

Standard licences for GBIF-mediated data

Proposed options for action

1. Purpose

This document outlines a range of significant issues and possible solutions to issues relating to licencing of data within the GBIF network. These issues significantly affect users of GBIF-mediated data and we need to find a workable long-term solution which will reinforce GBIF's position as a high-value, global open data infrastructure. We hope to find and implement such a solution as soon as possible, but wish first to explore the issue thoroughly with the GBIF network. This consultation is intended to enable us to understand how much agreement already exists and how much additional exploration may be required.

We are seeking input from all GBIF Participants and stakeholders on the following questions:

1. Do you have any comments on the plan to associate all GBIF-mediated data with a machine readable licence?
2. Do you have an opinion on the relative merits of *Creative Commons*, *Open Data Commons* or other licence types in the context of the GBIF network?
3. Which of the two options described in section 8 of this document should GBIF pursue? If you support "Option 2", would your position be modified if it resulted in a significant decrease in data published to the GBIF network?

These issues affect data publishers, data users and developers of tools using GBIF infrastructure. We therefore ask you to circulate this document and seek responses to these questions from all interested parties. The GBIF Secretariat will circulate this document to existing GBIF data publishers.

We ask that you provide responses, and any comments or suggestions, by **email to licensing@gbif.org by 5 September 2013**.

2. 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.

3. 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.

Unfortunately existing practice within the GBIF network is not sufficiently standardised and clear to support these goals. GBIF requires users to follow any restrictions specified in the metadata for each data set, even though finding, understanding and following these restrictions is difficult and may involve significant effort. Indeed the metadata may in some cases not be in a language that the user understands.

The difficulties become even more significant when users access data from GBIF through tools and web services or through other web sites using GBIF services. Automated tools and processes cannot interpret clauses in textual metadata. This means that many users of GBIF-mediated data today are likely to ignore some of the data publishers' requirements without even being aware of the fact.

As discussed below, the most significant (and frequent) restriction specified in metadata is that usage should somehow be limited to non-commercial contexts. This presents several problems. First, the informal way in which this restriction is specified means that there is no simple way for any user just to find data which could be used in a commercial context. Secondly, it is not clear what range of uses any particular data publisher would consider to be "commercial". This is providing a real problem for some significant users of GBIF data who offer not-for-profit services but receive payments from some or all users to offset the costs of running their own infrastructure. An example of this is the [Local Ecological Footprinting Tool](#) developed and operated by the *Biodiversity Institute Oxford*. This team is keen to act in good faith in regard to GBIF data publishers but has no way to be sure whether all GBIF data publishers would in fact support their not-for-profit model.

GBIF therefore urgently needs to clarify and simplify its processes so all users can have full confidence as to how they may use GBIF-mediated data and what obligations they should fulfil for the publishers. A consistent community approach to commercial/non-commercial use is also essential.

4. Current situation

Currently, the GBIF data portal requires data publishers to agree to the [GBIF Data Sharing Agreement](#) and for users to indicate assent to the [GBIF Data Use Agreement](#).

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 Use 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.

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](#) – 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. It should be as simple as possible for any user acting in good faith to retrieve only data which are licensed for their intended purpose.

For more detailed discussion, see the following:

- Hagedorn et al., 2011, [Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information](#) (ZooKeys 150: 127–149)
- [The Case for Free Use: Reasons Not to Use a Creative Commons -NC License](#)
- [The Definition and Future of Noncommercial](#)

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](#) are familiar and widely adopted, although not specifically designed for sharing data. [Open Data Commons licences](#) represent an attempt to create a family of open licences for this purpose, and many other licence types are in use.

All of these licence types seek to support simplified use of digital content across the Internet, by establishing a consistent basic understanding of how this content can be reused, without requiring direct communication between the publisher and the user.

Such licences offer a range of rights to users and may also impose a range of obligations on users of a database, data or other content.

The main categories of rights which may be offered under these licences are as follows:

- **Share:** The right to copy, distribute and use the content.
- **Create:** To produce works and products from the content.
- **Adapt:** To modify, transform and build upon the content.

Creative Commons combines the second and third of these as a right to “Remix”.

The main categories of obligations which may be imposed under these licences are as follows:

- **Attribute:** Users must give credit to the publishers of the content.
- **Share-Alike:** Users release any works or products or adapted versions based on the content under the same licence as the original content.
- **Non-Commercial:** Users must not use the content for commercial purposes.

Annex A: Creative Commons and Open Data Commons Licences shows how various commonly-used licences from *Creative Commons* (CC0 and all licences with names starting “CC-BY”) and from *Open Data Commons* (others) combine these options and how these relate to publishing of data and to current practice within the GBIF network. As can be seen, only a subset of these licences corresponds to GBIF’s needs. Licences with the No-Derivatives option would prevent most practical uses of data. Licences with the Share-Alike option could not be used within the GBIF network at the same time as other licences owing to conflicts in the obligations they place on users.

7. Licences within the GBIF network

7.1 Codifying existing GBIF practice

Within the GBIF network today:

- **Most** data publishers are in effect making their data available explicitly or implicitly under terms that are consistent with an “**Attribution-only**” licence that requires users to give appropriate attribution.
- **Some**, as with *Canadensys*, are more radical and adopt a “**Public-domain**” licence, while at the same time recognising that attribution is important and should be standard practice. See the [discussion on the Canadensys site](#) 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 a “**Non-commercial**” 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 “Public-domain”, “Attribution-only” or “Non-commercial”.

7.2 Using “Public-domain” and “Attribution-only” licences

Within the GBIF network, infrastructure, tools and users would be expected to handle “Public-domain” and “Attribution-only” licences in exactly the same way in almost all contexts:

- These data would be freely reusable by all users for any purpose.
- 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.
- The advantage of the “Public-domain” option would be that it would leave open the possibility of the obligation for attribution being relaxed in some contexts in which attribution-stacking becomes an issue.

7.3 Using “Non-commercial” licences

If GBIF supports data licensed under a “Non-commercial” licence:

- These should only be made available for download and use in contexts that are considered non-commercial.
- In non-commercial contexts, these data can be freely combined with other GBIF-mediated (“Public-domain” and “Attribution-only”) data, provided the relevant licence information is preserved and presented in any subsequent delivery of these data.
- Users seeking data for a purpose considered commercial should be enabled to access only the data licensed as “Public-domain” or “Attribution-only”.

7.4 Adopting licence standards

The *Creative Commons* and *Open Data Commons* licences listed in relevant rows in the table in *Annex A* would be possible choices for appropriate “Public-domain”, “Attribution-only” and “Non-commercial” licences.

The *Open Data Commons* licences are specifically designed for use with data and databases and so may be the most appropriate for GBIF to use for licensing aggregated data downloads, but individual data publishers could use *Creative Commons*, *Open Data Commons* or any other equivalent licences provided an agreement is in place as to how GBIF will map these to the standard licences.

There is no *Open Data Commons* licence for the “Non-commercial” case. If GBIF plans to support this option for its data publishers, the *Creative Commons* CC-BY-NC licence is the obvious choice.

8. Options for GBIF network

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. Given the variation in form and content in metadata, it is difficult to be completely sure how many existing data publishers seek to impose this restriction. However, a preliminary review suggests that up to one third of GBIF-mediated data may suggest at least some concerns about commercial use.

For this reason, GBIF should explore the possibility and implications of retaining such a restriction. There are two obvious and achievable options for how GBIF might solve its licensing problems at this time:

Option 1 – Support restrictions on commercial use

For this option, the following changes will be required:

1. The *GBIF Data Sharing Agreement* and the *GBIF Data Use Agreement* must be modified to include a clear statement of the range of uses that are considered “commercial” within the context of the GBIF network.
2. All GBIF-mediated data sets must be associated with a machine-readable “Public-domain”, “Attribution-only” or “Non-commercial” licence.
3. The Secretariat would review existing metadata provisionally to assign each current data set to one of these categories and would then communicate with data publishers to confirm the assignment.
4. GBIF must modify its data management and its download interfaces and web services to enable users to work with data in either of two different modes:
 - a. Access to all data records (no commercial-use intended)
 - b. Access only to records approved for commercial use

The advantages of this option are as follows:

- Most, or all, existing data publishers should be able to agree to one of the proposed licences
- The confused existing practice within the network would be clarified and simplified

The risks of this option are as follows:

- Not all data within the GBIF network would be fully free and open for use in all contexts
- Any attempt to define “commercial” and “non-commercial” is likely to be difficult
- GBIF would be committed to supporting a complex licensing model into the future

Option 2 – Only support fully free-and-open data

For this option, the following changes will be required:

1. All GBIF-mediated data sets must be associated with a machine-readable “Public-domain” or “Attribution-only” licence.
2. The Secretariat would review existing metadata provisionally to assign every possible current data set to one of these categories and would then communicate with data publishers to confirm the assignment.
3. For those data publishers with metadata restrictions that are incompatible with “Attribution-only”, the Secretariat would communicate with the publisher to determine whether such a licence is acceptable or the publisher prefers to withdraw the data.

The advantages of this option are as follows:

- All data within the GBIF network would be fully free and open for use in all contexts
- The confused existing practice within the network would be clarified and simplified
- It would be much simpler for others to build tools and solutions using the GBIF data and infrastructure

The risks for this option are as follows:

- A number of existing data publishers, and some potential future data publishers, may be excluded from the network if they or their institutions standardly require a “Non-commercial” clause in licences. In the worst case, this could represent a large subset of existing GBIF-mediated data.

Please send observations and responses, by 5 September, to:

licensing@gbif.org

For any inquiries or clarifications about this briefing, contact:

Donald Hobern

GBIF Secretariat

dhobern@gbif.org

Annex A: Creative Commons and Open Data Commons Licences

Licence type	Rights	Obligations	Notes on usage
"Public-domain" <ul style="list-style-type: none"> • CCO • PDDL 	Share Create Adapt	None	<ul style="list-style-type: none"> • Places no restrictions of any kind on use • Recommended by Canadensys (major biodiversity data repository managing GBIF-compatible data) • Corresponds to practice in parts of GBIF network today
"Attribution-only" <ul style="list-style-type: none"> • CC-BY • ODC-By 	Share Create Adapt	Attribute	<ul style="list-style-type: none"> • Only requires users to give credit to data publishers • Attribution can cause practical problems in cases where large numbers of data publishers have contributed content and/or when combined data are reused in other contexts ("attribution stacking") • Corresponds to practice in most of GBIF network today
"Non-commercial" <ul style="list-style-type: none"> • CC-BY-NC 	Share Create Adapt	Attribute Non-Commercial	<ul style="list-style-type: none"> • Limits use to non-commercial contexts • "Commercial"/"non-commercial" not well-defined (e.g. in context of not-for-profit organisations) • Not considered a "Free Culture License" • Corresponds to expectations in parts of GBIF network today
"Share-alike" <ul style="list-style-type: none"> • CC-BY-SA • ODbL 	Share Create Adapt	Attribute Share-Alike	<ul style="list-style-type: none"> • Requires derived works and adaptations also to be shared under an equivalent "Share-alike" licence (i.e. a licence neither adds additional rights not additional obligations) • Data shared under this licence cannot be combined with data shared under "Non-commercial", "No-derivatives", "Non-commercial Share-alike" or "Non-commercial No-derivatives" licences • Data from the Open Street Map project is shared under ODbL • Not compatible with practice in GBIF today
"No-derivatives" <ul style="list-style-type: none"> • CC-BY-ND 	Share	Attribute	<ul style="list-style-type: none"> • Forbids all modified products based on the source content • Not suitable for sharing data, since in principle it prevents all analysis or visualisation • Not considered a "Free Culture License" • Not compatible with practice in GBIF today
"Non-commercial Share-alike" <ul style="list-style-type: none"> • CC-BY-NC-SA 	Share Create Adapt	Attribute Non-commercial Share-Alike	<ul style="list-style-type: none"> • Same as "Share-alike" but with additional restriction on commercial use • Not considered a "Free Culture License" • Data shared under this licence cannot be combined with data shared under "Share-Alike", "No-derivatives" or "Non-commercial No-derivatives" licences • Not compatible with practice in GBIF today
"Non-commercial No-derivatives" <ul style="list-style-type: none"> • CC-BY-NC-ND 	Share	Attribute Non-commercial	<ul style="list-style-type: none"> • Same as "No-derivatives" but with additional restriction on commercial use • Not considered a "Free Culture License" • Not compatible with practice in GBIF today