## Options in response to GBIF Consultations

# Licensing of data AND endorsement and evaluation of datasets

## Background

The GBIF Secretariat organised two consultations, [*Endorsement and evaluation of data sets within GBIF*](http://imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=073594d5f7a4b7071d079da80f9101) and [*Licensing of data within GBIF*](http://imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=906e4c47b4e0052e703624a2ea1981), running from April to June 2014. The responses received from GBIF Participants and other stakeholders have been summarised in two reports, [*Summary of responses: Licensing of data within GBIF*](http://imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=368b7c3e812c52a405418fc2b0f500) and [*Summary of responses: Endorsement and evaluation of dataset within GBIF*](http://imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=004bfeff64eed0b6cc5e5ca55c9f97). Readers are advised to consult these reports to understand the range of views expressed.

During each consultation, many responses indicated significant support for proposed modifications to existing GBIF processes, but others identified major issues regarding the likely effect of these changes within their own or other countries.

This document follows these consultations and outlines options for ways that GBIF may address both the needs and the concerns identified.

The GBIF Governing Board is requested to consider the three sets of proposals outlined in this document**.** The main areas of broad agreement are presented as the **Core Proposal**.The remaining proposals, **Option 1** and **Option 2**,are offered as possible extensions to the **Core Proposal.**

**At the GBIF Governing Board meeting, GB21, in September 2014, Voting Participants will be asked to decide on the adoption of these proposals.**

## Licensing of data

The following key points arose from the consultation:

1. There is uniform recognition that GBIF must ensure that access rights for users to download and use biodiversity data must be clear and unambiguous and must support large-scale data integration and filtering. **GBIF must ensure that all data are associated with appropriate machine-readable licences.**
2. Both data publishers and data users require simple, scalable processes and tools to give clear attribution and reference to data used, in the form of citations. Even when data publishers do not require attribution, good scientific practice requires results and publications to remain linked to the underlying data. **GBIF must make rapid progress in delivering and promoting comprehensive and workable solutions to support data attribution**.
3. In many jurisdictions, intellectual property rights do not apply to factual data of the kind mobilised through the GBIF network (but other countries, such as South Africa, do legislate such rights). Requirements that users must give attribution when using data, or that users should observe other restrictions, are unlikely to be legally enforceable in most countries. **GBIF can promote a culture of citing data used and giving attribution to data publishers, but cannot guarantee that this will be uniform practice everywhere.**
4. Several countries and significant publishers of data within GBIF have already adopted unrestricted public domain tools (rather than licences), in particular [CC0](http://creativecommons.org/about/cc0), as the basis for data publishing within their networks, arguing that other options are not legally valid and/or that this option best supports open science. **GBIF should support data publishers adopting public domain tools wherever these are appropriate.**
5. A number of institutions publishing data within GBIF currently have policies that do not allow data to be shared without asserting intellectual property rights or defining restrictions on use. Several significant networks publishing data within GBIF have existing data sharing agreements with their publishers and cannot vary the terms associated with these agreements. **Under current circumstances, GBIF cannot move completely to adoption of public domain tools (e.g. CC0) without risk of serious disruption and loss of data.**
6. Several responses also indicated that, at least at present, some of their data publishers expect to be able to restrict data use to non-commercial activities. However, there is no clear and accepted definition of which activities are to be considered non-commercial. As with other restrictions on data use, it is unlikely that such restrictions could be enforced in a consistent way throughout the network. Creative Commons indicates that non-commercial restrictions do not contribute to [“Free Culture”](http://creativecommons.org/freeworks). **Under current circumstances, GBIF may face some disruption and loss of data if it excludes the option for data publishers to request exclusion of commercial uses.**
7. Several responses identified the importance of recent developments by Creative Commons in issuing version 4.0 of its licences, since these explicitly address data publishing (and also provide more scalable recommendations for attibution). **GBIF should consider CC 4.0 options as part of any final implementation of digital licences.**

## Endorsement and evaluation of data sets

The following key points arose from the consultation:

1. Many responses emphasised the importance of GBIF developing processes to allow institutions outside current GBIF Participant countries to publish their data. **GBIF should remove blocks which prevent data publishing from suitable institutions.**
2. A large number of responses however argued that the current process with GBIF Nodes endorsing data publishers within their country (or linked to their organisation) is critical for the visibility and health of national GBIF activity and should not be compromised by allowing data publishers to ‘bypass’ their Node. **GBIF should maintain the key role of national GBIF Nodes in developing and coordinating their data publisher networks.**
3. Within GBIF today, some data publishers are already added to the network following endorsement by a GBIF Node in another country or by a GBIF organisational Node. To date, this has not resulted in significant problems, but the process has sometimes been arbitrary and confusing. **GBIF must address existing practice as part of future processes.**
4. Several responses expressed concern that dataset evaluations might be confusing or contentious, particularly 1) if the dataset does not receive endorsement from the appropriate national Node, or 2) if endorsements are based on subjective or qualitative judgments. **GBIF must ensure that any process for endorsing datasets is clear, transparent and objective.**
5. Several responses expressed concern that the proposals could lead to significant quantities of unevaluated (and potentially low-quality) data within the network, possibly lowering user confidence in GBIF as a data resource. **GBIF must ensure that any changes to endorsement processes do not lower data quality for users.**
6. Several responses expressed concern that a shift to endorsement of individual datasets, particularly without a fixed role for national GBIF Nodes, may lead to significant additional work for the GBIF Secretariat. **GBIF must ensure that any changes to endorsement processes can be accommodated within existing resources and that expectations are realistic regarding processing speed for new datasets.**
7. Almost all responses expressed support for improved fitness-for-use indicators at the dataset level. Many responses identified the need for a clearer definition of indicators to be developed and of the means by which these would be presented to users. **GBIF needs a clear, detailed proposal for the processes and mechanisms for developing fitness-for-use indicators.**

## Proposals to Governing Board

GBIF’s mission is to make biodiversity data freely and openly available and accordingly should champion approaches that remove limitations on data accessibility. However, GBIF is a broad global network of collaborators and needs to ensure that this global network is strengthened and that its processes are appropriate in all regions and countries.

Given these reasons, the following proposals are presented to the Governing Board.

The **Core Proposal** is a set of recommendations, seeking to accommodate the elements which have received broad support. The GBIF Governing Board will be asked at its 21st meeting in September 2014 (GB21) to take a decision on adopting these recommendations.

If the Core Proposal is adopted, the Governing Board will be asked to take separate decisions on the remaining proposals (**Options 1 and 2**). These are optional additions to the Core Proposal and represent elements for which the support is more mixed or where support is not completely clear.

#### Core Proposal – Recommendations on endorsement, licensing and dataset evaluation

The GBIF Governing Board is asked to adopt the following recommendations, incorporating the key elements for which consensus seems high.

ENDORSEMENT

* GBIF will retain the **current role of** **national GBIF Nodes in endorsing new data publishers** located within their country before these are added to the GBIF network.
* Potential **data publishers located in countries that are not currently GBIF Participants** will be able to seek endorsement from any existing GBIF Participant. See also Option 1 below.
* The GBIF Secretariat will develop **decision criteria to assist GBIF Participants** in reviewing data publishers for endorsement.

LICENSING

* All new GBIF data publishers will be asked to select a **digital licence to be associated with each dataset**. The GBIF Secretariat will review recent developments in digital licences (particularly CC 4.0) and will support one or more digital licence options that are compatible with each of the following choices:
	+ Fully-open public access (CC0 or equivalent)
	+ Attribution required (CC-BY or equivalent)
	+ Attribution required, non-commercial use only (CC-BY-NC or equivalent) – but see Option 2 below
* The GBIF Secretariat will work with existing data publishers to **map their published data sets to these same licence categories over the next six months**. At the end of this time, any datasets for which the licensing is unclear will be associated with the most conservative licence option (CC-BY-NC or equivalent) and the data publisher will be advised to review and respond to this choice. Data publishers will be given the option to change the selected licence at any time.
* Any **datasets which cannot be assigned to any of the supported licence categories should be removed** from the GBIF network. GBIF should continue to review these licence choices to ensure that GBIF reflects best practice in open data delivery. Any significant recommendations to alter support for particular licence categories will be brought to the Governing Board for approval.
* The GBIF Secretariat will continue to work to **assign stable identifiers to all datasets** and to provide **comprehensive tools for users to cite datasets** and to give attribution to data publishers for data used. These tools will be released during 2015.
* The GBIF Secretariat will provide clear **filters for users to include or exclude data with restrictions on commercial use** from data views and downloads [unless Option 2 is supported by the Governing Board].
* The GBIF Secretariat will prepare and highlight **information materials to clarify expectations for both data publishers and data users** regarding use and citation of data accessed through the GBIF network and to promote standard citation of data as the recommended norm for all data users, regardless of whether data sets are licensed under fully-open public access (CC0 or equivalent) or attribution-required (CC-BY or equivalent) licences.
* The GBIF Secretariat, in consultation with Participants and the wider community, will additionally develop information materials for data publishers and data users to establish expected **norms within GBIF for distinguishing between commercial and non-commercial use** [Unless Option 2 is supported by the Governing Board] This will not be a legally-binding document, but will exist to ensure that all parties can share a common understanding.

DATASET EVALUATION

* The GBIF Secretariat will develop a **concrete set of recommendations for developing multiple fitness-for-use indicators** for each dataset and for presenting these to users.
* As part of these recommendations, the GBIF Secretariat will consider **options to include direct review or endorsement of datasets** as a component of these fitness-for-use indicators.
* Participants, data publishers and user groups will be **consulted further on these recommendations** prior to any decision to implement changes.

#### Option 1 – Enable data publishing from institutions not located in GBIF Participant countries

Some institutions outside GBIF Participant countries already secure endorsement for their data from another GBIF national Participant or from a relevant Associate Participant organisation. In some cases, there is no natural choice for an endorsing Participant. Recognising the concern of many Participants to maintain the publisher endorsement model, but also the importance of finding ways to address major gaps in GBIF’s data coverage, the Governing Board is asked to consider the following additional option.

ENDORSEMENT

* The GBIF **Nodes Steering Group will develop and implement a process for review and decision regarding endorsement** of potential data publisher institutions which do not have links to an appropriate GBIF Participant. This process may be organised at either global or regional level. Such endorsement will follow the same decision criteria developed for GBIF Participant Nodes (see core proposal on endorsement, final bullet point above).
* Endorsed publishers in countries not currently participating in GBIF will be encouraged to **build support for national GBIF participation**, and given the tools to promote participation among institutions and national authorities.
* If a new country becomes a GBIF Participant, and if institutions within the country have already been endorsed as GBIF data publishers, existing data sets will remain within the GBIF network, but the **new national Node will be asked to endorse the institution** as a data publisher before new data sets are registered.

#### Option 2 – Discontinue explicit support for restrictions other than attribution

Many responses indicated concern that restrictions on data use beyond the requirement for attribution, in particular the non-commercial use restriction, are not compatible with large-scale integration and open access to data. Creative Commons, for example, identifies only a subset of licences as “approved for free cultural works”. It is also difficult to reach agreement on whether any particular use can be considered commercial or non-commercial.

Given existing data use agreements in place with GBIF Participants and data publishing networks, there are good pragmatic reasons for GBIF to consider supporting non-commercial restrictions at least over the next few years. However, this is an area on which opinions are divided.

The Governing Board is therefore asked to consider the following optional modification to the Core Proposal.

LICENSING

* GBIF will support **only two digital licence options** that are compatible with each of the following choices:
	+ Fully-open public access (CC0 or equivalent)
	+ Attribution required (CC-BY or equivalent)
* Any data sets which include **restrictions other than the requirement for attribution should be removed** from the GBIF network.