GBIF-Sweden: Scientific reporting 2012-2017

2018-03-31

Introduction

The Global Biodiversity Information Facility came into being as an intergovernmental initiative in March 2001. Two years earlier, science ministers from the Organization for Economic Cooperation and Development (OECD) had endorsed a recommendation from a scientific panel to set up a body "to coordinate…the standardization, digitization and global dissemination…of the world´s biodiversity data." GBIF has completed its early development phases and is now an operational infrastructure, recognized as the global aggregator for species occurrence data and as a leader in development of globally connected solutions for biodiversity information.

In January 2001 the Swedish Research Council (VR) was instructed by the government to organize the national membership in GBIF. While reserving the position as Swedish "Head of Delegation" in the international GBIF consortium for themselves, VR suggested that a national node be set up at the Swedish Museum of Natural History (NRM). The office of GBIF-Sweden was thus established in January 2003, and work began by an inventory of Swedish natural science collections and observation databases, and by initiating the coordination of digitization efforts at the relevant Swedish institutions.

Following two successive contracts covering 2002-2007, and a third five-year funding period 2007-2011 the presently reported contract spanning 2012-2016 was for the first time approved after an application process in open competition with other research infrastructures. For the period 2017-2021 an application was completed and approved where GBIF-Sweden made a constituent part of a national consortium - Swedish e-Infrastructure for Biodiversity and Ecosystems Research. Due to the failure of some of the consortium members to finally agree on the terms of collaboration, however the funding awarded was withdrawn by VR and GBIF-Sweden left to be financially temporarily upheld internally by its principal organization, the Swedish Museum of Natural History (NRM) until another round of applications was opened in 2017. GBIF-Sweden has become part of the new research infrastructure Biodiversity Atlas Sweden (BAS) and will during 2018-2020 collaborate with the Swedish LifeWatch infrastructure to form a joint national biodiversity information infrastructure. To anyway create a coherent picture of the development within the GBIF organization and GBIF-Sweden *per se,* events taking place in the year 2017 is also covered in this report.

GBIF International

GBIF currently consists of 42 voting participant countries, 12 associate participant countries and 36 participant organizations (Fig. 1). Since the start of the funding period 2012-2016 GBIF has evolved from the end of a pioneering phase in which the global community explored and defined the processes and models for making biodiversity data available into a fully operational infrastructure where the number of voting participant countries, associate participant countries and participants organizations has increased by more than 25%, the number of publishing institutions and occurrence data published has more than tripled and the publications based upon data mediated by GBIF increased roughly by 300%.

A completely new web-site was launched late in 2014 and was considerably updated in the fall of 2017. Recent advances in the services offered by GBIF include improved citation and persistence of data (enabling users to cite datasets and downloads consistently using DOI:s), consistent licensing for datasets (ensuring that all data accessed via GBIF are associated with a machine-readable, Creative Commons license), introduction of event-based data (for mobilization and discovery of biodiversity monitoring data using defined sampling protocols, including measures of species abundance), and streamlining the process of registering and endorsing institutions wishing to publish data through GBIF. These improvements all contribute to the firm establishment of GBIF in the mindset of biodiversity researchers, stakeholders and policy makers in conservation and natural resource management.

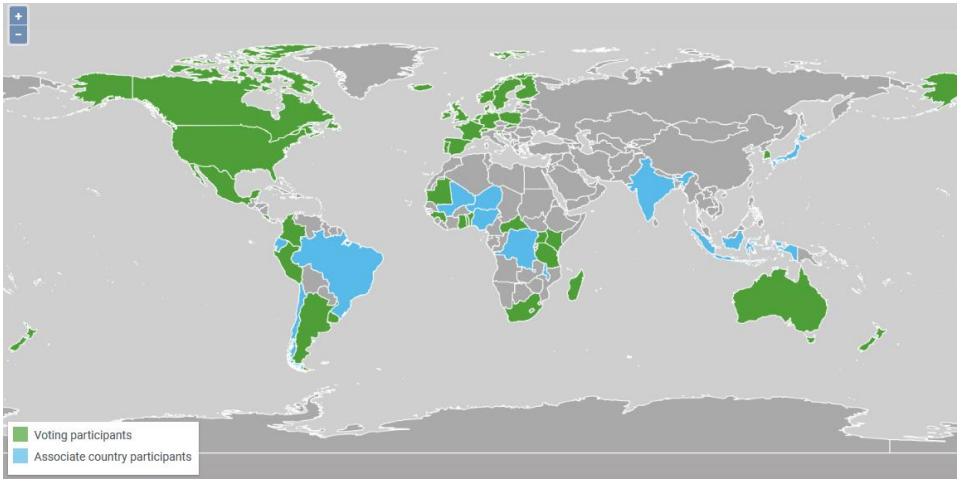

Figure1. The GBIF global network of participants (www.gbif.org).

GBIF-Sweden: Vision and ambitions

The overall vision is that GBIF-Sweden should serve researchers, public authorities and others interested in biodiversity as the Swedish hub for information and international exchange of data on the biological diversity of the world. The ambition of GBIF-Sweden is and has been to continuously harvest and publicly present up-to-date biodiversity data hosted by Swedish institutions and government authorities to stakeholders around the world through the international GBIF portal (www.gbif.org), in accordance with Swedish commitments under the international GBIF agreement. Furthermore, GBIF-Sweden has aimed to follow the strategic goals[1] set by GBIF international. More specifically these goals have been to strengthen the trust in GBIF data products, fill data gaps and broaden the evidence base both in terms of the completeness and coverage of GBIF data.

[1] http://bit.ly/2pnM3zy; http://bit.ly/2GGkTv7

GBIF-Sweden - By the numbers

GBIF-Sweden maintains and continuously extends a network of decentralized data publishers, from which information is collected and published through a machine interface, as well as for human consumption using a traditional web search interface at www.gbif.org and at the national web site http://bioatlas.se (in development). In 2015, the GBIF Secretariat began producing annual 'country reports' at the request of the GBIF Governing Board. After two years using this format, the Secretariat is now reevaluating the Country Reports concept in order to reduce the labor-intense production of these reports. The elaborate way by which the ´Country Reports´ represented statistics differs from the present mode of giving simplified figures. This is the reason for a lot of the statistics presented herein starting in the year 2016**,** which is when the last county report[2] for Sweden was constructed.

At the start of the funding period 2012, GBIF-Sweden provided ca. 33 million records to the global GBIF community. Today that number is more than doubled with ca. 78.6 million records available (March 2018). Sweden is currently the third largest provider of biodiversity data in the world (#1 United States offering ca. 345 million records, and #2 United Kingdom offering ca. 79.1 million records). This gives a record density of 208 records/km^2 (land and water surface), far above the goal set by GBIF of having at least 1 record/km^2. Altogether ca. 982 million records are presented by www.gbif.org (March 2018) and that number is expected to pass one billion during the spring of 2018.

The usage of GBIF-mediated data has drastically increased during recent years, both visible in the statistics of data downloaded from GBIF.org and from the number of scientific papers being published using GBIF-mediated data. Of the biodiversity data made available by GBIF-Sweden, an average of 1900 million records per month was downloaded during 2016 (2030 million per month during 2017). During 2017 the total number of downloaded records from GBIF.org was around 40 billion per month, an increase by ca. 40% from previous year. At the time of writing (March 2018) this number has further drastically increased to 127.6 billion downloads per month. During 2016 GBIF.org had on average 280 sessions per week originating in Sweden. During 2017 that number increased to a weekly average of 368 sessions originating in Sweden.

The number of peer-reviewed publications using GBIF-mediated data has steadily increased since 2008 when the metrics started being followed (Fig. 2a). During 2017 the number of publications took a big leap forward and two peer-reviewed publications per day are now being published (see the 2017 GBIF Science Review for further reading[3]). The number of peer-reviewed publications using GBIF-mediated data by Swedish author(s) has more than doubled during the period 2012-2017 (Fig. 2b). With the new literature tracking system of GBIF.org and the introduction of stable DOI:s for data downloads we expect the number of publications using GBIF-mediated data to increase further. Moreover, three of the 44 datasets mobilized by GBIF-Sweden are among the 10 most cited datasets at GBIF.org (38 504 datasets in total), highlighting the importance of our Swedish contributions.

[2] https://www.gbif.org/sites/default/files/gbif_analytics/country/SE/GBIF_CountryReport_SE.pdf
[3] https://www.gbif.org/science-review

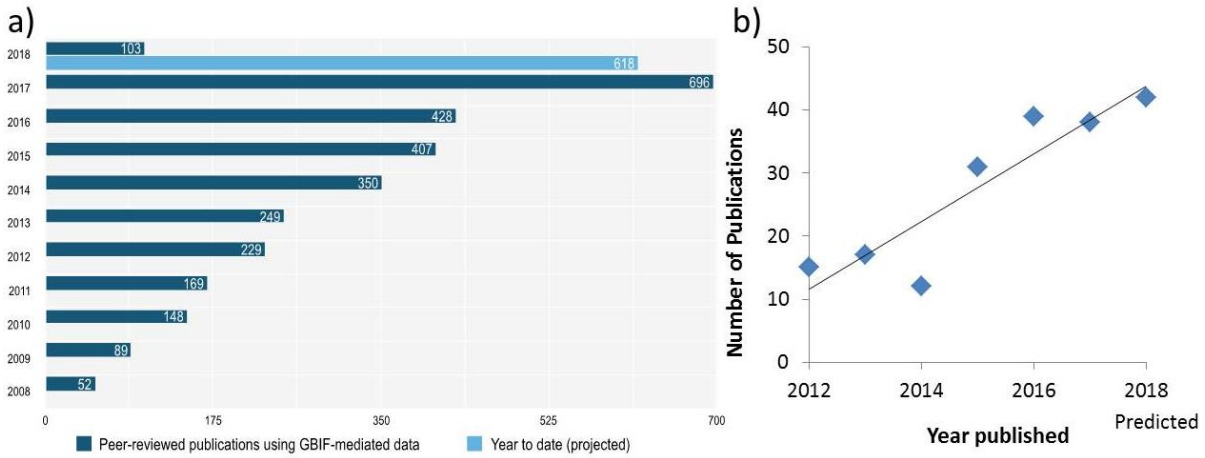

Figure 2. a) Peer-reviewed publications using GBIF-mediated data (through February 2018). b) Annual number of peer-reviewed publications using GBIF-mediated biodiversity data by Swedish author(s).

Focus areas during 2012-2017

1. Data publication

The general aim of GBIF is to create free and easy access to global biodiversity information and GBIF-Sweden has worked towards this goal by remaining a large provider of information to GBIF (Fig. 3). By tightening the links to fellow organizations, by running several developmental projects including potentially new data providers and users, and not the least by participating by its respective member organizations in the development, adoption and dissemination of the Atlas of Living Australia[4] (ALA) software in the Living Atlases Community[5], the prospect of gaining even stronger influence certainly looks promising.

During the five-year-period we have managed to make practically all digitally available Swedish natural history collection and observational data public at www.gbif.se and www.gbif.org – from late 2017 also at www.bioatlas.se. The time lag between digital data capture at the host institution and publication of the data at GBIF has decreased wherever possible. All major data provider´s datasets have been kept updated and published on a weekly basis. Furthermore, all datasets published by GBIF-Sweden have been fitted with appropriate Creative Commons licenses[6] (CC0, CC BY, CC BY-NC).

GBIF-Sweden has located and published a number of new datasets, assisting providers in data preparation, refining transport protocols and in data delivery. Machine harvesting using the Integrated Publishing Toolkit[7] (IPT) that allows efficient and frequent data transfer is employed and special technical solutions supporting efficient automatic data flow from DINA (Digital information system for natural history collections) developed at NRM have been implemented. We continue to support the development and adoption of web-based information systems allowing efficient and frequent data transfer to GBIF.

[4] www.ala.org.au/
[5] http://living-atlases.gbif.org/
[6] https://creativecommons.org/licenses/
[7] https://www.gbif.org/ipt

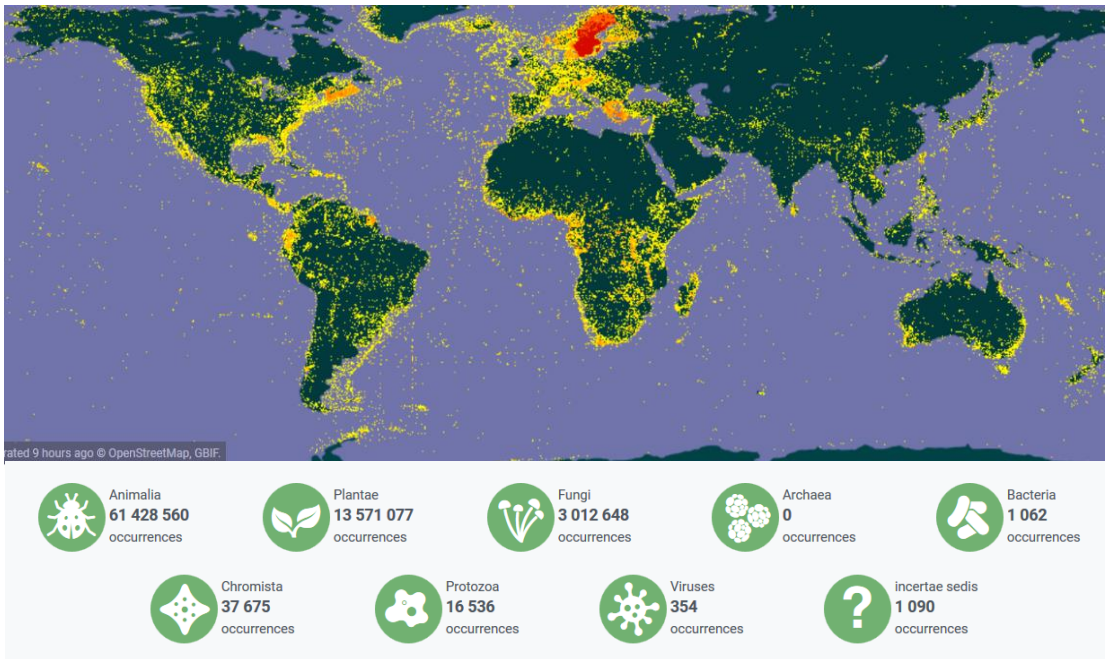

Figure 3. Top: Map of biodiversity data provided to the global GBIF community by GBIF-Sweden, currently consisting of ca. 78 million records covering 245 countries or areas (www.gbif.org). Bottom: Total data available for selected taxonomic groups in Sweden (https://www.gbif.org/country/SE/publishing).

In acting as a national partner in international biodiversity information standardization efforts (such as Biodiversity Information Standards, TDWG[8]), based on Swedish concerns and priorities, GBIF-Sweden has also supported standardization efforts and implemented international biodiversity information standards to Swedish data, which facilitate information exchange.

GBIF-Sweden has successively worked to introduce event-based data for mobilization and discovery of biodiversity monitoring data, a standard that was finally established in 2015. Several new event-based datasets await publication e.g. data from the Swedish Malaise Trap Project (SMTP) that represents a large-scale inventory of the Swedish insect fauna, handling an estimated total of 80 million specimens[9]. One of the general strategic goals of GBIF is to identify and fill taxonomic and geographic gaps in the data. From a gap-analysis performed of the data mobilized and provided by GBIF-Sweden we can conclude that insects and fungi are severely underrepresented in the repository (Fig. 4). Adding data from e.g. SMTP will thus partly address this goal and fill a knowledge gap in Swedish biodiversity data.

GBIF-Sweden has also worked towards introducing molecular based data, and as part of the national infrastructure BAS, extensive work has been devoted to adding various types of new data to the GBIF repository including massive-parallel sequencing (MPS) samples and tools which allow conversion of the sequence data into taxonomically identifiable data that can be imported into GBIF.

[8] www.tdwg.org
[9] http://www.stationlinne.se/sv/forskning/the-swedish-malaise-trap-project-smtp/

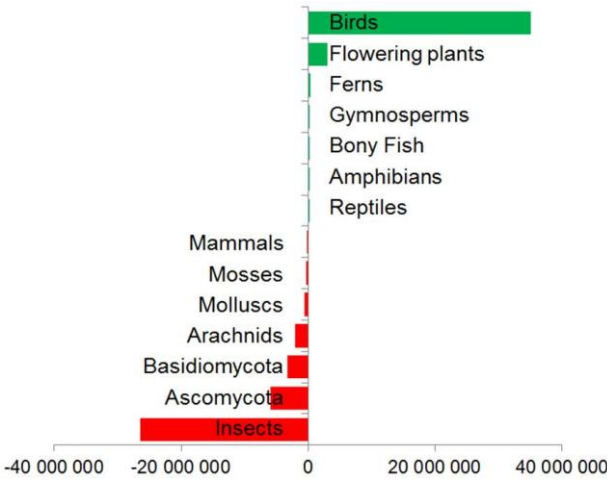

Figure 4. Taxonomic bias in biodiversity data mobilized by GBIF-Sweden. The vertical line at x = 0 depicts the 'ideal' number of occurrences per class, where each class is sampled proportionally to its number of known species. Green and red bars show the classes that are over- and under-represented compared to this 'ideal' sampling, respectively. Based on Troudet *et al*. 2017, doi:10.1038/s41598-017-09084-6.

2. Digitization effort

Specimen data from Swedish natural history collections are particularly valuable to scientists and to biodiversity management officials because they: (1) cover a historically very long time span; (2) offer more extensive coverage of species and collection sites than most other sources of biodiversity information; and (3) consist of actual preserved plants and animals, which provide material for further study. In total, it is estimated that the major Swedish natural history collection institutions hold some 33 million specimens, only just over 8 million of which have been digitized to some extent. Traditional digitization of museum specimens is labor-intense and unsatisfactorily slow.

To address the issue of rationalizing and increasing the speed of digitization of museum specimens, GBIF-Sweden has spent considerable effort during 2012-2016 at evaluating suitable ways to rationalize and speed up the digitization of museum specimens. Two subsequent proposals have been submitted (but rejected) for ambitious projects addressing this situation.

Suitable national and international partnerships with similar objectives in mind have been explored (The Swedish Media Conversion Centre, National Archives; Digitarium in Finland; Naturalis in the Netherlands) and the new BAS project will include technical innovation and ways to implement existing techniques for industry-scale mass digitization and crowd-sourcing, with the aim of digitizing the bulk of the backlog within a limited time period (15 years).

We have supported data providers with advice and assistance in minor but crucial digitization efforts. Since many old specimens (even those already presented by GBIF-Sweden) are not accurately georeferenced we have encouraged and supported the development of methods to add spatial accuracy to biodiversity data, e.g. during a national conference in 2016 on data validation tools organized in combination with the Biological Museum in Lund and the Swedish Botanical Society.

3. Building a national biodiversity research infrastructure

Driven by scientific opportunity and societal challenges, biodiversity and ecosystems research are rapidly developing into big-data sciences, focused on modeling processes that affect entire biotas and predicting system-wide effects of environmental change. GBIF-Sweden specifically, and the GBIF community in general span the entire spectrum of biodiversity data mobilization, harvest, presentation and analysis throughout its history.

Alongside preparing and publishing biodiversity data as described above, the process of connecting various providers and users has been a major effort during 2012-2016. Relating to this long-term work we are now in position to establish a working national community. During all previous project phases GBIF-Sweden has been the key driver and facilitator of all these activities in Sweden, activities which will now continue and become part of Biodiversity Atlas Sweden (BAS). Open data and open development has been the underlying principle for GBIF-Sweden since the start in 2001 and we have during the last couple of years been working on preparing a national shift to open-source software developed in international collaboration, especially within the Atlas of Living Australia (ALA) and the GBIF communities. This has now become a reality with funding secured for the national research infrastructure BAS. GBIF-Sweden will as a part of BAS and in collaboration with Swedish LifeWatch continue to contribute to the Swedish biodiversity research in providing a platform for collaboration, rich quality data and the best possible toolbox for visualization and analysis of biodiversity data.

Swedish contributions to ALA/GBIF development will focus on streamlining collections and observational occurrence data, but in adding systematic monitoring data, tracking data, molecular biodiversity data and data on microdiversity (prokaryotes, unicellular eukaryotes, and microscopic fungi) the national Swedish GBIF node will remain instrumental in the biodiversity landscape nationally and internationally.

4. Web portal

The Swedish web portal www.gbif.se has during the main project period been the national site for users of services provided by GBIF-Sweden. Developed by GBIF-Sweden, the earliest version of the portal was launched in 2007, and a new portal was published in November 2014. During 2017 the portal was run in parallel to the portal currently in development within the BAS consortium (https://bioatlas.se/), which will come to replace gbif.se during 2018. The new web portal of GBIF-Sweden (as part of BAS) is extended with components and functions offered by Atlas of Living Australia (ALA). The ALA system is already in use a several other national GBIF nodes worldwide and additional countries are currently considering implementation[10].

The new portal already includes core functionality similar to that of gbif.se. There is also a web-based front-end for the research community (Mirroreum[11]) enabling reproducible research. This front-end support usage of a wide range of tools used in the research community including a curated set of widely used R packages and tools for biodiversity analysis, such as ALA4R, rgbif and taxize. ALA further comes with powerful tools e.g. a graphical web user interface for integrated analyses (Spatial Portal) and a crowdsourcing tool (DigiVol) to recruit the help of citizen scientists in digitizing Swedish natural history collection as well as a mobile app for capturing biodiversity data (BioCollect).

5. Education and outreach

The usage of GBIF services continues to increase and at present (March 2018) the number of scientific publications based on data provided by GBIF surpasses 600 per year. However, the potential is much larger and Swedish scientists are still poorly represented among GBIF users, especially given the large amount of data contributed by GBIF-Sweden. The continued success of GBIF rests heavily upon presentation of data and services to potential users.

[10] https://www.gbif.org/programme/82953/living-atlases
[11] https://bioatlas.se/mirroreum/

To promote the use of GBIF data we have approached university institutions and other potential user groups to inform about the services GBIF-Sweden provides, taken part in and organized education events. GBIF-Sweden has also participated in biodiversity informatics courses, workshops and symposia for graduate students, researchers, and other potential users of GBIF resources. Other activities have included participation in national or international discussion fora such as within the GBIF organization, in TDWG and in the Ecology Education Forum of the European Ecological Federation.

6. Support to biodiversity informatics in developing countries

A large fraction of the data handled by GBIF-Sweden consists of information on organisms observed or collected in foreign countries, particularly in the tropics. In many countries, access to data and information on indigenous biodiversity is restricted and repatriation of such information is among the most important tasks of GBIF internationally. Simply by making data on foreign biodiversity hosted by Swedish institutions available through the GBIF portal, GBIF-Sweden is contributing to this goal. GBIF internationally and GBIF-Sweden have the ambition to transfer knowledge technology in biodiversity informatics to developing countries, where requested. In addition, representatives of GBIF-Sweden have on several occasions participated in training and exchange programs connected to the BID-program[12] and mentorship discussions have been initiated with colleagues from Ethiopia.

[12] https://www.gbif.org/programme/82243/bid-biodiversity-information-for-development