|  |  |
| --- | --- |
| Standard licences for GBIF-mediated data  Proposed action | GBIFwww_2.gif |

**1. Introduction**

GBIF exists to facilitate the publishing, organisation, discovery and reuse of biodiversity data. Most uses of GBIF-mediated data depend on bringing together data records from many different publishers and data sets. GBIF aims to support publishers to make data accessible under well-defined terms of use and with appropriate citation and feedback. Achieving these goals requires a simple, clear and consistent approach to licensing data.

2. Why do we need standard licences?

GBIF currently organises access to more than 400 million data records from over 12,000 data sets. Users need to be able to filter, download and use combinations of these records from many different data sets. They may do so through the user interface of the GBIF data portal, through web services, or through other interfaces offered by GBIF Participants and other collaborators. In all cases, they need clarity as to how the data may be used and as to their obligations in giving credit to data publishers. This is only achievable if every data set, and every individual data record, is associated with a licence that can be interpreted by software in the GBIF portal and in end-user tools to ensure that the data are handled appropriately.

3. Current situation

Currently, the GBIF data portal requires data publishers to agree to the [GBIF Data Sharing Agreement](http://data.gbif.org/tutorial/datasharingagreement) and for users to indicate assent to the [GBIF Data Use Agreement](http://data.gbif.org/tutorial/datauseagreement).

The Data Sharing Agreement includes the following terms:

* ”Biodiversity data accessible via the GBIF network are openly and universally available to all users within the framework of the GBIF Data Use Agreement and with the terms and conditions that the Data Publisher has identified in its metadata.”
* “GBIF Secretariat may cache a copy and serve full or partial data further to other users together with the terms and conditions for use set by the Data Publisher.”
* “GBIF promotes the free dissemination of biodiversity data and, in particular … should respect conditions set by Data Publishers that affiliate their databases to GBIF.”

The Data Publishing Agreement includes the following terms:

* “Users must publicly acknowledge, in conjunction with the use of the data, the Data Publishers whose biodiversity data they have used. Data Publishers may require additional attribution of specific collections within their institution.”
* “Users must comply with additional terms and conditions of use set by the Data Publisher. Where these exist they will be available through the metadata associated with the data.”

In effect, the current situation is that data are openly and universally accessible through the GBIF network, provided that the user follows any terms and conditions specified in the associated (free-format, text) metadata.

4. Terms and conditions in metadata

The metadata for GBIF-mediated data sets includes many different statements of such terms and conditions. For example:

* Data may be used for non-commercial purposes only, with appropriate credit given to the Atlas project partners. For more detailed data requests, inquire at the Ontario Atlas website: <http://www.birdsontario.org/atlas/atlasmain.html>
* No part of this data base may be copied or reproduced without written permission from the legal owner.
* Data are provided for personal research only. Data may not be used for any commercial purposes. Data may not be transferred to another database for distribution to others without prior, written permission from the Canadian Museum of Nature.
* The copyright for any material created by the DSMZ is reserved. The duplication or use of information and data such as texts or images is only permitted with the indication of the source or with prior approval by the DSMZ.
* All rights reserved. Use by terms of GBIF Data Use Agreement. If these data form a qualitatively or quantitatively substantial part of data used in a work, permission for their use must be obtained.

In some cases, the exact requirements expressed in such statements may imply that all use of the data requires separate and specific permission from the publisher. In other cases, significant interpretation or legal knowledge may be required to determine whether a particular use is considered acceptable. In other cases, it may be that reporting obligations can only be met by seeking additional detail not presented in the metadata.

5. Non-commercial usage

A particular issue arises around the most frequent special restriction included by data publishers, the requirement that use should only be for “non-commercial purposes” or “not for any commercial purposes”.

As has been [made clear by Creative Commons](http://wiki.creativecommons.org/Defining_Noncommercial) – even though one of the licence types that they offer is for non-commercial use – there is no agreed definition (and significant international variation in the interpretation) of the term “non-commercial”. For a large proportion of uses of licensed content (and hence of licensed data), it may be impossible to get universal agreement as to whether the use has a “commercial purpose”.

GBIF recognises that many holders of biodiversity data wish to make these freely available for uses that they consider worthwhile but have concerns about others selling these data or selling products based upon these data. In such cases, the assignment of a “non-commercial” restriction seems natural, but this does present problems for users and for automated processing of data.

6. Machine-readable licences

In recent years, the use of standard machine-readable licences has become the norm on the web to support free and open discovery and reuse of resources, particularly text and multimedia materials, but increasingly also databases and data sets. [Creative Commons licences](http://creativecommons.org/licenses/) are familiar and widely adopted, with six different licence types, including [CC-BY](http://creativecommons.org/licenses/by/3.0/) (requiring only attribution for the publisher) and a series of alternatives adding restrictions on 1) commercial use (NC – “Non-Commercial”), 2) modification of the original resource (ND – “No Derivatives”), and 3) licensing for any derived products(SA – “Share-Alike”). Creative Commons also offers the [CC0](http://creativecommons.org/publicdomain/) licence for those who wish to dedicate their resources to the public domain. Similarly the [Open Data Commons](http://opendatacommons.org/) offers [attribution-only](http://opendatacommons.org/licenses/by/) and [public-domain](http://opendatacommons.org/licenses/pddl/) licences for data and databases, as well as a [share-alike](http://opendatacommons.org/licenses/odbl/) version.

All of the available optional restrictions on these cause problems when combining resources, as happens whenever users search the GBIF network for data. The “No Derivatives” option is probably never appropriate for data since it would limit any visualisation or analysis. The “Share-Alike” option limits use to those users who are in a position, and willing, to make their own products, including visualisations, analyses and publications, available under an equally open licence. Some of the issues with the “Non-commercial” option have already been discussed, but it is not permissible to mix resources that include both a “Non-commercial” and a “Share-Alike” clause with other resources that include the “Share-Alike” option without a “Non-commercial” clause (since the “Share-Alike” clause in the second case would not allow the addition of the “Non-Commercial” restriction for derived products, whereas the same clause in the first case would require it).

For these reasons, only some of the available licence options can be considered fully to support a culture of free and open access. See <http://freedomdefined.org/Definition> and, in the context of Canadensys, a major biodiversity data repository managing GBIF-compatible data, <http://www.canadensys.net/2012/why-we-should-publish-our-data-under-cc0> for more information.

7. GBIF and machine-readable licences

Most data publishers within the GBIF network today are in effect making their data available explicitly or implicitly under terms that are consistent with the Creative Commons CC-BY licence that requires users to give appropriate attribution.

Some, as with Canadensys, are more radical and adopt the CC0 public-domain licence, while at the same time recognising that attribution is important and should be standard practice. See the Canadensys document above for some discussion of the benefits of leaving this as a norm rather than a binding term for a licence.

Other publishers, as noted above, share data under terms that seem equivalent to the CC-BY-NC licence which requires attribution and excludes use for commercial purposes, with the corresponding problem of defining this more exactly.

It would seem that existing practice within the GBIF network could be codified by assigning one of three licence types to all data sets, corresponding to either CC0 or CC-BY or CC-BY-NC.

Within the GBIF network, infrastructure, tools and users would be expected to handle CC0 and CC-BY in exactly the same way in almost all contexts. These data would be freely reusable by all users in all contexts. Users would be expected to provide appropriate attribution and the GBIF infrastructure should be developed to make this obligation as easy as possible to meet.

Data licensed under a CC-BY-NC licence would be available only in contexts that were considered non-commercial. In those circumstances, these data could be freely combined with other GBIF-mediated (CC0 and CC-BY) data, provided the relevant licence information was preserved and presented in any subsequent delivery of these data.

8. GBIF and restrictions on commercial use

The model discussed in the previous section would codify, clarify and simplify the existing situation. However this is also a good time to consult with the GBIF community and to consider whether GBIF should adopt a more radical approach to promoting free and open access to data.

The non-commercial restriction makes sense to many data publishers, even if it presents practical difficulties. For this reason, GBIF should explore the possibility and implications of retaining the option for such a restriction.

Two approaches seem possible:

1. GBIF could adopt a policy of only mobilising data which offer the fullest possible support for free-and-open use, i.e. data licenced under conditions equivalent to CC0 or CC-BY, with the following implications:
   1. Data publishers would be required to select one of these licences or to withdraw their data from the network.
2. GBIF could adopt a policy which allows for data to be licensed under conditions equivalent to any of CC0, CC-BY and CC-BY-NC, with the following implications:
   1. A working definition would be required within the context of the GBIF network for how our community will interpret the non-commercial restriction (although it must be noted that the CC licences state explicitly that no conditions other than those defined by the licence itself are binding on the user).
   2. GBIF infrastructure would need to be modified to enable all user searches and downloads (via user interfaces, web services, and other associated web sites) to operate in either a non-commercial (i.e. all data available) or a commercial (i.e. only CC0 and CC-BY data available) mode
   3. Users would need to be informed and to make an appropriate choice of which mode corresponds to their use of GBIF-mediated data.

We are seeking input from GBIF Participants and stakeholders on the relative merits of these approaches.

|  |
| --- |
| **Please send observations and responses, by 1 September, to:**  [**licences@gbif.org**](mailto:licences@gbif.org)  **For any inquiries or clarifications about this briefing, contact:**  XXX  GBIF Secretariat  [eotuama@gbif.org](mailto:eotuama@gbif.org) |