

Developing and Maintaining a National Biodiversity Data Infrastructure - An example from Norway

Knut Anders Hovstad[‡], Stein Hoem[‡], Eveliina Päivikki Kallioniemi[‡], Arild Lindgaard[‡], Stine S Markussen[‡], Toril Loennechen Moen[‡], Tove Rimestad[‡]

[‡] Norwegian Biodiversity Information Centre (NBIC), Trondheim, Norway

Corresponding author: Knut Anders Hovstad (knut.hovstad@artsdatabanken.no)

Abstract

Biodiversity data infrastructures are fundamental to halting the ongoing loss of species and habitats. Here we provide an overview of the national biodiversity data infrastructure developed and implemented in Norway by The [Norwegian Biodiversity Information Centre \(NBIC\)](#). Key elements and properties of this infrastructure are highlighted and directions for future development are outlined.

The overarching objective for the infrastructure is to make data on habitats and species in Norway available for policy and decision makers, researchers and the general public. Here we will focus on data on species and their distribution. The infrastructure is built as a modular system but with a high level of integration between the components. NBIC has the main responsibility for developing and managing the infrastructure in collaboration with natural history museums, research organizations, private companies and non-governmental organizations (NGOs), including the Norwegian node of the Global Biodiversity Information Facility ([GBIF](#)). The infrastructure includes a citizen science portal that gives both amateurs and professionals the possibility to report species sightings. The data from this portal together with data from natural history museums and other data providers are then made available through the [Species Map Service website](#). The infrastructure also includes the Norwegian taxonomic backbone database and a trait bank that is under development. The trait bank is planned to contain both ecological traits but also other information about species such as [Red List](#) status.

The infrastructure builds on a few simple principles. The exchange of data is based on the [Darwin Core Standard](#) and its extensions, or other open data standards. Each data owner is responsible for quality control, secure storage and management of their own data. The data owners publish the data using the [Integrated Publishing Toolkit](#) (IPT) developed by GBIF. NBIC then harvests all observations with spatial coordinates that fall

within Norway and adds information about, for example, status on the Norwegian Red List before the data are made available in the map.

The core parts of the infrastructure only handle data that are open and adhere to the [FAIR principles](#), i.e. that the data are Findable, Accessible, Interoperable, and available for Reuse. An important exception to this principle is observations of threatened species that are particularly vulnerable to human disturbance. Sensitive data are managed in a separate, secure system and with a restricted access portal for data viewing.

Data in the Species Map Service are also available through a public API, which can be used to harvest and import the data or subsets of it into other systems and services. To give an example, the forestry industry integrates data on Red Listed species into their planning systems tools and has even an option to display a subset of species occurrences on their forestry machine computers.

NBIC has in cooperation with NGOs established a system in which experts validate observations that are reported through the citizen science portal. As it is not possible to validate all observations, species of particular interest for environmental management and conservation are given priority. A future solution could be to develop a system that can identify observations with a low likelihood of being correct using, for example, statistical models that describe the likelihood of an occurrence in a given space and environment. If photos are available, image recognition based on machine learning can be used to spot species that are likely to be misidentified. It is also possible/conceivable to use a more heuristic approach based on an ontology for organisms and ecological traits (for example, "this is a fish, fish are always aquatic organisms, this observation was made on dry land and is thus unlikely to be true"). The development of improved tools and systems to improve data quality is a major task for future development of the infrastructure.

In Norway, spatial planning processes and decisions are required to integrate information made available through the infrastructure. If new information about the occurrence of Red Listed species or other species of special interest for environmental management and conservation are discovered through the planning process, the developer is responsible for archiving these observations in a publicly available database.

Key factors for the success of the biodiversity data infrastructure described here are: (i) the close cooperation between NBIC, data providers, NGO's and data users, (ii) the modularity of the infrastructure, (iii) the use of open and flexible standards for data exchange, and (iv) the integration of the infrastructure in the legal framework for spatial planning requiring that data from e.g., environmental impact assessments are made available through the infrastructure.

Keywords

citizen science, Darwin Core, data exchange

Presenting author

Knut Anders Hovstad

Presented at

TDWG 2022

Conflicts of interest