Swedish Biodiversity Data Infrastructure (SBDI): Insights from the Swedish ALA installation

Margret Steinthorsdottir[‡,§], Veronika A Johansson[‡,|,§], Manash Shah[‡,|,§]

‡ Swedish Museum of Natural History, Stockholm, Sweden
§ Swedish Biodiversity Data Infrastructure (SBDI), Stockholm, Sweden
| Global Biodiversity Information Facility (GBIF), Stockholm, Sweden

Corresponding author: Margret Steinthorsdottir (margret.steinthorsdottir@nrm.se)

Abstract

The Swedish Biodiversity Data Infrastructure (SBDI) is a biodiversity informatics infrastructure and is the key national resource for data-driven biodiversity and ecosystems research. SBDI rests on three pillars:

1. mobilisation and access to biodiversity data;
2. development and operation of tools for analysing these data; and
3. user support. SBDI is funded by the Swedish Research Council (VR) and eleven of Sweden’s major universities and research government authorities (Fig. 1).

SBDI was formed in early 2021 and represents the final step in an amalgamation of national infrastructures for biodiversity and ecosystems research. SBDI includes the Swedish node of the Global Biodiversity Information Facility (GBIF), the key international infrastructure for sharing biodiversity data.

SBDI's predecessor Biodiversity Atlas Sweden (BAS) was an early adopter of the Atlas of Living Australia (ALA) platform. SBDI pioneered the container-based deployment of the platform using Docker and Docker Swarm. This container-based approach helps simplify deployment of the platform, which is characterised by a microservice architecture with loosely coupled services. This enables scalability, modularity, integration of services, and new technology insertions.

SBDI has customised the BioCollect module to remove region-specific constraints so that it can be more readily improved for environmental monitoring in Sweden. To further support this, there are plans to develop services for the distribution of terrestrial map layers, which will provide important habitat information for artificial intelligence and machine learning research projects.

The Amplicon Sequence Variants (ASVs) portal, an interface to sequence-based observations, is an example of integration and new technology insertion. The portal developed in SBDI and seamlessly integrated with the ALA platform provides basic

functionalities for searching ASVs and occurrence records using the Basic Local Alignment Search Tool (BLAST) or filters on sequencing details and taxonomy and for submitting metabarcoding dataset Fig. 2.

Future developments for SBDI include a continued focus on eDNA and monitoring data as well as the implementation of procedures for handling sensitive data.

Keywords

SBDI, Docker, containers, microservice architecture

Presenting author

Margret Steinthorsdottir

Presented at

TDWG 2023

Hosting institution

Swedish Museum of Natural History / Naturhistoriska riksmuseet

Conflicts of interest

The authors have declared that no competing interests exist.

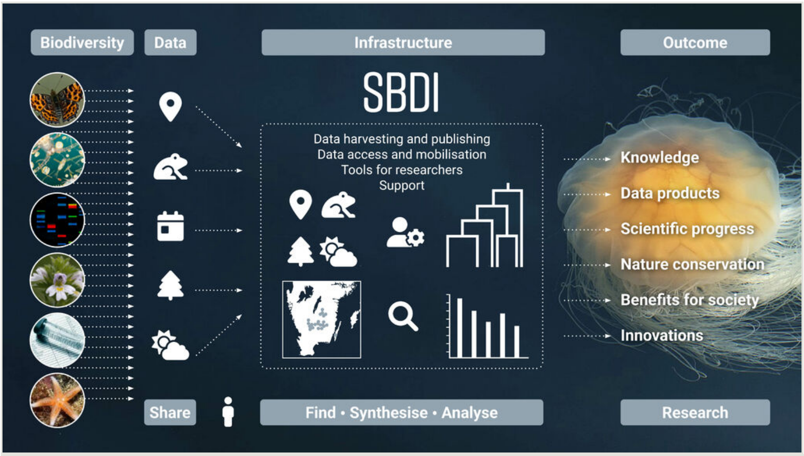

Figure 1.

SBDI includes tools for mobilising and working with a wide range of biodiversity data. The figure shows examples of data types and data mobilisation tools, as well as examples of web-based graphical front ends and software wrappers relying on SBDI APIs to provide data access, analysis and visualisation capabilities. Illustration by Johan Samuelsson for the Swedish Biodiversity Data Infrastructure.

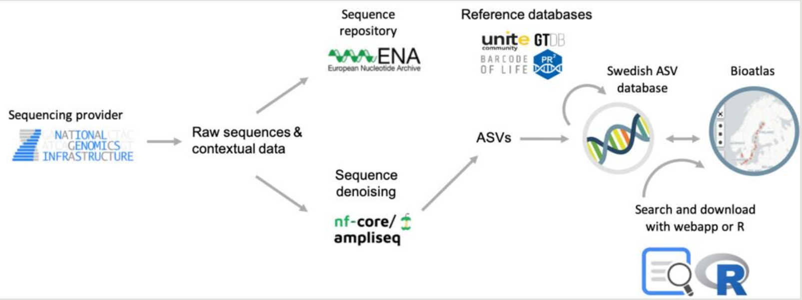

Figure 2.

Flow of metabarcoding data in SBDI: A sequencing service provider, such as the National Genomics Infrastructure (NGI), delivers sequencing data to the user. The user submits raw sequencing data and contextual data to the European Nucleotide Archive (ENA), denoises the data using e.g., nf-core/ampliseq, and submits the denoised data (ASVs and their counts in different samples, and contextual data) to the SBDI ASV portal. Metabarcoding data in the ASV database can be searched and downloaded using the portal or using the SBDI4R package (wrapped on the ALA galah package). Figure created by Manash Shah.