Margret Steinthorsdottir,
Veronika Johansson, Manash Shah,
Mats Bovin. May 2025.

SBDI
SWEDISH BIODIVERSITY DATA INFRASTRUCTURE

Swedish Museum of Natural History

ORGANIZATIONAL AND TECHNICAL OVERVIEW

Department of Bioinformatics and Genetics

Box 50007

104 05 Stockholm

Visiting address: Frescativägen 40

www.biodiversitydata.se

SBDI Organizational Overview

The Swedish Biodiversity Data Infrastructure **SBDI**, including the **Swedish node** of the Global Biodiversity Information Facility **GBIF**, is the national research infrastructure that enables the research community and other stakeholders to tackle the **biodiversity crisis** by providing Open Access biodiversity data and innovative tools for data analysis and visualization. SBDI services are delivered by a consortium of 11 universities and government agencies in Sweden: the Swedish Museum of Natural History (NRM), the Swedish University of Agricultural Sciences (SLU), Karolinska Institute (KI), KTH Royal Institute of Technology (KTH), Linnaeus University (LnU), Lund University (LU), Stockholm University (SU), Swedish Meteorological and Hydrological Institute (SMHI), Umeå University (UMU), University of Gothenburg (GU), and Uppsala University (UU). SBDI is jointly funded by this consortium and by the Swedish Research Council (VR).

1. SBDI Geographical and Organizational Distribution

The Swedish Biodiversity Data Infrastructure **SBDI** is a distributed e-infrastructure, where each partner contributes data and/or tools according to the specific specialized biodiversity data types and systems developing skills represented at their home institutions. The Executive Office (ExO) is located at NRM (Director, Project Coordinator) and SLU (Deputy Director). The SBDI Coordination Group comprises representatives from all partner organizations and is responsible for planning and reporting SBDI activities, including deliverables. SBDI's technical governance rests on a team of developers, data managers, and IT specialists. This team is responsible for the technical architecture, infrastructure resource planning and allocation, operational management, implementation of necessary developments, customizations and integrations, while considering evolutionary paths and new technology insertions. The technical staff are distributed among the consortium members with their responsibilities assigned by their home institution and planned deliverables. The consortium brings together a diverse range of expertise and SBDI's commitment to guiding principles like **FOSS** and **FAIR** enables better collaboration to achieve the common objectives. As part of the **Living Atlases Community**, SBDI collaborates with other national and international biodiversity infrastructures to share knowledge, tools, and best practices. This collaboration enables SBDI to remain at the forefront of biodiversity informatics and to integrate the latest developments in the field. SBDI pioneered the implementation of the ALA platform in a container-based environment and this approach of deployment has been adopted by a number of Living Atlases installations.

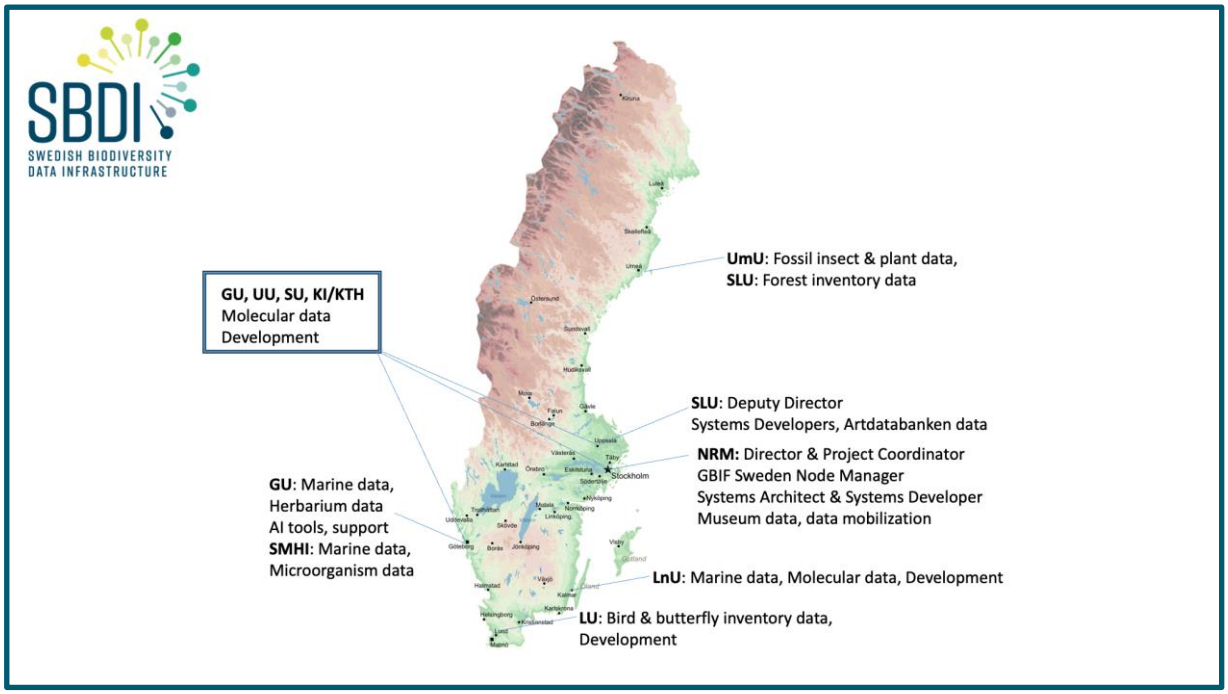

Figure 1: The figure shows the geographical as well as organizational distribution of SBDI in Sweden, with all important operations and intra-consortium collaborations indicated. Image: Margret Steinthorsdottir

SBDI operations are organized in five modules: 1) Management, 2) Communication and User Support, 3) Data Mobilization and Interoperability, 4) Integrated Analysis Service and 5) International Collaboration (GBIF Sweden), with activities in modules 1-4 comprising the bulk of SBDI operations.

Specialization in biodiversity data and related tasks varies across the consortium. Each partner organization is represented by a member in SBDIs Coordination Group. A brief description of partners' core competences and contributions is outlined below:

- **Swedish Museum of Natural History (NRM)**
  The Swedish Museum of Natural History (NRM) serves as the principal institution and fund manager for the Swedish Biodiversity Data Infrastructure (SBDI). NRM hosts key leadership roles for SBDI, including the Director, Project Coordinator, and the Node Manager of GBIF Sweden. The Director of SBDI reports directly to the SBDI Steering Committee and is responsible for coordinating consortium activities, including communication and outreach.

  A significant portion of SBDI's IT development is led and coordinated from NRM. While NRM contributes to SBDI in various capacities, it is particularly through GBIF Sweden that NRM plays a leading role in biodiversity data mobilization. GBIF Sweden, hosted at NRM, is responsible for supporting institutions and researchers across Sweden in publishing and sharing biodiversity data in alignment with international standards, and mobilizing this data to the global GBIF network. The **biodiversity data** curated at NRM primarily includes museum specimen data, survey-based observational data, and environmental DNA (eDNA) datasets.

- **Swedish University of Agricultural Sciences (SLU)**
  SBDIs Deputy Director at SLU works together with the Director, Project Coordinator and GBIF Sweden Node Manager in coordinating and organizing consortium activities. SLU is the host of **SLU Swedish Species Information Centre** (Artdatabanken), contributing a wealth of data to SBDI, including occurrence data, marine and forest inventory data, and habitat data. SLU staff also support users with data mobilization. Developers at SLU continuously develop and maintain the Swedish Observation System **SOS**, mobilizing data in standardized format to SBDI.

- **University of Gothenburg (GU)**
  GU is a leading SBDI partner in developing AI tools for marine biodiversity observation data, through the **SUBSIM** project. Gothenburg Museum of Natural History (GNM) contributes museum collection data (entomological, invertebrate and vertebrate). GU's Herbarium is the host of **Sweden's Virtual Herbarium**. GU contributes marine metabarcoding genetic observations in collaboration with SBDI molecular group partners (see below). GU also handles user support, together with NRM and various additional SBDI partners.

- **Lund University (LU)**
  LU provides important data from two long-term monitoring schemes: **Swedish Bird Survey** and **Swedish Butterfly Monitoring Scheme**, and contributes IT development to mobilize this data into SBDI and GBIF Sweden. LU's Biological Museum contributes biodiversity data from its botanical, insect and zoological collections.

- **Swedish Meteorological and Hydrological Institute (SMHI)**
  SMHI hosts the **Shark** National Environmental Monitoring project, providing biodiversity data on seals, porpoises, jellyfish, pico-, phyto- and zoo-plankton, as well as epi-and zoo-benthos, together with physical and chemical measurement data. SMHI-hosted **Nordic Microalgae** is a source of information about microalgae and related organisms in the Nordic area, i.e. the Baltic Sea, the North East Atlantic and lakes, rivers and streams in the area.

- **Linnaeus University (LnU)**
  LnU contributes data to SBDI on marine ecosystems, microbiomes and zoonoses. LnU was additionally involved in the Molecular Group metabarcoding efforts as detailed below.

- **Umeå University (UmU)**
  UmU contributes biodiversity data to SBDI from the world's largest database of Quaternary fossil insects, **BugsCEP**, a part of the Strategic Environmental Archaeology Database: **SEAD** and Swedish National Infrastructure for Digital Archaeology: **Swedigarch**. UmU also contributes fossil occurrence data for plant macrofossils and pollen from SEAD. Over time, this will be expanded to include fossil data for other organism groups.

- **SBDI Molecular Group**
  Karolinska Institute (KI) and **KTH Royal Institute of Technology** (KTH), both via **SciLifeLab**, **University of Gothenburg** (GU), **Uppsala University** (UU), as well as **Stockholm University** (SU), work together on linking genomic data with species occurrences and providing access to sequence data. The flagship **Swedish ASV Portal** is a novel, bespoke service developed within SBDI, consisting of a database of amplicon sequence variants (ASVs) and a web interface that manages observations based on DNA sequences. It includes a pipeline for denoising and taxonomic annotation of metabarcoding data.

2. SBDI Services, Data Mobilization and Support

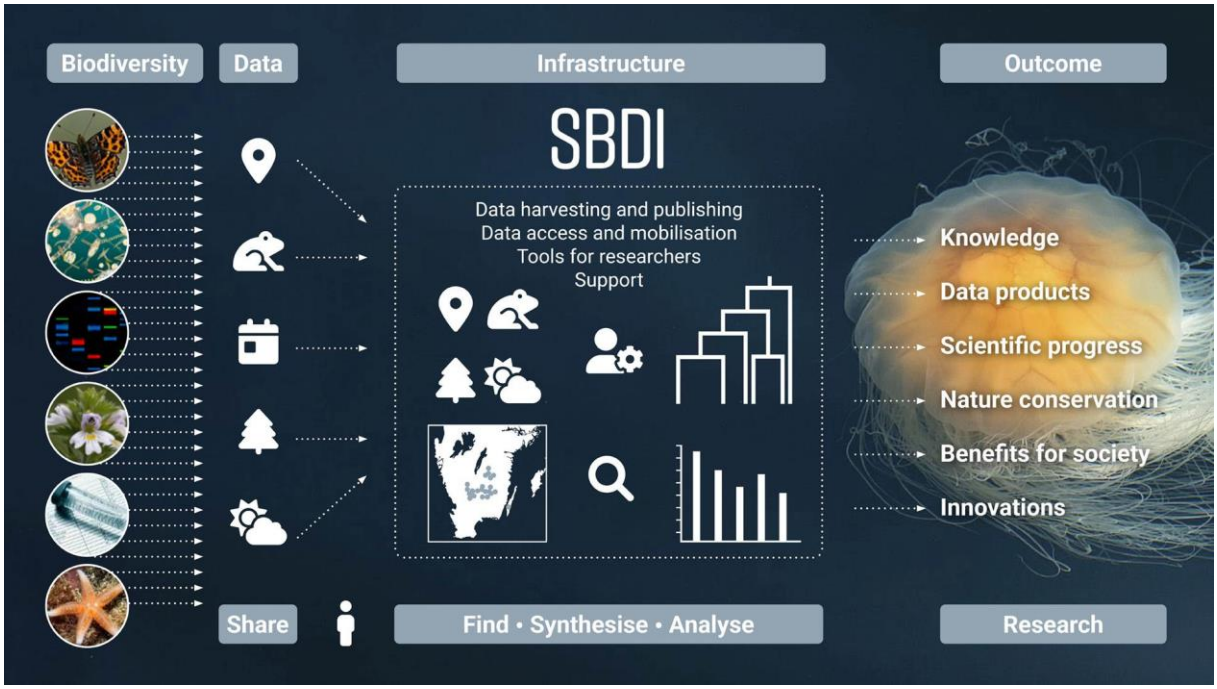

Figure 2: SBDI includes tools for mobilizing a wide range of biodiversity data. The figure shows examples of data types and data mobilization tools, as well as examples of web-based graphical front ends and software wrappers relying on SBDI APIs to provide data access, analysis and visualization capabilities. Image: Johan Samuelsson.

The core of **SBDI** is a set of interoperable **data and analysis services** which can be combined by users into custom-made pipelines. The technical platform is developed and maintained through international collaboration within the **GBIF** and **Living Atlases** (LA) communities, with SBDI being a key partner in both. The Swedish LA instance, the **Bioatlas**, includes several SBDI contributions to the core development effort, and a set of unique extensions for analysis of Swedish ecosystems. **SBDI provides user support and training**, both through its website and through the SBDI Support Center. Outreach activities and products include news feed, online manuals and training materials, training events at scientific meetings, workshops, and hackathons. More extensive support is provided to projects of high scientific merit. **Data mobilization is a core focus of SBDI**. The efforts focus on data sources of high scientific value, with SBDI services being vital in helping excellent research projects

and Swedish data providers comply with open-data and open-science requirements. Important target users include **Swedish systematic monitoring programs**, which provide some of the longest and most valuable biodiversity datasets in the world. SBDI staff provide consulting and other types of support to assist with data standardization, validation, and development of automated services and pipelines, following international standards and best practices. SBDI also contributes to the development of international data standards in several domains, facilitating novel analyses that integrate these data with other biodiversity and environmental data across national borders.

3. SBDI Solutions and Benefits for Users

Access to Data and Tools
SBDI provides seamless access to a vast collection of biodiversity data from Sweden and beyond. Users can explore species observations, environmental data, and genetic information through powerful tools for analysis, visualization, and modeling. The infrastructure supports both researchers and decision-makers in making data-driven conclusions about biodiversity and ecosystem changes.

Sharing Data for Increased Impact
SBDI offers a robust platform for institutions, researchers, and citizen scientists to publish and share their biodiversity data in a structured and **FAIR** (Findable, Accessible, Interoperable, Reusable) manner. By contributing data, users increase visibility, enable collaboration, and enhance the impact of their research while ensuring long-term data preservation.

The Benefits for Users of SBDI
Publishing Open Access and FAIR biodiversity data through SBDI provides access to a national and international platform to advance biodiversity knowledge and conservation. Users gain access to a collaborative network, state-of-the-art digital infrastructure, and shared expertise. SBDI fosters innovation, seeks to influence policy, and strengthens Sweden's position as a leader in biodiversity data management and research.

4. SBDI in the Swedish Biodiversity and Research Data Landscape

SBDI is an integral part of the Swedish landscape of Research Infrastructures and Centres that support research and publish data related to biodiversity, bioinformatics, ecosystems and the Earth's climate system. **SBDI** and **GBIF Sweden** focus mainly on open access to **FAIR** Biodiversity Data. **SciLifeLab**'s **NBIS** and **Planetary Biology** study life on Earth on multiple scales and provide data science support. **SITES** is the Swedish infrastructure for Ecosystem Science, generating and publishing data through field access. **ArchLab** creates new palaeobiodiversity data from environmental archaeological and palaeoecological investigations, and **Swedigarch** makes these data available to SBDI and other systems and users. The **Bolin Centre for Climate Research** includes research on the climate impacts on ecosystems and biodiversity, while **SND**'s **Researchdata.se** collects all kinds of research data into a single portal to easily find, share and reuse this data. **InfraVis** is the national infrastructure that offers the expertise and tools to produce state-of-the-art visualizations of research data. Commonalities and overlaps between these infrastructures and centres include focus on Open Access and FAIR research data, biodiversity and ecosystems, Earth science and climate, as well as operational components such as research communication, outreach and user support. SBDI

has initiated a community effort to collaborate more closely across these national RIs and centres, to mutual benefit for all.

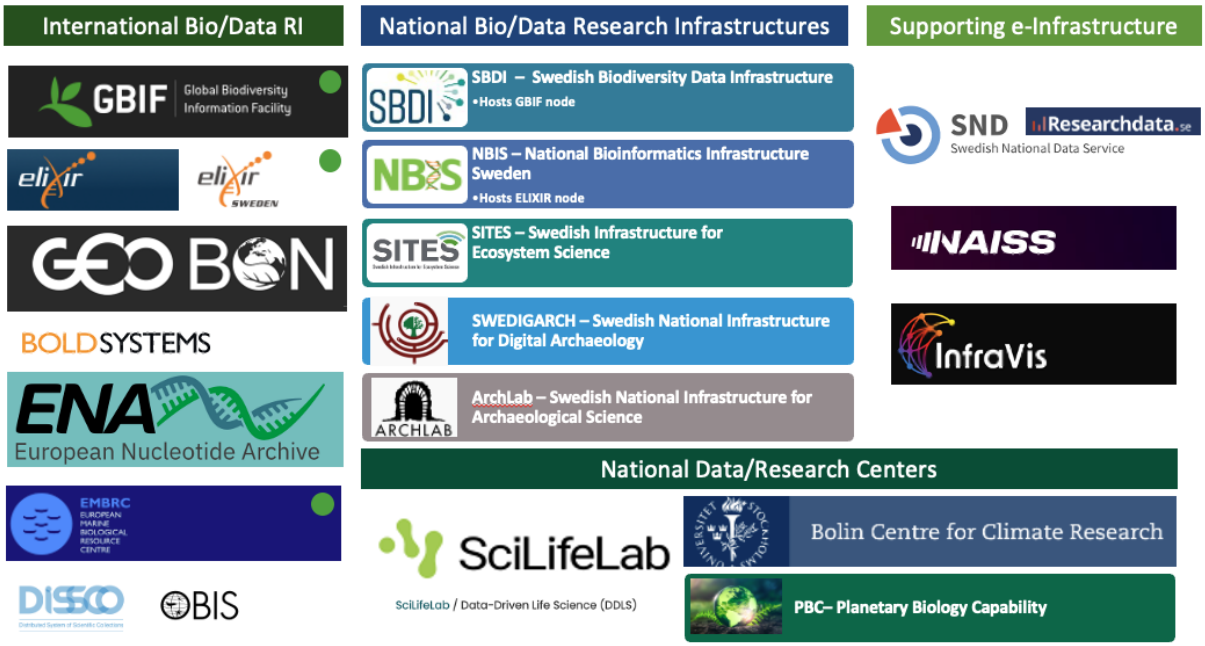

Figure 3: SBDI in the National and International Biodiversity and Research Data Landscape. Green dots indicate formal Swedish membership in International RI. Uniting key-themes include FAIR Research Data, Biodiversity, Earth, Life and Ecosystem Science, as well as Climate Impacts. Image: Margret Steinthorsdottir & Veronika Johansson.

5. SBDI in the European and Global Biodiversity and Research Data Landscape

The **Swedish Biodiversity Data Infrastructure (SBDI)** is Sweden's national research infrastructure for biodiversity data, and belongs to a landscape of **European** and **Global networks** that facilitate access to, sharing of, and analysis of **biodiversity data—including genetic data**. SBDI integrates data from a wide range of Swedish institutions, including museums, universities, private sector, environmental agencies, and provides tools and services for researchers, policymakers, and the public. SBDI also aligns with the **ELIXIR** infrastructure (the European life sciences data infrastructure) through **ELIXIR Sweden**, with a particular focus on **biodiversity genomics, metagenomics, and molecular data management**. Globally, SBDI is part of the **GBIF network**, promoting FAIR (Findable, Accessible, Interoperable, Reusable) data practices and supporting the integration of species occurrence data, taxonomic information, ecological observations, and DNA-derived data (such as barcoding and metagenomics). It also aligns with initiatives like **GEO BON** (Group on Earth Observations Biodiversity Observation Network), focusing on **global biodiversity monitoring**. In the **marine biodiversity domain**, SBDI contributes to and utilizes data from **OBIS** (Ocean Biodiversity Information System), helping integrate Swedish marine data into this global marine biodiversity platform. This enhances international efforts to **monitor and understand marine life**, in coordination with GBIF. Regarding **biodiversity genetic and genomic data**, SBDI is connected with infrastructures such as **EMBRC** (European Marine Biological Resource Centre), particularly through its support for environmental genomics and access to marine model organisms and reference data. EMBRC's **EMO BON** and

European Blue BioBank (EBB) initiatives align with SBDI's goals for integrating genomic data into biodiversity monitoring and research. SBDI links traditional species occurrence records with sequence-based biodiversity data by integrating tools and databases for **environmental DNA (eDNA)**, **DNA barcoding** of individual taxa, and **metabarcoding** of mixed-community samples. This **bridges phenotypic and genotypic approaches to biodiversity monitoring and research**. This further links to broader international initiatives such as **BOLD** (Barcode of Life Data System) and **EMBL-EBI's ENA** (European Nucleotide Archive), with which **Data Interoperability** is essential.

SBDI envisions future involvement in **DiSSCo** (Distributed System of Scientific Collections), which aims to **unify access to natural science collections across Europe**. SBDI can contribute digitized specimen data and support efforts toward **standardization, persistent identifiers, open data** practices in line with DiSSCo's vision. Thus far, Sweden has an Observer status in DiSSCo, with full membership as aim.

SBDI Technical Overview

1. Introduction

The Swedish Biodiversity Data Infrastructure **SBDI** is a national research infrastructure dedicated to managing, analyzing, and disseminating biodiversity and ecosystem data in Sweden. SBDI empowers users with seamless access to biodiversity data and analytical tools, enabling informed research and decision-making. By providing a platform for sharing data in a structured and FAIR-compliant way, it fosters collaboration and increases the impact of biodiversity research. As part of the consortium, users gain access to cutting-edge infrastructure, a strong expert network, and the opportunity to drive innovation in biodiversity science and conservation. As a key component of Sweden's environmental research landscape, SBDI caters to a wide range of stakeholders, including researchers and policymakers, through its diverse functionalities and data types.

This section provides a technical overview of SBDI, detailing below the implemented modules, their functionality, and potential future developments.

2. Technical overview

SBDI´s technical infrastructure is primarily built on the open-source platform, **Atlas of Living Australia (ALA)** developed in Australia and supported by an active global community, the **Living Atlases Community**. It is designed in microservices-based architecture. As a platform providing integrated data services of curated, quality-assured data, it consists of a number of modules loosely coupled together to form a unified infrastructure.

2.1 Data types

The infrastructure aggregates diverse datasets, including species occurrence records, sampling-event data, and environmental DNA (eDNA) data. These datasets are subsequently processed through an enrichment pipeline, and integrated with complementary data sources, such as dataset metadata, checklists, taxonomic lists, geographical and environmental maps, image and multimedia content, and sensitive species lists. This enrichment process augments the original datasets by providing

taxonomic, spatial, and environmental context, while also enhancing data quality through validation and the application of sensitivity and data quality flags. The result is a more comprehensive, accurate, and reliable dataset, suitable for advanced analysis and biodiversity research. Following is a non-exhaustive list of data types that constitute a core element of the infrastructure:

- Species Occurrence
- Sampling-event data
- Environmental and molecular DNA data
- Checklists & National Taxonomic List
- Dataset Metadata
- Geographical and Environmental Maps
- Image and Multimedia

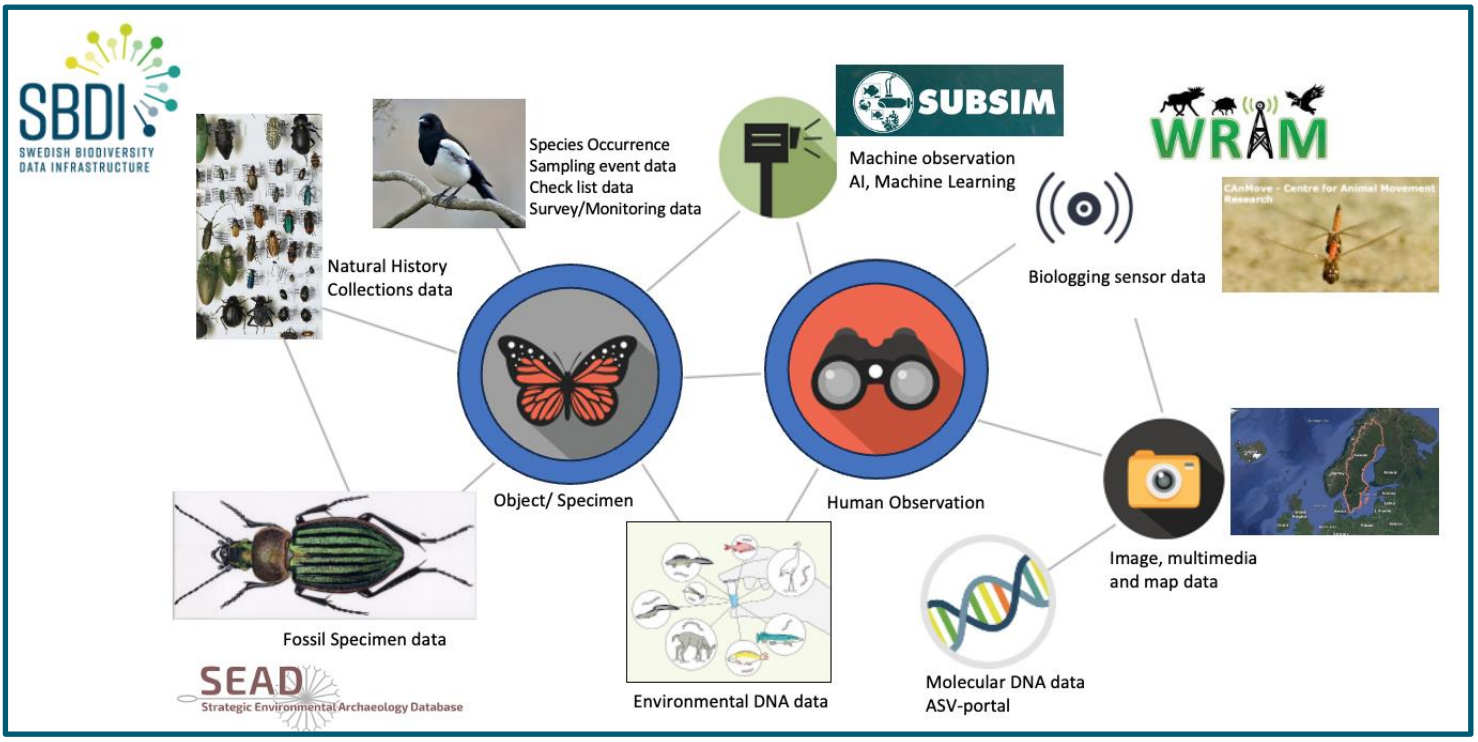

Figure 4: An overarching conceptual model illustrating the relationships between observations, various data collection methods and different data types in SBDI. Edited and adapted from an Atlas of Living Australia diagram. Source: ALA.

2.2. Data standards

SBDI mobilizes data from a multitude of sources, such as natural history collections, environmental monitoring programs, eDNA, research projects, citizen science projects, and more. Each source offers invaluable data on biodiversity. However, this data may be collected and managed in diverse systems and formats, varying widely depending on the details captured and stored for each individual record. In order to mobilize and publish the data according to the Open Science and FAIR principles, data needs to adhere to global standards of data formatting. SBDI/ GBIF Sweden staff aid data providers in transforming their verbatim data into standardized data formats.

SBDI primarily uses the **Darwin Core Standard (DwC)** as a cornerstone in mobilizing biodiversity data. This standard provides a stable, straightforward, and flexible framework for integrating data from

varied and variable sources. Developed and maintained by the Biodiversity Information Standards (TDWG) community, Darwin Core is an ever-evolving, community-driven standard that underpins the sharing, use, and reuse of open-access biodiversity data. It plays a critical role in SBDIs efforts, contributing to the hundreds of millions of species occurrence records accessible through SBDI as well as GBIF.

Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information. Within the TDWG community, **Darwin Core Extensions** are continuously developed to allow data owners to express data elements that go beyond the current version of the Darwin Core. Swedish biodiversity data providers are encouraged to actively participate as domain experts in this process. SBDI data publishers are strongly recommended to adhere to the established and stable Darwin Core Extensions where appropriate.

The main standard for metadata used by SBDI is the **Ecological Metadata Language** (EML), which was originally developed to support the description of ecological research datasets.

The data curated and published by SBDI adheres to global community standards and hence complies with the FAIR data principles.

2.3 SBDI Services & Modules

The infrastructure comprises a suite of essential services designed to facilitate the aggregation, management, processing, analysis, and dissemination of biodiversity data. The functionalities provided by the SBDI services are accessible both via user interfaces (UI) and Application Programming Interfaces (**API**).

Below is a list of the key services provided by the platform, sorted by importance and introduced by the key user benefit of each:

- **Search and Download Species Occurrences :**
  User benefit: Enables users to efficiently search and download species occurrence data using advanced filters, making biodiversity research more accessible and customizable.

  The occurrence search service based on the SOLR search index provides advanced search functionalities to the user applications. The user applications can send requests with filters to refine data queries enabling users to search and download Species occurrences.

- **Species Information:**
  User benefit: Provides comprehensive species data, integrating taxonomy, ecology, and conservation status, helping researchers and conservationists make informed decisions.

  The Biodiversity Information Explorer (BIE) module aggregates and provides comprehensive details about species linking the taxonomy, ecological characteristics, conservation status, and occurrence information in a search index.

- **Taxonomic Classification:**
  User benefit: Ensures accurate and up-to-date species identification with support for exact and fuzzy name searches, improving data reliability in biodiversity studies.

  The service provides taxonomic classification using the **GBIF Backbone Taxonomy**, with the Swedish Taxonomy Database **Dyntaxa** constituting one of the >100 resources used to build

the GBIF Backbone Taxonomy, and thus ensures alignment with up-to-date global taxonomy. The service supports taxonomic search based on both scientific and vernacular names and with exact and fuzzy name matching logic, as well as species hypotheses.

- **Usage Analytics:**
  User benefit: Tracks and visualizes service usage to improve platform functionality and user experience through tools like Metabase and Matomo.

  In order to better understand the usage of the various services and tools SBDI recently started using Metabase as a data visualization and analytics tool. Currently work is underway to collect and aggregate usage data into this tool. SBDI also uses Matomo for tracking user activity.

- **Monitoring and Observability**
  User benefit: Ensures platform reliability and performance by using open-source monitoring solutions to track system metrics, logs, and alerts.

  SBDI leverages leading open-source solutions for system and service monitoring. Prometheus is used for gathering metrics from servers and applications and Grafana provides the user interface for viewing and monitoring. Loki is used for gathering logs from all services. The logs are also viewable in Grafana.

- **Sensitive Data Management:**
  User benefit: Implements conservation sensitivity protocols to protect endangered species by controlling data precision and access levels.

  The Sensitive Data Service (SDS) implements protocols for managing sensitive species data using Conservation Sensitivity, whereby occurrence records for sensitive species are generalized according to country rules, e.g. by reducing their spatial precision or restricting access.

- **Sequence-based Observations Portal:**
  User benefit: Connects DNA sequencing data with species occurrences, facilitating genomic-based biodiversity research and simplifying the submission and analysis of sequence data. Enables easy record searches using Basic Local Alignment Search Tool (BLAST).

  The Swedish ASV Portal, a bespoke service developed within SBDI, consists of a database of amplicon sequence variants (ASVs) and a web interface manages observations based on DNA sequences, linking genomic data with species occurrences and providing access to sequence data. It includes a guide to submitting raw sequence data to ENA (European Nucleotide Archive) and a pipeline for denoising and taxonomic annotation of metabarcoding data.

- **Modeling and Predictive Analysis Tools:**
  User benefit: Supports biodiversity forecasting and decision-making by integrating species occurrence data with environmental datasets for visualization, tabulation, and predictive modeling.

  The service provides functionalities to explore relationships between species, location and environment. It includes modules to manage environmental datasets, such as climate and

land cover data, enabling integration with species occurrence data. Further the tools for tabulations, visualization, modeling and predictive analysis support biodiversity forecasting and decision-making.

- **Swedish Biologging Portal:**
  User benefit: Provides a platform for accessing and analyzing animal movement and sensor data, aiding ecological and behavioral studies with open and controlled data access options.

  The Swedish Biologging Portal is a novel platform developed within SBDI to support the inclusion of biologging data from animal sensor systems. Biologging research involves the collection of time-series sensor data. The data can be used to create visualisations of e.g. migration tracks, foraging and home range trajectories and/or to characterize behaviour or physiology of animals. The system enables management of data collected using animal-attached electronic tags or sensors and retrieved or remotely transmitted, or collected from radar/optical sensors. Through the Swedish Biologging Portal one can find sensor data from animals, including movement trajectories and data characterising behaviour or physiology of animals. The first version of the portal was launched in autumn 2024. The Swedish Biologging Portal is developed in collaboration between Lund University (CAnMove Centre of Animal Movement) and the Swedish University of Agricultural Sciences (**WRAM** Wireless Remote Animal Monitoring). There are three data access levels: Datasets can either be full open access, partial open access or without open access. Find a description of the access right levels here. The open data can be downloaded in json format from the web client, with files containing information according to the biologging data model, contained in a zip archive. The data can also be retrieved directly from the API.

- **Systematic Monitoring Project Management:**
  User benefit: Streamlines biodiversity monitoring projects, enabling standardized data collection, storage, and sharing for long-term species population tracking.

  The service caters to biodiversity monitoring projects and provides functionality to collect, store, manage and share data collected in the field in a standardized manner. The service manages the **Swedish Butterfly Monitoring Scheme**, a national program for monitoring butterflies in Sweden and **Swedish Bird Survey**, a standardized method for detecting and reporting changes in the abundance of birds and their distributions in Sweden over time. The data from the Swedish Butterfly Monitoring Scheme is maintained in a spatially enabled PostgreSQL database and is converted into DwC-Archives through GBIF's **IPT** (Integrated publishing Toolkit). In the instance of the Swedish Bird Survey, several LA modules are used to achieve required tasks :
  - BioCollect : the front-end Java client that allows surveyors to enter the data, and the admin to monitor the surveys.
  - Ecodata : the instance in charge of communicating with the Mongo database.

  These two modules interact extensively with other LA modules within the SBDI cloud, such as the authentication module (CAS) and the lists module (species management). SBDI uses internal tools to convert the data produced by BioCollect/Ecodata into DwC-Archives through GBIF's IPT.

- **API and Data Interoperability Services:**

User benefit: Allows seamless integration with external biodiversity platforms through APIs, facilitating automated data retrieval and analysis via e.g. the R package.

The services within SBDI offer APIs for programmatic access to the platform's data, supporting interoperability with external biodiversity systems and applications. This includes programmatic access of the data using the R packages **sbdi4r2** and galah.

- **Metadata Management:**
  User benefit: Ensures proper dataset curation, attribution, and citation, maintaining data transparency and integrity for research and publication.

  The Collectory module provides the functionality for dataset curation, ensuring management of dataset metadata, correct attribution, licensing, and visibility of data providers' contributions. It ensures proper dataset citation and tracks version history to maintain transparency and integrity in data usage, through GBIF services, such as IPT and DataCite.

- **Image and Multimedia management:**
  User benefit: Supports the storage and organization of species-related media, making it easier to document and visualize biodiversity information.

  The service supports the storage and management of images, videos, and other multimedia linked to species occurrences.

- **User Management:**
  User benefit: Provides secure access control, managing user roles and permissions to safeguard data integrity and streamline platform usage.

  The service manages users and roles within the platform. It controls user permissions and access to data, ensuring that different roles have appropriate rights along with managing API access to the different modules of the platform.

- **Knowledge Base and Helpdesk:**
  User benefit: Offers documentation, best practices, and expert support, ensuring users can effectively utilize the platform's tools and services.

  SBDI comprises a comprehensive knowledge base and a dedicated support service. The knowledge base provides access to resources and information on the services and tools along with best practices, guidelines, policies and case studies. The support system provides access to dedicated assistance and guidance.

- **Citizen Science Project Management:**
  User benefit: Engages the public in biodiversity research through crowdsourced tasks, expanding data collection and community involvement.

  The service built using Pybossa, a crowdsourcing platform, presents georeferencing as microtasks to volunteers. It provides a platform for management of citizen science projects, users allowing public participation. This service is developed and maintained by SBDI.

- **Mirroreum:**

User benefit: Offers a portable, pre-configured research environment with R Studio and biodiversity analysis tools, enhancing reproducibility and efficiency in data analysis.

Mirroreum is a portable dockerized software stack consisting of a web-based R studio environment developed and maintained by SBDI. It includes a pre-installed set of various assorted packages supporting reproducible research efforts within biodiversity informatics.

Table 1: SBDI Modules. Table outlining the SBDI services and modules listed above - their development origin, presence/absence of customization, as well as current development status.

SBDI Services and modules	Developed by	Customized	Dev status
Search and Download Species Occurrences	ALA	Yes	Active
Species information	ALA	Yes	Active
Taxonomic classification	ALA	No	Active
Usage Analytics	SBDI	-	Active
Monitoring and Observability	SBDI	-	Active
Sensitive data management	ALA	No	Active
Sequence-based observations portal	SBDI	-	Active
Modeling and Predictive Analysis Tools	ALA	No	Active
Swedish Biologging Portal	SBDI	-	Active
Systematic Monitoring project management	ALA	Yes	Active
API and Data Interoperability Services (sbdi4r)	SBDI	Yes	Active
Metadata management	ALA	Yes	Active
Image and Multimedia management	SBDI	Yes	Active
User management	ALA	No	Active
Knowledge Base and Helpdesk	SBDI	-	Active
Citizen science project management	SBDI	-	Inactive
Mirroreum	SBDI	-	Inactive

2.4 Integrated SBDI Services

The following services hosted by SBDI consortium partners are tightly integrated with the core SBDI LA-based services and are integral to SBDI's functionalities:

- **Integrated Publishing Toolkit (IPT):**
  User benefit: Facilitates standardized data sharing by enabling researchers and institutions to publish biodiversity data in globally accepted formats, ensuring interoperability and accessibility.

A core component of SBDI's data publishing workflow is the **IPT**—an open-source Java-based web application developed by GBIF and operated by **GBIF Sweden**. The IPT enables data providers to publish biodiversity datasets in structured formats using the **DwC** standard, packaging tabular data (e.g. occurrence, checklist, sampling-event records) alongside rich **EML** metadata. The tool supports mapping of local data schemas to Darwin Core terms, enabling semantic standardization and alignment with GBIF and other global infrastructures. It offers version control, persistent identifiers (DOIs), validation tools, and automated data indexing pipelines, ensuring datasets are **FAIR-compliant**, machine-readable, and ready for integration into global biodiversity research efforts, databases, and infrastructures.

- **Species Observation System (SOS):**
  User benefit: Provides efficient species observation data management, integrating raw and processed data from multiple sources while offering flexible export formats and API access for biodiversity research and conservation.

  The Species Observation System (**SOS**) is developed in .NET 9 and C#. The system runs on the Kubernetes container platform and uses a MongoDB database to store harvested observations in raw data format. Processed observations are stored in an Elasticsearch database, providing efficient access for search queries. To handle observation search and exports, SOS provides a REST API available to consumers via Azure API. The system supports exporting data to SBDI in several file formats, including CSV, GeoJSON, Excel, and DwC-A (Darwin Core Archive). The DwC-A format is specifically used to synchronize SLU's databases, such as Artportalen, with **IPT** (Integrated Publishing Toolkit). More information on SOS on GitHub. In summary, the SOS system harvests verbatim species observations from a number of data sources which it integrates and enriches with supplementary information, including taxonomic classifications, species traits, geographic distributions, and conservation statuses. The system curates the data, exports it in standard DarwinCore Archive format, and shares it via API with SBDI.

- **SUBSIM:**
  User benefit: Supports marine ecosystem research through AI-powered image and video analysis, enabling efficient data management, automated species classification, and large-scale collaboration, including citizen science participation.

  SBDI partners have developed an open-source platform for subsea image analysis called **Swedish Platform for Subsea Image Analysis** - **SUBSIM** to support research and monitoring in marine ecosystems. SUBSIM provides essential functions to conduct research and automated monitoring with image and video surveys, including data management, machine learning, digital collaboration, citizen science, and high-performance computing. SUBSIM was developed at the University of Gothenburg. The platform was officially launched in April 2024 and has until today been used by more than 30 research groups in marine, social, and data science projects. Visit "User stories" for data products and models, and visit "Publications" for examples. SUBSIM is designed to support scientists and investigators in academic institutes, environmental agencies, public bodies as well as environmental consultancy companies. The services for management and analysis of marine data are open-source and can be used by anyone. SUBSIM has a modular architecture, which currently consists of the following components: (1) Classification module: A compilation of annotation functions provided through third-party open source tools (e.g. Biigle, Zooniverse). The citizen

science component of SUBSIM has so far generated >260.000 image classifications from >7,200 public contributors (up to 2500 image classifications per day). (2) AI/ML module for model training and testing. (3) Infrastructure module: “Cloudina” is the central data management and orchestration infrastructure providing temporary data storage, authentication-authorisation services (AAI), scaling operations, high performance computing (HPC), data federation and provenance functions, as well as interoperability with external systems. The latter is enabled through data exchange scripts to other popular computer vision platforms (e.g. Fathomnet, Roboflow, BIGGLE). This module is fully integrated into the **National Academic Infrastructure for Supercomputing in Sweden** (**NAISS**), which allows for storage and processing of large data volumes in an environment with the highest data security standards and best possible conditions for implementing FAIR data principles. SUBSIM has access to extensive storage capacity (5 TiB), computational servers (700 GPU-h/month), as well as cloud services at the **Swedish Science Cloud** (15.000 coins). These resources are sufficient for running 5-15 projects in parallel and can if needed be upgraded with short notice. (4) Publication module: Individual script-based functions to publish models (e.g., to Zenodo or Swedish National Data Service, SND) as well as species observations to the Global Biodiversity Information Facility (GBIF) via the Integrated Publishing Toolkit (IPT). Documentation: https://github.com/ocean-data-factory-sweden

- **SEAD:**
  User benefit: Bridges environmental archaeology and biodiversity research by linking historical fossil (especially insect) data with contemporary biodiversity records, enabling interdisciplinary studies in archaeology, palaeoecology, paleontology, and climate research.

  Strategic Environmental Archaeology Database, or **SEAD**, is a national research infrastructure for Swedish archaeological science and part of the **National Infrastructure for Digital Archaeology**: **Swedigarch**. Whilst hosting Swedish content, it is an international database for environmental archaeology and Quaternary science data. Swedigarch is a partner infrastructure to SBDI and the representative of partner organization UmU. SEAD contains data from archeological excavations and palaeoecological investigations, including data on the past occurrence of plants and animals. The dataset and service most relevant to SBDI is the Quaternary Insect **Coleopteran Ecology Package** **BugsCEP**. BugsCEP is maintained as a downloadable MS Access database. Technical support is through SBDI or Swedigarch at UmU, with data collation and single point data entry managed by an archaeologist in Sheffield, UK. SBDI links data from SEAD to contemporary biodiversity data sources, including GBIF, using the scientific names of species and other taxa. The taxonomic checklist is a hybrid of UK, Swedish, German and European checklists, which is now maintained in GBIF through the SEAD-SBDI cooperation. Data is periodically ingested from BugsCEP into SEAD through a bespoke import routine with numerous automatic checks. Export to SBDI is through a DarwinCore export API in SEAD's backend, which can be run on demand. New data is released to the online browser and API's in SEAD's periodic releases. Other types of data are curated by partners in the Swedigarch consortium and international partner projects.

2.5 IT Infrastructure, Deployment Architecture and Security

SBDI is hosted on a cloud platform procured from a private cloud service provider (Blue **Safespring** AB) with physical data centers in Sweden. It is hosted on OpenStack with the facility of self-

provisioning of necessary resources. NRM has assessed the security of the cloud service provider Safespring according to established internal routines and approved the use of its services.

The Service Level Agreement (SLA) with Safespring covers the following services:

- **Compute** Safespring Compute service, including local storage and block storage.
- **Storage** Safespring object storage.
- **Backup** Safespring cloud backup service.

Service Level Guarantee is 99,99% - 24/7/365. The SLA further covers calculation of Uptime and Downtime, using support, support process, as well as critical and major incidents.
The provisioning of the resources, Virtual Machines, Memory, Disk Storage, Network topology, Security policy is done using Terraform, adhering to Infrastructure as Code practice. This enables a reproducible and scalable environment with the declarative definition files checked into the version control system. The deployment and configuration of applications is automated using Ansible, ensuring consistency and reliability. This results in faster deployments, reduced manual intervention, and uniformity in application environments.
The applications are packaged as Docker images and deployed in a containerized environment. The applications deployed as microservices are hosted on a cluster of host machines where Docker Swarm is used as the orchestrator. This setup provides isolated, scalable, and lightweight deployments with resource allocation, load balancing, and service scaling while maintaining high availability and fault tolerance.

In order to **enhance security and limit potential attack vectors**, the infrastructure is designed with a minimal publicly exposed surface. The Docker swarm cluster hosting the applications are housed within an internal, private network that is not directly accessible from the internet. External access to these host machines are permitted via secure, redundant gateway hosts to authorized users only.

The external web traffic is directed to one of two gateway machines, configured in a high-availability (HA) setup. HAProxy is used for load-balancing and to ensure redundancy. Requests are secured through SSL encryption at the gateway level, providing end-to-end encryption and routed by HAProxy to the appropriate service on the internal network.

A Network File System (NFS) server with an NFS volume is used to share persistent storage space between Docker services within the Docker Swarm cluster. The use of object storage is being explored for storing immutable data like multimedia files in addition to the block and file storage.

For the periodic data ingestion process, a cluster of machines is provisioned on demand.
These machines are instantiated only when the data ingestion task is scheduled and destroyed once the ingestion process is completed. The root disks of these machines are persisted for rapid re-provisioning of new instances in the future. This approach of using ephemeral machines helps in optimization of resources and costs.

Data files and periodic database dumps are transferred and synced to NRM for backup. At NRM, these files are further backed up to an off-premise location providing extra layers of data protection and redundancy.

3. Future Development and Innovation

SBDI's technical team continuously explores new features and modules that could be integrated into the platform in addition to improvement and customization of existing ones. This includes assessing the feasibility of implementing modules like MERIT and Profiles, as well as expanding existing tools like Spatial Portal and BioCollect to accommodate more diverse data collection, analysis and visualization efforts.

3.1 Enhancements

SBDI has identified several ALA modules that could enhance its capabilities if implemented or further improved in the future:

- **MERIT** (Monitoring, Evaluation, Reporting, and Improvement Tool):
  This tool supports environmental projects by providing a platform for managing, reporting, and evaluating conservation and restoration efforts. Integrating MERIT could allow SBDI to better support environmental monitoring projects in Sweden.

- **Profiles:**
  The Profiles service offers a platform for creating species profile syntheses, allowing users to explore expert insights and Indigenous Ecological Knowledge on species groups. It organizes taxon profiles according to taxonomy, with each profile featuring authoritative descriptive content from experts and dynamically linked data. This data includes distribution maps, images, taxonomy, nomenclatural information, and diagnostic keys. The built-in editorial access management ensures the information remains up-to-date. Implementing Profiles would enable SBDI to offer more detailed species information, especially on flora and fauna.

- **Spatial Portal:**
  The **SBDI Spatial Portal** offers tools to visualize and analyze biodiversity data. SBDI can further improve or adapt the portal to better meet national needs. For example, the Austrian Living Atlas Spatial Portal is customized for management stakeholders and includes specific features related to the EU Habitats Directive, such as tools for reporting and compliance. Adapting the SBDI Spatial Portal to include similar features would make it more useful for conservation management and policy implementation in Sweden, aligning the platform with national and EU environmental directives (**CORDIS**).

- **Language and localization:**
  Currently, the SBDI portal is available in English only. Translations to Swedish can however be facilitated using the AI powered collaborative translation tool Crowdin, which has been successfully used for other languages in the **ALA Internationalization** project. About 30% is already translated to Swedish, with some modules like the BioCollect being fully translated. Since all partners of SBDI are Swedish authorities, we are obliged to provide translations according to the law of accessibility. However, exemptions apply to research and map

services, which has deprioritized this work. To engage a broader audience, including authorities and the private sector, we may need to reconsider the present approach.

4. User Interface / User experience

The SBDI web presence is organized across the main domain **biodiversitydata.se** and a number of subdomains. The services described in the previous sections are assigned and accessible at individual subdomains. The main landing page is served via a Wordpress instance accessible at https://biodiversitydata.se which is complemented by two other separate instances of Wordpress, one for Knowledge base (https://docs.biodiversitydata.se) and the other focused on the functionalities and tools available in the infrastructure (https://tools.biodiversitydata.se). A common wordpress theme implementing SBDI's graphic profile is used across all three wordpress instances. A common theme compatible with the wordpress theme is developed and used by the ALA based modules and services for a harmonized look and feel across the services. However, the current setup of the three WordPress instances compounded with modules/services on individual subdomains presents several challenges despite best efforts, some of which are listed below:

- **Navigation Issues**: Users find it difficult to navigate between the separate wordpress instances. The hyperlinks pointing to different subdomains complicate the user experience.

- **Limited Search Capabilities**: The search functionality is limited to individual wordpress instances. Users cannot search across all SBDI-related content from a single point, which hampers the ability to quickly find relevant information. The lack of a unified search functionality across all instances makes it harder to locate specific information or resources.

- **Inconsistent Information**: Maintaining three separate instances can lead to inconsistencies, where updates made on one site might not be reflected on the others. This can create confusion and have adverse effects on user experience.

- **Indexing Challenges**: Search engines may struggle to index content spread across multiple subdomains effectively and correctly, potentially reducing the visibility of certain resources and the infrastructure.

- **User Uncertainty**: With content split across different subdomains, users may be uncertain about where to go for specific information. The division between general information, documentation, and tools is not always clear.

Suggestions for UI/UX Improvement:

- **Single WordPress Instance with Subsections**: Consolidating the three instances into a single WordPress installation with clear subsections (e.g., /docs /tools) would streamline the user experience. This would reduce redundancy and ensure that updates are consistently applied across all sections. This would also simplify the process of maintaining the site and keeping information up-to-date. Work in progress - consolidation completion planned in 2025.

- **Global Search Feature**: A global search index that can search for all content across the site, allowing users to find documentation, tools, and general information from a single search query. A single wordpress instance would allow for a unified search function.

- **Breadcrumb Navigation**: A clear breadcrumb navigation across all the components and wordpress pages can be helpful to users in understanding their location and help with navigation.

- **Redesign the Landing Page**: The landing page should act as a comprehensive gateway providing clear user pathways, directing users to the data, documentation, tools, or general information sections based on their needs.

- **Consistent Design**: Ensure that all sections share a consistent design and layout, making it easier for users to recognize that they are within the same interface.