

Minutes

SVNHC-Workshop, SwissCollNet

Date: Thursday, July 7th 2022, 13:30 - 16:30

Present:

Alex Bernhard (Swissbryophytes, IT)	AB	Lukas Wotruba (WSL)	LW
Alexis Beck (NHM Geneva, IT)	ABe	Marc Limat (MBL)	ML
Bärbel Koch (NHM Lugano)	BK	Michael Greeff (Ent.Coll ETH, BoE)	MG
Beda Hofmann (NHM Bern)	BH	Philippe Juillerat (Info Flora, IT)	PJ
Donat Agosti (Plazi)	DA	Pascal Tschudin (GBIF)	PT
Glenn Litsios (InfoFauna)	GL	Pierre Dèzes (SCNAT, Geos.)	PD
Hao Wang (SSEF)	HW	Rachel Walcott (NMS UK, GeoCase)	RW
Jean-Luc Gattolliat (MZ LS)	JG	Silvia Stofer (InfoSpecies)	SS
		Silvan Thüring (NMSo)	ST

Excused: Peter Wandeler, NHMF PW

Staff Ana Petrus, FH GR AP
 Yvonne König, SCNAT YKo
 Jessica Joaquim, SCNAT JJo
 Pia Stieger, SCNAT PSt

Agenda

Time	Topic	Form	Who
13:15	1. Opening of the meeting, introduction round	I	MG/all
13:20	2. Today's goals, introduction to the break-out groups	I	AP
13:30	3. Work in parallel break-out groups on main topic <ul style="list-style-type: none"> • G1: technical aspects for the realization of SVNHC • G2: definition of use cases 	Di	All AP PSt
14:15	Short Break – changing rooms		
14:20	4. Work in parallel break-out groups on secondary topic <ul style="list-style-type: none"> • G1: definition of use cases • G2: technical aspects for the realizastion of SVNHC 	Di	All PSt AP
14:50	Break		

15:05	5. Present results and discuss topic "technical aspects for the realization of the SVNHC"	I	orally
-------	---	---	--------

15:40	6. Present results and discuss topic “definition of use cases”		
16:15	7. Conclusions, next steps		all
16:30	Apéro		

I = Information, Di = Discussion, De = Decision

1. Opening of the meeting, introduction round

MG opens the meeting. The participants introduce themselves.

The Swiss Virtual Natural History collection (SVNHC) has the aim, to display all Swiss specimen curated in Swiss natural history museums and institutes. Therefore, a data aggregator (SVNHC aggregator) and an online-portal (SVNHC) have to be put in place. By defining the technical requirements and use cases for the SVNHC in this workshop, it can be discussed, if a collaboration with InfoSpecies and the linking to their data aggregator PICTIS is an appropriate technical solution for natural history collection data.

2. Today’s goals, introduction to the Break-out groups

The goal of the workshop is to discuss and agree on **key technical aspects** and **use cases** for the data aggregator. These agreements can be used for further discussion and the constitution of project bodies. As time is the most limiting factor during the workshop, the moderator controls the time very strictly. Deeper discussions can be continued in the breaks.

3. Discussion and results of the topic: technical aspects for the SVNHC

Paragraphs 3 and 5 of the agenda are summarised in this section.

Flipcharts produced during the discussions of the two break-out groups see paragraph 6.

3.1. Data model used and number of elements per specimen

- The data model discussed in DiSSCo should be used with the specimen in the center (digital specimen), thus a specimen centric view with at least minimal data set as for GBIF.org. More ideally, the data aggregator should have the same detail as data are in specify (CMS).
- The core information of a specimen needed are: “what, where, when, who”. Fields for the original data and interpretation of the data should be available.
- For observation data, georeferencing is very important, and geographic information of the specimen (observation of a specimen) is a key information for InfoSpecies. The start with georeferenced as part of primary information is proposed.
- The minimal requirements for data information in the SVNHC aggregator are to have the same elements recorded as in GBIF.ch, which are about 30 elements. The number of elements needed depends on the use cases defined. It will be a tradeoff between the content of the CMS in the museum and the defined use cases for the SCN data aggregator as long as compatibility with the model used in DiSSCo is guaranteed.
- Photos should be stored at a long-term repository (FAIR-principles applied). A link to the SVNHC data aggregator is enough. MIDS (minimal data standards of metadata for images) should be applied.

- To incorporate a collection management system (CMS) in the data aggregator is beyond scope.

3.2. Data richness for the data aggregator

Three different levels of data richness have been identified as layers forming a pyramid. The data richness in the different layers has been described:

- **Basal level:** this is the richest data level and is representing data in the local collections. The basic level can be subdivided in two layers. One layer is with all data including sensitive data. The second layer **filters out all the sensitive data**. The data from this layer can be transferred to data aggregators and made available for use.
- **Middle level:** this is the centralised data layer composed of decentralized data from basic layers (at the moment PICTIS for InfoSpecies and GBIF.ch.). The aggregator level is relevant for the network in Switzerland, the homogeneity of the data is key. This middle layer should be further developed to make GBIF.ch stronger by including the data and requirements of collections.

Note: The data layer called "InfoSpecies" in the drawing during the workshop is misleading. It should be named differently in the future.

- **Top level:** this is the global aggregation like in GBIF.org.

Sensitive data, an example from the biodiversity area:

The occurrence of bats. Bats use to live in roofs of private houses. Cantons need very precise data about the location and they will get the exact coordinates, but this information should not be open for the public. Not everybody should see where the bats are living. For the public the geographic location will only be visible in a resolution of 5 to 5 km square.

Discussion about the "one specimen entity"

Challenge (question to follow-up): In biology a specimen may be composed of more than one organism, or an organism of different organs (vegetative part of plant, fruit etc.), in geosciences, a specimen can be composed of different samples. How to deal with a specimen, which contains several species or compounds. Thus, such organisms should be divided in single samples and information linked together. With the **unique identifier per sample or compound**, linking is possible.

Discussion about unidentified, incomplete or imperfect data

How to handle data reporting for unidentified species, where only family information is available?

Can the genus level be recorded at GBIF?

In GBIF this kind of information can as well be added and recorded accordingly.

Where would the collection data be accessible?

On a web application.

3.3. Data flow:

Important to have an automatised feedback mechanism from data users to data providers. This feedback should be standardised, easy and automatised (two-way flow). A comment from a data user should get transferred to the data provider. It could be made visible in the system as a "pending" question for the curator. Answers should be given in realistic time. Corrections should be made and doubling of data avoided. Correction mechanism is solved at GBIF-level. In this sense, there would be a **benefit** as the result would be **common enriched data**.

The whole discussion about an **object**, with all the comments should be visible. As example "inaturaliste", <https://www.inaturalist.org> for observational data (InfoSecies), is mentioned. A similar mechanism could be built for collections data. Then, the question of how and by whom the update to collections data should be made has to be solved.

It is also important to know, which **determination key** has been **used** for the determination of a species.

3.4. Standards

Standards should cover both Darwin Core and ABCDEFG. As internationally the focus is on Darwin Core and it is rapidly expanding, the aggregator should concentrate on Darwin Core with possibilities of adaptation along with evolution of DiSSCo and GBIF. It can include up to 117 basic data, that is fine. ABCD-EFG, which can include 1500 basic data gets complicated. This may not be required for the SVNHC data aggregator.

For Geosciences, DwC has minimal fields of basic information, but it is evolving. In about two years' time, Darwin Core should include all necessary standards for geosciences. In Geosciences the standards are indeed less developed than for Biology. The **Paleobiology** data are handled with existing standards like DwC. The next step planned by GEOCASE and the TDWG (Biodiversity Information Standards) is to extend DwC for **Mineralogy**. For **meteorites** there is still no action planned. Geosciences has also an overlap with several other disciplines like biology, gemology, crystallography. With the **unique identifier** of an object, linking can be done.

For the time being, data from geosciences are integrated at international level in GeoCase. Geosciences could however be included into the SVNHC data aggregator solution. All disciplines should be under the same umbrella. Geosciences data could be incorporated in the SVNHC data aggregator, as they are available at the moment. RW thinks that geosciences are slower in the process of data aggregation, thus these data might be aggregated later on.

3.5. Vocabularies and taxonomic backbone

As backbone for taxonomic names the catalogue of life respectively the GBIF taxonomic backbone should be used.

As vocabulary, the biodiversity standards of TDWG should be used.

3.6. Persistent Identifiers for specimens

The persistent identifiers for specimens are central to make data findable, accessible, interoperable and reusable. They are also important to create metrics for the use of the specimen digitized. With that specimens can be properly cited in scientific publications and the use of them can be measured. In the further step it has to be defined, who is issuing the persistent identifiers for specimens.

3.7. Conclusions:

Many arguments have been brought up for an aggregator at the national level with all data aggregated in a Swiss data layer (this might be at GBIF.ch with PICTIS as start to be expanded).

GBIF.ch could be enforced with a strong webpage and different access portals to extract information. To do so, either GBIF hosted webpages could be used or GBIF.ch could do fitting the needs of the SVNHC.

For data standards, it should be started with Darwin core, ev. Expanded to ABCDEFG or have Darwin core covering information like ABCDEFG in the future.

One aggregator with many different outputs should be built. For communication, the technical background and the output should clearly be separated. This often causes confusion.

4. Results and discussion of the topic: definition of use cases

Paragraphs 4 and 6 of the agenda are summarized in this section.

Flipcharts produced during the discussions of the two break-out groups see paragraph 6 and 7.

4.1. SVNHC users and data access

Users of the SVNHC would be essentially researchers/curators. Researchers include researchers from sciences in a broad way, such as applied and fundamental sciences, citizen science, administration, industry etc. Scientific publishers might also be added to this list. All these users might need a high level of details about collections and / or specimens but they might need fairly different kind of data e.g., for population studies (distribution modeling), conservation (red lists), localization of invasive species, monographic works, etc. To map sequences or genomic data back to specimens might also be a use case. It would be important to extract measurements of scientific output made possible thanks to the SVNHC aggregator. AP informs that the University of Applied Sciences together with SWITCH have deposited a grant proposal to measure scientific output. This could be interesting for SwissCollNet purposes. **Second and third priority are the education and the media or more generally the public** where probably less detailed and more general data is required (for instance material for lectures such as pictures).

From a data user point of view, **one aggregator with many different outputs /displays** should be built. For a better communication, the technical background and the output should clearly be separated. If **most of the data should be open-access** some of the data might be sensitive (geographical location) or protected (copyright). Therefore, **in certain cases, data access should be partly restricted** but a contact person or owner (data provider) should be indicated to possibly give more detailed information (cf. section 3.1 Data model). Also, misuse of information should be avoided e.g., if data accessibility is important, actual physical availability of specimens should be available too.

Added value of the SVNHC and SVNHC data aggregator

- Accessibility / Visibility / Transparency

The SVNHC would not only enhance the visibility and accessibility of specimens but would also **enhance visibility of collections and institutions themselves** which could be especially valuable for small collections and institutions. The SVNHC would also **give to data users a single point of entry** instead of having to gather data out of each institution in Switzerland as they look for specific collections / specimens. If feedback possibilities are set up (cf. section 3.3 Data flow), the SVNHC would also give the opportunity to data users (citizen scientist, taxonomic group specialist) to **enrich or even correct the data** from collections.

- Networking and Standardization

As no data aggregator between collection holding institutions exists in Switzerland, the SVNHC aggregator would enhance the possibilities to launch new projects. The **networking between institutions at national (and international) level** would indeed be **enhanced** as it would be easier to know which institution has which collection, which reference specimens, etc. If a national portal would for instance be set up, the enhanced visibility would **self-dynamize the network**. The more data is shared, the more the data can be used, the more data will be shared.

The SVNHC aggregator would also increase the harmonization of the data, coordinate the flow of data (and material) facilitating data sharing (for instance to InfoSpecies and its network). This standardization would **enhance the understanding of the relevance of the data of collections** and might allow, depending on questions asked by data users, to get a quicker overview of the data and material available.

However, it has to be noted that curators as data providers are until now confused about the purpose of the SVNHC and about what is expected from them. **The added value of a data aggregator for data providers themselves has to be explained, supported and better communicated so that they invest time to provide their data.** It should also be considered that data have to be transferred directly from local CMS to the aggregator. Some money might have to be invested in local interfaces.

Moreover, **for the success of the SVNHC aggregator towards users and providers, it should not be an additional GBIF.ch.** Therefore, the aggregator should be developed at the national level with all data aggregated in GBIF.ch (cf. section 3.1 Data model). Then, it should be enforced with a strong webpage and different access portals to facilitate the extraction of specific information for specific use cases.

4.2. Added value for Geosciences?

The use cases for geosciences in Switzerland are not clear yet and have not been discussed sufficiently. The earth science community in Switzerland should discuss what are their needs in terms of collection data from earth sciences and what needs to be aggregated. Chemical and mineral composition of specimens are for instance important data. Storage and accessibility of specimen are others.

A project proposal will be deposited to SCN 2nd call for project proposals to integrate geosciences data into Specify at the CMS level. The question is then whether financial resources should be invested for geosciences in the middle layer of SVNHC data model (cf. section 3.1 Data model) and then the aggregated data exported to GeoCase or should the data be directly transferred to GeoCase?

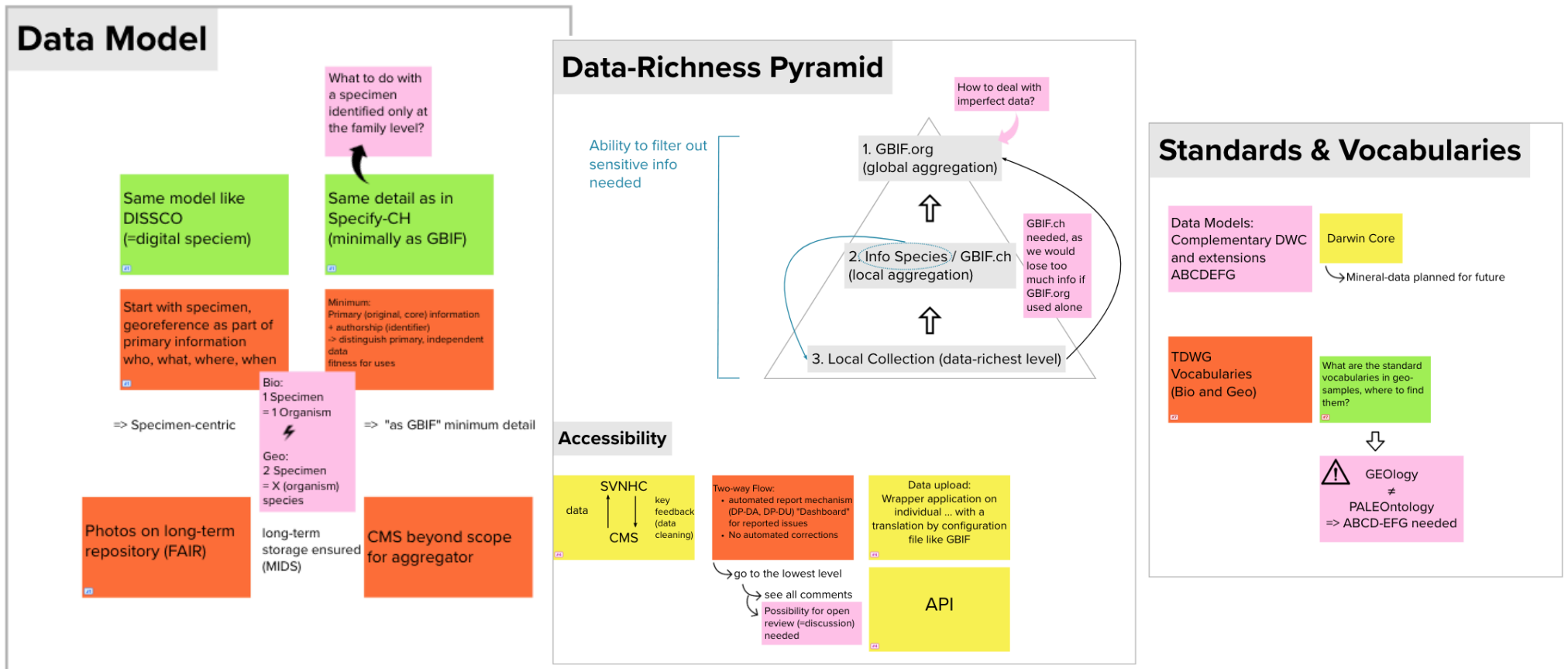
5. Conclusions, next steps

- Constitute a small strategic group with representatives of InfoSpecies and SwissCollNet. The group should meet soon (before the end of August) to work out a project plan and establish organization of the project (work packages). Members who have contributed so far (data management group, exchange meeting with InfoSpecies) can be involved in the project. For the financing of SVNHC and/or aggregator there is 1 Mio CHF available, an additional Mio as matching funds has to be found.
- Constitute a Geoscience Working group, to find out the needs and the technical solution for the standards. For geosciences, the corresponding governmental administrative office

to BAFU for biosciences is SWISSTOPO, which is in the department of defence. They might contribute as soon as data vocabularies and standards will be further developed.

Start 13h15/ End 17h00 JJo/YKo/PSt

6. Flipcharts: topic technical aspects for the SVNHC



7. Flipcharts: topic definition of use cases

