Swedish LifeWatch ─ A Biodiversity Infrastructure Integrating and Reusing Data from Citizen Science, Monitoring and Research

ULF GÄRDENFORS, Swedish Species Information Centre, Swedish University of Agricultural Sciences (SLU)

MARI JÖNSSON, Swedish Species Information Centre, Swedish University of Agricultural Sciences (SLU)

MATTHIAS OBST, Department of Biology and Environmental Sciences, University of Gothenberg

ANNA MARIA WREMP, Swedish Species Information Centre, Swedish Univ. of Agricultural Sciences (SLU)

OSKAR KINDVALL, Swedish Species Information Centre, Swedish University of Agricultural Sciences (SLU)

JOHAN NILSSON, Swedish Species Information Centre, Swedish University of Agricultural Sciences (SLU)

ABSTRACT

With continued pressure on biodiversity and ever-growing conflicts with human development, qualified systems for scenario modelling, impact assessment and decision support are urgently needed. Such systems must be able to integrate complex models and information from many sources and do so in a flexible and transparent way. To that end, as well as for other complicated and data-intensive biodiversity research purposes, the concept of LifeWatch has emerged. The idea of LifeWatch is to construct e-infrastructure and virtual laboratories by integrating large data sources, computational capacities, and tools for analysis and modelling in an open, service-oriented architecture. To be efficient and accurate, a continuous inflow of large quantities of data is essential. However, even with new techniques, government-funded monitoring data and research data will not feed the system with up-to-date species information of sufficient scale and resolution. To fill this void, skilled amateur observers (citizen scientists) can contribute to a very valuable extent.

After a preparatory phase, a Swedish LifeWatch (SLW) consortium was initiated in 2011. Swedish LifeWatch developed an infrastructure where all components are accessible through open web services. At the SLW Analysis portal, different formats of species and environmental data can be accessed instantly, and integrated, analysed, visualized and downloaded at selected

temporal, spatial or taxonomic scales. Swedish LifeWatch currently provides 46 million species observations from eight different databases, all harmonized according to standardized formats and the Dyntaxa taxonomic backbone database. Almost 40 million of these observations were provided by citizens through the online reporting system named the Species Observation System (SOS) or Artportalen. This paper describes this system, as well as the incentives that make it so successful. The citizen science data in the SOS are accessible, together with data from research and monitoring, in the SLW infrastructure, making the latter a powerful instrument for large-scale data extraction, visualization and analysis.

1. **INTRODUCTION**

Ecosystems are under enormous pressure worldwide and are experiencing rapid declines in biodiversity (Rockström et al., 2009, Pimm and Raven, 2000). In response to these changes, a large number of national and international conservation and management programmes are being launched to assess ecological integrity and help establishing sustainable ecological conditions. These initiatives are central instruments for the implementation of international commitments and legislations, such as the Convention on Biological Diversity and its associated Aichi biodiversity targets (http://www.cbd.int/sp/targets/), or the European Biodiversity Strategy for 2020 (COM(2011)0244). At the same time, numerous science-policy bodies have been founded (e.g. the Intergovernmental Platform on Biodiversity and Ecosystem Services or IPBES[1], the GEO Biodiversity Observatory Network or GEO BON[2]) to support governmental decisions for halting biodiversity loss, improving the status of biodiversity and weighing human expansion and vital biodiversity through societal planning, and remediating the effects of biodiversity loss. To be effective, however, any natural resource management programme or decision support body must depend on their underlying system for critical information on patterns and processes of biodiversity, which ultimately allows for analysis of biodiversity data from all relevant sources.

Today, there are many successful attempts to gather biodiversity data. Globally, there are already more than 680 known bioinformatics projects[3]. Species observation records for example are typically collected in national or regional databases and then further aggregated on a global scale, such as through the Global Biodiversity Information Facility (GBIF)[4]. In this process, national research councils, international directives (Infrastructure for Spatial Information in the European Community or INSPIRE[5]) and initiatives, such as the Research Data Alliance (rd-alliance[6]) or the

---

[1] http://www.ipbes.net
[2] http://www.earthobservations.org
[3] http://www.tdwg.org/biodiv-projects/projects-database
[4] http://www.gbif.org
[5] http://inspire.ec.europa.eu
[6] https://rd-alliance.org

Bouchout declaration[7] play important roles by enforcing data standards as well as controlled and open access to biodiversity information, which in turn enables contextualization of national and regional data in a global context. Although large amounts of valuable information is already organized in this way, the majority of biodiversity data resources are still not accessible through national reporting systems that can be globally connected. Notably, research data which are typically of high quality and expensive to produce often reside in small and local databases with inaccessible and incompatible interfaces, and often with a limited lifetime. Capturing such data offers considerable advantages as these data can be re-used in a different context as well as integrated with data from biodiversity monitoring programs. Web-services that facilitate access and sharing of large collections of data from primary resources, in a common structure, is an active and evolving field (Constable et al., 2010, Hardisty et al., 2013).

However, even if all existing data from biodiversity monitoring and research programs would be accessible to scientists for analysis today, there would still be large gaps in the spatial, temporal and taxonomic coverage across the globe (EU BON 2014). This creates a principal problem that obstructs our holistic understanding of the processes that connect all actors in an ecosystem. Therefore, new and cost-effective technologies need to be developed, based on automated and human resources, to increase the quality and amount of biodiversity data in natural and cultural ecosystems. Here, keen amateurs, so-called "citizen scientists", have the potential to become a valuable resource in gathering and processing large amounts of relevant biodiversity research data. In the context of cost-effective biodiversity monitoring systems, a citizen system is especially helpful where automated observation systems and biodiversity monitoring programmes cannot be used, or where they are too expensive (e.g. Silvertown, 2009). Citizen science projects have expanded in number and scope as a cost-effective way to undertake large-scale and/or high-resolution surveys and to connect scientific research to public outreach and education (e.g. Dickinson et al., 2010, Dickinson et al., 2012, Bonney et al., 2014). Access to the Internet and advances in information technologies have also made citizen science projects more visible and accessible, allowing interested citizens to participate in data collection, storage and application (e.g. Silvertown, 2009, Hochachka et al., 2012, Bonney et al., 2014). Many of the barriers relating to data quality control (e.g. sampling bias and measurement errors) during the compilation and analysis of large complex data sets can be handled through novel statistical and computational solutions (Hochachka et al., 2012, Bird et al., 2013). Besides issues such as data quality control, there is a challenge in motivating people, not only to spend a lot of time observing species in the field, but also to enter their data in a repository – without payment. For citizen science to reach its full potential, strategic investments and coordination efforts are needed aiming to organize, synthesize and analyze already available repositories of citizen science data (Bonney et al., 2014).

In this paper we will describe an effort to (i) substantially enhance the amount of accessible biodiversity data by activating keen amateurs to report their species observations through a win-win biodiversity informatics system, the Species Observation System (SOS) (also called "Artportalen"); (ii) develop and coordinate web services which make biodiversity data of

---

[7] http://www.bouchoutdeclaration.org/

different origin and purpose (i.e. citizen science, research, government monitoring, museum collections) accessible in coherent formats; and (iii) provide tools for exploring, analysing and visualizing biodiversity data. The latter two initiatives are done in the context of the Swedish LifeWatch (SLW) project.

2. COMPONENTS OF SWEDISH LIFEWATCH

2.1 The LifeWatch Concept

The central mission of the European initiative LifeWatch[8] is to develop an infrastructure that facilitates the integration of large and complex data sources and provides advanced capabilities for research on complex biodiversity systems (Los and Wood, 2011). The intention is to achieve this through interconnecting primary data repositories and creating e-services and virtual laboratories in conjunction with these. The concept is part of the European Strategy Forum on Research Infrastructures (ESFRI) roadmap (http://ec.europa.eu/research/infrastructures) and was established through a Seventh Framework Programme for Research and Technological Development (FP7)-funded preparatory phase (2008–2011). At the European level, the project is still in the beginning of its construction phase. A few countries, like Belgium[9] and Greece[10] have started technical development on a national level, while several countries await the formation of a formal European Research Infrastructure Consortium (ERIC). At the end of 2014, eight European countries are expected to submit a proposal to the European Commission to create such an ERIC, where Spain will uphold the statutory seat, Italy the Service Centre, and the Netherlands the lead development of virtual laboratories. Sweden will initially participate as an observer, but plans to join the ERIC as soon as possible.

Sweden started the construction of its national LifeWatch infrastructure in 2010, connecting four universities (the Swedish University of Agricultural Sciences, Uppsala, and Lund, Gothenburg and Umeå Universities), the Meteorological and Hydrological Institute, the Board of Fisheries, the National History Museum, and the Swedish GBIF node. These bodies formed an SLW consortium[11] with the Council for Research Infrastructures at the Swedish Research Council as the main financier. The conceptual construction of the SLW largely follows the Reference Model developed during the LifeWatch preparatory phase.

2.2 Web Services and Taxonomic Backbone

A central feature of the SLW infrastructure is providing seamless access to biodiversity data through web services, enabling all national information resources to be integrated and analysed in

---

[8] http://www.lifewatch.eu
[9] http://www.lifewatch.be
[10] http://www.lifewatchgreece.eu
[11] www.swedishlifewatch.se

many different contexts. When it comes to species data, SLW gives access to a number (currently eight) of data repositories through the web service Swedish Species Observation Service (not to be confused with the SOS, being one of the eight repositories). For example, all (non-sensitive) data are continuously provided to the Global Biodiversity Information Facility through this web service. The data stored in the underlying repositories range in content and scope, from research and monitoring data to citizen science data (Fig. 1).
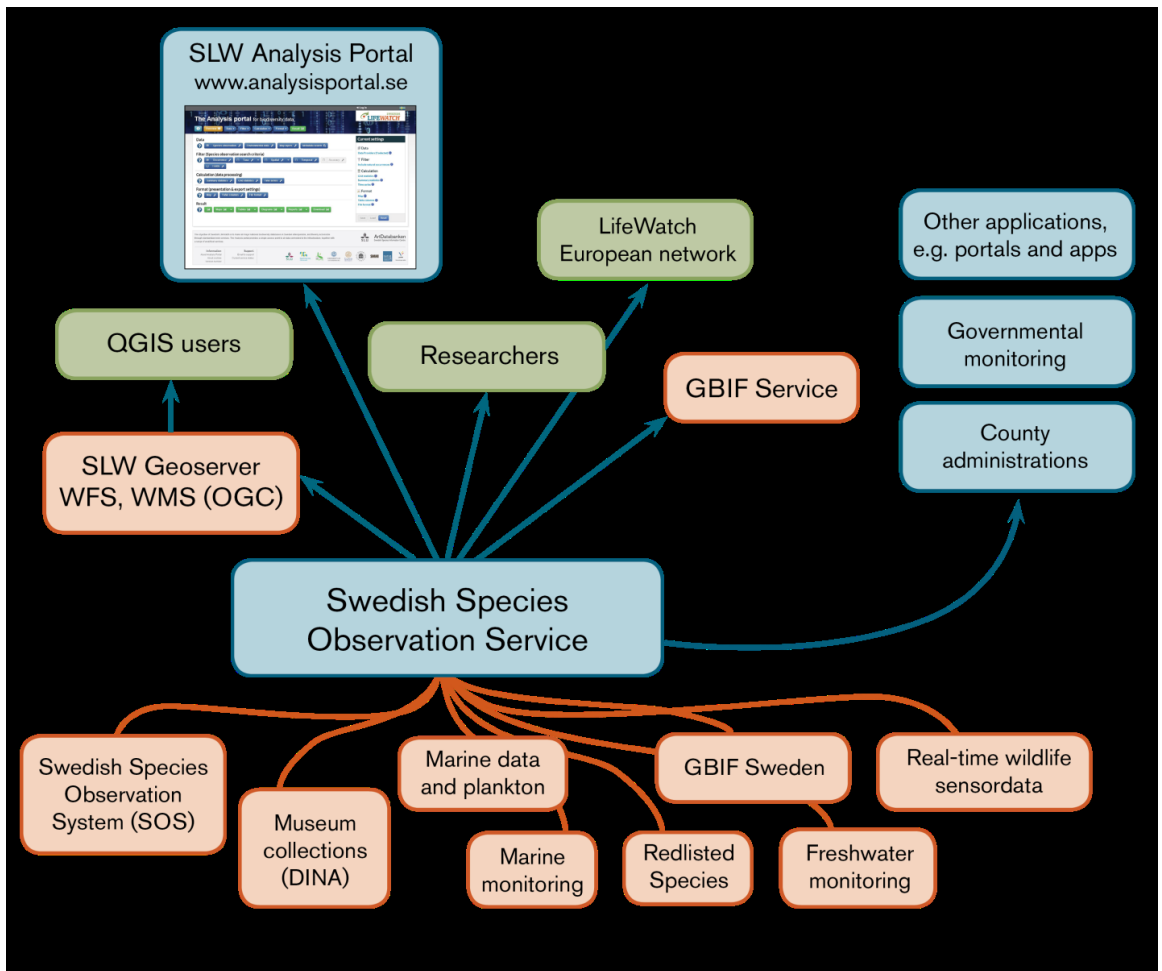

Figure 1. *Data repositories and web services dealing with species observations in the Swedish LifeWatch infrastructure*

Integration of the data requires not only standardized tables (using, e.g., Darwin Core) and web service formats (OGC, WFS, SOAP, REST), but also a taxonomic backbone for all taxa whereby every record of species data can be referenced (Patterson et al., 2010). Therefore, we developed a

taxon concept-based taxonomic database including an application called Dyntaxa[12]. This taxonomic backbone builds on globally unique identifiers (GUIDs) and safeguards that the names presented are unambiguously connected to a defined taxon and, further, that each taxonomic entity is represented by only one valid name. It keeps track of and stores historical taxonomic changes, such as splitting and lumping, over time. Therefore, there is no need to taxonomically clean the material presented by web services. Dyntaxa today covers more than 95% of the estimated 60,000 multicellular species known to occur or to have been observed in Sweden. Therefore, it is by all comparisons the most comprehensive source of taxonomic identities for Sweden, and Dyntaxa can supply GUIDs with data on almost all taxa.

Even though it is a national database with initially nationally created life science identifiers (LSIDs) for the taxa, the aim is to successively match the concept of each taxon to the concepts ─ and consequently GUIDs ─ of the Pan-European Species Directories Infrastructure (PESI). Thereafter, PESI GUIDs will be used as the recommended taxon identifiers. As a consequence, the Swedish taxonomic backbone will successively merge into the European database, and hopefully eventually into a global one, such as the Catalogue of Life[13]. Still, the content of a taxonomic backbone can never be static, but must continuously be updated. Currently, the Swedish Species Information Centre has three full-time employees curating content.

2.3 Species Data Sources

An array of data repositories are already included in the SLW infrastructure (Fig. 1), covering terrestrial, fresh water and marine habitats, and organisms ranging from microalgae to plants, marine invertebrates to fish, and collembolans to large mammals. In November 2014, the Analysis portal provided over 46 million species records of almost 29,000 different species. The Species Observation System (section 3.4) alone provided >43 million of these. The numbers will rise substantially in the near future when more data and further data providers become harvested by SLW. These repositories provide access to large heterogeneous data of mixed origin and quality, from monitoring and research data to museum collections and citizen science data.

One important data source is the common database for all Swedish natural history museums, currently under development, named the Digital Information System for Natural History Museum Collections (DINA). This system will provide a major source of legacy data from all museums as collections that are digitized, geo-referenced and imported to the infrastructure. It will allow researchers to compare and analyse current biodiversity data against historical data. Such comparison and analysis is especially helpful for assessing cumulative and long-term anthropogenic impacts on the environment and to develop realistic conservation targets in the face of growing environmental change.

---

[12] http://www.dyntaxa.se
[13] http://www.catalogueoflife.org

2.4 **The Species Observation System – An important Citizen Science Tool**

The SOS (or "Artportalen", art meaning "species" in Swedish)[14] is by comparison the most important repository of SLW. It set out in the year 2001 as a reporting system for bird observations, at that time administered by the Swedish Environmental Protection Agency. In 2004, the system was transferred to the Swedish Species Information Centre and successively developed into a reporting system for all multicellular organisms. In 2007, the Norwegian Biodiversity Information Centre set up a copy of the system in Norway where it also quickly became a very popular tool for reporting species observations.

The majority of the reports are submitted by skilled amateur naturalists (herein rapporteurs). However, the system is also used as a repository for county administration inventories and, to a growing extent, for government monitoring programmes and research-generated data. For instance, the Swedish Environmental Protection Agency and the county administrative boards today demand contractors of inventories to deposit their data in the SOS. From 2012, the Swedish Research Council Formas requires that every scientist receiving financial support must present a plan for data storage, with the SOS being one possible repository.

The mandatory variables in the SOS databank are species name, geo-referenced location, time of observation and who reported the observation. Besides these fields, there are a number of optional variables that can be reported, such as the number and age of individuals observed, activity of the species, habitat, observation methodology, who else observed the species, who has confirmed the identity, whether it has been incorporated into a collection, and description in free text, etc. The restricted number of mandatory variables means that contributing to the system is not perceived as arduous. Still, most rapporteurs choose to enter data for several more variables than those that are mandatory.

The SOS builds on the principle of "publish first, validate afterwards". Verification is conducted in several ways. Firstly, the name cannot be misspelled since it must follow the taxonomic backbone (Dyntaxa). The geolocation is decided through identifying the spot on an online map or by entering GPS coordinates. Secondly, through exposure on the Internet to tens or even hundreds of thousands of eyes, obvious mistakes or dubious species identifications are normally quickly spotted and reported back to the person who uploaded the observation. For many species groups, experts also work voluntarily as validators, checking in particular records of species that are rare or hard to identify or that fall outside the expected occurrence pattern, spatially or temporally. Here, uploaded pictures are an important source for checking the identification.

With the exception of data concerning sensitive species, according to a classification regulated by the Swedish Species Information Centre, all data are totally open to view and download at no cost. It is only the rapporteurs who can delete or change any primary data. However, validators of

---

[14] http://www.artportalen.se

the system can classify and flag the observations upon verification and approval. When a validator finds a dubious identification or an obvious misidentification, the rapporteur is asked to correct the observation. The validator can also hide low-quality observations for others to see and download. This type of citizen science data, verified by experts or community peers through human computation processes, has been shown to be a cost-effective way to provide accurate estimates of species richness and diversity values (Gardiner et al., 2012).

Each day, 10−15 000 new species observations are submitted to the SOS. It has about 20,000 rapporteurs and 370,000 unique visitors every year (in a country of less than 10 million inhabitants). About 95% of the uploaded observations concern common and widespread species. About 80% of all data (92% of bird data) reported to SOS are entered into the system within 24h of the observation. Therefore, the site provides news about interesting or rare species that other naturalists can follow and experience themselves.

The SOS also provides lists of which among the rapporteurs have reported the highest number of species of various groups in different areas and time periods. The success of the system has led to increasing reports of older observations, which have been extracted from observation books and collections. Even though the system started for birds in 2001 and for other species in 2003, 8.6 million records prior to 2001 have been registered, providing a valuable reference of earlier distribution and relative abundance. Taxonomically, the system today covers most organism groups and altogether 28,660 different species from different terrestrial, fresh water and marine environments reported thus far in the SOS.

Many rapporteurs also upload pictures to the system. Currently, the system contains 600,000 pictures of 14,400 different species. These can be used for verification but also as reference for other users regarding species appearance. Therefore, the large and growing collection of pictures serves as a reference library for the identification of different species. The system also offers a selection of various visualization tools, such as thematic maps, diagrams and summary statistics. Users can use these to explore, e.g., species' patterns in space and time, and also keep a personal diary about their observations.

2.5 Analysis and Visualization – The Analysis Portal

Since all data and tools of SLW are presented as web services, SLW's facilities can be combined into different forms of e-services and virtual laboratories. For example, we developed an analysis portal[15] (Fig. 2.) where species data can be filtered and combined with environmental data (as web feature services) and thematic maps (as web map services). Through catalogue services for metadata, the user can find and explore further available data in various areas and formats. The first version of the portal was launched in December 2013 and is continuously extended with additional functionalities.

---

[15] https://www.analysisportal.se/

Figure 2. *The Swedish LifeWatch (SLW) Analysis portal (https://www.analysisportal.se)*.

The Analysis portal allows the user to select one or several taxa in any combination. Taxa can be selected by entering the name or be selected based on their classified traits, such as life history, habitat preferences and pressuring factors, according to a traits database maintained at the Swedish Species Information Centre. Taxa can also be selected by drawing a polygon on a map querying all databases about what taxa have been registered within them, or by asking what taxa from a certain time period or a certain habitat are represented in them, or any combination of such variables.

Based on selected taxa, area and/or time, species and environmental data can be accessed and downloaded. Species data are selected from the eight currently connected data repositories. Environmental data are accessible through web feature services or presented in web map service format. Data can be aggregated on any grid scale (Fig. 3), set of map polygons and time period. Differences in sampling efforts can be adjusted for using specific abundance indices. All data can be explored or visualized in various forms of tables, diagrams, maps, or be downloaded for further analyses, using any statistical software and geographic information system (GIS) tools. Besides, data can feed analytical workflows at servers in the global infrastructure, such as the

Biodiversity Virtual e-Laboratory[16] and the upcoming e-laboratories at the European Biodiversity Observation Network (EU BON)[17] and European LifeWatch[18].
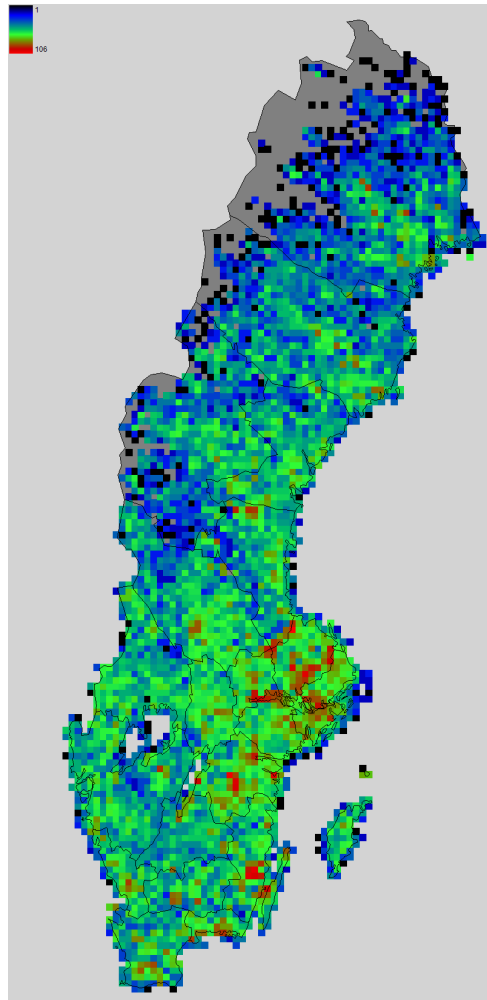

Figure 3*. A heat map showing the number of species connected to aspen (Populus tremula L.) as a substrate in Sweden, presented per 10x10 km2, according to data in the Swedish LifeWatch (SLW) Analysis portal.*

As examples, the user could (1) investigate which species have been observed within a specific area (polygon) or location; (2) produce grid cells (heat maps) where the highest number of

[16] http://www.biovel.eu
[17] http://www.eubon.eu/
[18] http://www.lifewatch.eu

species have been observed; (3) analyse the correlation between the occurrence of red-listed species and protected areas; (4) analyse the temporal correlation between species occurrence and weather factors; (5) analyse the correlation between spatial occurrence and habitat or soil characteristics; (6) analyse the phenology of different species, including activities over the year (e.g. shown by week or month); and (vii) investigate how the average arrival day for migrating birds has changed over time. Users could also develop predictive models or workflows, answering more complex questions concerning species-environment relationships. Finally, because the entire infrastructure is based on open web services, skilled users could develop their own applications for targeted research including apps for smartphones, e.g. showing what species currently can be seen or expected at a particular place.

2.6 Costs

Substantial resources have been and are continuously invested by the Swedish government and Research Council into the SLW biodiversity e-infrastructure. The Species Observation System has a yearly budget of c. 1.3 M€ which includes development, maintenance, and support. The yearly budget of the taxonomic database Dyntaxa is c. 0.3 M€, of which most is used for upholding and enhancing the content quality. The SLW web-services and Analysis portal, including development, maintenance, support and communication, has a yearly cost of c. 0.7 M€. Augmenting these resources are additional resources invested by the Swedish Species Information Centre into upholding a traits database (e.g. Fig. 3). Still, from a societal perspective this infrastructure is quite cost-effective. Not only are huge quantities of species observations gathered for free into SOS, but county administration boards and municipalities that used to uphold local databases today use the infrastructure for data storage and harvesting in their daily work, e.g. extracting critical information during conservation planning, environmental management and impact assessments. Researchers get easy and free access to large amounts of novel biodiversity data and tools for exploring and visualizing such data together with other environmental data. Dyntaxa is used as a central taxonomic backbone by different databases (including the Swedish Wikipedia).

3. DISCUSSION

The prime aim of the SLW infrastructure, including its analysis portal, is to promote biodiversity research and contribute to Swedish biodiversity research having world class quality. Nevertheless, the SLW infrastructure is designed to make biodiversity data and information accessible also to other user groups, such as practitioners, policy makers, and ordinary citizens. To address the urgent questions around biodiversity conservation, environmental change and land-use, we need an open and accessible systems approach that moves beyond taxonomy and species observations (Hardisty et al. 2013). The SLW infrastructure is such a novel bioinformatics approach; it allows users to address not only species, but also their interaction with a broad set of environmental variables and other species, as harvested from multiple primary data providers.

Representing a novel effort to organize, synthesize and analyse already available repositories of large heterogeneous biodiversity data, the development of the SOS and the SLW infrastructure offers a unique opportunity for sharing and discussing "lessons learned". In this section, we focus foremost on lessons learned during the development of the SLW infrastructure for integrating and reusing citizen science data. To some extent, we also discuss our experiences of working with the development of new web services and our approaches to data handling and citizen engagement. These experiences can be transferrable to other citizen-based infrastructure projects.

The SLW concept is of general applicability in other countries or geographical regions. To transfer or enlarge the infrastructure to other geographical areas would require solutions for handling taxonomy, maps and geospatial issues, administering the authorization of admission to certain sensitive taxa, as well as provisioning of user and technical support in such regions. These challenges are achievable, but should not be underestimated, and are preferably handled by a body with long-term funding and a stable organization.

3.1 Enhancing the status of biodiversity data through the Species Observation System

Skilled and passionate amateurs can contribute both data and information to the SOS and make use of it for their own interest and/or educational benefit. Even if the majority of rapporteurs use the SOS for their own benefit, many are also enthusiastic about contributing to both research and conservation (e.g., environmental impact assessments, protected area planning). For example, amateurs who are passionate about biodiversity conservation issues can locate a relevant citizen science project (e.g., become so-called "flora and fauna guardians" and monitor the presence of rare plants and animals in Sweden), follow its guidelines, submit data directly to online databases, and join a community of peers. Already, some 90% of all species occurrence data in the SLW database infrastructure are uploaded by amateurs. This suggests that non-professional citizens are extremely important biodiversity data providers. In this respect, Sweden is fortunate in being able to provide a coherent system where citizen science data can be stored, maintained and accessed, rather than entered into casual databases or files, or not stored at all.

We believe that the success of the SOS, not the least among amateur naturalists, is a result of a combination of factors:

1. The observations are instantly visible and accessible on the Internet upon submission. Observations are published first and validated afterwards, resulting in up-to-date, newsworthy reports.
2. The system does not require a load of mandatory parameters or a particular observation methodology, meaning both that the threshold for reporting is comparatively low and that the user apprehend that the reporting is on the observers premises.

3. More than 90% of all records are totally open and free for viewing and downloading. The user can easily find study locations for both rare (including occasional vagrant birds or insects) and more common species.

4. The system keeps track of how many species each observer has seen over different time periods and regions, and who has observed the greatest number of species. This element of personal record keeping and competition keeps the rapporteurs motivated and triggers further participation.

5. Photos of species observations can be uploaded and commented on by the community. Photos can be used for data validation and act as a reference on species appearance. Progressively, this large and growing collection of photos serves as a reference library for the identification of different species.

6. The system offers a selection of visualization tools, such as thematic maps, diagrams and summary statistics. The user can explore a species' patterns in space and time or get statistics for personal observations.

7. The system provides a personal diary where the user can make notes, e.g. noting excursion details and weather conditions.

8. The system is apprehended as a tool for the user, rather than being a system where you put a lot of energy and time for somebody else's purpose.

Taken collectively, the SOS clearly provides benefits both to individual citizens (such as recording and keeping track of own observations, providing an arena for personal exposure and networking, providing status and credit within the community) and the society at large by being an online database for biodiversity data used by government organizations, researchers, etc.

The vast number of biodiversity informatics projects that have been implemented worldwide share many characteristics, such as open access, taxonomic backbones, integration of resources, and service orientation (e.g. providing maps, information, lists, summary statistics), but differ substantially in their architectures and technological approaches (e.g. Hardisty et al. 2013). A challenge for the future is to develop a common standard for web services that integrates ever increasing volumes of biodiversity data and at the same time deliver immediate benefits to the user, hiding the majority of the technology behind user interfaces (Hardisty et al. 2013). Common to successful citizen-science programs is that they facilitate long-term community-level engagement and activities, use appropriate cyberinfrastructure, have diverse goals and evaluation strategies, maintain financial stability, and effectively disseminate results (Bonney et al. 2009).

In cooperation with the Norwegian Biodiversity Information Centre, and with funding from the Swedish Environmental Protection Agency and the Norwegian Ministry of Environment, a new generation of the SOS has now been developed. Hitherto, the SOS has been comparatively static in the number and formats of variables that can be reported with a specific observation. In order to entice and help the scientific community to use the SOS and to view it as "their own" data repository, the flexibility of data entry will successively be expanded in the new generation. The first improvement is that an unlimited number of variables related to an observed individual can be defined and added by the user, such as size or shape of particular body parts, host affiliation,

feeding, behaviour, or speed of flight. In the next step soon to be taken, it will be possible to define new variables related to all observations of a specific sampling occasion (e.g. temperature, pH and methodology), study location (e.g. vegetation, edaphic conditions and altitude) and project (e.g. general methodology and study aims), cf. Newman et al. (2011). Within the frame of SLW, a number of tools for automatic quality control will also be developed.

3.2 Integrating and Reusing Large Heterogeneous Biodiversity Data

All inventories and monitoring activities funded by Swedish governmental organizations are required to deposit their data in the SOS. Hence, the SOS supports very large and heterogeneous data sets of mixed quality and origin – from opportunistic data collected by amateurs to systematic data gathered within monitoring or research programmes. Through the SLW infrastructure, all these data are synthesized and made accessible together with biodiversity and environmental data from several other sources. Within this process, the taxonomic database Dyntaxa is essential for harmonizing and continuously keeping the taxonomy accurate and up to date within the infrastructure. Together, this makes a very powerful platform for large-scale analysis and modelling.

In the SLW Analysis portal, the user can to a certain extent delimit the categories of data to view and analyse. When structured monitoring data are not available, or are limited in space or time, citizen science data represent an alternative or complementary source of information. These citizen science data naturally present several analysis-related challenges relating to sampling bias, observer variability, and measurement errors (see, e.g., Yoccoz et al., 2001, Jeppsson et al., 2010, Kery et al., 2010). None the less, novel statistical and computational approaches exist to account for possible biases (see, e.g., Kelling et al., 2009, Jeppsson et al., 2010, Kery et al., 2010, Hochachka et al., 2012, Bird et al., 2013). Initial analyses based on SOS bird data have shown good potential for birds (Snäll et al., 2011) and longhorn beetles (Jeppsson et al., 2010, Snäll et al., 2013).

The coherent Analysis portal has been in place for less than a year, and functionality has been successively added during this time. Still, several research groups are already using the facility to explore issues on phenology over time of migratory birds, long-term changes of biodiversity in the Kattegat Sea, predicting invasive species, exploring effects on biodiversity of restoring wetlands, identification of optimal forest management from a biodiversity perspective given detailed biogeographic zone, and identifying biodiversity hot-spots along roads. The 21 County Administration Boards of Sweden, responsible for, e.g., environmental impact evaluation and provisioning of permissions for exploration activities, have for a few years had direct access in their GIS-system to the SOS data. Right now this service is extended to provide the entire SLW web-service of species observations, giving the officials access to even more data. This will not only enable better evaluations of exploitation applications, but also serve as valuable base for planning nature reserves and other conservation actions.

When it comes to taking advantage of citizen science data, the tradition in Sweden has been not to steer people directly towards a specific scientific enquiry, but rather to let the observers report opportunistically. This has been a success factor in terms of generating large amounts of data. At the same time, such liberty results in less structured and less representative data that pose some challenges when it comes to specific analyses. We believe that new developments within the SOS and SLW will aid the establishment of a number of more structured citizen-based reporting schemes and projects, linking directly to specific biodiversity conservation issues and trans-disciplinary research questions. Such science-driven projects will benefit from implementing well-designed and standardized collection methods and explicit scientific hypothesis testing (e.g. Silvertown, 2009). With appropriate guidelines, amateurs can indeed collect data of quality equal to those collected by experts (Crall et al., 2011, Danielsen et al., 2014).

Another way to improve the quality of citizen science data is to educate interested citizens in species identification. Citizen success with species identification has been shown to increase with self-identified comfort level (e.g. Crall et al., 2011). Species identification training furthermore needs to be extended to those organism groups for which there is a lack of existing literature and identification keys. To this end, the Swedish Species Information Centre is currently developing an online system with multi-entry identification keys and species descriptions. In the future, these tools will likely also be integrated with the SOS into a coherent species information system. With continuously more informed and engaged people, in hand with a fast developing field of biodiversity informatics, the prospects of engaging citizens for data collection and monitoring biodiversity are promising.

4. CONCLUSION

Swedish LifeWatch has developed an operative e-Infrastructure for biodiversity research with free access to environmental and biodiversity data from keen citizens, monitoring programs, inventories and research. Citizens get incentives to report their observations in the Swedish Species Observation System through win-win functionalities and properties of the system. This has appeared to be an effective way of gathering large amounts of data over spatial and taxonomic ranges, invaluable for biodiversity research, monitoring and nature conservation.

5. ACKNOWLEDGEMENT

6. REFERENCES

Bird, T.J, Bates, A.E., Lefcheck, J.S, Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F., Pecl, G.T., Barrett, N. & Frusher, S. 2013. Statistical solutions for error and bias in global citizen science datasets. Biological Conservation 173: 144–154.

Bonney, R., Cooper, C.B., Dickinson, J., et al. 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience 59: 977–84.

Bonney, R., Shirk, J., Phillips, T.B., Wiggins, A., Ballard, H.L., Miller-Rushing, A.J. & Parrish, J.K. 2014. Next steps for citizen science. Science 343: 1436–1437.

COM(2011)0244. Communication from the Commission to the European Parliament, the Council, the Economic and social committee and the Committee of the regions. Our life insurance, our natural capital: an EU biodiversity strategy to 2020 http://www.ipex.eu/IPEXL-WEB/dossier/document/COM20110244FIN.do

Danielsen, F., Jensen, P.M., Burgess, N.D., et al. 2014. A multicountry assessment of tropical resource monitoring by local communities. Bioscience 64: 236–251.

Constable, H., Guralnick, R., Wieczorek, J., Spencer, C., Townsend Peterson, A. & the VerNet Steering Committee. 2010. VertNet: A new model for biodiversity data sharing. PlosBiology 8 :e1000309

Crall, A.W., Newman, G.J., Stohlgren, T.J., Holfelder, K.A., Graham, J. & Waller, D.M. 2011. Assessing citizen science data quality: an invasive species case study. Conservation Letters 4: 433–442.

Devictor, V., Whittaker, R.J. & Beltrame, C. 2010. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. Divers. Distrib. 16: 354–362.

Dickinson, J.L., Zuckerberg, B.J. & Bonter, D.N. 2010. Citizen science as an ecological research tool: challenges and benefits. Annu Rev Ecol Evol S 41: 149–72.

Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R. Crain, R.L., Martin, J., Phillips, T. & Purcell, K. 2012. The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment. 10: 291–297.

Gardiner, M.M., Allee, L.L., Brown, P.M.J., Losey, J.E., Roy, H.E. & Rice Smyth, R. 2012. Lessons from lady beetles: accuracy of monitoring data from US and UK. Frontiers in Ecology and the Environment 10: 471–476.

Hardisty, A., Roberts, D. & The Bioinformatics Community. 2013. A decadal view of biodiversity informatics: challenges and priorities. BMC Ecology 13:16 http://www.biomedcentral.com/1472-6785/13/16

Hochachka, W.M., Fink, D., Hutchinson, R.A., Sheldon, D., Wong, W.-K. & Kelling, S. 2012. Data-intensive science applied to broad-scale citizen science. Trends in Ecology and Evolution 27: 130–137.

Jeppsson, T., Lindhe, A., Gärdenfors, U. & Forslund, P. 2010. The use of historical collections to estimate population trends: a case study using Swedish longhorn beetles (Coleoptera: Cerambycidae). – Biological Conservation 143: 1940–1950.

Kelling, S., Hochachka, W.M., Fink, D., et al. 2009. Data-intensive science: a new paradigm for biodiversity studies. BioScience 59: 613–620.

Kery, M., Gardner, B., Monnerat, C. 2010. Predicting species distributions from checklist data using site-occupancy models. Journal of Biogeography 37: 1851–1862.

Newman, G., Graham, J., Crall, A. and Laituri, M. 2011. The art and science of multi-scale citizen science support. Ecological Informatics 6: 217-227.

Los, W. & Wood, J. 2011. Dealing with data: Upgrading infrastructure. Science 331 (6024): 1515–1516.

Patterson, D.J., Cooper, J., Kirk, P.M., Pyle, R.L. & Remsen, D.P. 2010. Names are key to the big new biology. Trends in Ecology and Evolution 25: 686–691.

Pimm, S.L. & Raven, P. 2000. Extinction by numbers. Nature 403:843–845.

Rockström, J., Steffen, W., Noone, K. et al. 2009. Planetary boundaries: exploring the safe operating space for humanity. Ecology and Society 14: 32. [online] URL: http://www.ecologyandsociety.org/vol14/iss2/art32/

Silvertown, J. 2009. A new dawn for citizen science. Trends in Ecology and Evolution. 24: 467–471.

Snäll, T., Kindvall, O., Nilsson J. &, Pärt, T. 2011. Evaluating citizen-based presence data for bird monitoring. Biological Conservation, 144 (2): 804–8103.

Snäll, T., Forslund, P., Jeppsson, T., Lindhe, A. & O'Hara, R.B. 2013. Evaluating temporal variation in citizen science data against temporal variation in the environment. Ecography 37: 293–300.

Wetzel, F. and 42 co-authors. 2014. Gap analysis and priorities for filling identified gaps in data coverage and quality.   Report EU BON, M22, http://www.eubon.eu/news/11688_gap-analysis-and-priorities-for-filling-identified-gaps-in-data-coverage-and-quality/

Yoccoz, N.G., Nichols, J.D. & Boulinier, D. 2001. Monitoring of biological diversity in space and time. Trends in Ecology and Evolution 16: 446–453.