but are not to be passed on to third parties or otherwise redistributed.
- Proper acknowledgment of the data network and source institutions is expected; in textual form, the following format should be used: "Biodiversity data were drawn from the distributed data network entitled Species Analyst, a project of the North American Biodiversity Information Network, and participating institutions." Acknowledgment of individual institutions should also be included.
- Data are provided under the understanding that their use will be generally consistent with the philosophy of not causing detriment to wild species.
- Data are provided in this facility under the general, overarching principle of free and open access to biodiversity information and technology.

CHRONOS.ORG

http://www.chronos.org/





CHRONOS

CHRONOS Home | Portal

CHRONOS | Cyberinfrastructure | Community Involvement | Education & Outreach

- CHRONOS
  - CHRONOS Index
  - CHRONOS Sitemap
  - Calendar & Events
  - Contact CHRONOS
  - Data Feedback
  - FAQs
  - Logos and Templates
  - Strategic Plan (pdf)
  - Management Plan (pdf)
  - K-12 CHRONOS Services (pdf)
  - Instructors CHRONOS Services (pdf)
  - Scientists CHRONOS Services (pdf)

Google site search | Search

CHRONOS

PDF

Geologic time is the intellectual theme that connects a wide variety of research endeavors in geoscience - missing is the corresponding cyberinfrastructure that allows the resources of all these endeavors to be pooled. CHRONOS's *(Greek: time)* purpose is to transform Earth history research by seamlessly integrating geoscience databases and tools.

Mission Statement

CHRONOS is a team of geoscientists and information technology specialists creating a cyberinfrastructure that will deliver open access to a global federation of Earth history databases, tools, and services, thus providing:

- For **academic, government, and industrial scientists** - access to multiple, disparate databases on Earth history, data evaluation and conversion services, and powerful analytical tools.
- For **autonomous databases, affiliated science initiatives, and data and tool contributors** - a larger user community, greater visibility and acknowledgment, and access to tools and best practices, without the cost and burden of reproducing interoperability.
- For **educators, students, and policy makers** - a convenient source of Earth history data, visualization tools, expert opinion, and educational materials.

Last Published: 08/10/2006 13:18:48

COSTA RICAN BIODIVERSTY INFORMATION FACILITY

http://www.inbio.ac.cr/attaing/index.htm

Terms of use of information provided in INBio´s website

General Policy

The information provided in INBio´s database belongs to The National Biodiversity Institute (INBio). We do consider, though, that its´use can help achieve a greater assessment of our biodiversity, and help with the decision-making concerning its´sustainable use. Consequently, INBio provides access and allows its´ free use to all public, just as long as it isn´t used for any commercial purpose, and the entities that use it clearly indicate the intelectual property rights of INBio. It is assumed that users that consult INBio´s database in the web accept all terms of use that are specified below.

1. Acknowledgements: Every user of the information must mention in his/her acknowlegments and bibliography the sources of the data, that is, he/she must indicate the name of the information system used, the name INBio, and the date of the query, for example: Source : Atta, INBio´s information system, July 14th 2001. This acknowledgement must be done in every subsequent use of these data. Every photograph, illustration and map taken from INBio´s website must be accompanied by INBio´s logo and the author´s name.

2. Commercial use of the information: Every user agrees not to use the obtained information for commercial uses. If the intention is to use the information for commercial uses, then he/she must contact otus@inbio.ac.cr . Examples of commercial uses are, but are not limited to: digital or printed publications, incorporation of the information to other databases, etc. Examples of non-commercial uses are: School homework, academic investigations in high schools and colleges, school material prepared by educators.

3. Ethical use of the information: Every user agrees not to use the obtained information to the detriment of biodiversity conservation (for example illegal traffic of species), nor in processes that would irreversibly alter the ecological equilibrium (for example undiscriminated extraction or the introduction of exotic species).

4. Responsibility: INBio keeps high quality standards in the biodiversity inventory process it conducts. There exists, though, no responsibility on INBio´s behalf nor its´scientists about the accuracy of the data, or of the use the data is given.

The persons who do not respect the terms of use above mentioned will be turned over to the corresponding authorities. Hereby, the reproduction, edition, addition or total or partial reduction of INBio´s database information (that do not comply to the terms above mentioned) are strictly prohibited.

DANISH BIODIVERSTY INFORMATION FACILITY

http://www.dk.gbif.net/

Home » Information » Copyright

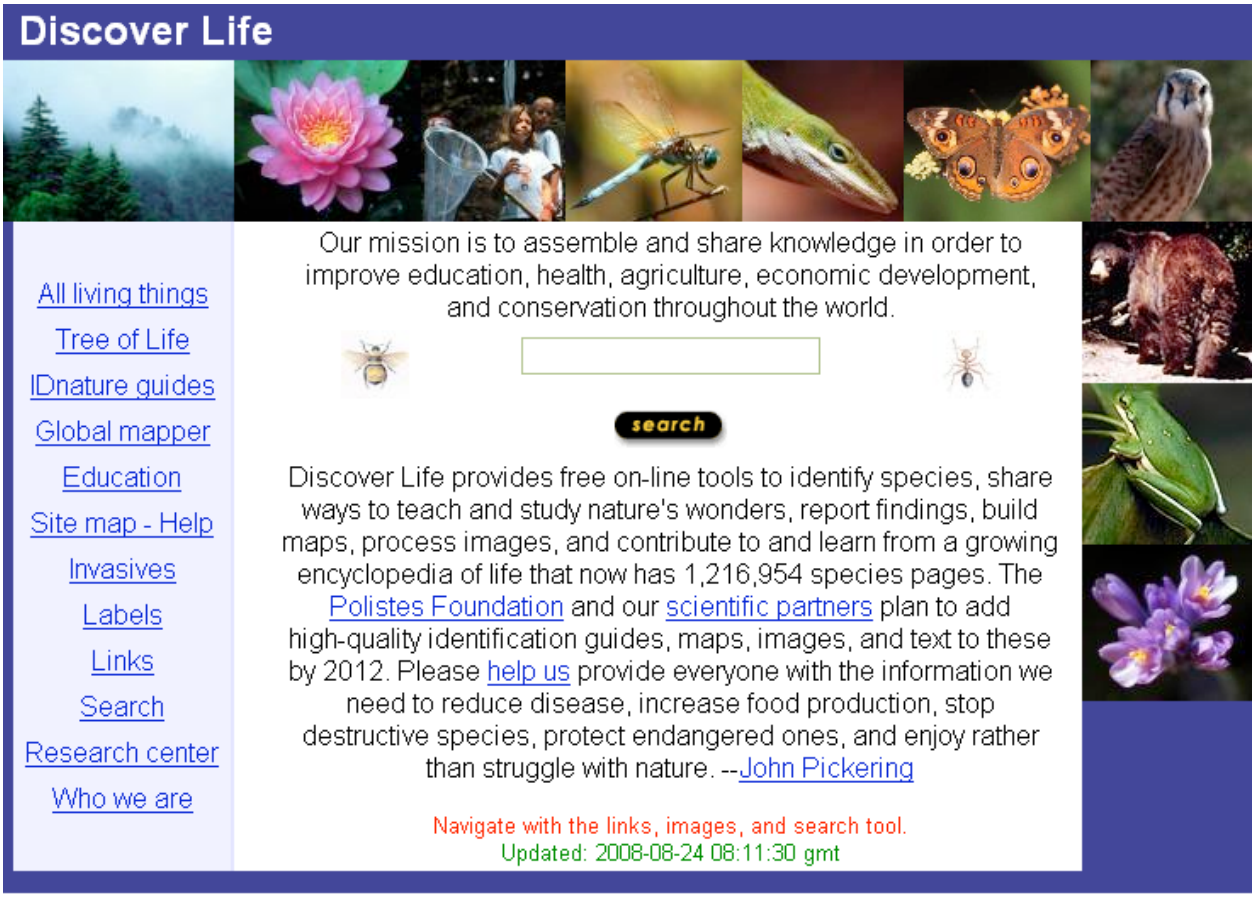
Copyright

Released	23/11/2006
Contributor	mccarausu

Page last modified: 05/04/2007

Discover Life and Discover Life in America

http://www.discoverlife.org/ http://www.dlia.org/



Discover Life in America

Great Smoky Mountains National Park All Taxa Biodiversity Inventory

Home
DLIA
ATBI
Park Species
Scientist's Information
Photo Galleries
Search
Contact Us
Site Map

Volunteer
Internship Info
Corporate Sponsorship
Support

Buy DLIA Products

Disclaimer

- This web site and the information it contains are provided as a public service by Discover Life in America (DLIA) and the All Taxa Biodiversity Inventory (ATBI).
  - This system is monitored to ensure proper operation, to verify the functioning of applicable security features, and for comparable purposes.
  - Anyone using this system expressly consents to such monitoring.
  - Unauthorized attempts to modify any information stored on this system, to defeat or circumvent security features, or to utilize this system for other than its intended purposes are prohibited and may result in criminal prosecution.

Timber rattlesnake (*Crotalus horridus*)

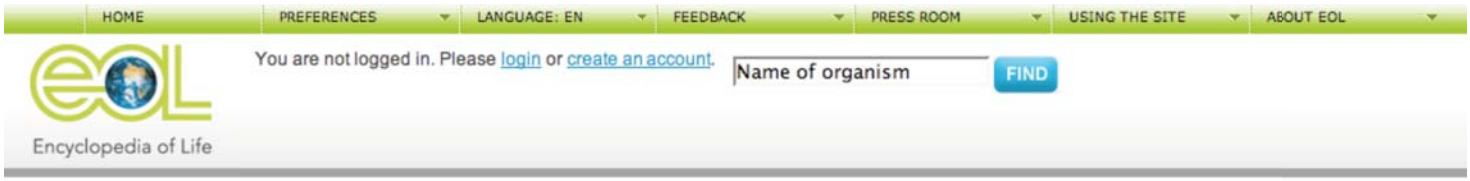
Click photo to enlarge.
Photo by Charles Wilder.

- DLIA/ATBI makes no claims, promises or guarantees about the accuracy, completeness, or adequacy of the contents of this web site and expressly disclaims liability for errors and omissions in the contents of this web site.
  - No warranty of any kind, implied, expressed or statutory, including but not limited to the warranties of non-infringement of third party rights, title, merchantability, fitness for a particular purpose and freedom from computer virus, is given with respect to the contents of this web site or its hyperlinks to other Internet resources.
  - Reference in this web site to any specific commercial products, processes, or services, or the use of any trade, firm or corporation name is for the information and convenience of the public, and does not constitute endorsement, recommendation, or favoring by DLIA/ATBI.

- Information presented on this web site, unless otherwise indicated, is considered copyrighted by DLIA/ATB.
  - It may only be be distributed or copied as permitted by applicable copyright law.
  - Not all information on this web site has been created or is owned by the DLIA/ATBI.
    - If you wish to use any non-DLIA/ATBI material, you must seek permission directly from the owning (or holding) sources.

ENCYCLOPEDIA OF LIFE

http://www.eol.org



Terms Of Use

Content available on this site has been provided by EOL's data partners, who may have placed restrictions on how you can re-use that content.

Please note that a single page may be made up of many different data elements, each covered by a different license. You are required to check to see which license applies to any portion(s) of the page you wish to re-use and to abide by any restrictions on that content.

The Creative Commons has set up a robust, free, easy-to-use set of protocols by which content owners can easily mark their creative work with the freedoms or restrictions that they want the information to carry.

In most cases, EOL data partners have made content available for re-use under one of the following Creative Commons licenses:

- CC-BY (Attribution), which allows users to copy, transmit and reuse the information, and to remix or adapt the information, as long as attribution regarding the source of the information is maintained. (see http://creativecommons.org/licenses/by/3.0/)
- CC-BY-SA (Attribution-ShareAlike), which has the additional constraint on top of CC-BY that if you alter, transform, or build upon this work, you may distribute the resulting work only under the same, similar or a compatible license. (see http://creativecommons.org/licenses/by-sa/3.0/)
- CC-BY-NC (Attribution-NonCommercial), which has the additional constraint on top of CC-BY that you may not use this work for commercial purposes. (see http://creativecommons.org/licenses/by-nc/3.0/)
- CC-BY-NC-SA (Attribution-NonCommercial-ShareAlike) which has the additional constraint on top of CC-BY that if you alter, transform, or build upon this work, you may distribute the resulting work only under the same, similar or a compatible license and that you may not use this work for commercial puroposes. (see http://creativecommons.org/licenses/by-nc-sa/3.0/).

To identify the terms of re-use of a photograph or drawing, click on the green information button on the bottom left corner of the picture.

Re-use terms for text items are located immediately above the text.

If you wish to re-use any content for purposes other than those allowed by the associated Creative Commons license, you must get permission to do so from the copyright holder (Source).

If there is no re-use information associated with an item (e.g. most of the literature provided by the Biodiversity Heritage Library), then there are no re-use restrictions on this content.

FISHBASE

http://www.fishbase.org

References Citing FishBase

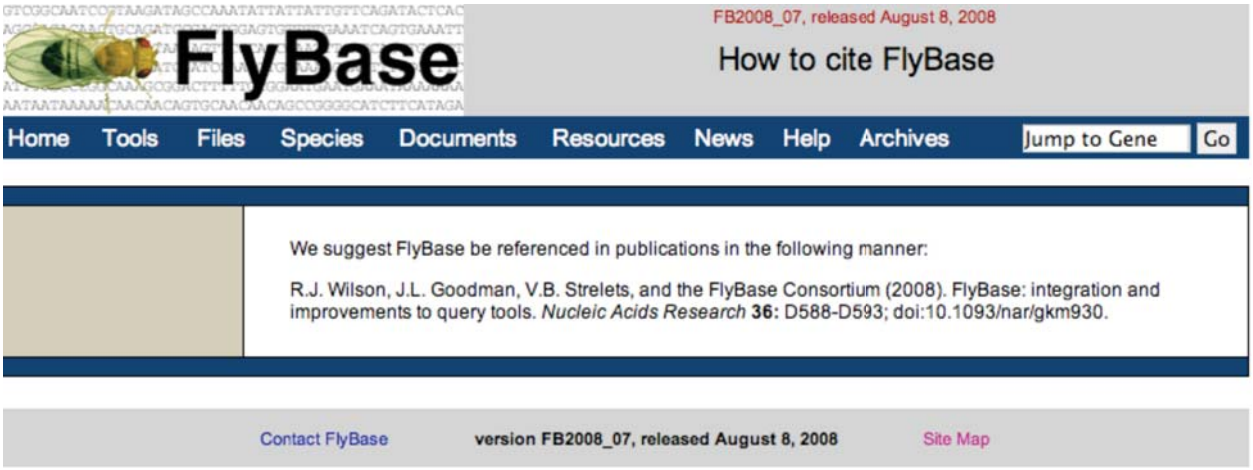
How to cite FishBase
To give due credit to the original authors, please cite data taken from FishBase by Main Ref. and/or Data Ref. of the respective record.

Cite FishBase itself as
Froese, R. and D. Pauly. Editors. 2008.FishBase.
World Wide Web electronic publication.
www.fishbase.org, version (06/2008).

Note: FishBase is also available on CD-ROM, with detailed information on population dynamics, genetics, morphology, trophic ecology, physiology, ecotoxicology, reproduction, etc. See the FishBase homepage or the FishBase book for more information. FishBase was assembled with the help of many partners and with the support of the European Commission and other sponsors. Contact us if you want to provide pictures, data or reprints.

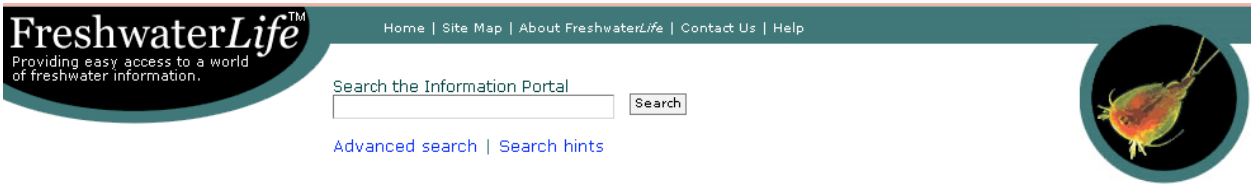
FLYBASE

http://flybase.org/



FRESHWATER LIFE

http://www.freshwaterlife.org/fwl/aboutfwl.html



2 Copyright statement [top]

Please also refer to the FreshwaterLife General Terms and Conditions and Content submission policy, which give further guidance on copyright issues.

This copyright statement is held jointly by the FreshwaterLife project partners and owners. The direct assignable body is the Freshwater Biological Association (FBA), Far Sawrey, Ambleside, Cumbria, LA22 OLP, United Kingdom. This assemblage of project partners will from herein be referred to as FreshwaterLife.

FreshwaterLife assembles, constructs and administers several databases, systems, softwares and websites with the aim of providing a freshwater information service. These systems, and the data and content held by these systems, will be referred to collectively as the FreshwaterLife website. This should be taken to include other FreshwaterLife products, such as CD-ROMS, documents and publications, derived from these systems and their content.

In addition to this general copyright statement, some of the data may be subject to specific terms that the owner or supplier of that data has imposed on its use. (We call these Specific Terms of Access.) [top]

2.1 Copyright notice

FreshwaterLife has a compilation copyright on the FreshwaterLife website. All material, data and information on this website are protected by copyright, database rights and other intellectual property rights. Any attempt to copy the whole FreshwaterLife website or any section of the website without the permission of the Chief Editor will be regarded as an infringement of copyright. Copying and downloading of individual items, pages, data, meta-data or other materials (referred to collectively as Knowledge Objects – KOs from herein) will be subject to any copyright conditions laid down in the Specific Terms of Access for that KO or other materials. Any use or inclusion of FreshwaterLife KOs or other materials in other documents or media must include an appropriate copyright and permissions notice. Use of any material copied from FreshwaterLife is restricted to informational, non-commercial or personal use only and no modifications to the information are permitted. Furthermore, you may not republish any of the data made available to you, or exploit it for commercial purposes. Reproduction and dissemination of material in this electronic product for educational or other non-commercial purposes are authorized without any prior written permission from the copyright holders provided the source is fully acknowledged. Where details of are given, you must acknowledge the original author/owner of data in any derived information product or publication, whether printed, electronic or broadcast, that is based wholly or in part on that record. Reproduction of material in this product for resale or other commercial purposes is prohibited without written permission of the copyright holders. Applications for such permission should be addressed to the FreshwaterLife Chief Editor. You must acknowledge and correctly cite FreshwaterLife in any derived information product or publication, whether printed, electronic or broadcast, that is based wholly or in part on the resources and information collated by FreshwaterLife. FreshwaterLife requests to be informed of any such acknowledgements and citations. Any information product or publication that you make which contains any part of the data made available to you via this website must contain a statement that the data supplier, recorder [where identified], and FreshwaterLife bear no responsibility for any further analysis or interpretation of that data. [top]

2.2 Copyright of submitted material

Knowledge of any type that is available (or convertible) in commonly used electronic formats (for example, pdf, text, html, jpeg, gif, WORD, EXCEL) format can be linked to, uploaded, or referenced in the FreshwaterLife website, where it becomes known as a Knowledge Object (KO) and is primarily listed under the heading of Relevant Knowledge within each topic. The person, institute or organisation that contributes a KO or other materials to FreshwaterLife is deemed either to be the owner or copyright holder of the material, or to have the permission of the copyright holder to contribute the material to FreshwaterLife. This does not in any way prevent third parties from submitting the details and location of any KO or other materials accompanied by their own review, abstract or summary of the materials. The Chief and the Topic Editors have the right to refuse and remove any materials from FreshwaterLife, including materials that have been improperly submitted, or for which there is any doubt about their ownership and legality. The ownership of an individual FreshwaterLife KO or other materials remains with the registered members who have contributed them. It is their responsibility to ensure that information they contribute is accurate and up to date, to secure and retain permissions to link to websites and to use copyrighted materials. [top]

GLOBAL BIODIVERSITY INFORMATION FACILITY

http://www.GBIF.org/

2. Citing Data

Use the following format to cite data retrieved from the GBIF network:

Biodiversity occurrence data provided by: (Accessed through GBIF Data Portal, www.gbif.net, YYYY-MM-DD)

For example:

Biodiversity occurrence data provided by: Field Museum of Natural History, Museum of Vertebrate Zoology, University of Washington Burke Museum, and University of Turku (Accessed through GBIF Data Portal, www.gbif.net, 2007-02-22)

Data Use Agreement

Background

The goals and principles of making biodiversity data openly and universally available have been defined in the Memorandum of Understanding on GBIF, paragraph 8 (see the relevant excerpts in the attached Annex).

The Participants who have signed the MoU have expressed their willingness to make biodiversity data available through their nodes to foster scientific research development internationally and to support the public use of these data.

GBIF data sharing should take place within a framework of due attribution.

Therefore, using data available through the GBIF network requires agreeing with the following:

1. Data Use Agreements

1. The quality and completeness of data cannot be guaranteed. Users employ these data at their own risk.
2. Users shall respect restrictions of access to sensitive data.
3. In order to make attribution of use for owners of the data possible, the identifier of ownership of data must be retained with every data record.
4. Users must publicly acknowledge, in conjunction with the use of the data, the data providers whose biodiversity data they have used. Data providers may require additional attribution of specific collections within their institution.
5. Users must comply with additional terms and conditions of use set by the data provider. Where these exist they will be available through the metadata associated with the data.

2. Citing Data

Use the following format to cite data retrieved from the GBIF network:

Biodiversity occurrence data provided by: (Accessed through GBIF Data Portal, www.gbif.net, YYYY-MM-DD)

For example:

Biodiversity occurrence data provided by: Field Museum of Natural History, Museum of Vertebrate Zoology, University of Washington Burke Museum, and University of Turku (Accessed through GBIF Data Portal, www.gbif.net, 2007-02-22)

3. Definitions

- GBIF Participant: Signatory of the GBIF-establishing Memorandum of Understanding (MoU).
- GBIF Secretariat: Legal entity empowered by the GBIF Participants to enter into contracts, execute the Work Programme, and maintain the central services for the GBIF network.
- GBIF network: The infrastructure consisting of the central services of the GBIF Secretariat, Participant Nodes and data providers. Making data available through GBIF network means registering and advertising the pertinent services via the GBIF central services.
- Node: A data provider designated by a GBIF Participant that maintains a stable computer gateway that makes data available through the GBIF network.
- Participant node: An organisational unit designated by the GBIF Participant to coordinate activities in its domain. It may also provide data.
- Biodiversity data: Primary data on specimens, observations, names, taxonomic concepts, and sites, and other related data on biological diversity.
- Metadata: Data describing the attributes and combinations of biodiversity data.
- Data: Biodiversity data and metadata.
- Data provider: A custodian of data making it technically available. This may or may not be the data owner. If not they will have declared to GBIF that they have permission to make the data available.
- Data sharing: The process of and agreements for making data freely and universally available on the Internet.
- User: Anyone who uses the Internet to access data through the GBIF network.
- Owner of data: The legal entity possessing the right resulting from the act of creating a digital record. The record may be a product derived from another, possibly non-digital product, which may affect the right.
- Sensitive data: Any data that the Node does not want to make available, e.g. precise localities of endangered species.

Also see the GBIF Data Sharing Agreement for the data providers.

Accept terms | Cancel

Depending on your browser settings, a cookie may be stored to acknowledge your acceptance of these terms.

GBIF-SPAIN

http://www.gbif.es/

Copyright Notice

How to cite this resource

Usage of the data hosted in this web server implies in all cases the explicit recognition and indication of the source, as follows:

> **Coordination Unit of GBIF Spanish Node (*Unidad de Coordinación de GBIF en España*)**, [Resource or web page consulted], [Date of request].

An example of this kind of citation could be:

> **Coordination Unit of GBIF Spanish Node (*Unidad de Coordinación de GBIF en España*)**, 'Report on Natural History Collections in Spain', http://www.gbif.es /informecolecciones_in.php, 2005/03/08.

To cite data fron the GBIF Network (like those obtained from the Coordination Unit Data Server), check the instructions provided in **this page**.

Herbario de Criptogamia

CONDICIONES Y RESERVAS DE USO DE LOS DATOS

1. **Responsabilidad:** Ni el Real Jardín Botánico ni sus conservadores se responsabilizan de la exactitud, fiabilidad y grado de cobertura de los datos, ni del uso que se les dé.
2. **Uso comercial de los datos**: Se prohíbe el uso comercial de los datos obtenidos, o que conlleve beneficio económico, sin el consentimiento explícito del Real Jardín Botánico
3. **Redistribución:** Queda estrictamente prohibido la edición, reproducción, redistribución, adición o reducción total o parcial de la información, incluida explícitamente su incorporación a otras bases de datos de acceso libre o restringido.
4. **Reconocimientos:** El uso de los datos conlleva la obligación de reconocer y citar su uso explícitamente (ver "como citar este recurso"). Esta atribución se debe mantener en cualquier uso subsecuente de los datos.
5. **Información obtenida:** el usuario se compromete a que la información obtenida no sea utilizada en detrimento de la conservación de la flora y fauna, ni en procesos que alteren irreversiblemente el equilibrio ecológico (extracción indiscriminada o introducción de especies).
6. **Uso de la información negativa:** los datos de biodiversidad constatan generalmente presencias; es decir, la falta de datos de una especie en un punto no implica que no se encuentre ahí. El muestreo biológico es normalmente irregular, incompleto e influido por factores ajenos al objeto de muestreo. Es fundamental tomar en cuenta este extremo para un uso apropiado de los datos.

El Real Jardín Botánico, y subsidiadamente el CSIC, se reserva el derecho de actual legalmente contra los infractores de estas condiciones de uso.

CONDITIONS AND PROVISOS FOR USE OF CONTENT

1. **Data are provided on an "as is" basis**- no responsibility by the Real Jardín Botánico and their curators of the accuracy, reliability, or completeness of the data or information provided data is accepted for uses or misuses.
2. **Data are not expected to be used on a for-profit basis**- use of the data for commercial or for-profit applications are permitted only via written permission from the Real Jardín Botánico
3. **Data are provided on an end-user basis**- that is, data are provided to users, but should not be passed on to third parties or redistributed. It is explicitly forbidden to incorporate these data into other databases of free or restricted access.
4. **Proper acknowledgment of the data source institutions is expected**; The data use implies explicit acknowledgment (see "how to cite this resource") whic should be kept with any subsequent use of the data.
5. **Data obtained:** Data are provided under the understanding that their use will be generally consistent with the philosophy of not causing detriment to wild species and ecosystem balance.
6. **Use of negative information** - Biodiversity data generally constitute positive information only; that is, absence of a record does not always mean absence of a particular species from particular point in space and time. Because biodiversity sampling is notoriously spotty, biased, and incomplete, understanding the positive-only nature of the data is critical to their proper use.

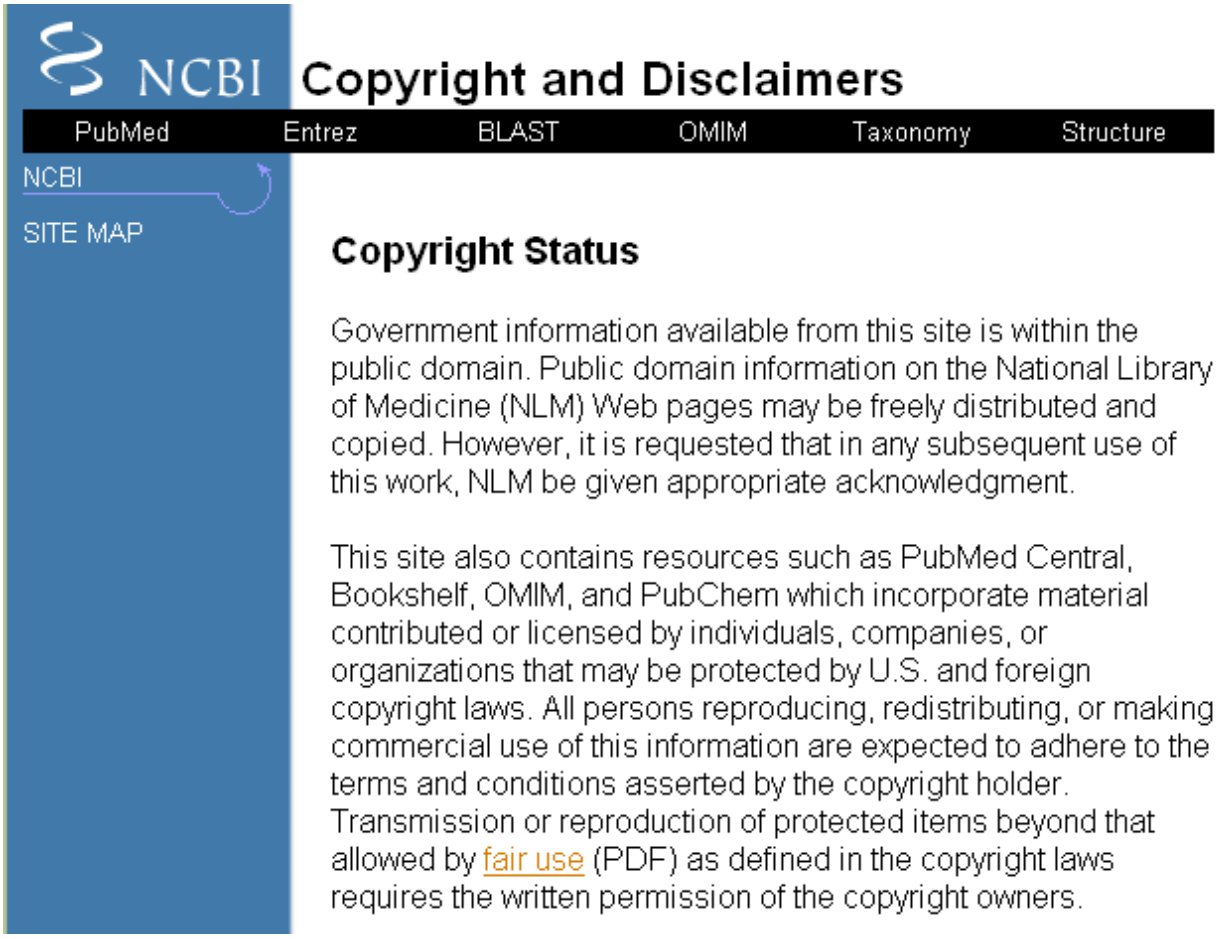
The Real Jardín Botánico, and subsidiarily the CSIC, reserves the right to pursue legal actions against those infringing these conditions of use.

How to cite this resource

Pando, F. (2000). MA Cryptogamic collections online databases. http://www.rjb.csic.es/herbario/crypto/crydb.htm. (date when consulted).

GenBank

http://www.ncbi.nlm.nih.gov/



Molecular Database Availability

Databases of molecular data on the NCBI Web site include such examples as nucleotide sequences (GenBank), protein sequences, macromolecular structures, molecular variation, gene expression, and mapping data. They are designed to provide and encourage access within the scientific community to sources of current and comprehensive information. Therefore, NCBI itself places no restrictions on the use or distribution of the data contained therein. However, some submitters of the original data (or the country of origin of such data) may claim patent, copyright, or other intellectual property rights in all or a portion of the data (that has been submitted). NCBI is not in a position to assess the validity of such claims and, therefore, cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in the molecular databases.

Disclaimer

External Links: Some NCBI Web pages may provide links to other Internet sites for the convenience of users. NCBI is not responsible for the availability or content of these external sites, nor does NCBI endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Users cannot assume that the external sites will abide by the same Privacy Policy to which NCBI adheres. It is the responsibility of the user to examine the copyright and licensing restrictions of linked pages and to secure all necessary permissions.

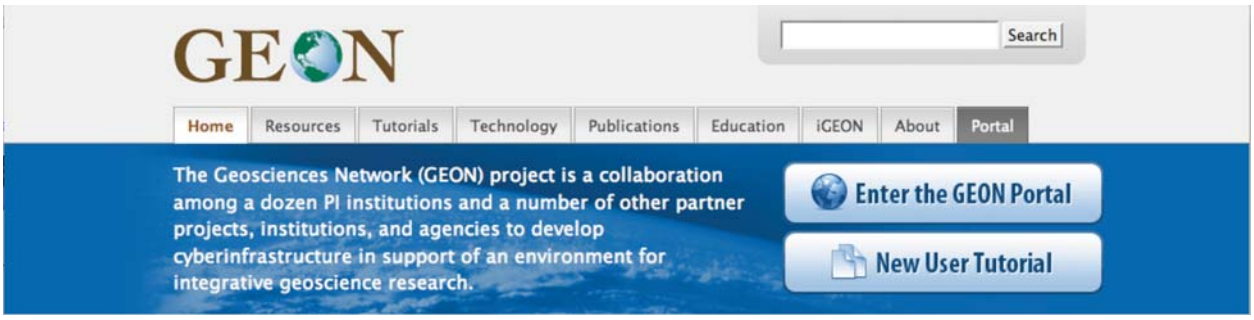
Conditions of Use

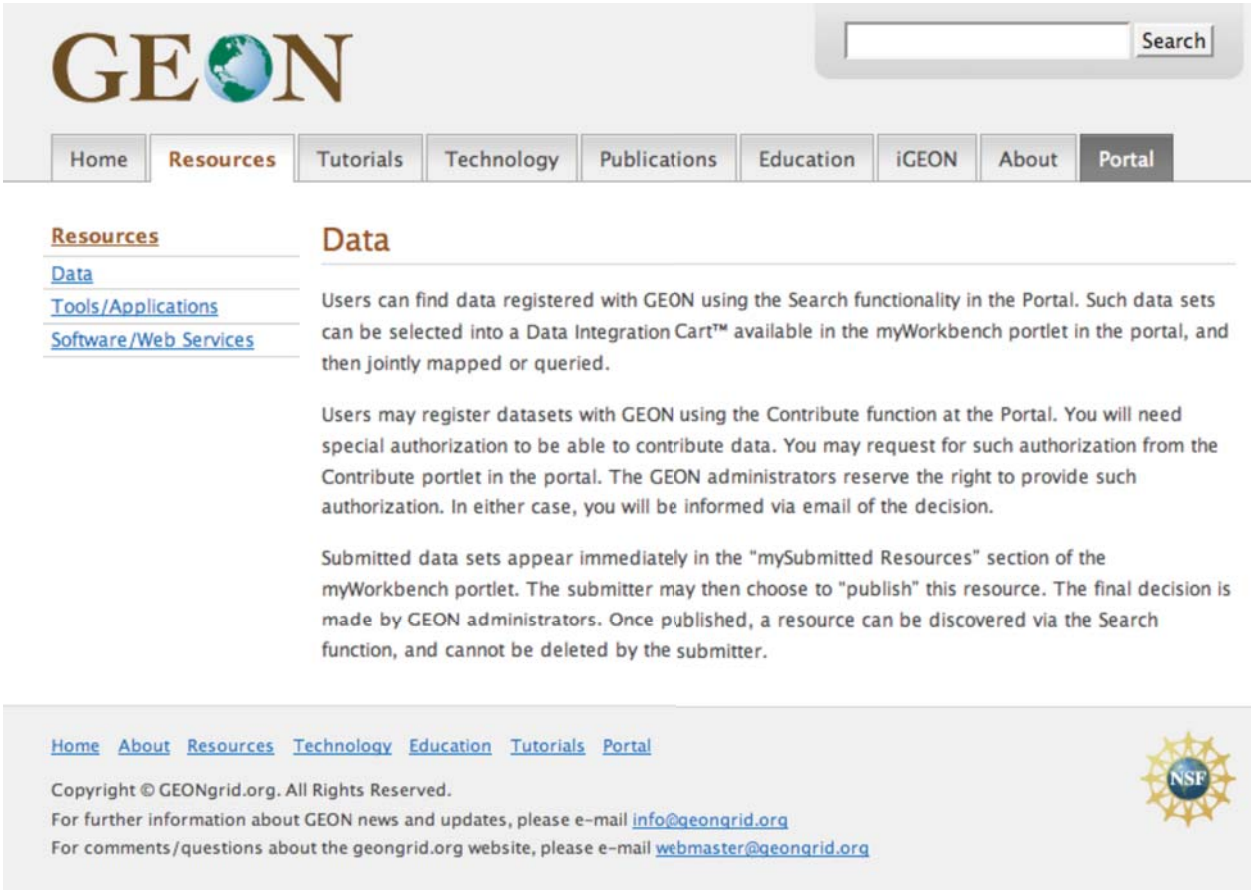
This site is maintained by the U.S. Government and is protected by various provisions of Title 18 of the U.S. Code. Violations of Title 18 are subject to criminal prosecution in a federal court. For site security purposes, as well as to ensure that this service remains available to all users, we use software programs to monitor traffic and to identify unauthorized attempts to upload or change information or otherwise cause damage. In the event of authorized law enforcement investigations and pursuant to any required legal process, information from these sources may be used to help identify an individual.

Revised: March 21, 2008.

GEON

http://www.geongrid.org/





HERPNET (followed by ORNIS, MANIS, and implicitly by FishNET2)

http://www.herpnet.org/

Museum data records may be used by individual researchers or research groups, but they may not be repackaged, resold, or redistributed in any form without the express written consent of a curatorial staff member of the participating institution. If any of these records are used in an analysis, report, presentation, etc., the provenance of the original data must be acknowledged and the specific institutions' curatorial staff and HerpNET notified. The individual institutions, HerpNET, and their staff are not responsible for damages, injury or loss due to the use of these data.

Publications and presentations resulting from the use of HerpNET data should list the names of **each** institution whose data were used, and HerpNET (http://www.herpnet.org) as the portal by which the data were accessed or downloaded. The citation also should list the date that the data were accessed via HerpNET. Users should refer to individual institutional policies on proper citation of data from those providers.

Example: Data were obtained from records held in the following institutions and accessed through the HerpNET data portal (http://www.herpnet.org) on 15 February 2006: California Academy of Sciences, San Francisco; Museum of Comparative Zoology, Harvard University, Cambridge; Museum of Vertebrate Zoology, University of California, Berkeley, etc.

By downloading and/or viewing data on the HerpNET portal, you are agreeing to these conditions.

SEARCH MUSEUM DATA

If your search does not work, select fewer institutions at a time. Hold down the Control or Shift buttons to select specific collections at once.

Home | Data Portal | Documents | Listserv | HerpNET Participants | Gazetteer | Georeferencing Resources | Join HerpNET

Updated 14 May 2008 CLS

INTERAMERICAN BIODIVERSITY INFORMATION NETWORK

http://www.iabin.net/

THE INTERAMERICAN BIODIVERSITY INFORMATION NETWORK (IABIN)

ACCESS TO INFORMATION AND INTELLECTUAL PROPERTY REGULATION

CHAPTER 3. ACCESS TO THE INFORMATION

3.1 Public nature of the data. The information obtained through IABIN will be considered as public and may be used at no cost, provided that it is consulted for informational purposes only.

3.2 Intellectual property of the information. Researchers, collections, institutions, legal and natural persons and IABIN users in general will assume the following obligations:

3.2.1 Access to data: IABIN will be an open-access facility. IABIN users in general will have equal access to the data in databases affiliated with or developed by IABIN, through an access code when required by IABIN.

3.2.2. IABIN will make efforts to strengthen the free distribution of data on biodiversity, for informational purposes only. However, the use of all information obtained through IABIN is subject to proper attribution.

3.2.3. Intellectual property rights of databases developed by other organizations is subject to their terms of use as well. Any condition established by the Data Provider will be respected. The source of the data should be acknowledged and such attribution be maintained in any subsequent use of the data.

KOREAN BIODIVERSITY INFORMATION FACILITY

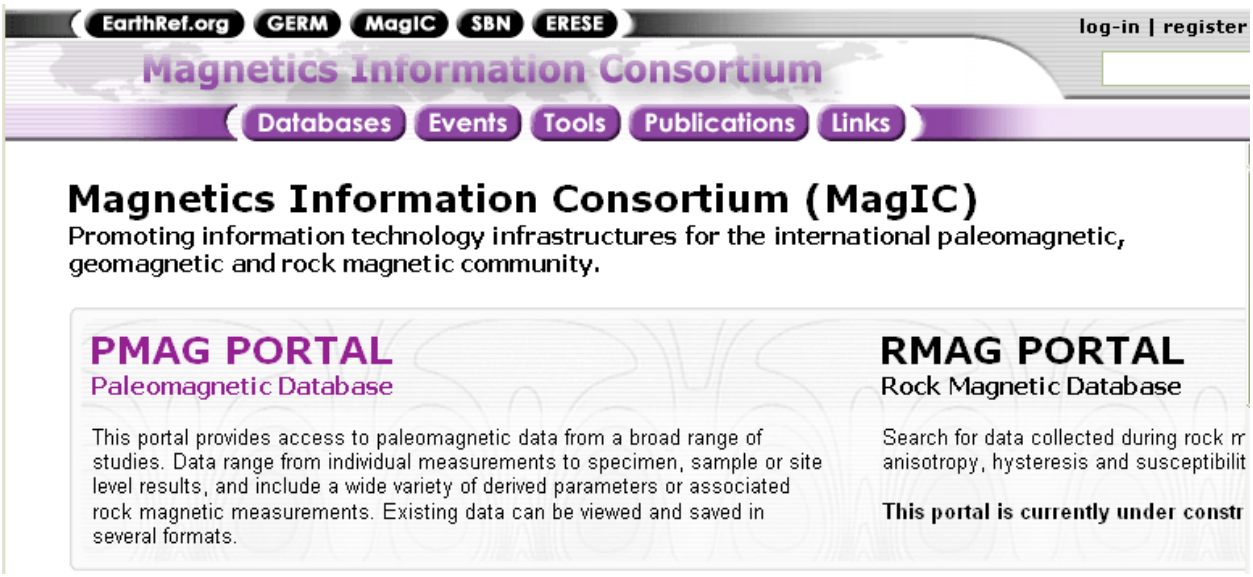
http://www.kbif.re.kr/chm2

Copyright

Released	21 Apr 2006
Contributor	ssahn

MAGNETICS INFORMATION CONSORTIUM

http://earthref.org/MAGIC/



Disclaimer

EarthRef.org - Privacy Policy

EarthRef.org respects the privacy of all its collaborators, contributors, users of its services, and visitors to its web site. Several functions and services or EarthRef.org require the user to supply identification and address information, specifically e-mail address, postal address of the home institution and/or business telephone numbers. This information is used exclusively for documentation and attribution via metadata, except for contributors of data or other materials displayed on EarthRef.org. The identification of contributors and their addresses is an integral part of scholarly data archiving and this information is continuously updated in the EarthRef.org address register. EarthRef.org does not distribute address information to third parties.

Copyright

EarthRef.org is a non-commercial database and an internet resource that supports a quantitative understanding of the Earth as a chemical, physical and biological system. All materials posted at EarthRef.org are permanently archived and freely available for not-for-profit scientific or educational purposes. Contributors to EarthRef.org agree to this copyright policy, but they retain the right to distribute these materials themselves for scientific or educational purposes. Contributors retain the exclusive right for commercial dissemination, but EarthRef will automatically gain these rights if the contributor cannot be located. Any use of EarthRef.org materials should be properly attributed to the original resource or contributor, in accordance with copyrights as defined by the original source.

Links to Other Sites

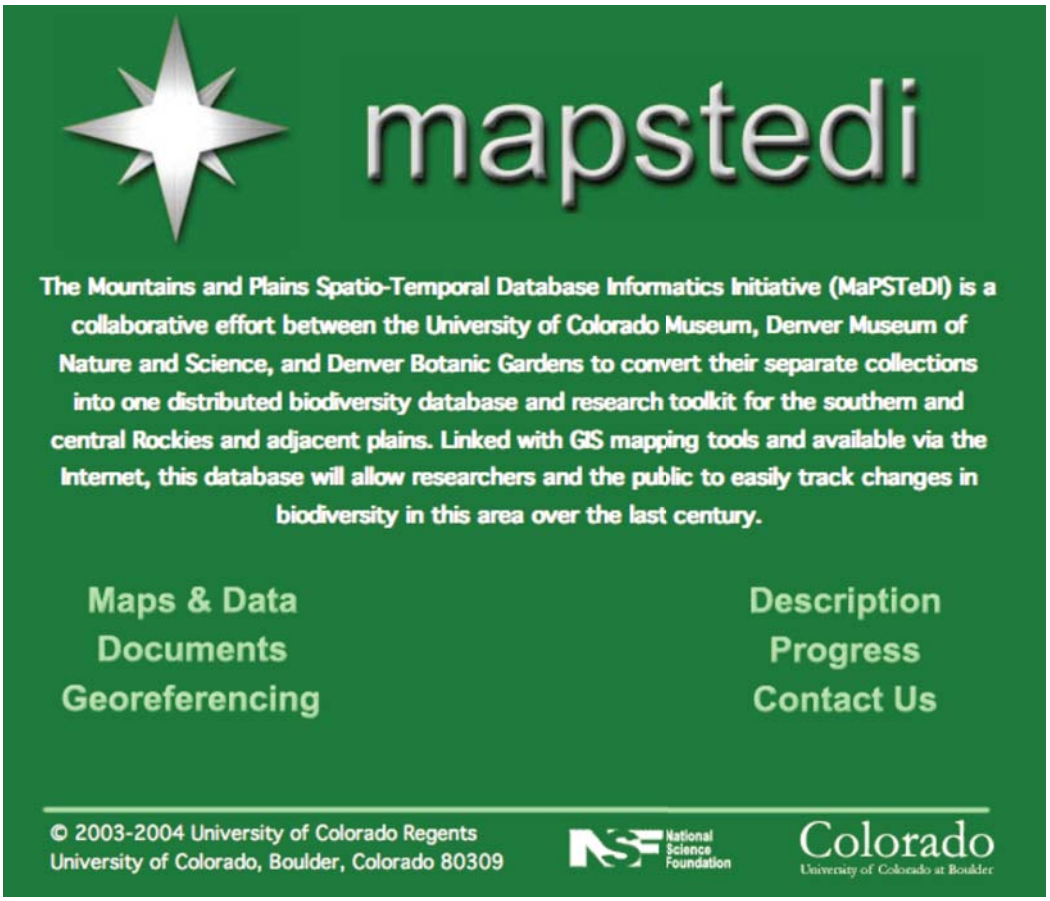
EarthRef.org links to other internet sites but it is not responsible for contents or privacy practices of those sites.

Liability

EarthRef.org is a community initiative and follows the principles and ethics of scholarly science publications. All contributors are expected to follow these principles or ethics guidelines. All users of EarthRef.org are encouraged to help maintain the highest scientific standards of its web contents through rigorous and critical review, and reporting these reviews to EarthRef.org. EarthRef.org does its best to incorporate these reviews and update its contents. However, all use of its services is at the user's risk. EarthRef.org does not take responsibility or liability for its service and the correctness of information on its websites.

MOUNTAIN AND PLAINS SPATIO-TEMPORAL DATABASE-INFORMATICS INITIATIVE

http://mapstedi.colorado.edu/

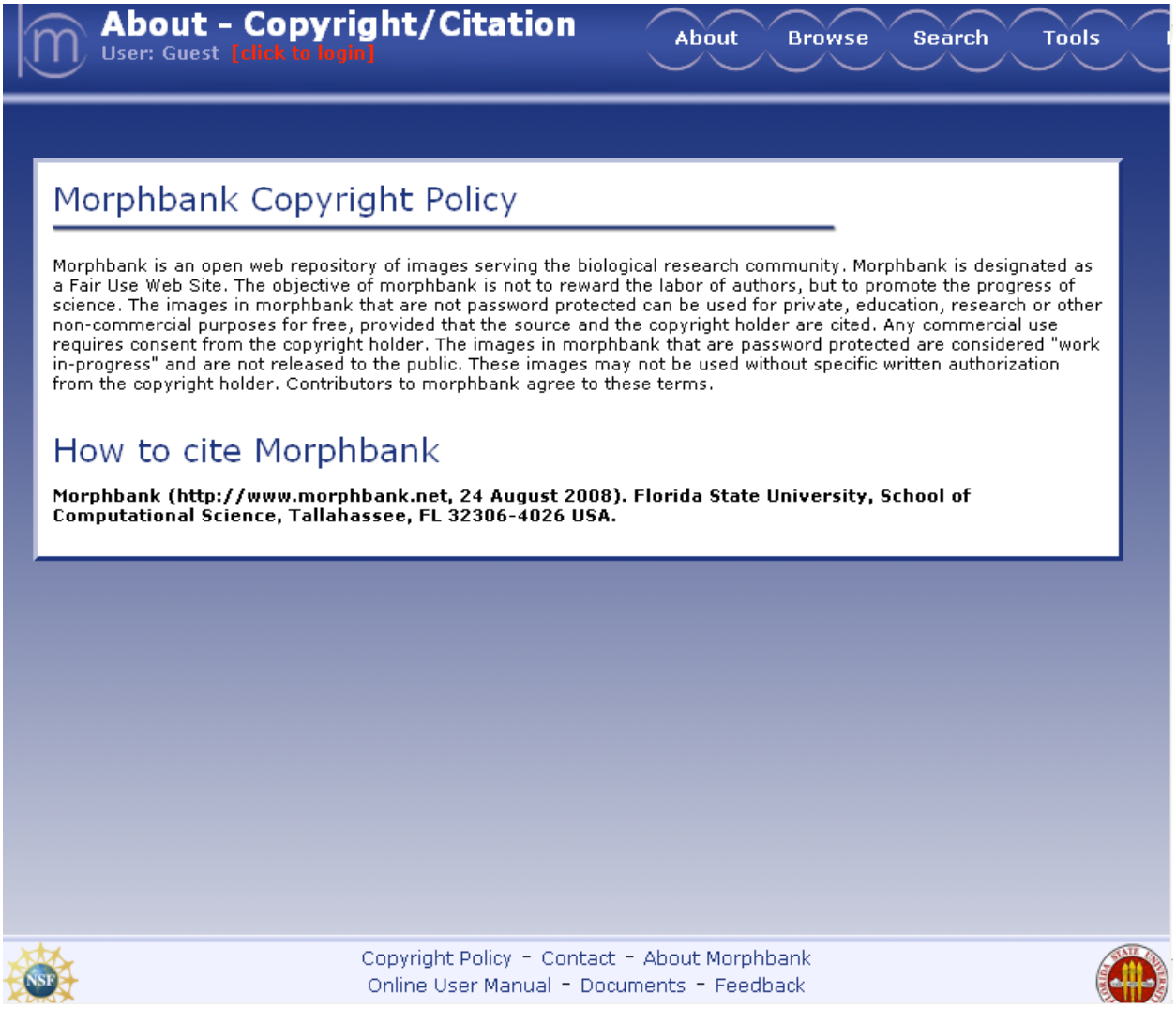


“Regarding the question at hand, yes, Mapstedi "trusts the good will of the user community". We do hope that anyone using the site for-profit would alert us, because we don't really want to give away data to those who are going to use it to make money themselves.”

Robert P. Guralnick, pers. comm.
25 Aug 2008

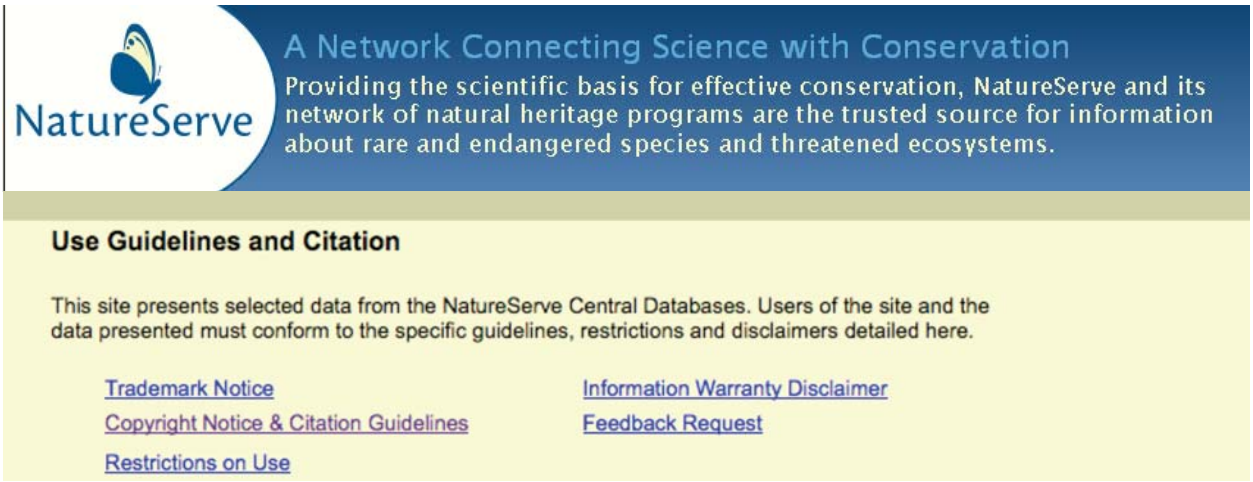
MorphBank

http://www.morphbank.net/



NATURESERVE

http://www.natureserve.org/



Citation for data on website including Watershed and State Distribution maps:
NatureServe. 2008. NatureServe Explorer: An online encyclopedia of life [web application]. Version 7.0. NatureServe, Arlington, Virginia. Available http://www.natureserve.org/explorer. (Accessed: August 28, 2008).

Citation for Bird Range Maps of North America:
Ridgely, R.S., T.F. Allnutt, T. Brooks, D.K. McNicol, D.W. Mehlman, B.E. Young, and J.R. Zook. 2003. Digital Distribution Maps of the Birds of the Western Hemisphere, version 1.0. NatureServe, Arlington, Virginia, USA.

Acknowledgement Statement for Bird Range Maps of North America:
"Data provided by NatureServe in collaboration with Robert Ridgely, James Zook, The Nature Conservancy - Migratory Bird Program, Conservation International - CABS, World Wildlife Fund - US, and Environment Canada - WILDSPACE."

Citation for Mammal Range Maps of North America:
Patterson, B.D., G. Ceballos, W. Sechrest, M.F. Tognelli, T. Brooks, L. Luna, P. Ortega, I. Salazar, and B.E. Young. 2003. Digital Distribution Maps of the Mammals of the Western Hemisphere, version 1.0. NatureServe, Arlington, Virginia, USA.

Acknowledgement Statement for Mammal Range Maps of North America:
"Data provided by NatureServe in collaboration with Bruce Patterson, Wes Sechrest, Marcelo Tognelli, Gerardo Ceballos, The Nature Conservancy-Migratory Bird Program, Conservation International-CABS, World Wildlife Fund-US, and Environment Canada-WILDSPACE."

NOTE: Full metadata for the Bird Range Maps of North America is available at:
http://www.natureserve.org/library/birdDistributionmapsmetadatav1.pdf.

Full metadata for the Mammal Range Maps of North America is available at:
http://www.natureserve.org/library/mammalsDistributionmetadatav1.pdf.

NETHERLANDS BIODIVERSITY INFORMATION FACILITY

http://www.nlbif.nl/

Liability

The NLBIF Web site has been assembled with the utmost care. The contents of this site are regularly checked and updated. NLBIF however cannot guarantee that all information is correct, complete and up-to-date. NLBIF does not assume responsibility for any possible damage or consequences arising from direct or indirect use of the contents of this Web site.

Information originating from third parties

Information obtained by means of the search options of the "bio-diversity portal" originates with third parties, as does the "biodiversity news" generated by means of RSS feeds. Although NLBIF strives to offer quality information, the gathered information has not been assessed for validity and accuracy. NLBIF expressly rejects any liability on this point.

Copyrights

The copyrights for the contents of this Web site rests with NLBIF. The contents of this site cannot be duplicated, modified and/or made public without prior authorisation. The copyrights of information originating from third parties that is available on the NLBIF site rest with the owners and generators of this information. Use of information obtained by means of the "biodiversity portal" for publication purposes is allowed only if a written authorisation has been obtained in advance from the owners of the information.

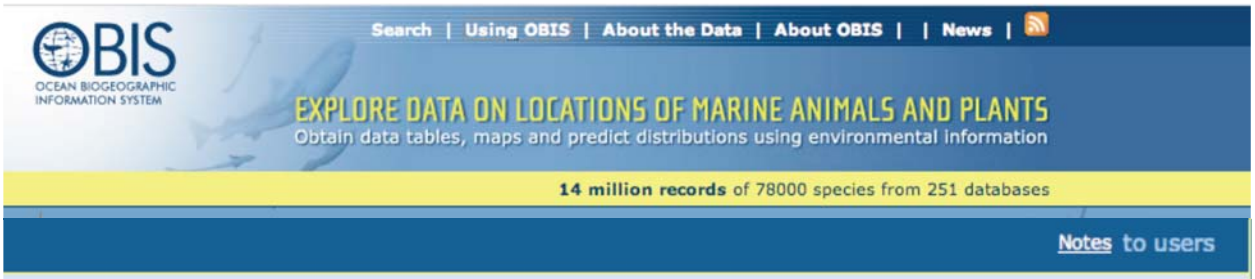
Hyperlinks

Hyperlinks to the homepage are permitted without prior authorisation of NLBIF. Links to underlying pages require prior authorisation from NLBIF.

Last modified: March 8th, 2007

OBIS

http://www.iobis.org/tech/



Notes to users

How to cite OBIS

By using data, software, or other information accessed through the OBIS Portal, I agree that, in any publication or presentation of any sort based wholly or in part on material so accessed, I will:
1. Acknowledge the use of specific records from contributing databases in the form appearing in the 'Citation' field thereof (if any); and acknowledge the use of the OBIS facility in one of the following prescribed forms:

For OBIS website:

Ocean Biogeographic Information System. [date accessed] www.iobis.org

For data used:

Author, initials. *Database title.* Retrieved [date accessed] from www.iobis.org

Examples:

(a) Single level

Stocks, K. *SeamountsOnline: an online information system for seamount biology. Version 3.1.* Retrieved [date] from www.iobis.org.

(b) multiple level (for example, to cite a database published on a Compact Disc and accessed through OBIS)

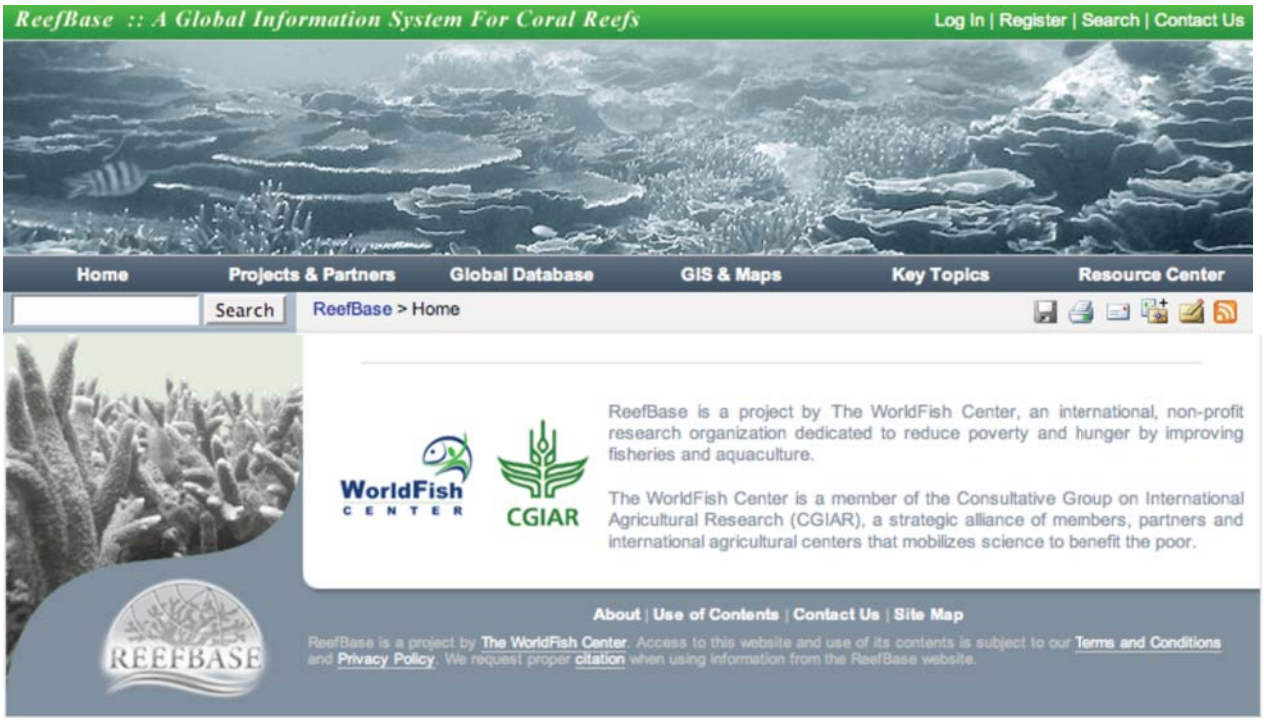
Picton, B.E., Emblow, C.S., Morrow, C.C., Sides, E.M., Tierney, P., McGrath, D., McGeough, G., McCrea, M., Dinneen, P., Falvey, J., Dempsey, S., Dowse, J. and Costello, M. J. 1999. Marine sites, habitats and species data collected during the BioMar survey of Ireland. In: Picton, B.E. and Costello M. J. (eds), *The BioMar biotope viewer: a guide to marine habitats, fauna and flora in Britain and Ireland*, Environmental Sciences Unit, Trinity College, Dublin. Retrieved [date] from www.iobis.org.

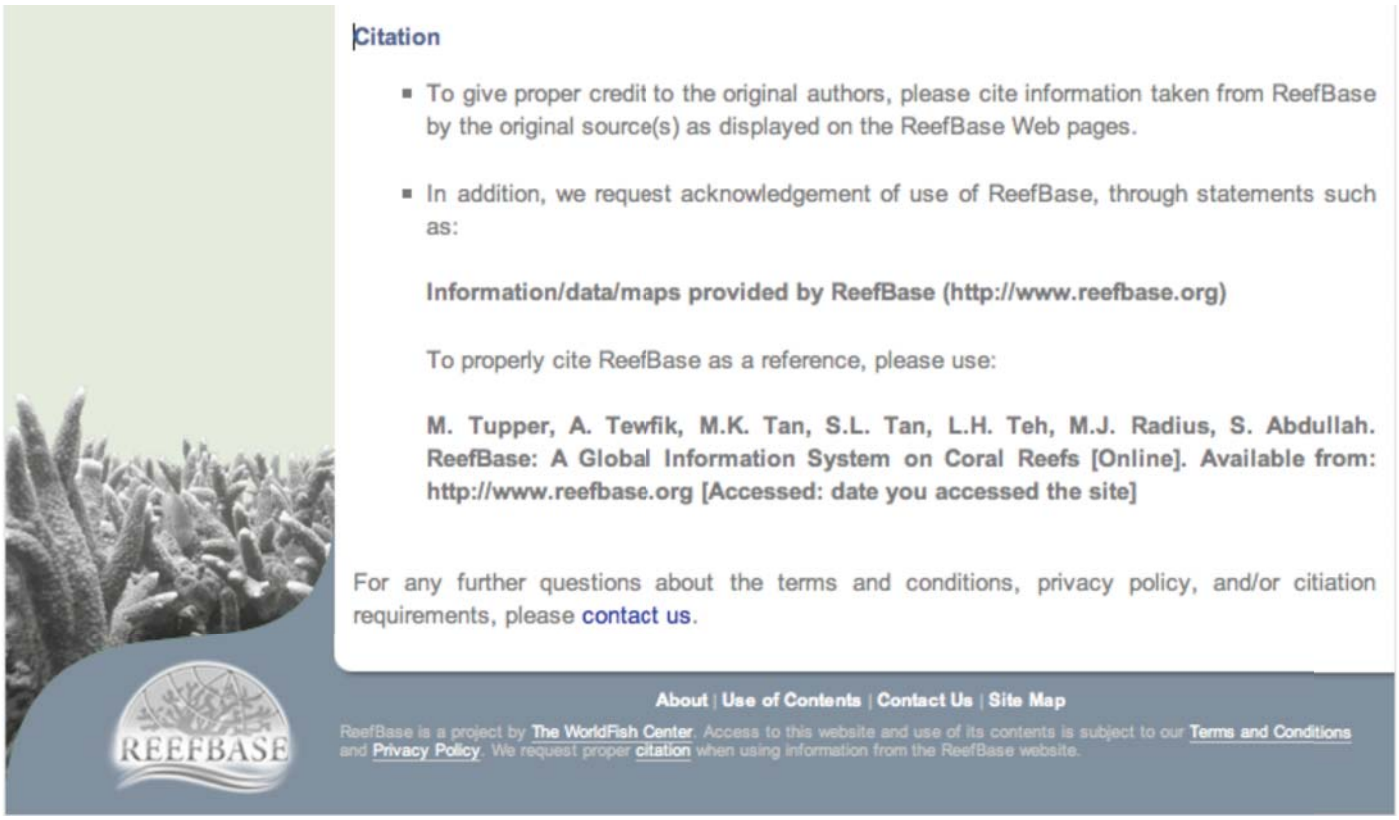
2. For information purposes, provide to OBIS Portal Support <obissupport(at)marine.rutgers.edu> the full citation of any publication I make (printed or electronic) that cites OBIS or any constituent part.

3. recognize the limitations of data in OBIS

REEFBASE

http://www.reefbase.org/main.aspx





Use of Contents

Terms and Conditions

* By accessing and using any of the datasets distributed by ReefBase, you acknowledge that you have read, understood, and agree to comply with the terms and conditions specified on this page. ReefBase may amend these terms and conditions at any time. Your continued use of this data constitutes acceptance of the terms and conditions stated at the time of your use. You should contact ReefBase to determine the current terms and conditions to which you are bound.

* You may use these datasets for non-commercial purposes, including research, education, presentations, and non-commercial publications. You understand and agree that you may not copy, reproduce, distribute or create derivative works for other purposes, without prior, written authorization from the ReefBase Project Leader. When using or publishing any parts of these datasets, proper acknowledgement of ReefBase and original source(s), is required.

* ReefBase makes reasonable efforts to provide accurate data and information through its Web site, database, and other channels. Nevertheless, you understand and agree that ReefBase shall not be liable for any loss or damage resulting from any use of, or inability to use, this data, or resulting form any errors or omissions in the data content.

* Unless expressly stated otherwise, the information, interpretations, and conclusions expressed on the ReefBase Web site are those of the various authors of the work and are not necessarily those of the ReefBase Project or the WorldFish Center.

* As a condition of use of the ReefBase Web site, the user agrees to indemnify the ReefBase Project, the WorldFish Center and its affiliates from and against any and all actions, claims, losses, damages, liabilities and expenses arising out the user's use of this site. If the user is dissatisfied with any material on this Site or with any of the terms and conditions of use of this Site, the user's sole and exclusive remedy is to discontinue using this Site.

* The ReefBase Web site contains links to third-party Web sites. The linked sites are not under the control of the ReefBase Project and the ReefBase Project is not responsible for the contents and/or privacy policies of any of the linked sites or any link contained in a linked site. The inclusion of a link does not imply endorsement of the linked site by the ReefBase Project.

Privacy Policy

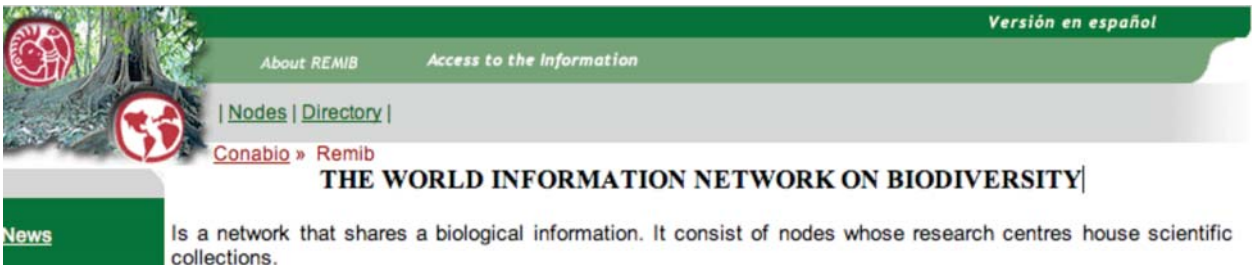
* ReefBase collects personal information when you register, when you visit the ReefBase website and when you contact us via email. When you register we ask for information such as your name, email address, occupation, and your educational/professional background.

* ReefBase does not rent, sell, or share personal information about you with other people or nonaffiliated companies except to provide products or services you've requested (such as our E-Newsletter), when we have your permission (as specified by the user during the registration process). Users can request and/or make changes to, and/or deactivate all their User Registration Information by using the tools on the Web site available after login.

* ReefBase automatically records information on its server logs from the user's browser, including IP address, ReefBase cookie information, the page(s) requested, and files downloaded. ReefBase uses this information to customize the content the user sees, fulfill the user's requests for products and services, improve its services, contact the user, conduct analysis and research, and provide anonymous reporting for internal and external purposes.

RED MUNDIAL DE INFORMACIÓN SOBRE BIODIVERSIDAD

http://www.REMIB.org/



A NETWORK THAT HAS 5,756,136 RECORDS

CONDITIONS FOR DATA USE

Queries can be made under the following conditions:

a. Queries: the user querying information on REMIB is obliged to notify those projects for whom they require information, and the form of academic acknowledgement required in relation to the intellectual ownership of the collectors, determinators or institutions that have generated the information.
b. Acknowledgements: each user consulting the database is obliged to mention in their acknowledgements and bibliographic citations the original source of the data. i.e., they must cite REMIB, the nodes consulted, as well as the suppliers of the data, and this acknowledgement must also be given in any subsequent use of the data.
c. Responsibility: neither REMIB nor the curator of the database takes responsibility for the accuracy and/or reliability of the data or of the uses that the data might be put.
d. Information obtained: the user has the responsibility that the information supplied will not be used to the detriment of national flora and fauna conservation (*i.e.* species trade) or in processes that irreversibly alter the ecological equilibrium (*i.e.* indiscriminate extraction or the introduction of alien species).
f. Commercial use of the data: the user must inform the REMIB if the data extracted is going to be used for commercial gain or profit.
e. Production of materials: if the central node or any other node of REMIB produces material for a user (such as, data summaries or syntheses, graphics, statistics, etc.) the node should:
- **a.** acknowledge through citations, the original sources of all data used.
- **b.** acknowledge the curator of the source of data used and inform him about the queries made to his data.

g. Redistribution: data cannot be redistributed.
h. Availability of the information:
- **NA.**- the field "Not available" was entered by the curator of the database.
- **999,99.**- for the longitude and latitude data, indicates that the data is "Not available".
- **Null.**- data was not entered in the field.
- **Restricted.**- the institutional node has restricted access to the information.

Note: National Authorities will be informed about anyone who does not respect the conditions mentioned above. Any unauthorized editing, reproduction, redistribution, addition or reduction of all or part of the information in REMIB or whatever other form of commercial exploitation for gain undertaken without the consent of those responsible for said data and REMIB is strictly forbidden.

Warning: The information contained in REMIB's data base could present some geographic inconsistencies, taxonomic errors or could be incomplete. If you detect inconsistencies please contact remib.conabio.gob.mx.

Accept | Reject

Last update: August 15 2006

SCAR-MARBIN

http://www.scarmarbin.be/

SCAR-MarBIN Data - Terms of use

By downloading or consulting data from this website, the visitor acknowledges that he/she agrees with the SCAR-MarBIN data policy, and agrees to the following:
I. If data are extracted from the SCAR-MarBIN website, the SCAR-MarBIN network should be cited, as appropriate:

For the **Website**:

- De Broyer C and Danis B (Editors). SCAR-MarBIN: The Antarctic Marine Biodiversity Information Network. [date accessed]. World Wide Web electronic publication. Available online at http://www.scarmarbin.be/

For the Register of Antarctic Marine Species (RAMS):

- De Broyer C, Clarke A, Koubbi P, Pakhomov E, Scott F, Vanden Berghe W and Danis B (Editors). The SCAR-MarBIN Register of Antarctic Marine Species (RAMS), [date accessed]. World Wide Web electronic publication. Available online at http://www.scarmarbin.be/scarramsabout.php

For **Data** used:

- Author, initials. Database title. Retrieved [date accessed] from http://www.scarmarbin.be/

Examples:

- De Broyer, C. Ant'Phipoda: biodiversity of Antarctic Amphipods. Retrieved [date accessed] from http://www.scarmarbin.be/
- David B., Chone T., Festeau A., De Ridder C.. Antarctic Echinoids, an Interactive Database. Editions Universitaires Dijon. Cd Rom. Retrieved [date accessed] from http://www.scarmarbin.be/.

II. For information purposes, please provide to SCAR-MarBIN (info(at)scarmarbin.be) the full citation of any publication (printed or electronic) that cites SCAR-MarBIN or any constituent part.
III. The user recognizes the limitations of data in SCAR-MarBIN
Depending on your browser settings, a cookie may be stored to acknowledge your acceptance of these terms.

Accept Decline

SCIDEV.NET

http://www.scidev.net/en/

SciDev Net

Science and Development Network
News, views and information about science, technology and the developing world

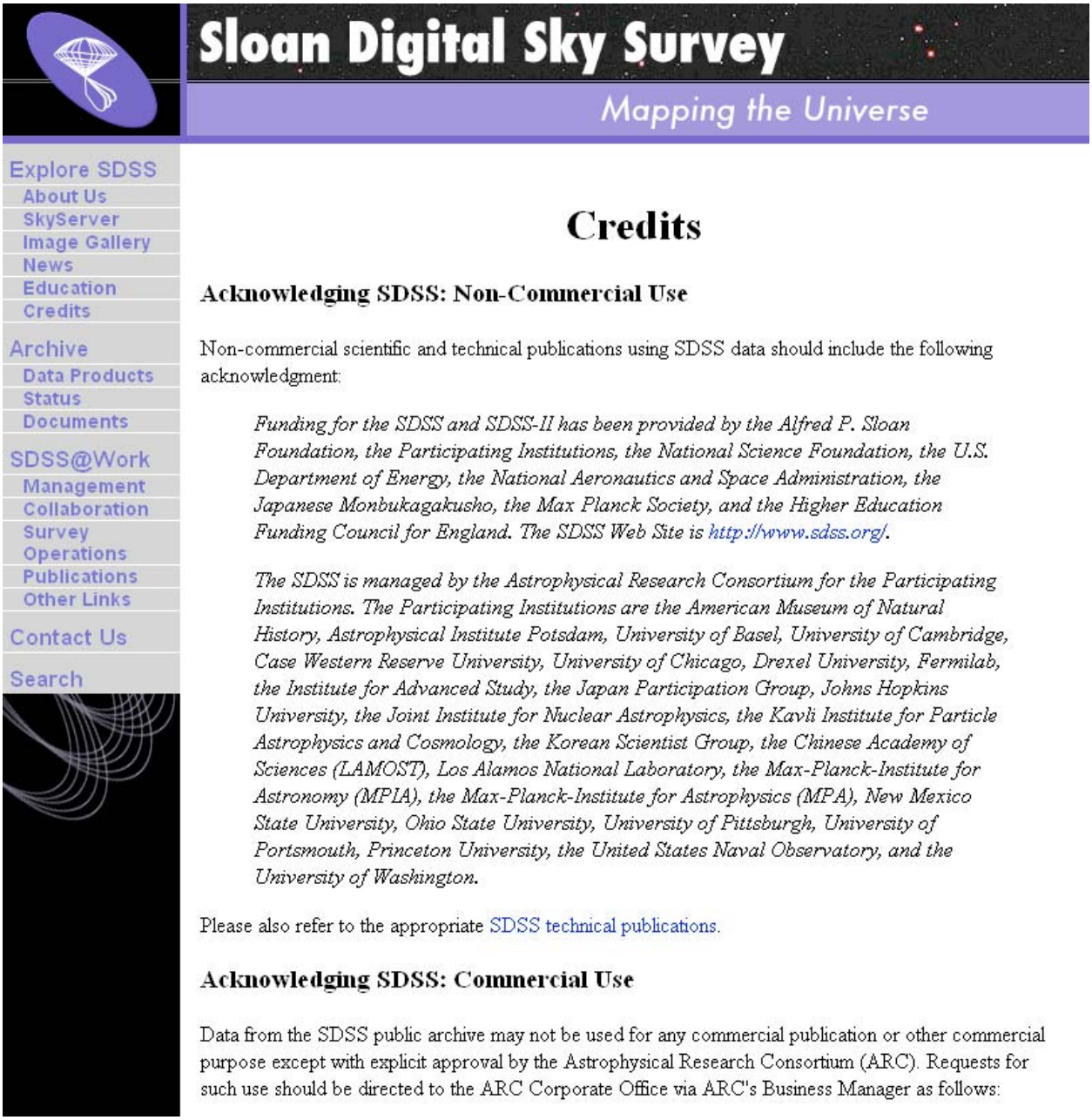
SEARCH Advanced Search

Using SciDev.Net material

Copyright

SLOAN DIGITAL SKY SURVEY

http://www.sdss.org/collaboration/credits.html



Sloan Digital Sky Survey

Mapping the Universe

Explore SDSS
About Us
SkyServer
Image Gallery
News
Education
Credits
Archive
Data Products
Status
Documents
SDSS@Work
Management
Collaboration
Survey Operations
Publications
Other Links
Contact Us
Search

Credits

Acknowledging SDSS: Non-Commercial Use

Non-commercial scientific and technical publications using SDSS data should include the following acknowledgment:

> *Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.*
>
> *The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.*

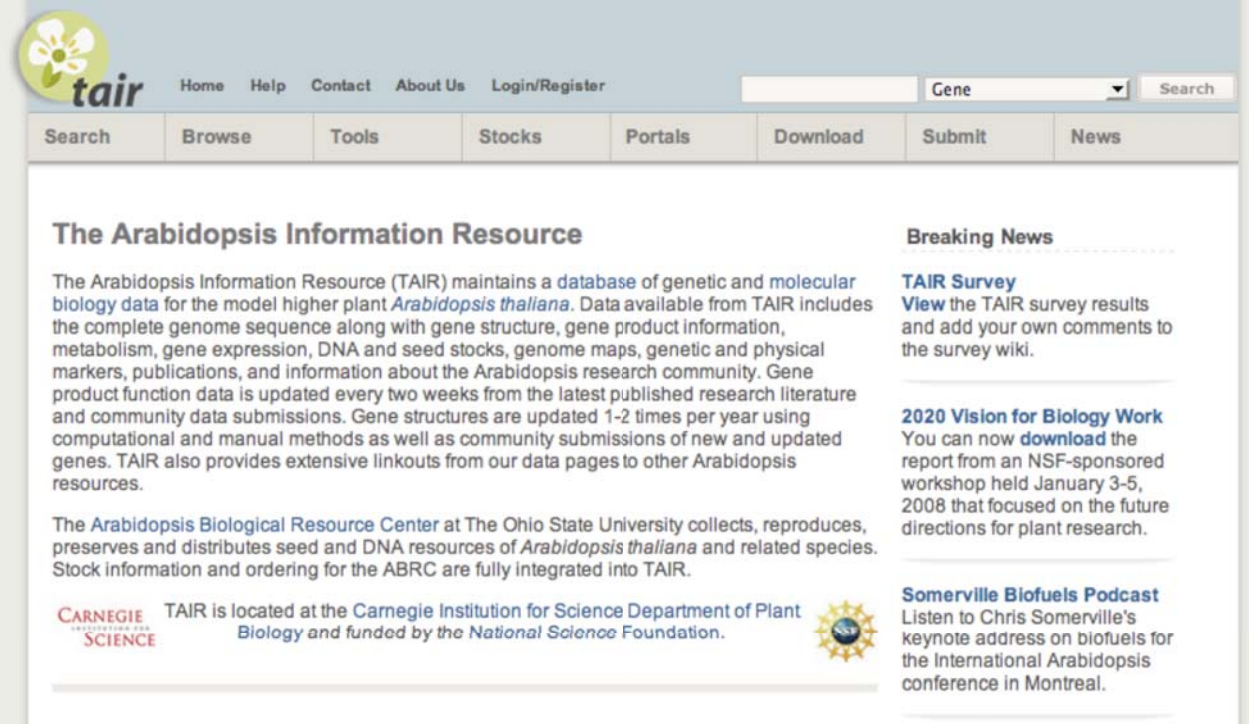
Please also refer to the appropriate SDSS technical publications.

Acknowledging SDSS: Commercial Use

Data from the SDSS public archive may not be used for any commercial publication or other commercial purpose except with explicit approval by the Astrophysical Research Consortium (ARC). Requests for such use should be directed to the ARC Corporate Office via ARC's Business Manager as follows:

THE ARABIDOPSIS INFORMATION RESOURCE

http://www.arabidopsis.org/index.jsp



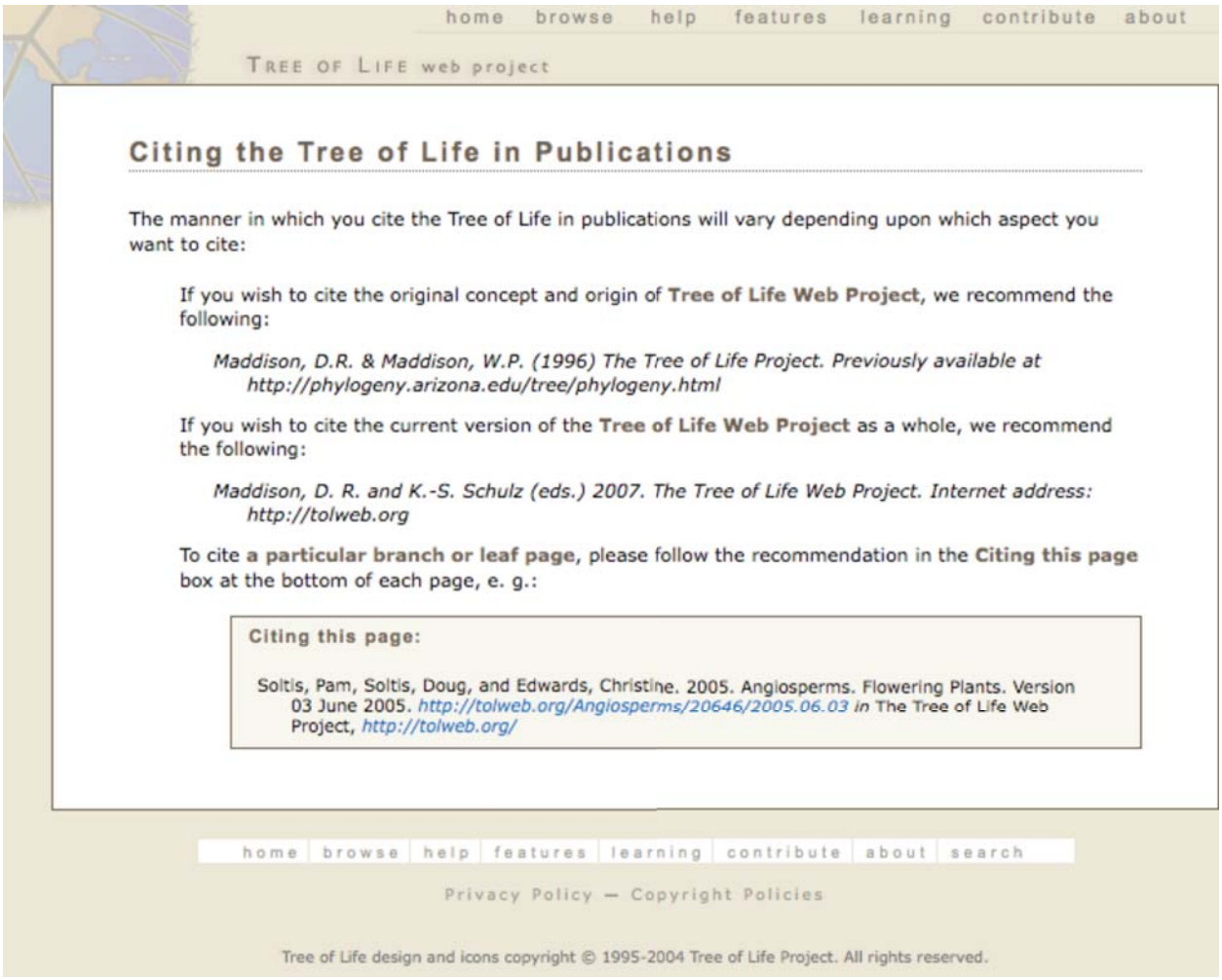
printer-friendly version

General comments or questions: curator@arabidopsis.org
Seed or DNA stock questions (donations, availability, orders, etc): abrc@arabidopsis.org

CARNEGIE INSTITUTION FOR SCIENCE ABRC

TREE OF LIFE WEB PROJECT

http://www.tolweb.org/tree/phylogeny.html



Tree of Life Copyright Policies

The Tree of Life is a collaborative project, featuring materials provided by hundreds of contributors. Most Tree of Life content is protected by copyright, and visitors to the ToL site may view, browse, and/or download these materials for temporary copying purposes only, provided these uses are for noncommercial, personal purposes. Except as provided by law, this material may not be further reproduced, distributed, transmitted, modified, displayed, published, or sold in whole or in part, without prior written permission from the copyright owner.

Note that there is no single copyright owner for all the materials displayed on the ToL web site. Rather, ToL contributors usually retain the copyright for their contributions (see below). Therefore, **Tree of Life project staff cannot generally grant permission for the use of materials displayed on ToL branch pages, leaf pages, articles, notes, or treehouses.** If you would like to use ToL content for your own projects, you need to check the associated copyright notices to find out who owns the copyright for the things you are interested in. If a copyright owner's home page url or e-mail address is published on the ToL site, you can contact that person directly. If no such information is provided, you have to assume that the copyright owner is not interested in receiving permission requests for alternative uses.

Educators: Under certain circumstances, you and your students are entitled to use portions of copyright protected works in educational multimedia projects without asking copyright owners for permission. For more information about this topic, please refer to the Fair Use Guidelines For Educational Multimedia. If you would like to use other people's materials for your own projects but aren't sure whether you need to ask for permission first, please take the time to brush up on your copyright knowledge with Georgia Harper's Crash Course in Copyright.

Overview of the Copyright Ownership for Different ToL Materials

Tree of Life Scientific Content and Learning Materials

The copyright for the text, images, and other media on ToL branch pages, leaf pages, other articles, notes, and treehouse pages generally belongs to the authors, photographers, and other artists who have created these materials. Except for the purposes of fair use as defined by copyright law, this content cannot be used without prior permission from the copyright owner.

Tree of Life General Design and Documentation

The Tree of Life Project owns the copyright for the design of the web site along with any graphics that define the general look of the site. We also own the copyright for all the text and images provided on the ToL documentation pages (the pages providing background information about the project). Reference to the Tree of Life Project and its web site may not be used in any advertising or publicity to endorse or promote any products or commercial entity.

Tree of Life Glossary

The Tree of Life Project owns the copyright for the ToL Glossary. The content of the ToL Glossary is made available to the public under the terms of the Attribution-NonCommercial Creative Commons License. This license allows people to copy, modify, and redistribute ToL Glossary definitions, subject to some restrictions.

Tree of Life Tree Structure XML File

The Tree of Life Project owns the copyright for the ToL Tree Structure XML File. This file is made available to the public under the terms of the Attribution-NonCommercial Creative Commons License. This license allows people to copy, modify, and redistribute the ToL tree structure, subject to some restrictions.

Copyright Infringement on the Tree of Life Website

Tree of Life contributors are expected to respect the intellectual property rights of others and agree to use materials according to the instructions of copyright owners, or get permission in writing for their specific use. If you believe that your work has been used on the ToL web site in a way that constitutes copyright infringement, you have the option of filing a Copyright Infringement Complaint.

TREEBASE

http://www.treebase.org/treebase/index.html

Hi Meredith,

We should probably write something explicit and upfront. For the most part I communicate this on an individual basis.

I guess we've always conceived TreeBASE as having two separate roles: one as a digital library where we just store data for published research, and two, as a data source for meta-analyses and phyloinformatics.

In the first role, which is by far the more common one, we don't expect people to cite us for downloading someone else's data (just like Genbank doesn't insist that people include a special citation in the event that a sequence is downloaded and reused).

Nonetheless, each downloaded file comes with a standard statement embedded as a nexus comment, like so:

> This data set was downloaded from TreeBASE, a prototype relational database of phylogenetic knowledge. TreeBASE has been supported by the NSF, Harvard University, and UC Davis. Please do not remove this acknowledgment from the Nexus file.
> TreeBASE © 1994-2008.

... with the idea that people who want to build their own, specialty database using our data are encouraged to show the source of data if it came from us as opposed to digitizing it or getting it directly from the author. Perhaps the "©" is a misnomer here -- and (cc) would be better, specifically allowing reuse with attribution.

In the second role, as a source for meta-analysis, we explicitly ask people to cite TreeBASE. The rationale here is that a collection of records, as a whole, has emergent properties that are specifically attributable to TreeBASE; whereas a record taken individually is more attributable to the original author (even if TreeBASE, in both cases, serves as a "publisher" of sorts). Since these kinds of meta-analyses invariably require that I send part or all of the database (as a compressed data dump) to the researcher, I normally make use of the email exchange to request that they cite us. Typically, I suggest this:

> Piel, W. H., M. J. Sanderson, and M. J. Donoghue. 1998. TreeBASE: a database of phylogenetic knowledge. Beta version. http://www.treebase.org

But sometimes they use citations like these:

Piel, W. H., M. J. Sanderson, and M. J. Donoghue. 2003. The Small-World Dynamics of Tree Networks and Data Mining in Phyloinformatics. Bioinformatics, 19(9): 1162-1168.

Piel, W. H., M. J. Donoghue, and M. J. Sanderson. 2002. "TreeBASE: a database of phylogenetic knowledge." Pp. 41-47. In: Shimura, J., K. L. Wilson, and D. Gordon, eds. To the interoperable "Catalog of Life" — with partners Species 2000 Asia Oceanea. Research Report from the National Institute for Environmental Studies No. 171, Tsukuba, Japan.

Sanderson, M. J., M. J. Donoghue, W. H. Piel, and T. Eriksson. 1994. TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. Amer. J. Bot., 81(6): 183.

Regards,

Bill

Wiliam Piel, pers. comm
25 Aug 2008

UK NATIONAL BIODIVERSITY NETWORK

http://www.nbn.org.uk/


NBN
National Biodiversity Network

NBN Gateway Terms and Conditions

IMPORTANT: By using this site you are agreeing to be bound the following *Gateway Terms & Conditions* and our *Privacy Policy* . The *Gateway Terms & Conditions* govern the way you can use the NBN Gateway Website Service and any *material, data and/or information* made available to you through it. Please read them carefully.

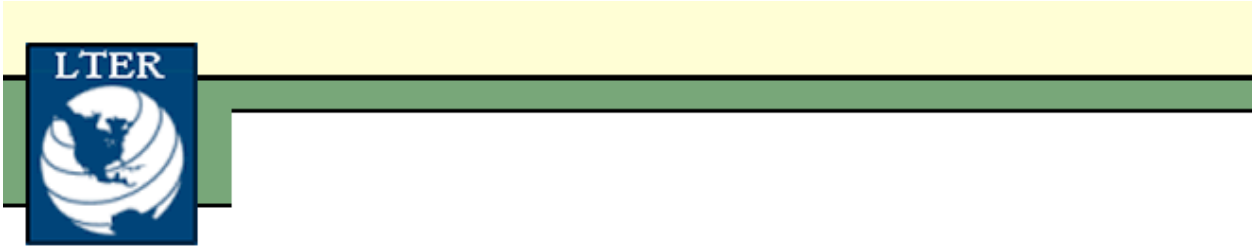
We may change these *Gateway Terms & Conditions*, or our *Privacy Policy* at any time without giving you notice, so please check them each time you visit this website. Additionally, any *Data Provider* may add, change or remove *Specific Terms of Access* for a dataset they administer on the NBN Gateway at any time without giving you notice, so please check these within the *Metadata* provided for any dataset each time you access and use them.

2. Use of *material, data and/or information*

2.1	All *material, data and/or information* on this website are protected by copyright, database rights and other intellectual property rights. If you access and use any of the *material, data and/or information* made available to you on this site, you do so subject to these *Gateway Terms & Conditions*.
2.2	The *material, data and/or information* you access may be subject to *Specific Terms of Access* imposed on its use by the Data Provider. Where *Specific Terms of Access* have been identified for a dataset they are provided within the 'Use Constraint' field of the *Metadata* for that dataset. If there is any conflict or inconsistency between these *Gateway Terms & Conditions* and the *Specific Terms of Access*, the latter will prevail, but only in relation to the particular *material, data and/or information* restricted by those *Specific Terms of Access*, and only to the extent necessary to resolve that conflict or inconsistency.
2.3	Under these *Gateway Terms & Conditions* you are permitted to view and download any *material, data and/or information* that *Data Providers* allow you to access on this website for your own private use or for use in the ordinary course of your business provided that use is in accordance with these *Gateway Terms & Conditions*.
2.4	You must acknowledge the contribution of the relevant *Data Providers* in any derived information product or publication, whether printed, electronic or broadcast, that is based wholly or in part on the *material, data and/or information* they make available to you. Where you make specific use of or reference to a particular wildlife record you must acknowledge the Original Recorder [where their name is made available].
2.5	*Data Providers* may restrict the availability of some or all parts of the *material, data and/or information* they supply for a variety of reasons using the Access Administration Controls on the NBN Gateway. If you are granted access to restricted information you must not disseminate, deliberately compromise or otherwise infringe the confidentiality of such information without first obtaining written permission from the relevant Data Provider. Any published interpretations or representations you make based on the information must be equivalent to the level of information available to non-registered users on this website.
2.6	You may not make any financial profit from use of the *material, data and/or information* on this website or from any products you derive without first obtaining written permission from the relevant Data Provider.
2.7	You may not republish wholesale the *material, data and/or information* made available to you, or exploit it for commercial or academic research purposes without first obtaining written permission from the relevant *Data Provider* [the contact details of which are provided within the metadata of the relevant dataset].
2.8	Any information product or publication that you make which contains any part of the *material, data and/or information* made available to you via this website must contain a statement that the Data Provider, Original Recorder [where identified], and the NBN Trust bear no responsibility for any further analysis or interpretation of that *material, data and/or information*.

US LTER

http://www.lter.edu



LTER Network Data Access Policy, Data Access Requirements, and General Data Use Agreement

approved by the LTER Coordinating Committee April 6, 2005

Long Term Ecological Research Network Data Access Policy

The LTER data policy includes three specific sections designed to express shared network policies regarding the release of LTER data products, user registration for accessing data, and the licensing agreements specifying the conditions for data use.

LTER Network Data Access Requirements

The access to all LTER data is subject to requirements set forth by this policy document to enable data providers to track usage, evaluate its impact in the community, and confirm users' acceptance of the terms of acceptable use. These requirements are standardized across the LTER Network to provide contractual exchange of data between Site Data Providers, Network Data Providers, and Data Users that can be encoded into electronic form and exchanged between computers. This will allow direct access to data via a common portal once these requirements have been fulfilled. The following information may be required directly or by proxy prior to the transference of any data object:

Registration
1. Name
2. Affiliation
3. Email Address
4. Full Contact Information

- Acceptance of the General Public Use Agreement or Restricted Data Use Agreement, as applicable.
- A Statement of Intended Use that is compliant with the above agreements. Such statements may be made submitted explicitly or made implicitly via the data access portal interface.

Conditions of Use

The re-use of scientific data has the potential to greatly increase communication, collaboration and synthesis within and among disciplines, and thus is fostered, supported and encouraged. Permission to use this dataset is granted to the Data User free of charge subject to the following terms:

1) *Acceptable use*. Use of the dataset will be restricted to academic, research, educational, government, recreational, or other not-for-profit professional purposes. The Data User is permitted to produce and distribute derived works from this dataset provided that they are released under the same license terms as those accompanying this Data Set. Any other uses for the Data Set or its derived products will require explicit permission from the dataset owner.

2) *Redistribution*. The data are provided for use by the Data User. The metadata and this license must accompany all copies made and be available to all users of this Data Set. The Data User will not redistribute the original Data Set beyond this collaboration sphere.

3) *Citation*. It is considered a matter of professional ethics to acknowledge the work of other scientists. Thus, the Data User will properly cite the Data Set in any publications or in the metadata of any derived data products that were produced using the Data Set. Citation should take the following general form: *Creator, Year of Data Publication, Title of Dataset, Publisher, Dataset identifier*. For example:

> McKee, W. 2001. Vascular plant list on the Andrews Experimental Forest and nearby Research Natural Areas: Long-Term Ecological Research. Corvallis, OR: Forest Science Data Bank: SA002. [Database]. http://www.fsl.orst.edu/lter/data/abstract.cfm?dbcode=SA002. (21 October 2004)

4) *Acknowledgement*. The Data User should acknowledge any institutional support or specific funding awards referenced in the metadata accompanying this dataset in any publications where the Data Set contributed significantly to its content. Acknowledgements should identify the supporting party, the party that received the support, and any identifying information such as grant numbers. For example:

> Data sets were provided by the Forest Science Data Bank, a partnership between the Department of Forest Science, Oregon State University, and the U.S. Forest Service Pacific Northwest Research Station, Corvallis, Oregon. Significant funding for collection of these data was provided by the National Science Foundation Long-Term Ecological Research program (NSF Grant numbers BSR-90-11663 and DEB-96-32921).

5) *Notification*. The Data User will notify the Data Set Contact when any derivative work or publication based on or derived from the Data Set is distributed. The Data User will provide the data contact with two reprints of any publications resulting from use of the Data Set and will provide copies, or on-line access to, any derived digital products. Notification will include an explanation of how the Data Set was used to produce the derived work.

6) *Collaboration*. The Data Set has been released in the spirit of open scientific collaboration. Data Users are thus strongly encouraged to consider consultation, collaboration and/or co-authorship with the Data Set Creator.

By accepting this Data Set, the Data User agrees to abide by the terms of this agreement. The Data Owner shall have the right to terminate this agreement immediately by written notice upon the Data User's breach of, or non-compliance with, any of its terms. The Data User may be held responsible for any misuse that is caused or encouraged by the Data User's failure to abide by the terms of this agreement.

USA NATIONAL BIOLOGICAL INFORMATION INFRASTRUCTURE

http://gbif.nbii.gov/portal/server.pt

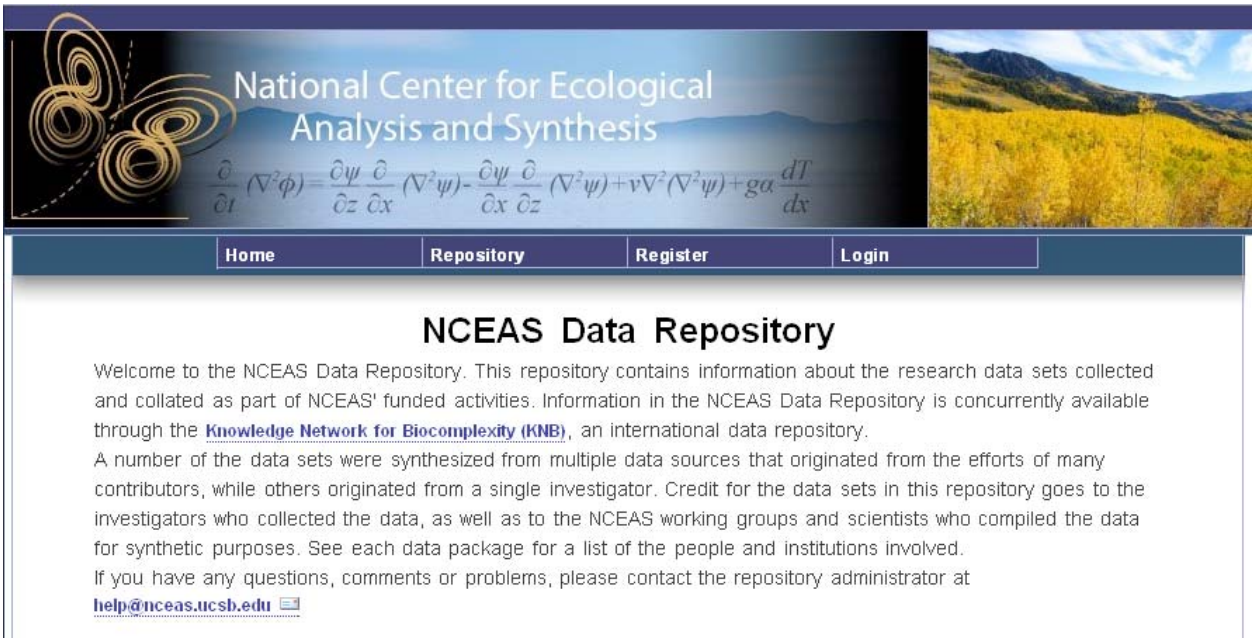
Copyright

Attribution and Citing of NBII Data and Information

As a condition of use, those who refer to data and information found through the NBII portal in their publications and presentations must formally cite both the original author(s) and data authority for any cited item(s) as well as the NBII itself, as the source through which the data or information was found. Such citation may be in any format appropriate to the publication or presentation, or in any format required by the original creator or provider of the data/information. Identifying the source for citation is the responsibility of the user, and users should be prepared to provide a copy of the citing publication or presentation to the cited authors or data providers, upon their request. The NBII itself should be referred to in full as " **National Biological Information Infrastructure (NBII) at http://www.nbii.gov.** "

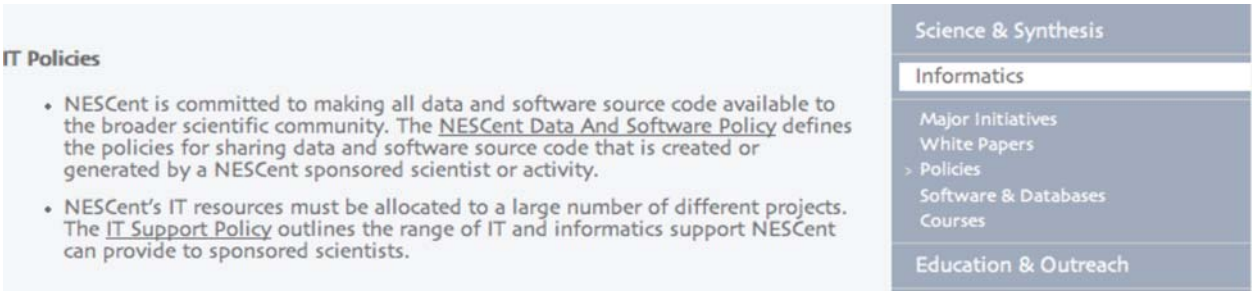
US NATIONAL CENTER FOR ECOLOGICAL ANALYSIS AND SYNTHESIS

http://www.nceas.ucsb.edu/



US NATIONAL EVOLUTIONARY SYNTHESIS CENTER

http://www.nescent.org/



Data and Software Policy

Introduction

NESCent strongly believes that sharing all data and software source code with the research community is a key ingredient towards fostering synthetic research at NESCent and elsewhere, and therefore contributes to advancing evolutionary science as a whole.

In order to accomplish its mission, and to best support the scientific community in advancing evolutionary science as a discipline, NESCent expects that all data and software created through NESCent-sponsored activities be made publicly available.

Scope

This document defines the policies for sharing data and software source code that is created, generated, used, hosted, replicated, sponsored, or made available by a NESCent sponsored scientist, project, or activity, or by a collaboration in which NESCent participates.

In the remainder of this document data and software source code within the scope of this policy is collectively referred to as NESCent-sponsored data and software source code, respectively.

Data or software used that have more restrictive terms of use and dissemination than those set forth here are exempt from this policy. However, NESCent strongly encourages its participants and collaborators to consider alternatives that do not restrict sharing and free dissemination.

NESCent-sponsored Data

NESCent is committed to making all evolutionary biology and any other data within the scope of this policy readily available to the broader scientific community.

To meet this goal, NESCent-sponsored data or datasets are to be accessible from a web-based interface, with no restrictions for use and dissemination, or at the choice of the principle investigator under a Creative Commons license or compatible terms of use, are to be made available in a timely manner, generally no later than one year after the conclusion of the NESCent award, or immediately upon publication of an associated article, whichever comes earlier, and are to be kept current, if they undergo continuing updates, and are to be sufficiently documented using appropriate standards and conventions agreed upon by the evolutionary biology community, including registration and deposition at a public meta-data registry and data repository if such a registry or repository exist. This documentation must include appropriate acknowledgment and attribution of any data used from other researchers or sources, as required by the copyright, license, or terms of use of such data.

NESCent recognizes that there are many questions associated with the open licensing of scientific data collections and encourages sponsored scientists to familiarize themselves with the issues (http://sciencecommons.org/data/dbfaq).

Copyright and Acknowledgment

Data or datasets, software, published books and scholarly articles generated or created by a NESCent fellow or through a project or meeting funded by NESCent must appropriately and conspicuously acknowledge NESCent and the National Science Foundation.

WORLD DATA CENTER FOR MICRO-ORGANISMS

http://www.wfcc.info/datacenter.html

Directory, Databases and Dictionaries compiled/indexed by WDCM

The use of the databases, including sending query, down loading or even partly copying data into your work, has to be requested proper citation of each database in your publication.

APPENDIX 2

In 2005, the GBIF Secretariat considered what citation format it should recommend to its users. The pages that follow are the results of that consideration.

Guidelines for citing specimen and observation data obtained via the GBIF Data Portal

DRAFT 0.4 (2005-04-08).

1. The citation of data obtained via the GBIF portal should include reference(s) to both GBIF and the data provider(s) from whom the data came. The GBIF citation should include the URL of the data portal.

<GBIF Data Portal, www.gbif.net>

2. Because the number of records obtained as the result of a query may vary over time, the date of the query in question must be included in the citation.

<GBIF Data Portal, www.gbif.net>. <yyyy-mm-dd>

3. Each of the data providers should be cited, with the number of records coming from that provider indicated.

<GBIF Data Portal, www.gbif.net>. <yyyy-mm-dd>. <Data Provider 1, x records>; <Data Provider 2, x records>; ...; <Data Provider n, x records>.

Special note: If the number of records is quite small, the actual specimen identifiers may be listed rather than merely number of records, but in the case of larger datasets, this may be prohibited by space available in the publication.

4. In case the provider metadata states a requirement of recognition of the individual resources (= databases, collections), they should be mentioned after the data provider.

<GBIF Data Portal, www.gbif.net>. <yyyy-mm-dd>. <Data Provider 1, Resource 1.1, x records>; <Data Provider 2, Resource 2.1, x records>; ...; <Data Provider n, Resource n, x records>.

EXAMPLE using data for *Puma concolor*. Please note that the last data provider here also includes reference to the individual resource, but the others do not.

GBIF Data Portal, www.gbif.net. 2005-03-31. Field Museum of Natural History, 10 records; Museum of Vertebrate Zoology, 204 records; Royal Ontario Museum, 1 record; University of Washington Burke Museum, 36 records; University of Turku, WWF Peru, 10 records.

EXAMPLE using all (5 records from 2 providers) data for *Amphiachyris amoena*.

GBIF Data Portal, www.gbif.net. 2005-03-31. New York Botanical Garden, 278126; University of Kansas Natural History Museum, 119894, 119895, 119896, 313664.

It should be noted that the form that these citations will take in the future will evolve from the format recommended here, once GBIF technologies have settled on type of globally unique identifiers, means of electronic recording of query results, etc. For a white paper that explains some of the challenges that must be resolved and the options available in this area, see www.gbif.org/DataProviders/Cite

These guidelines are compatible with GBIF Data Sharing Agreement and GBIF Data Use Agreement http://www.gbif.org/DataProviders/Agreements/

Guidelines for citing names data obtained via the GBIF Data Portal

BORRADOR 0.5

1. The citation of data obtained via the GBIF portal should include reference(s) to both GBIF and the data provider(s) from whom the data came. The GBIF citation should include the URL of the data portal.

<GBIF Data Portal, www.gbif.net>

2. Because the number of records obtained as the result of a query may vary over time, the date that the query was made must be included in the citation.

<GBIF Data Portal, www.gbif.net>. <yyyy-mm-dd>

3. If the records downloaded were only some subset (e.g. unabiguous synonyms or subordinate taxa) of all the data available, that fact should be indicated.

<GBIF Data Portal, www.gbif.net>. <yyyy-mm-dd>. <data subset>.

4. Then, each of the providers of name data should be cited.

<GBIF Data Portal, www.gbif.net>. <yyyy-mm-dd>. <data subset>. <Data Provider 1, Database 1.1, Taxonomist Name 1.1.1, x records>; <Data Provider 2, Database 2.1, x records>; ...; <Data Provider N, x records>.

Special note: As GBIF currently serves names data obtained from one large provider only, and by agreement with that provider, it is also required to cite the individual database(s) and the taxonomist(s) who made the intellectual contribution. Such a citation is constructed as the one from Data Provider 1 above. The other providers here require less detail.

EXAMPLE for *Acacia burttii*

GBIF Data Portal, www.gbif.net. 2005-03-31. Acacia burttii Baker f. Catalogue of Life Partnership, ILDIS World Database of Legumes, Rico ML, 1 accepted name.

EXAMPLE using subordinate taxa for genus *Salmo*

GBIF Data Portal, www.gbif.net. 2005-03-31. Subordinate taxa of genus Salmo. Catalogue of Life Partnership, Fishbase, 25 records.

EXAMPLE using synonyms for *Gutierrezia sarothrae*

GBIF Data Portal, www.gbif.net. 2005-03-31. Synonyms of Gutierrezia sarothrae. Catalogue of Life Partnership, Integrated Taxonomic Information System, 13 unambiguous synonym records.

It should be noted that the form that these citations will take in the future will evolve from the format recommended here, once GBIF technologies have settled on type of globally unique identifiers, means of electronic recording of query results, etc. For a white paper that explains some of the challenges that must be resolved and the options available in this area, see www.gbif.org/DataProviders/Cite

These guidelines are compatible with GBIF Data Sharing Agreement and GBIF Data Use Agreement http://www.gbif.org/DataProviders/Agreements/ and the Memorandum of Cooperation between GBIF and the Catalogue of Life Partnership.

How to cite GBIF data

White paper

1. Introduction

GBIF integrates millions of data records from hundreds of rather heterogeneous different sources (resources) and providers. Users of that data are required by GBIF Data Use Agreement [Annex 1] to recognise the efforts of those who make the data available. Making data available involves a value chain of IPR where each party has contributed something, and should be acknowledged as appropriate. Those who make data available should get scientific credit for doing so.

In general, citations are meant to be used in publications as references to other information resources. They should facilitate accessing these resources, checking the facts, and reproducing materials and experiments. Citations of GBIF data are no different. However, technical challenges arise from the fact that datasets served on Internet often grow, records can change, and the data providers can withdraw their data at any time.

For many reasons, including citations, individual data records and arbitrary datasets should be possible to identify. GBIF is interested in globally unique identifiers (GUIDs), comparable to GenBank's accession numbers, that would be used in biodiversity literature in similar fashion. However, a solution for the GUIDs is still being worked on and is not available at this writing. A mechanism is discussed here that would give citations a local unique reference within the GBIF data portal.

This document defines the formats and structures for citing GBIF data. It is supported by two short guidelines for occurrence and names data, respectively. The paper also discusses some technical issues that must be considered.

Goals in this design include the following:

1. Consistency in the form of citation regardless of the circumstances behind the selection of the data.
2. Avoidance of reformatting what the providers have made available (too much guesswork).
3. Independence of what particular unique identifier scheme may be adopted later by GBIF.
4. Compatibility with machine interfaces of Open Archives Initiative Protocol for Metadata Harvesting.
5. Compatibility if the existing data sharing agreements of GBIF.

2. Format of the citations

Scientific citations normally take the form of Author(s), Year, Title, Reference, Publisher. This might work for data as well, but reconstructing such a citation of very heterogeneous data sources is probably going to fail. Therefore a simplified form is sought for below.

Mapping between the above classic form of citations and the components of the GBIF information model is as follows:

- GBIF Data Portal www.gbif.net is not semantically an "author", but an "editor" or "compiler". Such entities can be in first position in traditional references.
- The sourced Data Provider is clearly the publisher.
- Names of the Resources might match titles, but not quite. Title is rather a phrase like "Data records 0000, 0001, ..., from <resource name>".

2.1. Individual record

For occurrence data the form is like this.

<GBIF citation>. <Datetime>. <Provider citation>, <Resource citation>, <Record citation>.

In cases where several records are concerned individually, the last element can be repeated.

<GBIF citation>. <Datetime>. <Provider citation>, <Resource citation> (Records: <Record citation>, <Record citation>, …).

For names data, the form is similar but includes the name. Also the taxonomist who made the revision is recognised when that is known.

<GBIF citation>. <Datetime>. <Name>. <Provider citation>, <Resource citation>, <Taxonomist name>.

2.2. Data from a single resource

All data from a resource is like above, but without specifying any records.

<GBIF citation>. <Datetime>. <Provider citation>, <Resource citation>.

2.3. Data from an entire provider

All data from a provider is like above, but without specifying any resources or records.

<GBIF citation>. <Datetime>. <Provider citation>.

2.4. Set of records from many resources and many providers (dataset)

Unlike the above, which also could be retrieved independently from a provider, this is a result of an integrative query to GBIF data portal. The result of the storage would be like this (in HTML or XML)

<GBIF citation>. <Datetime>.
<Provider citation>, <Resource citation> (, Records: <Record citation>, <Record citation>, …);
<Provider citation>, <Resource citation> (, Records: <Record citation>, <Record citation>, …);
…

Such a citation can get quite long and may not necessarily be publishable. In those cases, and where the exact dataset must be available over a longer period, it would be desirable that the query and the result be stored and referenced as one entity. Such a reference would simply be

<GBIF citation>. <Datetime>. Archived dataset <GBIF identifier>.

3. Individual elements

The question then is what needs to be included in <GBIF citation>, <Datetime>, <Provider citation>, <Resource citation>, <Record citation>, and <GBIF identifier>. The elements <Name> and <Taxonomist name> are as written and not elaborated further below.

3.1. GBIF citation

This is a simple static description of the fact that these data were accessed through GBIF, like "GBIF Data Portal, www.gbif.net".

3.2. Datetime

This would be simply a timestamp in ISO 8601 format when the query was issued, with or without time of day, except for individual records the current value of the DateLastModified field. For example:

2005-03-31T21:57:00Z

3.3. Provider citation

This is the name of the provider as retrieved by DiGIR or BioCASe:

Example: Australian Antarctic Data Centre

3.4. Resource citation

Resource name can in most case be included as the main Title. This is typically the name of a collection or database.

Example: Elephant Seal Sightings, Heard Island

We should note here that data provider and resource metadata does contain names of their custodians that possibly could be used as authors. If an Author identity was attainable, it would be the resource "administrative" contacts where these are specified. If there are no "administrative" contacts, "other" contacts would be used, and default to "technical" contacts otherwise. The names here can have either surname first or last, and probably could be formatted correctly in most cases. However, these contacts semantically do not correspond to Authors but rather to Editors or similar compilers. Therefore, these are not included in the citation. Authors are included only in citations of names data.

3.5. Record citation

There are elements in the data standards which are intended to guarantee uniqueness. For Darwin Core this will mean that we construct the citation from the InstitutionCode, CollectionCode and CatalogNumber elements, like (Records: Institution A, Collection B, Catalogue numbers ABC, DEF, GHI, JKL; Institution C, Collection D, Catalogue numbers MNO, PQR)

Example: InstitutionCode AADC, CollectionCode Ellie-Heard, CatalogNumber 1000

As the institution and collection are normally identified as the data provider and resource, respectively, only the catalog number needs to be given as the record citation. For names data, the name itself would be given as title.

If the number of records is large for the publication targeted and/or the individual identification of the records is not necessary, only the number of records may be mentioned. If the all records from a provider or resource are included, the record citation can be omitted.

3.6. GBIF identifier

In this section we discuss the issues related to storing and citing archived datasets. Archiving GBIF data is a controversial issue as it potentially removes from data providers their capability to withdraw data. This would be very problematic in cases where sensitive data on endangered species was accidentally shared. Therefore we must note that that no decision on building such an archiving mechanism has been made.

However, large arbitrarily constructed datasets are being used for analysis and there is a need to store, document, and cite them. Such citations can become very large and unpublishable using the other mechanisms discussed above. Even using the original query parameters, the resulting dataset is not likely to be identical to what was obtained at the time when the query was issued. Archiving the dataset can be done in many ways, but storing the incoming and resulting XML stream may be the simplest solution.

Archived datasets would be referenced using the <GBIF identifier>. In order to be compatible with the Open Archives Initiative, the format of the <GBIF identifier> must correspond to that of the URI (Uniform Resource Identifier) syntax. We call it GBIF_URI. It must be made clear that this is not a globally/universally unique identifier but a local one.

GBIF_URI must be simple and short, but should be able to produce to perform the same request

again (even if it is a query which may return different results in the future). It must also be future-proof because users will be publishing them in printed publications, so it must always return something sensible when users request it, even if "sensible" means a clear error message). These URIs cannot probably be rationally created for each page view of the GBIF data portal, but should be possible to generate using a specific request, i.e., a button that the user can push or a XML request generate. That event would then create and store in a database a persistent URI and return it to the user or requester.

A persistent GBIF_URI based on that model might like these examples:

http://www.gbif.net/record/1234567890
http://www.gbif.net/resource/123456
http://www.gbif.net/dataset/12345

4. Full examples

Now we can give some combined examples of static citations:

GBIF Data Portal, www.gbif.net. 2005-03-31. Museum of Vertebrate Zoology, Terrestrial Vertebrate Specimens, Record numbers 20045, 25678, 31098; University of Washington Burke Museum, 120 records.

GBIF Data Portal, www.gbif.net. 2005-03-31. Catalogue of Life Partnership, Integrated Taxonomic Information System.

GBIF Data Portal, www.gbif.net. 2005-03-31. Field Museum of Natural History, 10 records; Museum of Vertebrate Zoology, 204 records; Royal Ontario Museum, 1 record; University of Washington Burke Museum, 36 records; University of Turku, WWF Peru, 10 records.

The static citations given in examples can be tedious to construct manually, and therefore could best be generated by some appropriate tools.

5. Recommendation for DiGIR Citation metadata field

DiGIR metadata includes a Citation-field for the resources. Using the Citation-field would be good alternative way for handling citations, but it will take some standardisation work. At this writing the use of that field is very inconsistent across the providers and resources. To alleviate that problem, we discuss here what would be the best use of that that field.

First, the above formats cannot be applied for the Citation-field as a user can access the data provider directly without going via the GBIF Data Portal. Second, the provider and resource owners know exactly what are the roles of various people in producing these data.

Therefore an appropriate form for this citation is much closer to a traditional citation with authors, year, title and publisher. E.g. "Smith, A., Turner, B., 2003-2005. Institution X, Collection Y, Taxon Y Database". If this is available, with a flag denoting well-formedness, GBIF data portal could forward it and the constructed citations could be dropped.

GBIF data validation services could include a review of this text as part of the process each new provider is helped to connect, and existing providers should be advised on this as part of the regular process of giving them feedback.

6. Discussion

There are not many are examples on other portals how a citations of the primary data can be handled. Most of the other portals just cite to the portal itself. We think such a citation model would not fulfil the requirements of the GBIF Data Sharing and Data Use Agreements. The agreements are quite clear on the need to recognise the efforts of the data providers. Data providers are often just technical bodies who publish the data, but do not own the data the same way that the resource (=database, collection) custodians may. This point may have to be revisited

in the agreements.

The purpose of citations in general is to enable the reader/consumer to retrieve the source of information in question. In the situation of live databases, this poses challenges. It is not expected that GBIF data providers keep the data available indefinitively. Quite the contrary, they can withdraw it any time. Static references as presented above do not therefore always enable retrieval.

Dynamic reference like the <GBIF identifier> can potentially enable access to original material stored under such reference. However, guaranteeing the persistence of such references and the underlying material has to be planned carefully. Issues on sensitive data and data provider authority have to be clarified and agreed on. These are clearly conflicting requirements which may require revising the GBIF Data Sharing Agreement. This paper does not assume that such a service is yet in place, or will be built by GBIF. Such service could perhaps be offered by external archiving services.

In other communities there are examples of direct references to information sources. In particular electronic publishing of scientific articles has touched the issue how to identify electronic content. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), a standard for retrieving metadata from digital document repositories (Lagoze & al. 2004). Adding an XML interface onto GBIF Data Portal that implements an OAI-PMH repository of citations is attractive as it could enable handling datasets the same way as publications, and hence pave the way for getting scientific merit for publishing data.

References

Lagoze, C., Sompel, H. van de, Nelson, M. & Warner, S. 2004. The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0 of 2002-06-14. Document Version 2004/10/12T15:31:00Z
http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

Annex 1. Excerpt from GBIF Data Use Agreement.

3. In order to make attribution of use for owners of the data possible, the identifier of ownership of data must be retained with every data record.

4. Users must publicly acknowledge, in conjunction with the use of the data, the data providers whose biodiversity data they have used. Data providers may require additional attribution of specific collections within their institution.

Version 0.2, Draft, Hannu Saarenmaa 2005-02-10
Version 0.3, Draft, Hannu Saarenmaa 2005-03-08, based on input by Donald Hobern
Version 0.4, Draft, Hannu Saarenmaa 2005-03-29, based on Open Archives Initiative materials
Version 0.5, Draft, Hannu Saarenmaa 2005-03-31, based on comments by Donald Hobern, Jim Edwards, Per de Bjørn, Meredith Lane
Version 0.7, Draft, Hannu Saarenmaa 2005-04-01, based on comments from the staff
Version 0.8, Draft, Hannu Saarenmaa 2005-04-08, based on comments from the staff
Version 0.9, Draft, Hannu Saarenmaa 2005-04-08, grammatical corrections by Meredith Lane
Version 0.11, Draft, Hannu Saarenmaa 2005-04-13, comments by Jim Edwards

Data Use Agreement

Background

The goals and principles of making biodiversity data openly and universally available have been defined in the Memorandum of Understanding on GBIF -MoU; see the relevant excerpts in Annex at:

http://www.gbif.org/DataProviders/Agreements/GBIFdataIPRprinciples.html

The Participants who have signed the MoU have expressed their willingness to make biodiversity data available through their nodes to foster scientific research development internationally and to support the public use of these data.

GBIF data sharing should take place within a framework of due attribution.

Therefore, using data available through the GBIF network requires agreeing with the following:

1. Data Use Agreements

1. The quality and completeness of data cannot be guaranteed. Users employ these data at their own risk.

2. Users shall respect restrictions of access to sensitive data.

3. In order to make attribution of use for owners of the data possible, the identifier of ownership of data must be retained with every data record.

4. Users must publicly acknowledge, in conjunction with the use of the data, the data providers whose biodiversity data they have used. Data providers may require additional attribution of specific collections within their institution.

5. Users must comply with additional terms and conditions of use set by the data provider. Where these exist they will be available through the metadata associated with the data.

2. Definitions

- **GBIF Participant:** Signatory of the GBIF-establishing Memorandum of Understanding (MoU).
- **GBIF Secretariat:** Legal entity empowered by the GBIF Participants to enter into contracts, execute the Work Programme, and maintain the central services for the GBIF network.
- **GBIF network:** The infrastructure consisting of the central services of the GBIF Secretariat, Participant Nodes and data providers. Making data available through GBIF network means registering and advertising the pertinent services via the GBIF central services..
- **Node:** A data provider designated by a GBIF Participant that maintains a stable computer gateway that makes data available through the GBIF network.

- **Participant node:** An organisational unit designated by the GBIF Participant to coordinate activities in its domain. It may also provide data.
- **Biodiversity data:** Primary data on specimens, observations, names, taxonomic concepts, and sites, and other related data on biological diversity.
- **Metadata:** Data describing the attributes and combinations of biodiversity data.
- **Data:** Biodiversity data and metadata.
- **Data provider:** A custodian of data making it technically available. This may or may not be the data owner. If not they will have declared to GBIF that they have permission to make the data available.
- **Data sharing:** The process of and agreements for making data freely and universally available on the Internet.
- **User:** Anyone who uses the Internet to access data through the GBIF network.
- **Owner of data:** The legal entity possessing the right resulting from the act of creating a digital record. The record may be a product derived from another, possibly non-digital product, which may affect the right.
- **Sensitive data:** Any data that the Node does not want to make available, e.g. precise localities of endangered species.

Also see the GBIF Data Sharing Agreement for the data providers at:

http://www.gbif.org/DataProviders/Agreements/DSA