

GBIF Consultation Licensing of data within GBIF

Brief description:

Request for feedback on a proposal to require data publishers to waive rights when publishing data through the GBIF network (equivalent to Creative Commons CCO), and to establish a set of community-developed norms for publishing and using biodiversity data, including guidelines for citation and respecting restrictions on commercial use.

Target audience:

GBIF Heads of Delegation, GBIF nodes, data publishers, data users

Deadline for responses:

14 June 2014

Process:

Please complete the reply form (available here) and send responses to licensing@gbif.org by 14 June 2014.

Unless otherwise indicated, all responses will be made available through the GBIF portal.

The GBIF Secretariat will review responses and develop a report and recommendations for next steps within four weeks of the response deadline.

1. Background

The mission of GBIF is to promote free and open access to biodiversity data. It has been very successful in this role. However, there is broad recognition globally that e-research requires the foundations of 'big data' to be as free from restrictions as possible. Limitations introduced at the stage of data publication multiply as datasets are aggregated to address big questions. The *Global Biodiversity Informatics Outlook*¹ identifies the "Open access and reuse culture" as one of the essential foundations for biodiversity informatics. Global initiatives such as the Research Data Alliance² are addressing these same issues for all research data. There is an increasing drive in many countries to make data, particularly publicly-funded data, open and accessible. GBIF is in a position, and has a responsibility, to contribute to these trends since they are essential to its own mission.

In August 2013, the GBIF Secretariat conducted a consultation surrounding the licensing of data, proposing a more standardized approach to ensure that data users are informed of the publishers' decision. This call received 32 responses, some of which were very extensive and collaborative efforts. The Secretariat is very grateful to all those who replied.

¹ http://www.gbif.org/resources/2251

https://rd-alliance.org/

Within the responses, the following key messages were identified:

- 1. The proposed copyright licensing model (e.g. CC licenses) may not be a suitable mechanism to control or restrict use, as the type of data published through GBIF falls in a legal 'grey area'. These data may not in fact be eligible for copyright at all, at least in some jurisdictions. Deciding this point may also be complicated by the types of content offered, particularly when video, image or other multimedia content is included. GBIF were advised not to contribute to misunderstanding by promoting licences which may not be legally actionable³.
- 2. There is a strong desire by some (about one third of the respondents) to attempt to restrict access for commercial use. However, as some discussed in their responses, it is very difficult to agree or define what constitutes commercial use.
- 3. Many of the concerns expressed relate to the need for better attribution. It is suggested that, rather than asserting a copyright license, GBIF should focus with high urgency on improving citation and tracking data use.

2. Current situation

Users of GBIF data agree to abide by the <u>Data Use Agreement</u>, and data publishers separately agree to the terms of a <u>Data Sharing Agreement</u>. Both of these agreements refer to additional 'terms and conditions' to which data publishers may attach to datasets and include in their metadata.

As a result, many GBIF data publishers include their own licensing details in the metadata they provide for each dataset. These are submitted in free text and are expressed in a wide variety of non-standard formats. This makes it difficult or impossible for users of the GBIF portal to understand the rights relating to large data downloads including data from multiple institutions. Use of web services to build automated tools is particularly problematic where usage restrictions are proposed in purely human-readable text.

3. Proposed changes

Based on the initial consultation, GBIF Secretariat solicits feedback on the following proposed next actions:

- 1. The Secretariat, in collaboration with the community, compiles simple documentation explaining the ambiguous situation surrounding licensing, and the benefit of waiving of rights when publishing data (e.g. through a CCO license), which is the preferred approach of many respondents. Arguments for this approach are given at http://www.vertnet.org/resources/datalicensingguide.html and http://www.canadensys.net/2012/why-we-should-publish-our-data-under-cc0.
- 2. The Secretariat works with the community to document "publishing and usage norms", a community-defined code of conduct that defines what is considered correct behaviour when using data published through GBIF. The data publishing

³See for example http://figshare.com/articles/Response to GBIF request for consultation on data licenses/799766

network Canadensys has initiated such a definition at http://www.canadensys.net/about/norms. This can serve as a good basis for discussion. The norms would include elements defining what constitutes commercial use.

- 3. The GBIF publishing process is modified to allow data publishers to flag datasets that they wish to restrict from commercial use (as defined in the norms). While accepting that this is not enforced through a licence, the norms can document that this is the express wish of the publisher, and should be respected by users. The alternative proposed in the original consultation, to use the CC-BY-NC licence, was discouraged by some respondents, as this would perpetuates a misguided belief that such a licence has a strong legal basis. The use of such licences may also have a number of undesirable outcomes, as discussed at http://dx.doi.org/10.3897/zookeys.150.2189.
- 4. GBIF aims to be in a position that within six months following approval of this model, all data accessible through the portal and web services be available with a CCO waiver, with an additional flag and filter allowing users to limit results to content suitable for commercial use. All data access services would link to the norms document with clear signposting.
- 5. GBIF proceeds with the envisaged work plan to establish a citation model built upon the Digital Object Identifier (DOI) system, whereby each dataset has a DOI, and all citations reference the datasets used through DOIs.