Vienna, Austria. 18th December 2019

Österreich forscht
www.citizen-science.at

Working group on
Open Biodiversity Databases in Citizen Science Projects

Catalogue of Questions for Project Managers

Version 1.0

Authors: Florian Heigl[1], Daniel Dörler[1], Theresa Walter[2] and Linde Morawetz[3]

Working Group Members: Philipp Hummer[4], Christine Resch[5], Stefan Resch[5], Ferdinand Schmeller[6], Silke Schweiger[7], Alfred Stanzl[6], Norbert Teufelbauer[8], Richard Zink[2]

Institutions:
1 University of Natural Resources and Life Sciences Vienna
2 University of Veterinary Medicine Vienna
3 AGES (Austrian Agency for Health and Food Safety)
4 Spotteron GmbH
5 apodemus OG
6 Wiener Umweltschutzzabteilung (City of Vienna)
7 Natural History Museum Vienna
8 BirdLife Austria

Preamble

As part of the Citizen Science Network Austria (https://www.citizen-science.at/netzwerk), the working group *Open Biodiversity Databases in Citizen Science Projects was* established in February 2018. The objectives of this working group are (I) to formulate a catalogue of questions to help deciding about open publishing of research data collected in a citizen science biodiversity project, (II) to accompany and document the process of open publishing of research data from a concrete project and (III) to write and publish a so-called data paper in addition to publishing research results. This document is the product of point (I) of the objectives, the questionnaire.

The following points have led us to deal with this topic:

1. There is an ethical dilemma in storing data in a closed manner that has been collected jointly with citizen scientists.
2. In the future, certain funding programmes will require the data collected in the course of a project to be made publicly available.
3. The EU general data protection regulation requires us to reconsider and update our handling of personal data.
4. A consideration of the topic by the project coordinators shows the progressive thinking in the field of citizen science.

5. When the citizen science community jointly identifies the challenges/problems, it is easier for them to find arguments for/against opening the databases in citizen science projects.
6. Technical developments in the field of infrastructure offer new possibilities for the publication of research data (e.g. www.gbif.org/)

Just to underline it: the working group should not be a one-sided lobbying for opening up biodiversity databases. We want to objectively point out which problems/challenges arise, if the databases are opened and which ways there could be to protect one's own interests or sensitive data on objects of protection and still make data openly available.

In the following, we would like to clarify terms that we use in the document in advance:
By research data we mean all data/metadata related to the research object, i.e. it excludes personal or sensitive data of citizen scientists. Among others, the Working Group on Legal Aspects deals with personal data in citizen science projects. The data to which we refer in the document include, for example, the location of discovery, time of discovery, species, but also any photos of the research object that may be available, which are transmitted or uploaded by the participants, but not the e-mail address, postal address or name of the participating people.
Open publishing of research data means that after the publication these data are accessible and usable without restriction. This includes other researchers as well as citizens, politicians and companies. We would like to orient ourselves here in particular on the FAIR principles of FORCE11, i.e. the data are **f**indable, **a**ccessible, **i**nteroperable and **r**e-usable. A detailed description of these principles can be found on the linked websites of Force11.
The presented document is intended to point out the most important questions that project leaders must ask themselves if they want to publish the research data of their project in an open format. The questions provide orientation as to which aspects need to be considered during the publication process. All questions are supplemented and explained by short explanations. However, we make no claim to completeness.

In future, if required, the catalogue of questions will be adapted in order to meet new challenges and developments.

For the working group,

Florian Heigl, Daniel Dörler, Theresa Walter and Linde Morawetz

1. Is the publication of the data feasible?

Before publishing the research data, one should be aware that such a step also requires resources which must be available in a project for successful implementation. The following questions should therefore be asked:

Is there enough *time* within the project to publish the data openly?

Examples of necessary time resources:

- Obtaining the consent of the citizen scientists, project partners, funders (if not already done at the start of the project).
- Editing the database to make the data anonymous.
- Research on which data portal (repository) and in which format the data can be published.

Is a publication of the data *technically* feasible for the project?

Examples of necessary technical resources:

The complexity of technical resources varies depending on the type of data and publication. Exemplary scenarios would be:

- Not much effort: an Excel file, which is freely accessible on your own project website, requires hardly any additional technical resources, but continuous updating if the end of the project has not yet been reached.
- Very time-consuming: a download button on the project website, with the help of which the current version of the database can be called up from outside the project team at any time, means a great deal of extra work and probably also the involvement of an IT specialist.

Is a publication of the data *financially* feasible for the project?

Examples of necessary financial resources:

- Additional manpower for additional time expenditure.
- Purchased services for the implementation of complex technical solutions.
- Purchased storage capacity if a not freely available repository is used.

2. Are my data relevant/problematic in regard to research ethics?

Before publishing the data, it should be checked whether the publication of these data may have negative consequences for the species and habitats described with the data. Negative consequences can include among others:

- An increase in the number of visitors in the areas described may lead to the following effects:
    - Change and/or pollution of the habitat (e.g. compaction of the soil, food or garbage left behind, targeted feeding attempts).
    - Stress for the animals due to human activity, especially in the vicinity of places of retreat or rearing of young.
- (Illegal) Hunting or poison baiting for certain species (trophy hunting, resentment against individual species...).
- Targeted destruction of (not yet protected) habitats in order to counteract a future protection status.

A possible ethical problem with the data can be mitigated by making the ethically relevant data “unspecific”. For example, location information can be given in a coarser resolution to prevent accurate localization. Similarly, accurate time data can be 'mitigated' by specifying wider periods or by making the data public months or years after discovery.
Another possibility that some open data platforms offer is a gradation in the accuracy of the information depending on the logged-in user (e.g. Atlas of Living Australia). Here, the platform can distinguish between civil servants, scientists and private individuals with different access authorisations, for example, who thus also have access to data with different resolutions.

Detailed metadata is necessary to prevent wrong results caused by an unintentional misinterpretation of the data. These should include among others information on the type of data collection, completeness and quality controls. This should clarify which results are possible with the available data.

Another possibility would be to determine whether the ethically relevant data can be separated from other data, in order to publish only the non-ethically relevant data.

3. Are there legal aspects to be considered before the data are published?

Copyright is a complex subject and cannot be discussed in as much detail as it is often necessary in this context. It is also partly dealt with in the WG Legal Aspects in Citizen Science. Therefore, please consider the following paragraphs as food for thought and contact a specialized lawyer in case of concrete problems or ambiguities.

An important obstacle for the publication of open data is the copyright of the reporting citizen scientists. In most cases, this will concern two different types of copyright works: uploaded photographs and descriptive text. Depending on the individual case, it must be decided whether the data is still meaningful without this copyright-protected information. If this is the case, we recommend removing the data under copyright before publication. It is also possible to avoid conflicts with copyright law regarding descriptive texts through standardized entries (drop-down menus, etc.) when setting up the database.

Should it be necessary to publish the data under copyright, this publication must be approved by all involved citizen scientists.

Photographs are important as evidence for species identification in biodiversity databases. Therefore, in most cases it is important for the operators of the database to secure access to photos for an unlimited time and to make them openly available. At the same time, photographs are regarded as " Lichtbildwerke" in Austria (and thus as "eigentümliche geistige Schöpfung" and are subject to copyright protection (§ 3 UrhG; https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=BgblAlt&Dokumentnummer=bgl1936_0111_00131). The copyright is not transferable under Austrian law (§ 23 UrhG), it cannot be transferred from the photographer (the producer of the work) to the database operator.

You can help yourself by publishing the photo under a cc-license (Creative Commons), so that the producer/licensor does not lose the right to possess the photo, but it can still be used in the database. A possible solution would be to use a CC-0 licence (https://creativecommons.org/publicdomain/zero/1.0/deed.de), i.e. the licence with the least rights to the producer. All other Creative Commons licenses require the naming of the producer/licensor (cc-BY). This again conflicts with the EU-GDPR, which requires the right to delete personal data (e.g. the name) in a database. When deleting, the photo marked with cc-BY with the name of the person would also have to be deleted.

If the data contains personal information and the participants have not agreed to a publication, the data must be separated and removed from the research data prior to open-access publication.

4. Was consent obtained for publication of the data within the project?

Before publishing the data, it should be checked whether something speaks against the publication from the project's internal point of view. These include the following points:

- Depending on the funding body/cooperation partner, it is possible that publication of the data is not permitted, not desired or explicitly requested.
- Before publishing, check whether the participants in your project have agreed to the publication of data (e.g. within the framework of the general terms and conditions of your project). If this is not the case, obtain consent before publication (see also questions 3 & 5).
- Think about effects the publication of data could have on your participants, e.g. could participants be tracked with the data?

5. Is publication of the data compatible with the project aim?

Before publishing data, it should be checked whether this is also compatible with the project aim. These include the following points:

- Does publication of the data contradict the project objective? If yes, is publication possible after the project objective has been achieved?
- If the data are to be published in the context of a scientific publication, this may initially argue against the publication of the data. In this case, consider whether open publication of data is an option at a later date (e.g. after successful publication of results).