

# Swiss Natural History Collections Network

## Concept of data aggregation and publication for Swiss natural history collection data

### 1. Overall goal

The goal of SwissCollNet is to facilitate the access to standardized natural history collection data held in Swiss institutions and data of specimens collected in Switzerland and curated abroad for researchers, data providers, curators, collection-holding institutions and customers from the cantons and the federal administration in Switzerland. An online-portal - the Swiss Virtual Natural History Collection (SVNHC) - will inform on collections curated in Switzerland and containing biological and paleontological specimens and guide the customer to national and international data nodes for more detailed information. In addition, the collections will gain visibility for the public.

The digitally available natural history collection data of Swiss public and non-profit organisations (natural history museums and collections, botanical gardens, institutions of higher education) will be aggregated and interconnected. By complying with the FAIR data principles, adhering to defined international data standards and specified data vocabularies, the interoperability and transfer of data to other national and international research infrastructures will be facilitated. The possibility to integrate data from Swiss natural history collections, e.g. into international initiatives like GBIF and GeoCAsE is mandatory.

To maintain and increase accessibility of natural history collection data housed in Switzerland and data of Swiss specimens housed abroad is a goal, which has to be pursued in long-term and has to be carried on after the end of the SwissCollNet initiative, which is lasting from 2021 - 2024. Therefore, a long-term organisation will be set-up as an umbrella association, which will be responsible for the planning, development and future orientation of natural history collection data aggregation and diffusion. The long-term goal of the organisation is to provide the know-how and tools that natural history collections need for the digitisation, data management, exchange and presentation of their collection objects and data. It also promotes the standardisation of data exchange by jointly defining data standards by the specialist groups and declaring them binding for exchange. In addition to exchange formats, "controlled vocabularies" are also compiled and maintained. It provides a platform through which knowledge, methods and tools can be exchanged, thus reducing redundancies and idle time. Ideas are collected and disseminated through regular events.

## **2. Aggregation and publication of biological and palaeontological specimen data**

More than 90% of the specimens curated in Switzerland are biological or paleontological. For the aggregation and publication of digital information on these specimens, SwissCollNet will collaborate with InfoSpecies and build on the already existing data infrastructure of the Swiss node of GBIF.

InfoSpecies is the umbrella organisation of the national data centers and coordination offices for species promotion. One of the main goals of InfoSpecies is the provision, management and diffusion of species records, which includes natural history collection data.

The long-term objective is to achieve automatised data transfer of collection data to the national aggregator and publication of data on national and international online portals as well as an automatised update of data records. As a minimal goal, data generated in the frame of the SwissCollNet initiative have to be nationally aggregated and be publicly accessible by the end of the year 2024 through the Swiss node of GBIF.

## **3. Aggregation and publication of geological and anthropological specimen data**

Aggregation of digital information from collections of minerals, rocks, sediments, meteorites and palaeontology, SwissCollNet will collaborate with GeoCAsE.

For anthropological/archaeological collections, a decision still has to be taken. Out of the 68 approved projects, 3 projects deal with this type of data and will address the question how they will (can be) represented within SwissCollNet/SVNHC. Of special consideration is hereby the question of data protection.

## **4. Project concept for biological and palaeontological specimen data**

SwissCollNet has worked out a first concept for the collaboration with InfoSpecies, based on a preparatory study performed by the University of Applied Sciences Grison, a workshop with stakeholders of collections, the biodiversity observation community, geosciences etc. and several exchange meetings with persons from InfoSpecies, in particular with info fauna.

### **4.1. Stakeholders (providers of data, users of data)**

#### **4.1.1. Data providers**

The main data providers are public and non-profit natural history collection-holding institutions.

The following collections are concerned: preserved specimens in botany (cryptogams, phanerogams), mycology, lichens and zoology (vertebrates, invertebrates), paleontological, archaeozoological and archaeobotanical collections.

The following disciplines are out of scope: microbiology, virology, veterinary medicine, human medicine.

#### **4.1.2. Data users**

The data users which were identified are listed below.

- Researchers (national and international levels)
- Data collection-holding institutions
- Swiss Datacenters for Species Information
- National and international research infrastructures

- Citizen scientists
- General public

## 4.2. Data flow and accessibility of data

The data are openly published and follow the FAIR data principles unless considerations about protection of biodiversity or ethical considerations apply. Some data are sensitive and should not be made available to the general public. Therefore, it must be possible to control the publication of the data on several levels. The sensitivity of the data is determined by the data owner and the InfoSpecies deontology that is applied to Swiss data flowing into GBIF.

The system has to offer the possibility for data providers to easily update their data on their own.

The data aggregator redistributes the data to the Swiss online portal SVNHC and to the international open-source data nodes, like GBIF.org, GeoCAsE and the Swiss data centers for species information. Data flow to other national aggregators is possible if it does not conflict with the business model of the respective aggregators. A one-way data flow from the collection holding institutions through the data aggregator to the online portal SVNHC is planned. An automated ticketing system for data users will be installed to give stakeholders the possibility to interact with data providers for data corrections etc.. Updates of the specimen data will only be performed by the data providers (owners of the data).

Data have to be traceable. It must be possible to trace who downloads which data. This information is needed by the collections as an argument for entering projects.

For publication of collection specimen data at the Virtual Data Center (VDC), validation of data by the corresponding data centers for species information has to be carried out. Therefore, collection data information to be published in the VDC will first be transferred to the data centers and be published after validation and adaptation to specific requirements of the data users of the VDC (federal and cantonal administration, eco-offices).

The Swiss online portal SVNHC may be integrated in existing infrastructures like GBIF.ch or GBIF.org.

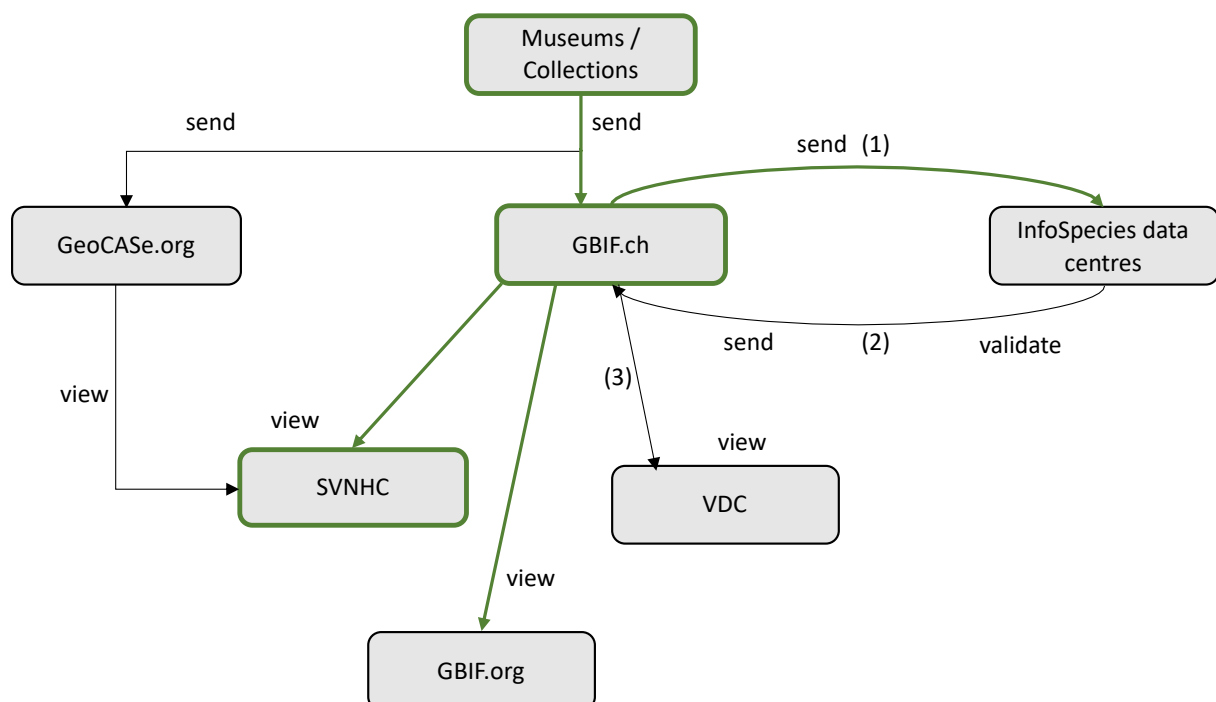


Figure 1: Data flow and publication of natural history collection data (i.e., biological and earth science collection data). The parts which will be carried out in collaboration with InfoSpecies are drawn in green.

### 4.3. Data structure and validation

#### 4.3.1. Data structure:

Whereas GBIF organizes data by the physical location of the finding, more inclusive data models have originated in recent years: DiSSCo developed the Digital Specimen concept and the Biodiversity Collections Network developed the Extended Specimen concept. As the latter two will be taken over by most aggregators in the near future, SwissCollNet will as well implement one of them or a combination of them. The data aggregator should include at least the details as used by GBIF.ch, ideally it should contain the details used in the collection management system SPECIFY.

The data model of SwissCollNet follows standards to guarantee interoperability with upstream data providers and downstream data aggregators. The two currently prevailing standards are Darwin Core and ABCD-EFG, Darwin Core is favored as it is used in the Collection Management System Specify and the aggregator GBIF. Darwin Core should be the main standard, the system may be expanded to ABCD-EFG or have Darwin core covering information of ABCDE-EFG in the future.

Vocabularies used in the aggregator of collection specimen data have to be internationally compatible, open and externally controlled and provide unique identifiers that link to the terms of the vocabularies. As vocabulary, the biodiversity standards of TDWG have to be respected and implemented. In biological species names, for instance, the GBIF Backbone Taxonomy has to be used.

Data sets have to be uniquely identifiable; this also applies to the 'controlled vocabularies. Persistent identifiers have to be used.

#### 4.3.2. Data validation:

The data providers are responsible for the correctness and quality of their data. User friendly reporting options should exist for feedback from users to the data providers.

### 4.4. IT-infrastructure

The IT-infrastructure for natural history specimen data is a standardised database operated by the Swiss node of GBIF. The IT infrastructure provides a central entry and exit point for biological and palaeontological data of the natural history collections. It allows data input via an API access with OAuth2 authentication and automated via a locally installed 'wrapper' e.g. GBIF. It allows automated or manual data output to GBIF.org, the SVNHC-online platform, the Swiss data centers of species information (e.g. info fauna, info flora etc.) and eventually GeoCAsE.

The current pipeline which takes care of collection data (GCH-DATA within PICT-IS) is run by info fauna in the frame of the GBIF Swiss Node mandate of the FOEN. It is not automatized and requires some freshen up. This update should allow the biodiversity data community to share and enrich data in an efficient and transparent way while respecting existing commitments (such as the publication of data on GBIF) and guidelines (use and diffusion of data: <https://www.infospecies.ch/de/daten/datennutzung.html>).

### 4.5. Data storage and long-term maintenance

In the scope of this project, a data management plan (DMP) will be generated that deals - amongst others - with data storage and long-term maintenance of the collection data in the data aggregator.

This DMP will be written by representatives of SCNAT, the SwissCollNet board(s), InfoSpecies and external consultants.

#### 4.6. Milestones

The milestones to be reached in the project are as follows:

- a. Project planning:  
Set up of the collaboration between SwissCollNet and InfoSpecies (organisational structure of the project, project milestones and contents, responsibilities, timeline, business plan, risks assessment).
- b. Specifications of project contents:  
Expert groups will further specify the requirements needed to fulfil the milestones defined in the project plan structured in work packages.
- c. Implementation and development of the data aggregation infrastructure (GBIF/PICTIS) and the SVNHC online portal:  
The infrastructure will be developed using an AGILE process.
- d. Aggregation and publication of natural history collection data:  
Data generated in the frame of the initiative SwissCollNet will be transferred to GBIF.ch and displayed in the SVNHC in 2024 and represent the prototype of the long-term infrastructure.
- e. Long-term development of the infrastructure:  
To further develop the prototype and keep-up with international developments, representatives of natural history collection institutions have to be part of the executive board of the Swiss node of GBIF and their role and responsibility be integrated into the rules of procedure of GBIF.ch.