

NBN TRUST - OPEN DATA POLICY

PURPOSE OF THIS DOCUMENT

The purpose of this document is to describe the NBN Trust's policy on Open Data in the context of the Trust's wider approach to data management and data aggregation on the NBN Atlas.

BACKGROUND

Many individuals and organisations collect vast amounts of data under the auspices of biological recording. This data is invaluable for multiple uses including conservation management planning and monitoring, informing planning and policy decisions and reporting on the changing state of nature and against national and international legislation. However, much of this data has historically not been easily accessible for a variety of reasons.

The concept of open data, whereby data is freely available to everyone to use for all purposes, has been gaining traction across all sectors in the UK. In the NBN Strategy 2015 – 2020, under strategic aim two – “Make biological data and information available to everyone” the NBN committed to “increasing the visibility and openness of UK biological data in perpetuity”. As the NBN Trust is a data aggregator, as opposed to a data collector or provider, we do not make the decisions on data licencing for the majority of the data available through our systems. However, we can encourage and facilitate data providers to provide more open data. This policy sets out the NBN Trust's open data principles which include how the NBN will facilitate more openness within the biological recording community.

DEFINITIONS

In order to have an effective Open Data policy, it is essential that all parties who provide and use the data have a common understanding of the definitions of Open, Shared and Closed data. The definitions below are from the Open Data Institute (ODI) and are widely accepted within the UK.

Open data

Open data is data that anyone can access, use or share for any purpose

Data with an OGL, CC0 or CC-BY licence are open data. By definition, it must be available to all without restrictions on what they do with it.

Shared data

The ODI states that there are three types of shared data:

1. Named access – data that is only shared with named people or organisations
2. Attribute-based access – data that are available to specific groups who meet certain criteria
3. Public access – data that is available to anyone under terms and conditions that are not open.

Data published under a CC-BY-NC licence is shared data because there are restrictions on the purpose that the data can be used for, i.e. it can only be used for non-commercial purposes.

Closed data

Closed data is data that can only be accessed by its subject, owner or holder.

It is important to note that it is only the licence under which data is available that determines whether data is open or not. The resolution of the data does not have any bearing on whether

Jo Judge

Version 1

13 October 2017

data is open, closed or shared, i.e. data available at 100m but with a CC-BY-NC licence is shared, whereas, data at 10km resolution with a CC0 licence is open.

PRINCIPLES FOR FACILITATING OPEN DATA

Principle 1: Open data by default

Any data generated by the NBN Trust, including development of the NBN Atlas infrastructure will be open by default.

We recognise that there are situations where data from our partners cannot be open e.g.;

- Some species data cannot be open at high resolution as doing so may lead to the harming or persecution of those species.
- When there is national and international legislation, in particular pertaining to intellectual property, personally-identifiable and sensitive information, which must be observed.
- some data cannot be released immediately to due commercial or academic sensitivities.

We will establish an expectation that all data on the NBN Atlas be published openly by default, while recognising that there are legitimate reasons why some data cannot be released.

Principle 2: Provision of a biological information management infrastructure

In order for data to be open, it needs to be accessible to all. Biological data is collected by many individuals and organisations across the UK. Many of these do not have the resources to make their data easily accessible to others. Furthermore, data users need to be able to easily access data from many sources in a cost-effective way. The NBN will facilitate this by **building and maintaining a fit-for purpose, world leading, biological information management infrastructure.**

Principle 3: Standardisation of formats and terminology

Open data is only useful if it is available in formats that people can understand. As a consequence of the UK's long history of biological recording and the diversity of the species that are recorded, there are multiple systems and formats for submitting records. This complexity can be difficult to navigate for data users and may result in data not being used appropriately or at all. The NBN will facilitate making data more open and usable by **promoting standardised, structured, data capture processes and consistent terminology across the network.**

Principle 4: Data Quality and traceability

Good open data should be of known quality and provenance. The NBN will contribute to the quality and traceability of data available through our systems by **implementing verification status by record** with standardised terminology, and **promoting the adoption of unique ids** for all records.

Schedule of review

This policy will be reviewed on an annual basis.