#### Open data in practice A quick guide to (re)licensing data from citizen science projects

Kyle Copas GBIF Secretariat ECSA Open Science Working Group ECSA Conference 2018 14 May 2018—Geneva



Global Biodiversity Information Facility

**BY THE NUMBERS** 1 May 2018

Species occurrence records

984,256,744

Country Participants Organizational Participants

56



Average records downloaded per month (2018)



Datasets 39,113

**Publishers** 

1,195

Average monthly user sessions (2018)



# What are the **three most feared words** for a citizen science project manager?

## 2. Call the lawyers

## Intellectual Property Rights Call the lawyers

## **Intellectual Property Rights**

## Intellectual Property Rights

## Intellectual Property Rights

#### Standardized data licensing Benefits to Citizen Science and Open Science

- Clarifies terms of use—what 'open' means in practice—for both data users and data publishers
- Enables wider reuse of best-available evidence
- Helps ensure reproducibility of scientific analyses
- By supports systems for tracking data use, demonstrates scholarly value & ROI for funders

#### Standardized data licensing Contribute to a greater good.

'Contrary to what many people assume, data sets from volunteers are among the most restrictive in how they can be used.'

Groom Q, Weatherdon L & Geijzendorffer IR (2016) Is citizen science an open science in the case of biodiversity observations? Journal of Applied Ecology 

Parks authority Citizen science Data publisher

Provider type

doi:10.1111/1365-2664.12767



## Standardized data licensing **Be patient.**

Three-year consultation led to full implementation of machine-readable licensing for 10K+ occurrence datasets

Since 2016, each dataset on GBIF.org now carries one of three Creative Commons designations: CC0, CC BY & CC BY-NC



#### GBIF www.gbif.org

#### 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 adreement already exists and how much additional exploration may be required.

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

- Do you have any comments on the plan to associate all GBIF-mediated data with a machine readable licence?
- 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 <u>licensing@gbif.org</u> 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

#### Standardized data licensing If you don't have permission, ask.



#### Standardized data licensing If you don't have permission, <u>ask</u> (UK edition).



## Standardized data licensing **Be open <u>and</u> FAIR.**

Wilkinson MD, Dumontier M, Aalbersberg IJ et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018. http://dx.doi.org/10.1038/sd ata.2016.18



Graphic by Association of European Research Libraries https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf

#### Be open <u>and</u> FAIR. Help human and computer users.



Happy Computer by roberto arenas from the Noun Project

#### Standardized data licensing Licences are restrictions, not protections.

