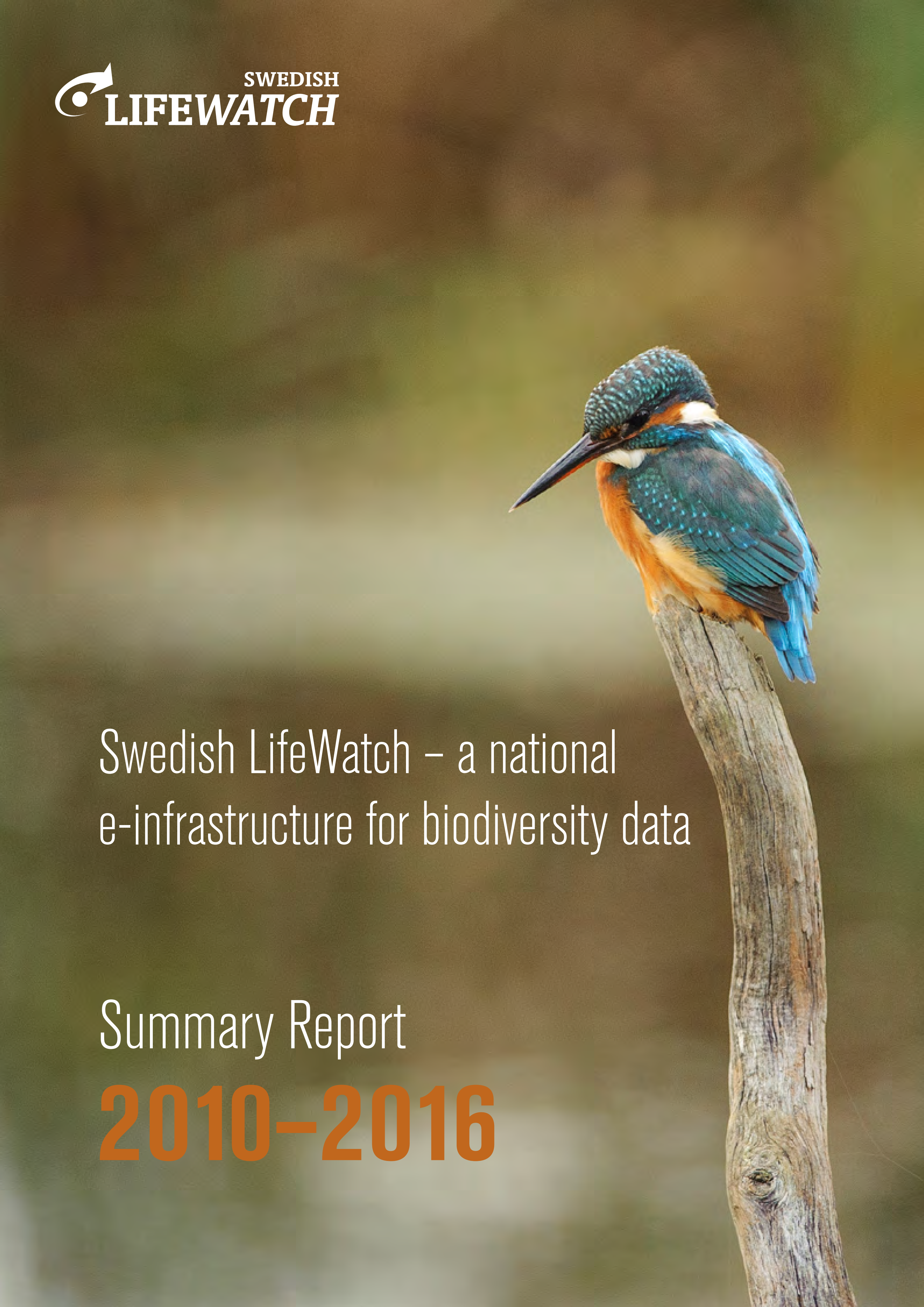


Swedish LifeWatch – a national e-infrastructure for biodiversity data

Summary Report

2010–2016

Suggested citation: Swedish LifeWatch – a national e-infrastructure for biodiversity data. Summary report 2010–2016. ArtDatabanken SLU 2017.

The pdf version of this publication provides interactive hyperlinks.
ISBN: 978-91-87853-16-6 (print), 978-91-87853-17-3 (pdf)
Print: Tabergs Media Group 2017

Layout and editing: Anna Maria Wremp, ArtDatabanken SLU
Illustrations: Katarina Nyberg, ArtDatabanken SLU
Cover photo: Frode Wendelbo
Photo of bear p 26: Lars Svensson/Mostphotos
Portrait photos p 17, 27, 59, 63: Johan Samuelsson.
Other photos: Johan Södercrantz (p 4 and 28); Magnus Dahlberg (p 24); Anna Maria Wremp (p 34, 48 and 52)

Download pdf at www.slu.se/lifewatch or contact lifewatch@slu.se for a printed version.

CONTENT

Photo: Johan Södercrantz

SUMMARY

The overarching mission of Swedish LifeWatch (SLW) is to make all Swedish biodiversity data openly available in standardised formats through interoperable web services, and to develop tools and virtual laboratories for advanced biodiversity and ecosystem analysis. SLW currently provides some 67 million Swedish species observation records relating to 35,000 different species from 15 primary databases. All data can be accessed, visualised and analysed in the SLW Analysis Portal. Datasets of interest can be assembled using sophisticated filtering tools (selecting taxonomically, spatially, temporally, or by accuracy, traits, Red List status and other attributes) and combined with environmental and climatic data from a wide range of providers. Results can be analysed and downloaded as refined data or maps, tables, diagrams and reports.

In 2010, the Swedish Research Council (VR) commissioned the Swedish University of Agricultural Sciences (SLU) to lead the design of an infrastructure for biodiversity and ecosystem research. An agreement was signed on 1 June 2011 between SLU, the University of Gothenburg, Lund University, Umeå University, the Swedish Museum of Natural History and the Swedish Meteorological and Hydrological Institute (SMHI) concerning formation of the Swedish LifeWatch consortium, led by a managing director at the Swedish Species Information Centre (ArtDatabanken, SLU) and an external and impartial Board. Thus the consortium has been operational for 6.5 years up to now.

The build-up has entailed a fantastic journey where we sometimes have had to invent new solutions. Even though the concept LifeWatch was established on the ESFRI roadmap already in 2006, and extensive and ambitious preparatory work took place over Europe the following years, Sweden was in reality the first country to implement a biodiversity informatics infrastructure of this kind. And we have received much appreciation and acknowledgment for what we have achieved.

When looking back at the 2009 application to VR, I can conclude with great satisfaction that not only have we been able to implement almost everything we planned back then, but we have delivered even more. Now I look forward with excitement and great enthusiasm to ongoing progress in a world of digital techniques evolving extremely rapidly and generation of huge amounts of biodiversity data of different types. This will give scientists undreamed-of opportunities to explore major questions and support society by providing well-informed recommendations on how to handle our vulnerable environment and restricted natural resources in a changing world.

Finally, I would like to take this opportunity to thank all the talented and enthusiastic people who have worked with the consortium and made it so successful: the Board, consortium partners and all their personnel, IT developers, researchers and other users, administrative staff at SLW and administrators at VR. I would also like to extend my thanks to everyone who has contributed to the compilation and production of this report.

Ulf Gärdenfors, Managing Director Swedish LifeWatch

Photo: Emmie Borgström / MostPhotos

FÖRORD på svenska

Svenska LifeWatch (SLW) är en e-infrastruktur för biodiversitets- och ekosystemforskning där biodiversitetsdata enkelt kan kombineras, utforskas, visualiseras och laddas ned. I SLW:s analysportal (www.analysisportal.se) ges också tillgång till miljö- och klimatdata liksom annan omvärldsdata som forskare, naturvårdare och övriga intressenter kan samanalysera med artdata för att utforska frågeställningar som tidigare vore helt orealistiska att angripa.

Vetenskapsrådet gav år 2010 i uppdrag till SLU att leda uppbyggnaden av denna infrastruktur och den 1 juni 2011 skrevs ett avtal mellan SLU, Göteborgs, Lunds och Umeå universitet, samt Naturhistoriska riksmuseet och SMHI att bilda konsortiet Svenska LifeWatch. I den här rapporten ger vi lite bilder av det vi uppnått, hur infrastrukturen används och den kontext som Svenska LifeWatch finns i.

Det har varit en fantastisk resa där vi bitvis har fått plöja ny mark. Även om konceptet LifeWatch etablerades på ESFRI:s (the European Strategy Forum on Research Infrastructures) vägkarta redan 2006 och omfattande förberedelser och diskussioner därefter försiggick i Europa, var Sverige först med att utveckla den här typen av infrastruktur. Och vi har fått många erkännande från runt om i Europa och världen för det vi byggt upp.

Grunden för SLW är ett betydande antal webbtjänster som sköter olika funktioner kring dataflöden, analyser, taxonomi, säkerhet, användarhantering etc. I dagsläget erbjuder SLW gratis och enkel tillgång till ca 67 miljoner artobservationer av 35 000 arter i enhetliga och kvalitetssäkrade format från 15 olika primära databaser, och detta växer kontinuerligt. Genom att alla observationer är kopplade till definierade artkoncept från den taxonomiska ryggraden Dyntaxa kan forskaren lita på vad en observation representerar och det behövs ingen datatvätt för att sortera bort synonymer, felstavningar, sammanslagna eller splittrade arter. Detta i sig är världsunikt.

I Analysportalen kan forskare, naturvårdare och alla andra intressenter utifrån sina frågeställningar skapa dataset baserade på utbredning (regioner eller egendefinierade polygoner), tidsperiod, taxonomisk grupp, rödlistestatus, artegenskaper, ekologiska preferenser, geografisk noggrannhet, m.m. Dessa data kan sedan aggregeras och/eller kombineras med omvärldsdata, analyseras och presenteras på en rad olika sätt i grafer, tabeller och kartor, eller i dataset som kan laddas ned och användas i fortsatta analyser eller modelleringar.

Även naturvården och övriga samhället drar stor nytta av SLW:s webbtjänster. Idag har handläggarna på alla landets länsstyrelser och många kommuner direkt tillgång till alla artdata rakt in sina geografiska informationssystem för planering och handläggning. Samma gäller skogsägare som via Skogsstyrelsens webb kan se vilka arter som finns på den egna fastigheten.

Det känns tillfredsställande, när jag nu blickar tillbaka på vår ansökan 2009, att kunna konstatera att SLW inte bara levererat det som utlovades utan totalt sett betydligt mer. Nu ser jag med spänning och stor entusiasm fram emot infrastrukturens fortsatta framåtskridande, med en kraftigt utvecklad digital teknik, med snabbt växande mängder och nya typer av biodiversitetsdata. Det kommer att ge forskarna oanade möjligheter att ta sig an stora ekologiska frågeställningar och även kunna ge välunderbyggda råd om hur vi ska handha vår natur och dess begränsade resurser i en föränderlig värld.

Slutligen vill jag passa på att tacka alla duktiga och entusiastiska människor som arbetat i konsortiet och på olika sätt bidragit till dess framgång: styrgrupp, konsortieparter med personal, IT-utvecklare, forskare och andra användare, SLW:s kansli samt handläggare på Vetenskapsrådet. Tack också alla som deltagit i skrivande och produktion av den här rapporten.

Ulf Gärdenfors, föreståndare för Svenska LifeWatch

The Swedish Government has adopted an open data policy. To promote greater openness and better service in the public sector, public authorities should make their public data accessible for reuse, either free of charge or on standardised and generous terms. The European Union has created drafts for an Open Science Policy Platform which is taking steps to make scientific results more accessible. In its definition of open science, the proposal also includes making the scientific process more open through citizen science and stakeholder interactions.

Photo: Clifford Shirley / MostPhotos

Open data – buzzword or virtual opportunities?

Open science transforms the ways in which science is performed and communicated. A number of elements of the research process can be made transparent to both scientists and society with the rapid digitisation of infrastructure.

CHRISTOPHER KULLENBERG, UNIVERSITY OF GOTHENBURG

Open science is an umbrella term that signifies a wide range of research practices aiming for greater transparency, access and dissemination of scientific results and methods. While the general notion of openness in science dates back to the scientific revolution and the first academic journals of the 17th century, the concept of open science has been reinvented with the rapid spread of digital technologies over the last couple of decades. Open science can be divided into three main areas.

The first aspect of open science concerns the dissemination of scientific findings and results. In this regard, open access is a key element in making journal publications accessible to all. However, source data – along with open source computer code to ensure replicability of published studies – is increasingly being made available to the public. Unrestricted dissemination of science pursuant to the three pillars of open access, open data and open source is steadily becoming a requirement for scientists working on projects funded by research councils. In this regard, another argument is often put forth alongside increasing the transparency and accuracy of science; namely that science funded by the government has an obligation to return its findings back to society, without limitations.

The second feature of open science concerns the inner workings of scientific practice, and opening this up presents more of a challenge. The notion of open methodology refers to the supplying of information about the practical and sometimes tacit elements of scientific lab work, field studies and handling of instruments. Even if data and instructions are attached to a research paper, reproducibility is often reliant on practical knowledge or conventions requiring prior knowledge. Open methodology aims to make such supplementary information more transparent.

Related to this issue is open hardware; scientific instrumentation that may be freely copied and rebuilt. The objective in this regard is to create patent-free alternatives to proprietary technologies, often in a quest to find cheaper solutions while still ensuring proper standardisation and data quality. Last but not least, open peer review is another drive towards opening up the practice of science, whereby a wider range of peers is invited to review studies and the peer review process – not just the final result – is made public.

Finally, open access to educational resources can be included as a third aspect of open science. In this concept, the materials needed for science education are made as accessible as possible, from reference works to software and databases, developing pedagogical resources along similar lines to open software projects.

The many aspects of open science will probably undergo many changes in years to come because structural implementation is happening right now, and also because many scientists are rethinking their research practices to make them more open.

One of the biggest open science resources in Sweden is the LifeWatch project, funded by the Swedish Research Council, which provides access to one of the world's largest biodiversity data repositories. The data is not only aimed at scientific colleagues, but has been designed to be useful to a wider range of societal stakeholders such as authorities, museums and the general public.

Christopher Kullenberg
is a researcher at the Department of Philosophy, Linguistics, and Theory of Science at the University of Gothenburg. His current research fields include Citizen Science and Open Data.

Big data, new opportunities and virtual laboratories

As biodiversity datasets grow and become more diverse, the system of environmental research infrastructures as a whole must support larger and more complex analyses to create the best possible understanding of how biodiversity responds to human activities. A system of physical and virtual biodiversity research infrastructures does not merely contribute towards addressing global sustainability challenges – it is fundamental to achieving them.

MATTHIAS OBST, UNIVERSITY OF GOTHENBURG

Ecosystems are under enormous pressure worldwide, while human activities are driving biodiversity loss on a planetary scale (Rockström et al., 2009). The anticipated impacts on our environment are so extensive that they define a new geological epoch – the Anthropocene. With the human population expanding beyond 7 billion this century, global environmental sustainability has become the defining challenge of our time. This is why research is urgently needed to provide knowledge on sustainable food production, endangered and invasive species, ecosystem services and the impact of climate change and other anthropogenic activities on the environment. Generating such knowledge relies increasingly on access to large data collections and the capacity to analyse contemporary changes in the structural composition of biodiversity, understand the factors causing biodiversity decline and predict the impact of biodiversity loss (Evans, 2013).

However, the physically diverse nature of biological data resources and the lack of advanced and universal analytical infrastructure prevent biodiversity and conservation research from keeping up with other fields of environmental sciences, such as climate change research.

Unlike many other scientific disciplines that address global challenges, biodiversity and ecosystem research is currently neither a major data discipline nor a globally united endeavour, while continuous or repeated processing of large and distributed datasets is still the exception rather than the rule. This is at a time when policy decisions dependent on knowledge about ecosystems have never been more pressing. Hence a system of physical and virtual biodiversity research infrastructures does not merely contribute towards addressing global sustainability challenges – it is fundamental to achieving them.

Nowadays more and more biological sensor systems are becoming operational and generating vast numbers of valuable biological datasets. Automated camera systems, radars or hydroacoustic sensors, environmental DNA sequencing machines, optical scanners such as continuous plankton recorders and human sensor networks deployed through citizen science programmes are all examples of these. These large datasets not only need to be structured, stored and made accessible, they also demand extensive data processing skills from scientists for mining and analysis purposes in order to unleash their potential. These requirements continue to constitute a major adoption barrier for many scientists and prevent ecological research of scale while making it difficult to integrate with other data-intensive fields in environmental science. Hence a robust e-Infrastructure that makes high-throughput data on biodiversity and ecosystems readily accessible and helps with basic processing tasks is a key component for sustainable economies in the future.

Photo: Service Mer-Station Biologique de Roscoff

Scientist from "Station Biologique de Roscoff" (France) are measuring different environmental parameters for the Ocean Sampling Day.

The LifeWatch concept

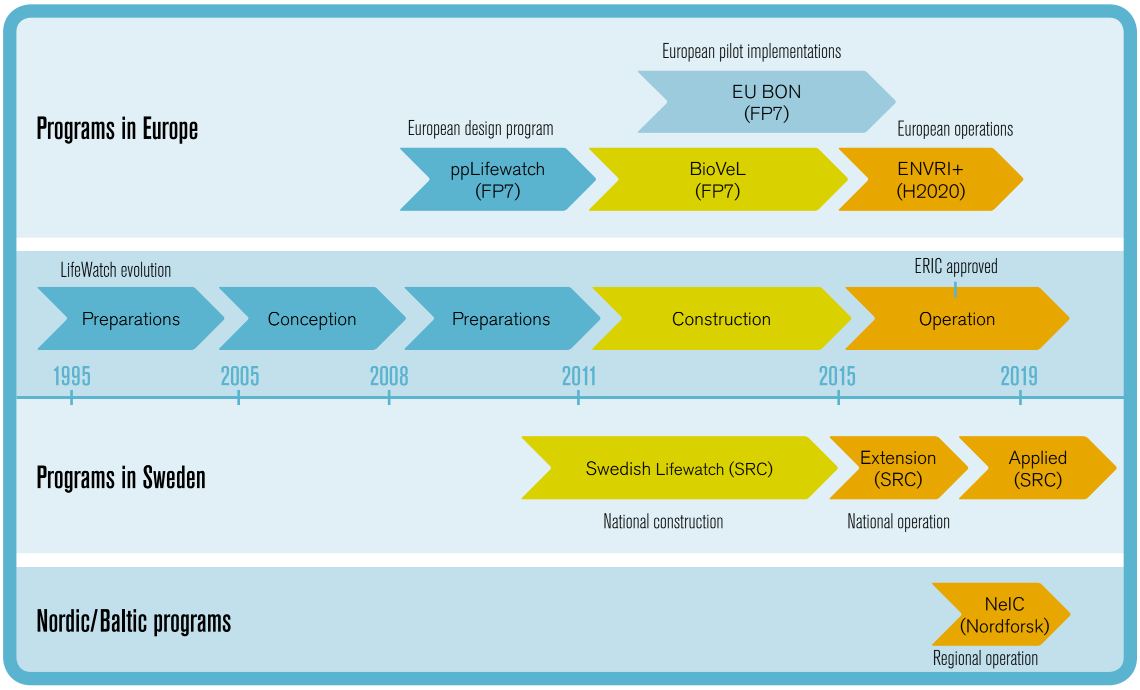
LifeWatch is a concept for a European distributed infrastructure developed under the auspices of the European Strategy Forum on Research Infrastructures (ESFRI). In short, the ambition of the Life-Watch infrastructure is to create a fully transparent system for global and Europe-wide exchange of biodiversity data following established standards, an interoperable systems architecture and a solid network of expertise. The LifeWatch concept also includes provision for computational power and functionality for data-intensive analysis and modelling of biodiversity scenarios (Basset & Los, 2012). The LifeWatch infrastructure aims primarily to support researchers with access to data repositories and virtual laboratories, which in turn should generate knowledge that supports decision-makers.

This is why there is strong dependence on a biodiversity data infrastructure among international conservation and management organisations or programmes such as the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) as well as international commitments and legislation, such as the Convention on Biological Diverstiy (CBD) and its associated Aichi biodiversity targets or the European Biodiversity Strategy for 2020 (COM(2011)0244).

Many key components of the infrastructure have already existed for a long time such as for genomic diversity (e.g. GenBank), species diversity (e.g. Global Biodiversity Information Facility, GBIF), ecosystem and habitat diversity (e.g. European Nature Information System), or taxonomic diversity (e.g. World Register of Marine Species). Others are still lacking at present, and all components need to be linked seamlessly to enhance interoperability.

European LifeWatch development

In 2008, the LifeWatch preparatory project started to develop the blueprint of a European e-Infrastructure for biodiversity and ecosystem research with a detailed technical, legal, financial and governmental framework. More than 20 European countries signed a Memorandum of Interest during that period concerning contribution to its development, and in many cases LifeWatch was included in national roadmaps for research infrastructure. Sweden played an important role during this period as the Swedish Museum of Natural History (NRM), the University of Gothenburg (GU) and the Swedish Research Council (SRC) participated in the preparatory work. In 2011, representatives from organisations in Italy, the Netherlands, Spain, Belgium, Hungary and Romania signed a Memorandum of Understanding so that building the Common Facilities of the LifeWatch infrastructure could commence. In November 2014, the Spanish government – acting on behalf of the proposed LifeWatch statutory seat in Spain – formally applied for LifeWatch ERIC (European Research Infrastructure Consortium) status. The ERIC was



Progress of the European LifeWatch development with reference to Swedish contributions on the national (blue) and European level (orange). Asterisk marks the anticipated formal acceptance of the European Research Infrastructure Consortium (ERIC).

Illustration: Katarina Nyberg

finally approved by the European Commission on17 March 2017. The countries signing the ERIC admission letter (or that are about to do so) include Belgium, Greece, Spain, Italy, the Netherlands, Portugal, Romania and Slovenia. An overview of the various national initiatives can be found at http://www.servicecentrelifewatch.eu.

National development

Apart from the ERIC preparations, an increasing number of European countries have independently developed their own national biodiversity e-Infrastructures over the past five years. In this regard, Sweden was the first country to build an independent national system for biodiversity data (fig above), which is now fully operational (Gärdenfors et al., 2014).

In Finland, the Finnish Biodiversity Information Facility (FinBIF) is developing a national service infrastructure for biodiversity research, where electronic biodiversity data and tools for visualisation and analysis are assembled (https://laji.fi). FinBIF also functions as a centralised gateway to international and in particular Nordic joint portals for biodiversity data. Norway and Denmark both have active consortia but are still waiting for dedicated funding for infrastructure development.

Regional development

For the years 2017–2019, Norway, Sweden, Denmark, Finland, and Iceland have initiated a Nordic e-infrastructure collaboration for biodiversity data financed by NordForsk (NeIC). This initiative explores the opportunities for enhancing cooperation and interoperability between existing Nordic (as well as Baltic) biodiversity information systems (Hanssen et al., 2014). Nordic and Baltic countries share many ecological issues, such as the impact of climate change on the Scandinavian Shield, sustainable use of the Baltic Sea and conservation of boreal forest ecosystems. Numerous common operations have already been established, such as species reporting systems in Sweden (Artportalen) and Norway (Artsobservasjoner).

FP7 and H2020 development

Many European programmes have started to deliver critical LifeWatch components over the last five years, financed largely through the European Commission's FP7 and H2020 programmes. Sweden played a major role in this process as well. The University of Gothenburg, for instance, led the scientific coordination of the Biodiversity Virtual e-Laboratory project (Hardisty et al., 2016) and is currently a partner to the Environmental Research Infrastructure project ENVRI+. The Swedish Natural History Museum in Stockholm participates in the European Biodiversity Observation Network EU BON where many foundations are being developed for a Europe-wide biodiversity e-Infrastructure.

Signs of payback on LifeWatch investments

Many valuable infrastructure components have come into being over the past five years as part of the LifeWatch vision. Representative examples of recently developed, consolidated or adapted European e-Infrastructure assets include taxonomic name systems (e.g. http://www.ipni.org/, http://www.catalogueoflife.org/), databases or aggregators (http://www.emodnet-biology.eu), service registries (https://www.biodiversitycatalogue.org/), data archives and publishers (http://biodiversitydatajournal.com/), Virtual Research Environments (http://scratchpads.eu/, https://portal.biovel.eu/), and supporting e-Infrastructure (http://www.egi.eu/, http://www.seek4science.org, https://zenodo.org/). These components are becoming more and more interlinked to form a federated, integrated system with common interfaces and improved access. We estimate that altogehter the users of these infrastructure components add up to at least 10,000 scientists across Europe, including in particular researchers in taxonomy, ecology and evolution (Dimitris Koureas, personal communication).

Cienpies Design / MostPhotos

> *" We estimate that altogether the users of these infrastructure components add up to at least 10,000 scientists across Europe, including especially researchers in taxonomy, ecology and evolution "*

This continuously growing and unifying research community is having tremendous effects on both the quality and the quantity of ecological research outputs. Scientists are not only beginning to investigate biodiversity on larger spatiotemporal scales, but are also studying ecological processes in greater depth. Recent examples include studies of anthropogenic effects on global genetic diversity (Miraldo et al., 2016) or global species invasions (Seebens et al., 2016), historical trends of continental-scale ecological features (Pereira et al., 2013; Kissling et al., in review) or studies of macroevolutionary and biogeographic patterns in large groups of animals and plants (Burin et al., 2016; Vilhena & Antonelli, 2015). Moreover, promising attempts have been made to gain a mechanistic understanding of the complex processes in an ecosystem based on data-intensive modelling approaches, including socioecological simulations of entire ecosystems (Cressey, 2015; Davies et al., 2016) or process-based modelling of entire parts of the biosphere (Purves et al., 2013).

Such research outputs are indicative of a dynamic, fast-moving scientific community attending to global sustainability issues and taking access to biodiversity e-Infrastructure for granted. While complex or global investigations were typically restricted to large scientific networks with extensive resources in the past, the majority of researchers – many of them young and working in smaller groups – can now set their scientific sights higher and address 'major' issues as the fast-growing body of globally assembled data becomes accessible to them.

Perspectives and challenges for the future

There has been tremendous progress in the development of e-Infrastructure over the past five years. Many scientists now not only work with these resources every day, but also assume that they are available for free. This creates a number of challenges that need to be addressed continuously and collectively.

One of these is the development of concepts for long-term sustainability and reliable user support. Another obstacle is the incomplete connection between research, data and computational infrastructures. The ESFRI landscape is still fragmented, which impedes scientific progress founded on collaborative and data-intensive research. This is particularly applicable to ecology and conservation research, where scientists seldom take advantage of infrastructures outside their own disciplines (Koureas et al., 2016). As biodiversity datasets grow and become more diverse, the system of environmental research infrastructures as a whole must support larger and more complex analyses to create the best possible understanding of how biodiversity responds to human activities.

Photo: Mikael Wallerstedt

Matthias Obst

is a researcher at the Department of Marine Sciences at University of Gothenburg. He has an active role in numerous European infrastructure projects as well as ESFRI program initiatives. He is also marine scientific coordinator for BioVeL and consortium chair of the Swedish LifeWatch initiative.

This is Swedish LifeWatch

ULF GÄRDENFORS, ARTDATABANKEN SLU

The overarching mission of Swedish LifeWatch (SLW) is to make all Swedish biodiversity data openly available in standardised formats through interoperable web services, and to develop tools and virtual laboratories for advanced biodiversity and ecosystem analysis. The SLW infrastructure uses a service-oriented architecture based on a flexible set of technology-independent units, allowing gradual expansion and easy replacement or updating of obsolete parts, and these elements can be combined to form several types of application.

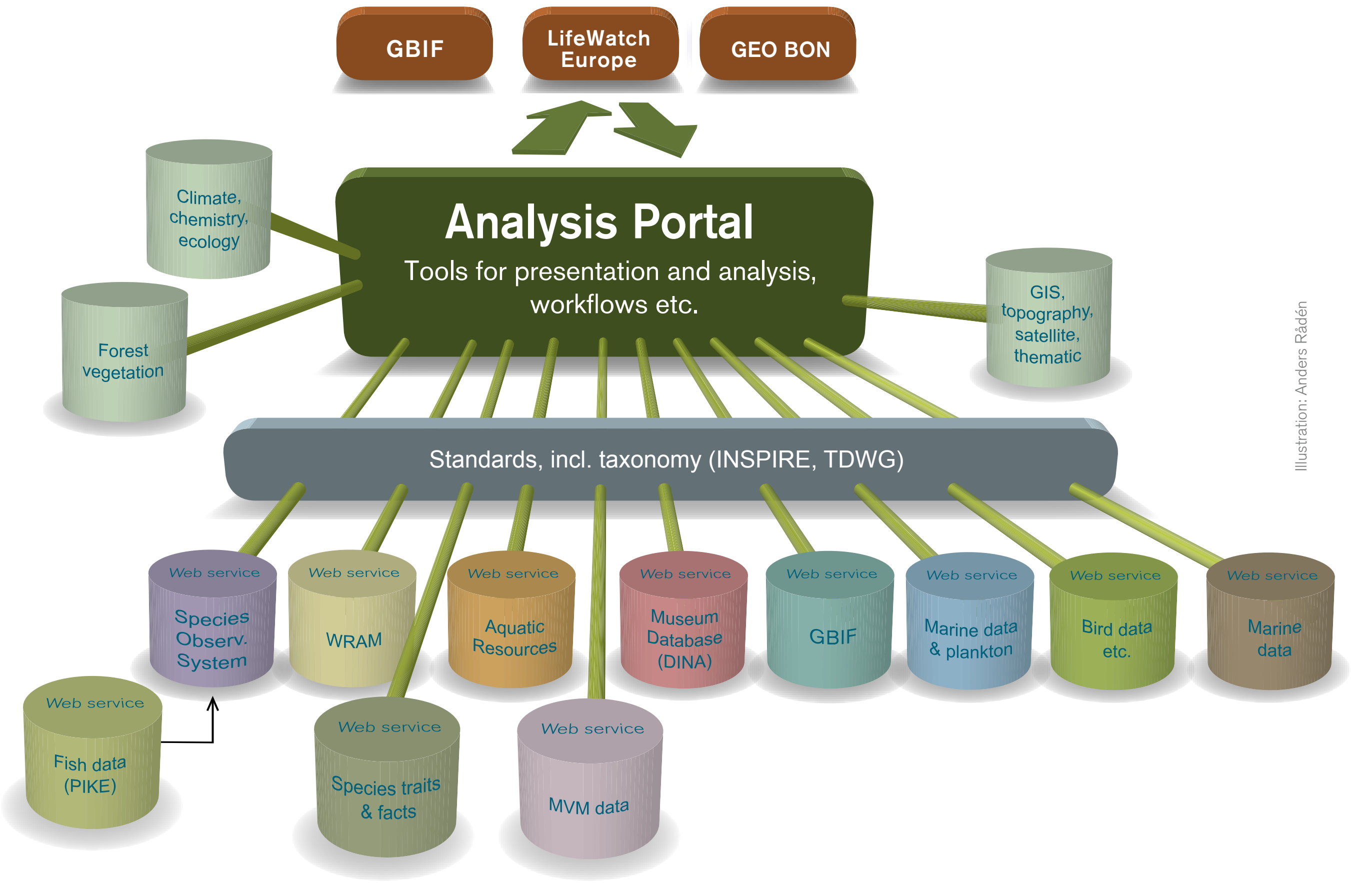
One important concept is to allow the primary databases to remain the repositories in which data can be corrected and supplemented, while LifeWatch incrementally harvests and presents the data in consistent and quality-controlled formats with no need to transform or clean observations prior to analysis. Therefore, much SLW activity has focused on building a biodiversity informatics infrastructure capable of achieving this. Web services have been implemented up to now for 15 primary databases supplying species observations from a wide range of

organisms and environments and currently encompassing 67 million Swedish species observation records of some 35,000 different species. The observation data harvested is also supplied by SLW to the global GBIF repository, making up the majority of the current Swedish contribution.

Research data storage service

The Swedish Species Observation System (Artportalen), hosted by ArtDatabanken, is by far the largest of the primary databases. It currently stores more than 57 million species observations (April 2017) from skilled citizen scientists in particular, but it is also the main repository for other species observations such as governmental monitoring data.

To ensure that research biodiversity data is retained, safely stored and reused, SLW has developed a function enabling the researcher to add any parameter of interest to Artportalen. This means that when a study is being planned and implemented, researchers can customise Artportalen to suit their own needs and define it as their data



Schematic overview of the Swedish LifeWatch infrastructure. SLW harvests data from a number of distributed data bases and makes all data available in the Analysis Portal, where it can be combined with environmental data. Data can also be exported or imported to/from other international initiatives such as GBIF or other European LifeWatch initiatives.

repository. This provides scientists with a database for their study which does not require data to be exported to a secure repository after completion, and also allows them to explore the data alongside the corpus of data in the Analysis Portal.

Data access and analysis

The Analysis Portal for Biodiversity Data is a key SLW product where all data can be accessed, visualised, analysed and downloaded. Datasets of interest can be assembled using sophisticated filtering tools (selecting taxonomically, spatially, temporally, or by accuracy, traits, Red List status and other attributes) and combined with environmental and climatic data from a wide range of providers. The latter are available as Web Map Services or Web Feature Services and can be accessed directly or via dynamically linked metadata catalogues. Results can be downloaded as refined data or maps, tables, diagrams and reports.

Analyses of various kinds can be performed in the portal. Users can explore how the occurrence of selected species correlates to environmental or climatological conditions or the distribution of other species. Which areas harbour the highest number of species, and how does this compare to environmental variables? Which species are found in protected areas or within certain habitats? How do population trends compare over a selected period with regard to species in a selected group? The assembled data can be easily downloaded, in original or aggregated form, and input into further workflows, models or simulations.

As the SLW infrastructure is modular, units communicating and remaining accessible via web services (APIs), data and tools can be used and combined in many contexts besides the Analysis Portal. Thus researchers (or enterprises, NGOs, governmental organisations, etc.) can have direct access to any part of the infrastructure for development of their own applications. For instance, SLW data now is fully integrated into the GIS handling systems of all 21 Swedish county administrative boards and many municipalities, enabling faster and better planning and decisions. Other examples are landowners signed up to the Swedish Forest Agency's website, allowing them to view which species are known from their characteristics, to acquire custom information about species with Red List status and to find advice on to manage their forests with respect to these species.

Taxonomic backbone

The content of the Analysis Portal and supporting web services is structured taxonomically throughout using the SLW taxonomic backbone Dyntaxa (now also the official Swedish taxon database) for reference. As far as is known, Dyntaxa was the first taxonomic database in the world to be based on the taxon concept with persistent identifiers (rather than unstable names) and dynamic tracking of taxon history, such as splitters and lumpers, leaving no uncertain taxonomic interpretations of SLW data. In combination with various validation tools, this ensures extremely high data quality exceeding that of comparable systems.

User administration

SLW has developed an authorisation and authentication system (A&A) called UserAdmin. Although all tools and the majority of SLW data are open for anybody to use and download without personal identification, an A&A system is essential; in particular for handling access to sensitive data such as the nesting sites of raptors or orchids exposed to a black market. ArtDatabanken is mandated to uphold the Swedish list of vulnerable species, and through UserAdmin it is capable of providing researchers or officials at county administrative boards (for instance) with access to sensitive observations delimited by species, region or time. Signed-in users can also personalise their workflows and save their settings in the Analysis Portal.

> *"The overarching mission of Swedish LifeWatch is to make all Swedish biodiversity data openly available in standardised formats through interoperable web services, and to develop tools and virtual laboratories for advanced biodiversity and ecosystem analysis."*

Other services and tools

Substantial resources have been allocated to fitting the infrastructure to national conditions and requirements. These include not only all custom web services harvesting and standardising heterogeneous data from all primary databases, the taxonomic backbone and the A&A system handling vulnerable species, but also a service managing administrative regions (polygons) at all levels and implementing the ArtDatabanken traits database of Swedish species, currently encompassing around 2 million classified states of 2000 parameters. These national adaptations require substantial cost, effort and know-how.

> *"The build-up has entailed a fantastic journey where we sometimes have had to invent new solutions. Even though the concept LifeWatch was established on the ESFRI roadmap already in 2006, and extensive and ambitious preparatory work took place over Europe the following years, Sweden was in reality the first country to implement a biodiversity informatics infrastructure of this kind. And we have received much appreciation and acknowledgment for what we have achieved."*

Ulf Gärdenfors

is professor of conservation biology at SLU Swedish University of Agricultural Sciences, Deputy Director at Swedish Species Information Centre and Managing Director of Swedish LifeWatch. He is also responsible for the scientific part of the Swedish Taxonomy Initiative and a member of the IUCN Red List Standards and Petitions Subcommittee.

SLW has also developed or supported the development of systems and tools targeting special user communities, such as Nordic Microalgae, the Plankton Toolbox and DINA. Nordic Microalgae is a system providing information such as illustrations relating to microalgae and aquatic protozoa. The

"One important concept is to allow the primary databases to remain the repositories in which data can be corrected and supplemented, while LifeWatch incrementally harvests and presents the data in constistent and quality-controlled formats, with no need to transform or clean the observations before analysis."

Plankton Toolbox is a free tool for aquatic scientists and others working with monitoring related to phytoplankton and zooplankton. It has been developed to aid researchers and conservationists working with and analysing plankton data. The workflow includes importing data, a quality control step, selecting data for specific purposes, functions for aggregating data, plotting tools and a number of statistical analysis tools. DINA (DIgital Information system for NAtural history data) is an information management system for natural history data supporting the compilation, management and sharing of data associated with natural history collections and their curation (collection management). DINA development is largely funded by the Swedish Taxonomy Initiative (Svenska Artprojektet), but SLW has also allocated resources to the system for developing web services so that its content can be easily harvested and accessible in the Analysis Portal, for instance.

Almost everything planned and promised in the original application has been developed and implemented throughout the period. Besides that, a number of features not included in the application have been developed and supplied for the infrastructure. The latter also includes advanced communication, education and support. Although there was no request for specific resources in the 2009 application, we soon realised the critical importance of outreach. Consequently, we employed a communications officer who is responsible for communication including the website, social media, exhibitions, information, etc. We have also developed training material and arranged a selection of workshops and courses at Swedish universities and other fora on how to utilise the SLW infrastructure.

International participation and contributions

SLW participates in a wide range of international initiatives. At a Nordic level, there is active participation in preparatory work on a Nordic LifeWatch and collaboration within an NeIC project on Nordic biodiversity informatics infrastructures. SLW has also contributed to the implementation of the first steps towards the European LifeWatch infrastructure, as well as the establishment of the European LifeWatch ERIC (although there has been no application for membership as yet).

What is Biodiversity Informatics?

Biodiversity informatics is a young discipline. The term was probably firstly used in 1996 by J.L Edwards in an unpublished document by the OECD Working Group on Biological Informatics. It encompasses information techniques for all kinds of biodiversity information and a framework for data sharing to address questions in the field of conservation management and ecosystem services. Bisby 2000 and Edwards et al. 2000 presented one of the first accessible databases, like GBIF (www.gbif.org), Species2000/ITIS Catalogue of Life (www.sp2000.org), FishBase (www.fishbase.org) and others, that they called the ´silent digital revolution` and which became accessible on everybody's desktop.

The terms *bioinformatics*, *eco-informatics* and *biodiversity informatics* are frequently used today, but often lead to confusion and misunderstanding. Traditionally, bioinformatics is connected to the molecular biology, while eco-informatics are dealing with environmental factors, such as climate and abiotic measurements. Globally, the LTER- network (Long Term Ecological Research Network; http://www.lternet.edu), DataONE (Data Observation Network on Earth, https://www.dataone.org) and GEO BON (Group on Earth Observations Biodiversity Observation Network, http://geobon.org) are more focused on eco-informatics, while LifeWatch and GBIF is more focused on biodiversity informatics. The term includes taxonomic databases as well as databases of species observation records and in a broader context even species distribution modelling.

From Leidenberger et al (2016).

Swedish LifeWatch in brief

The Swedish LifeWatch consortium was founded in 2010 as a national collaboration between four universities including the Swedish University of Agricultural Sciences (SLU), the University of Gothenburg (GU), Lund University (LU) and Umeå University (UmU), the Swedish Meteorological and Hydrological Institute (SMHI) and the Swedish Museum of Natural History (SMNH), including the Swedish GBIF node. The project is coordinated by the Swedish Species Information Centre (ArtDatabanken) at SLU and financed by the Swedish Research Council (VR) and the consortium partners.

The principal aim of Swedish LifeWatch (SLW) is to build and maintain a sustainable e-Infrastructure providing open access to data for biodiversity & ecosystem research. SLW's technical goal is to assemble a continuously growing body of data from terrestrial, freshwater and marine habitats through high-quality web services. SLW's scientific goal is to support large-scale, long-term, and holistic investigations on the status and change of Swedish ecosystems.

2006 LifeWatch is established on the ESFRI (European Strategy Forum on Research Infrastructures) roadmap

2008 LifeWatch Europe starts its preparatory phase

2009 A proposal to form a Swedish LifeWatch is submitted to the Swedish Research Council

2010 The Swedish Research Council grants SEK 36 million (2010–2014) to SLU for establishment of a Swedish LifeWatch infrastructure.

2011 A consortium is formally established which includes SLU, the University of Gothenburg, Centre for Animal Movement (CanMove) at Lund University, Umeå University, the Swedish Meteorological and Hydrological Institute, the Swedish Museum of Natural History and the former Swedish Board of Fisheries. The User-Admin authentication and authorisation system is produced. The Nordic Microalgae information system is launched.

2012 Dyntaxa, a new, concept-based taxonomic backbone, is released and web services are developed for data flow. An international interim evaluation of SLW is performed: in the opinion of evaluators, the SLW programme has in place an excellent infrastructure with the requisite computer and software resources to fulfil its aim of providing biodiversity data combined with GIS mapping, visualisations, modelling, metadata and support manuals with a view to producing world class research.

2013 The SLW Analysis Portal for Biodiversity Data is launched at a user and stakeholder conference. All Swedish county administrative boards are granted direct access to the SLW Species Observation Service via their GIS handling systems. A Nordic LifeWatch meeting takes place in Akureyri, Iceland, planning for future collaboration.

2014 New data sources and services are provided, including the SLW Geoserver. New tools are developed for analysis and visualisation in the Analysis Portal. The University of Gothenburg organises a Swedish LifeWatch–BioVeL hackathon with the aim of integrating BioVeL workflows into the Swedish LifeWatch Analysis Portal.

2015 A number of new data providers are added and are directly accessible via the Analysis Portal, now totalling 55 million observations. SHARKweb (containing marine species and environmental data) and the Plankton Toolbox are substantially updated. Several training workshops take place.

2016 A system is developed for flexible addition of Artportalen parameters, enabling researchers and other users to customise the reporting system to suit their own needs. More databases are added, resulting in the availability of 67 million observations of 35,000 species. The Graphical User Interface (GUI) of the Analysis Portal is substantially improved. Preparation for the new proposal to the Swedish Research council, due in March 2017.

2017 LifeWatch ERIC statutes are approved by the European Commission. New proposal to SRC for support 2018–2023.

SWEDISH LIFEWATCH

Data content

Artportalen

Artportalen (the Swedish Species Observation System), launched in 2000, is an Internet-based, freely accessible reporting system and data repository for georeferenced species observations of animals, plants and fungi. It now provides more than 67 million observations of more than 30,000 different species. Around 5 million new observations are added to the database each year. The data is provided by citizen science, environmental monitoring programmes and research projects.

wwww.artportalen.se

WRAM

Wireless Remote Animal Monitoring (WRAM) is a national e-Infrastructure for automatic reception, long-term storage, sharing and analysis of biotelemetry sensor data from animals such as mammals, birds and fish. Moose data is supplied to SLW in aggregated form.

www.slu.se/wram

Observations-databasen

The Swedish Species Observation Centre (ArtDatabanken) also provides data from the Observation Database of Red-Listed Species, with around 1 million occurrence data records. The database includes protected species only, and permission is required to access it. These datasets will eventually be integrated into Artportalen.

www.artdatabanken.se

Museum collections via GBIF

DINA is a web-based collection management system developed by the Swedish Museum of Natural History. DINA will be applied to all major Swedish natural history collections over the next few years, and also supply data to the Analysis Portal.

While waiting for the DINA system to be completed, several databases from a variety of Swedish museums (such as the herbariums at Oskarshamn Botanical Museum, Lund University and Umeå University) have been linked to the SLW infrastructure using the GBIF web service (www.gbif.org). The Swedish database on harbour porpoises is also linked to the SLW infrastructure, as well as the Swedish Bird Ringing Centre's species occurrence records, of which there are some 6 million.

www.gbif.se

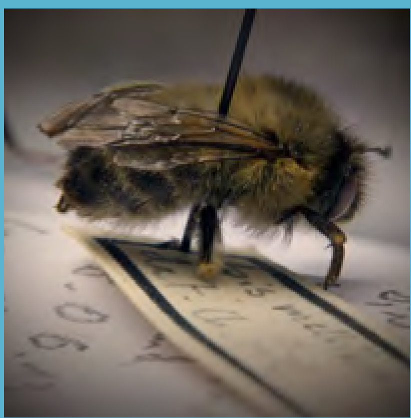
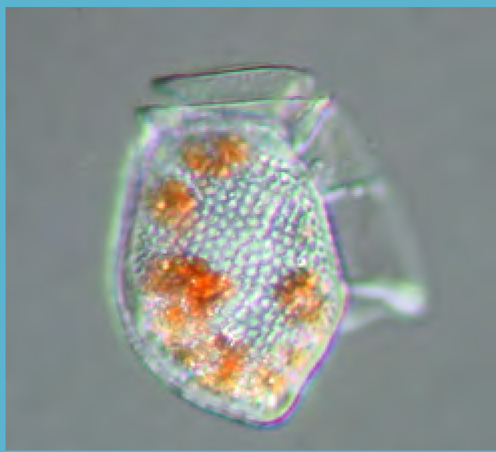


Photo: Johannes Hansen / MostPhotos

SHARK marine data

Marine data from environmental monitoring is made available by the Swedish Oceanographic Data Centre at the Swedish Meteorological and Hydrological Institute (SMHI). The SHARK database (Svenskt HavsARKiv) includes 3.2 million entries containing marine biological data including phytoplankton and zooplankton, benthic flora and fauna and seals. Data on environmental parameters such as nutrients, oxygen, salinity and temperature are also available.

https://sharkweb.smhi.se/
http://sharkdata.smhi.se



MVM benthic data

Freshwater species observations are represented by the environmental data web service MVM, linked with nearly 500,000 species occurrence data records. The Department of Aquatic Sciences and Assessment at the Swedish University of Agricultural Sciences is the data host for inland waters and the MVM database, which includes data on national and regional monitoring of phytoplankton, benthic diatoms, macrophytes, zooplankton and benthic fauna.

http://miljodata.slu.se/mvm

PIKE

PIKE is a freshwater fish database administered by Umeå University. Integration with Artportalen is planned, but is currently awaiting the migration of PIKE to PIKE 2.0.

Photo: Kav Kav / MostPhotos

NORS, SERS & KUL

The national register of survey test-fishing (NORS) includes more than 2 million records on fish caught, dating back to the 1950s, and provides a good overview of species occurrence and abundance of fish fauna in lakes. The Swedish Electrofishing RegiSter (SERS) provides data from electrofishing in rivers, while the Database for Coastal Fish (KUL) provides data relating to net and fyke net catches in Swedish coastal waters.

All three databases are hosted by the Department of Aquatic Resources at the Swedish University of Agricultural Sciences (SLU) on behalf of the Swedish Agency for Marine and Water Management and provide fish data

www.slu.se/aquatic-databases

Data content status

For detailed and updated figures relating to available observations, see Data Providers at:

https://www.analysisportal.se

Artportalen – world leading citizen science data repository

Artportalen (Swedish Species Observation System) is the biggest data provider to Swedish LifeWatch, and also one of the world's biggest providers of data to global GBIF. The database is a world-unique comprehensive repository for citizen science data as well as data from national environmental monitoring, public inventories and research data.

ANNA MARIA WREMP & STEPHEN COULSON, ARTDATABANKEN SLU

The Swedish Species Observation System called Artportalen (www.artportalen.se) is an Internet-based, freely accessible reporting system and data repository for georeferenced species observations of the major organism groups, i.e. animals, plants and fungi.

Private individuals as well as professionals and authorities contribute with data, making Artportalen a unique source of knowledge that is already delivering targeted conservation results and which supports in preventing and mitigating environmental problems.

Started in 2000 for birds

Artportalen set out in 2000 as a reporting system for bird observations and successively developed into a reporting system for all kinds of multicellular organisms. Even now the majority of reports are submitted by skilled amateur naturalists, but the system is also used as a repository for county administration inventories and – to a growing extent – for governmental monitoring programmes and data generated by research.

For nature enthusiasts, the system provides a simple and secure way to keep track of and analyse their own findings, as well as providing a forum for discussion of these observations. This is one of the main driving forces for citizen science contributions. Contribution to nature conservation and research is another strong incentive to use Artportalen. The many uses of Artportalen data make it unique in a citizen science context.

Artportalen is also the biggest database in the Analysis Portal web service established by Swedish LifeWatch, where it provides 84% of available records. Artportalen also supplies data to GBIF, where it provides almost 10% of all georeferenced data.

www.artportalen.se is an Internet-based reporting system for all major organisms (animals, plants, fungi). All data is freely accessible, with the exception for a few sensitive species.

Use of data

Artportalen data is used by data reporters to see how their own findings compare with the national picture of Sweden, but this data is also of major significance as regards conservation and societal planning.

Data from Artportalen is used by authorities, consultancies, and researchers, either directly or via the Analysis Portal or other interfaces. Authorities also use Artportalen for long-term secure storage of biodiversity data. Searches of Artportalen data by

Photo: Anna Maria Wremp

Artportalen – 90% of the data comes from skilled amateurs, but the system is also used as a data repository for observation data deriving from research, environmental monitoring and national inventories.

authorities and consultancies often form the basis for knowledge-based environmental management and informed decision-making. Researchers use the data to study ecology and biodiversity, often set against changes in either land use or climate processes, resulting in numerous publications.

Validation and security

The reported observations are validated by means of a combination of built-in probability tools that alert the data reporter if the reported species is rarely reported at a specific time of year or in a particular region, for instance. Many taxonomic groups also have validators that review and certify the observations. Last but not least, the openness of the system also provides validation in itself as errors in reports are very quickly identified and commented on by other users. With the exception of a list of vulnerable species, all data is freely accessible to view and download.

Data storage

Every observation in Artportalen has four obligatory data fields; taxon, reporter, date and location. In addition there are fields for extra information related to the find or the field visit. Together these enable a wide variety of data to be uploaded or searched, but for certain data sets these inbuilt default parameter fields are too limited or restrictive. A development of the platform that was initiated and funded by Swedish LifeWatch now enables the user to tailor new additional data fields for observations or field visits within a project and thereby extend the range of data types that can be entered. There are now three configurable data field types available that can be used in any combination; project specific value lists (a parameter that can have certain allowable values, or a yes/no choice), additional numeric values. This extended capability has provided Artportalen with enhanced flexibility and the ability to meet the often complex and unique requirements of research projects or inventories.

Artportalen – short facts

- Contains >58 million species observations of 30,000 different species (including plants, vertebrates, invertebrates and fungi)
- Over 1 million submitted photos of 18,000 species
- On average, a new observation is added every 4th sec
- Around 600 000 unique visitors to the website each year
- Around 10 000 reporters
- Provides 84% of the data to Swedish LifeWatch
- Provides 8% of all georeferenced data in global GBIF
- An observation must include species identity, location and date but may also include additional information such as habitat type or weather.
- Artportalen is developed and operated by the Swedish Species Information Centre (ArtDatabanken) at the Swedish University of Agricultural Sciences, on behalf of the Swedish Environmental Protection Agency.

Aquatic resources

SLU's Department of Aquatic Resources works on behalf of the Swedish Agency for Marine and Water Management to host fish data collected as part of both national and regional environmental monitoring. Fish data from three major aquatic databases is available through Swedish LifeWatch.

After the reorganisation of the Board of Fisheries in 2011, SLU's Department of Aquatic Resources became the national fish data host on behalf of the Swedish Agency for Marine and Water Management.

The hosting of fish data means that all the results collected as part of the national and regional environmental monitoring programmes are published in open databases available to the public.

"The Department of Aquatic Resources has a long tradition of collecting data on fish and fishing. Collection and storage of data are an important element in all our operations and give us the support we need to develop good management advice," says Anders Kinnerbäck, data hosting coordinator.

"The databases are also used as research material and reference data. Data collection is a requirement and a foundation for our national and international commitments and provides a valuable basis for our cooperation with the outside world."

SERS – Database for electrofishing in streams

The Swedish Electrofishing RegiSter (SERS) database includes results from surveys of electrofishing in running waters that have been carried out in Sweden. The majority of these electrofishing sessions have taken place within the county administrative boards' environmental monitoring and liming follow-up programmes, but other authorities and organisations have also made contributions. The database began in 1989 and now includes the results of more than 60,300 electrofishing sessions over more than 18,350 sites.

NORS – National Register of Survey test-fishing

The NORS (National Register of survey test-fishing) database consists of thousands of test-fishing sessions dating back to the 1950s. The Department of Aquatic Resources is responsible for collecting and checking test-fishing data generated by national and regional environmental programmes, and also collects test-fishing data from several other types of investigation in order to create a database that is as representative as possible. Its objective is to facilitate the acquisition of high-quality data for national investigations and reports.

Standardised test-fishing using Nordic multi-mesh gillnets is a common way of investigating fish fauna in lakes in Sweden. These nets, which include 12 mesh sizes, catch most of Swedish species in a representative manner, providing a good estimate of species abundance and size distribution. The results are used for environmental protection and fishery management purposes. The method became a European standard (EN 14757) in 2005.

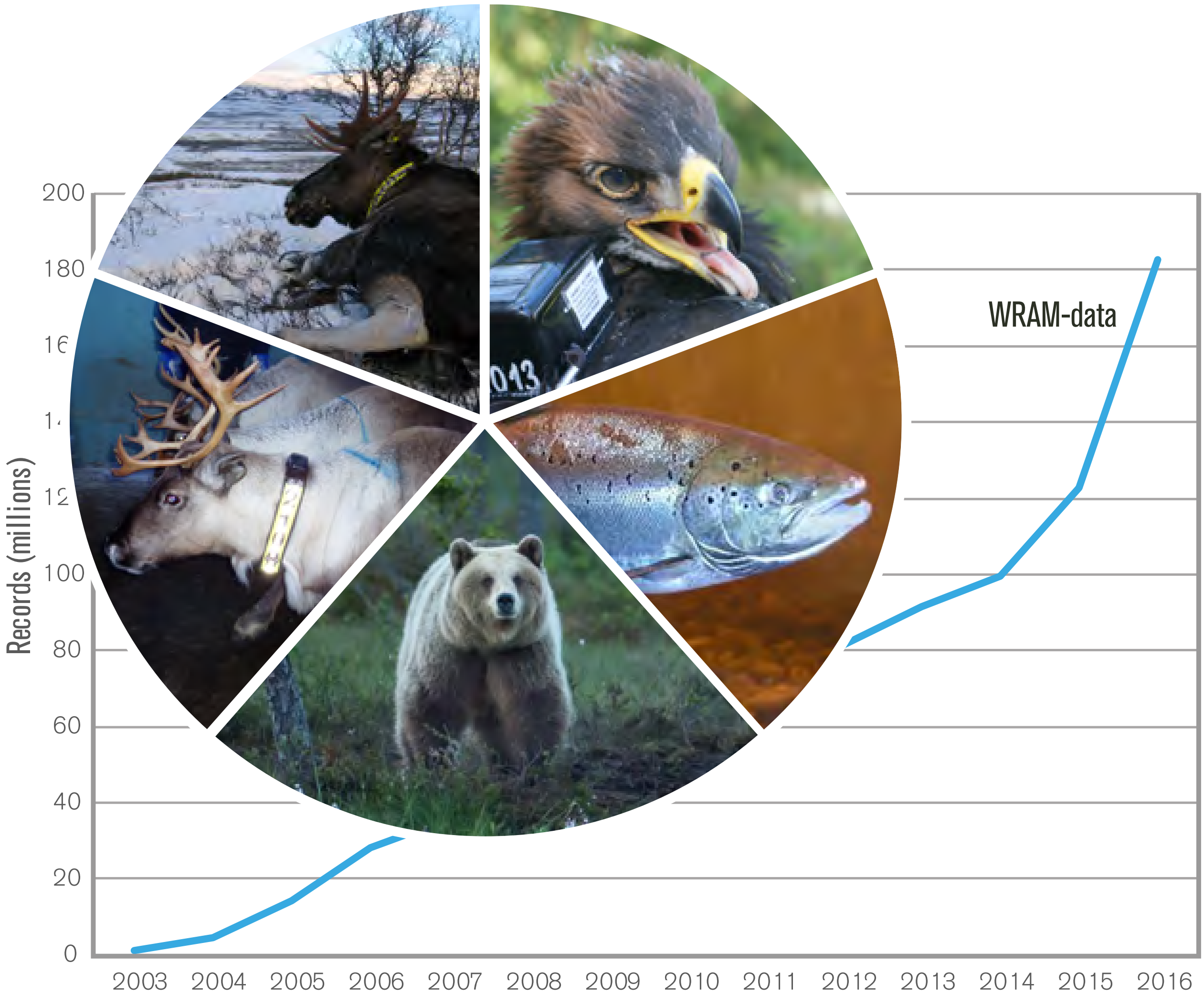
Database for Coastal Fish – KUL

The KUL database, which contains data from net and fyke net fishing in coastal waters, has been in use since 2006 and provides quality-assured coastal fish catch data. The database also includes information on fish gender, length, weight and age. Work on transferring older data to the database is constantly ongoing.

LifeWatch data harvest and pilot studies

Building web services to harvest data from NORS, SERS and KUL has been a Swedish LifeWatch priority, and preparations are being made to harvest data from a fourth aquatic database, Fishdata2, as well.

The Department of Aquatic Resources has also invested resources in case studies on the simultaneous use of fish species observations from different original databases. Examples include how to create and compare aggregated estimates of fish species richness and species lists for combined and original data sources at different geographical and temporal levels.



Wireless Remote Animal Monitoring – a Swedish LifeWatch data provider

HOLGER DETTKI & GÖRAN ERICSSON, SLU

Traditionally, species observations found in the Analysis Portal are performed by human 'observers' and are often uploaded manually to biodiversity repositories such as the Swedish Species Observation System (Artportalen). However, automatic or semi-automatic tracking sensors have been deployed on roaming animals since the 1960s in order to monitor their movements. Recent advances in animal-based GPS and biotelemetry sensors have allowed for remote data capture from a steadily increasing number of taxa, species and individual animals, resulting in the generation of enormous amounts of data ranging from spatial positions to physiological measurements. These animal-derived datasets are unprecedentedly detailed, allowing complex modelling of animal physiology, behaviour and ecology. Thus biotelemetry datasets represent a different type of data, albeit it a type related to 'classic' biodiversity data. Hence it is difficult to host this type of data directly in the Swedish Species Observation System, where it could be made accessible in the SLW Analysis Portal.

WRAM

The Wireless Remote Animal Monitoring (WRAM) e-Infrastructure is a national infrastructure for automatic reception, handling, storage and sharing of biotelemetry sensor data from animals. It was founded in 2003 and is run nowadays by the "Umeå Center for Wireless Remote Animal Monitoring" (UC-WRAM) national competence centre at SLU.

When the Swedish Research Council granted support for an upgrade in 2010, the infrastructure was redesigned to cope with enormous data volumes and the ever-growing numbers of new and different sensors developed and deployed on animals. It now features two major elements. The first is the backend WRAM Data Warehouse (WDW), a high-performance data warehouse for automatic reception and storage of real-time 'big data' such as position, acceleration or heartbeat data from fish and wildlife. Modern 'key/value pair' database techniques allow any type of current or future sensor data to be stored. Secondly, the WRAM Data Broker (WDB) – with its Client and Admin Portals – is a single sign-on web interface seamlessly federating the WDW with other similar database systems around the world, such as Movebank (Germany/US) and CanMove (Lund University, Sweden).

Funded by a grant from SLW, WRAM has developed integration tools to make suitable spatial biotelemetry data available in the SLW Analysis Portal. This spatial data is automatically aggregated over a certain period of time and displayed to users as the 'average position' of an individual over four weeks, for example. A specific web service was developed for this purpose to regularly collect and aggregate data from the WRAM Data Warehouse. Administrators can easily add and configure various aggregated transformations calculated on the sensor data held within the WRAM Data Warehouse. The same component is also used by external data providers such as CanMove – who are linked to the WRAM data broker – to aggregate their data.

This makes the aggregated data accessible in the Analysis Portal as 'observations'. This data resolution is already adequate for many large-scale scientific issues. However, as information about the original data resolution, data owner, sample method and other potential sensor types deployed on the same animal during the 'observation period' is made available as 'data attributes', the original data becomes discoverable and can be accessed and reused via the WRAM Client Portal, in consultation with the original data owner.

3 questions for Holger Dettki, WRAM coordinator at UC-WRAM

How much data is there in WRAM?

– The WRAM infrastructure currently (April 2017) contains around 185 million records from real-time biotelemetry sensors and is used by 38 user groups from eight countries, monitoring 24 species and more than 2800 individual animals. It is currently growing at a rate of around 40 million records each year.

How do you see WRAM developing in future?

– We are expecting WRAM to grow massively in the future due to new sensors and transmission methods becoming available for less and less money. Even though most of this new data will be non-spatial, e.g. heart rate sensors generating data at 20 Hz, for the most part every tagged animal will also be equipped with a tracking device to enable the animal to be handled for exchanging or removing the tags and sensors. Hence the amount of data available through the Analysis Portal will continue to grow as well.

– Increasing integration with environmental and climate data in the Analysis Portal will enable researchers to answer new questions regarding animal behaviour in an ever-changing world.

What part does WRAM play in SLW?

There is currently a huge push internationally to make genetic and sensor data from different repositories available together with 'classical' observation data to describe biodiversity and ecosystems. WRAM is here at the forefront of this development, providing a fully functional integration of this important source of information into SLW's infrastructure.

Photo: Johan Södercrantz

SHARK – oceanographic and marine biological data

SHARK is the name of the Swedish Ocean Archive (Svenskt HavsARKiv), a database hosted by the National Oceanographic Data Centre at SMHI. SHARK stores quality-controlled physical, chemical and biological data.

LISA SUNDQVIST, SWEDISH METEOROLOGICAL AND HYDROLOGICAL INSTITUTE (SMHI)

The Swedish Meteorological and Hydrological Institute (SMHI) has a long tradition of marine data collection and storage. SMHI has been the national host for physical and chemical marine monitoring data since 1998, and for marine biological data since 2007.

The SHARK database includes data on species ranging in size from bacterioplankton to seals and provides information on species abundance, distribution and traits. All data is searchable and available for download via SHARKweb (https://sharkweb.smhi.se). SHARK currently contains more than 9 million measurements from over 50,000 locations in the waters surrounding Sweden.

As part of Swedish LifeWatch, SMHI has developed a new SHARK web service that makes data available for machine downloading via a REST API. This service is used at Swedish LifeWatch to harvest data from SHARK for the Analysis Portal for Biodiversity Data, hosted by SLU. The service is also used by international organisations such as the International Council for the Exploration of the Sea (ICES) and the European Marine Observation and Data Network (EMODnet). Making data available for harvesting is a big step forward as it paves the way for a different kind of data usage. For instance, it facilitates the use of data for model validation. It also makes it possible for other databases to receive the latest version of SHARK content automatically.

"In future, monitoring data volumes will increase as image, video and genetic techniques are expected to be used frequently. This will generate huge amounts of data," says Patrik Strömberg, head of data and information at SMHI. "It is important to start preparing SHARK for these new data formats right now."

Photo: Michael Thorell / MostPhotos

Data included in SHARK:

- Physical and Chemical
- Chlorophyll
- Primary production
- Sedimentation
- Bacterioplankton
- Picoplankton
- Phytoplankton
- Zooplankton
- Zoobenthos
- Epibenthos
- Grey seal
- Harbour Seal
- Ringed Seal
- Seal Pathology

Dynamic taxonomic backbone

"If we do not know the names the knowledge of things becomes meaningless", said Carl Linnaeus, the world-famous father of taxonomy. Similarly, a taxonomic backbone is one of the most vital components of a biodiversity infrastructure.

OSKAR KINDVALL, SLU / CALLUNA CONSULTING

One absolutely crucial component in a biodiversity infrastructure is a taxonomic backbone with globally unique identifiers (GUID) for all taxa to which every item of species data can be referred (Patterson et al., 2010). When the Swedish LifeWatch infrastructure was being planned, there was a clear need for a system building consistently on taxon concepts (represented as Life Science Identifiers/GUIDs), defined by descriptions, pictures and/or observations, where scientific names – as well as vernacular names, hierarchies, distributions, etc. – are merely attributes to the taxon ID.

Such a taxonomic backbone needed to be constructed at a national level, rather than relying on the existing EU-nomen Pan-European Species directories Infrastructure, PESI, or a global level such as the Catalogue of Life. There are a number of reasons for this.

We needed a more or less complete list of all species occurring in Sweden right from the outset. In other words, we were unable to wait for years for this to be achieved on a larger scale. We also needed a system that tracks and stores historical taxonomic changes such as splitting and lumping over time, otherwise every analysis including old data would need initial, more or less manual taxon matching and data cleaning.

We needed a flexible system where we could define pragmatic taxon concepts for the country. Examples include hard-to-distinguish species pairs or species aggregates for which observations are recorded within monitoring programmes, for example, and which therefore need to be stored. Other examples include pragmatic, non-systematic groups such as lichens and algae, about which users would like to download data or other information.

We needed a system where we had the power to adjust both the content and the functions (software) more or less instantly, not having to compile a global wish list and wait for somebody else to fix it.

This resulted in the development of the taxonomic database Dyntaxa (http://www.dyntaxa.se). Nowadays it includes more than 95% of the 60,000 or so multicellular species known to occur in Sweden. Even though this is a national database with LSIs/GUIDs initially created nationally for the taxa, the aim is to gradually match the concept of each taxon to the concepts – and hence LSIs/GUIDs – of the European PESI, and subsequently to use PESI LSIs as the recommended taxon identifiers. As a consequence, the Swedish taxonomic backbone will gradually merge into the European backbone, and eventually – we hope – into a global backbone such as Catalogue of Life.

Dyntaxa is openly available at www.dyntaxa.se and can also be acquired from the Taxon Service web service.

Illustration: Anders Rådén

For a full description of Dyntaxa, see Kindvall, O., in collaboration with Roscher, S., Bailly-Maître, J. and Šípková-Gaudillat, Ž. 2015. Dyntaxa taxon concept administration and how to handle information related to taxa.TC/BD report to the EEA.

The Analysis Portal for Biodiversity Data

The Analysis Portal for Biodiversity Data provides an easy access point to all data provided by Swedish LifeWatch and a range of analytic services. The portal also makes it possible to combine observation data with other types of environmental data, creating a powerful tool for visualisation of ecological contexts and environmental challenges.

The Analysis Portal for Biodiversity Data (www.analysisportal.se) was launched in 2013, and now almost 70 million Swedish species observation records can be assessed, visualised and analysed via the portal. Datasets can be assembled using sophisticated filtering tools and combined with environmental and climatic data from a wide range of providers. Different validation tools, such as the official Swedish taxon concept database Dyntaxa, ensure high data quality. Results can be downloaded in a variety of formats as maps, tables, diagrams, Excel files and reports.

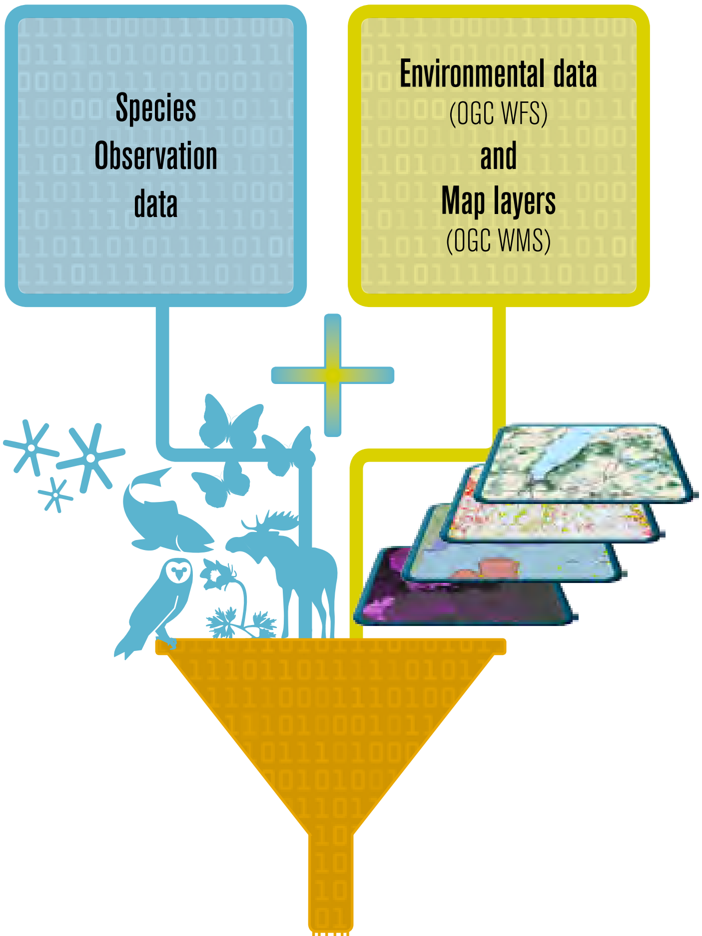
Observation data is harvested on a daily basis from a number of data providers. All connected databases are searched by default, but selected data sources can be chosen as well.

The various kinds of functions used include calculations of species richness and observations for defined sets of taxa, polygons, grids, time series or taxon traits. The Analysis Service can also perform calculations on grid-based summary statistics on specified WFS data layers.

Portal inputs include species observation data, climate and environmental data as layers (WFS) and maps (WMS), drawn polygons and uploaded GeoJSON files. Users can visualise and download different kinds of data tables, maps, histograms and diagrams on species observation, species richness and time series of species records as output (see examples on the next page).

The Analysis Portal uses the same user administration system as Artportalen. There is no login requirement, but users are provided with personalised settings options and certain search queries can be saved and reused. Although 90% of observations are publicly accessible, a few vulnerable species are excepted from this. Access to vulnerable species is subject to certain user requirements and permissions are handled via the user administration system.

The Analysis Portal undergoes constant development in response to users' feedback and requests, and Swedish LifeWatch is working actively on capacity building, trainee workshops and support.



Input options

– Access to species information data, environmental data and map layers

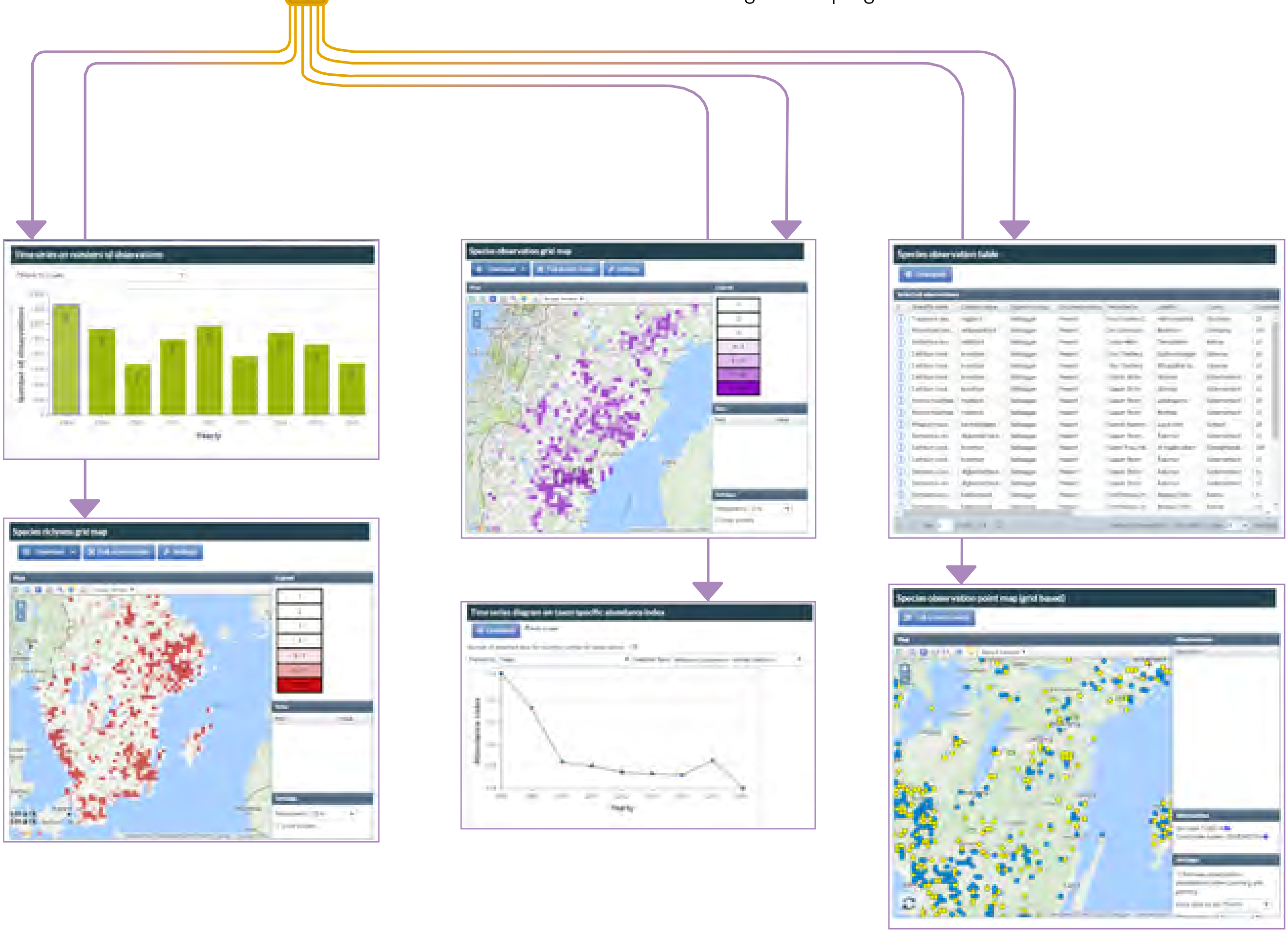
The Analysis Portal provides a single point of entry to all observation data within the Swedish LifeWatch infrastructure, but also allows the option of combining environmental data and map layers such as

- Topographic maps
- Climate data
- Hydrologic data, soil background data, etc
- Socioeconomic data

Many data hosts provide their data as WFS (Web Feature Service) or WMS (Web Map Service) layers which can be imported into the Analysis Portal for co-analysis with the observation data provided by Swedish LifeWatch.

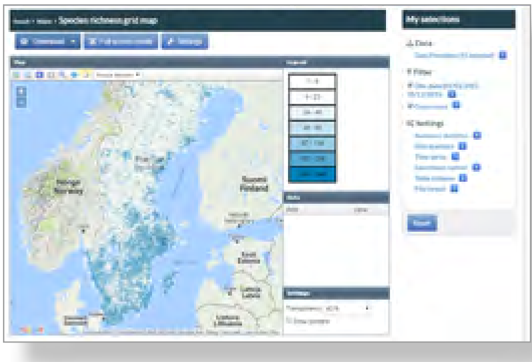
Output options

The results can be visualised in many ways; as species observation grid maps, species richness grid maps, tables or time series, or downloaded as XLS files. Data can also be downloaded to a GeoJSON file for importing to GIS programs.



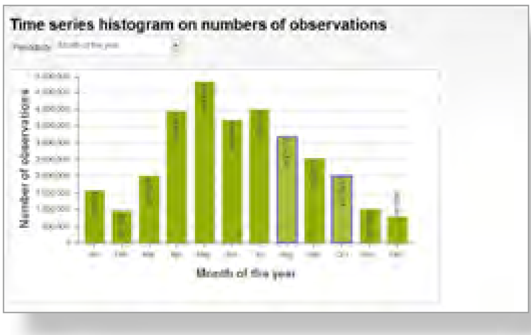
Analytical and visualisation services

The Analysis Portal provides a single point of entry to all data within the Swedish LifeWatch infrastructure. The portal also permits free access to biodiversity and environmental data and provides a range of analytical and visualisation services. Three examples of the opportunities it offers are presented below.



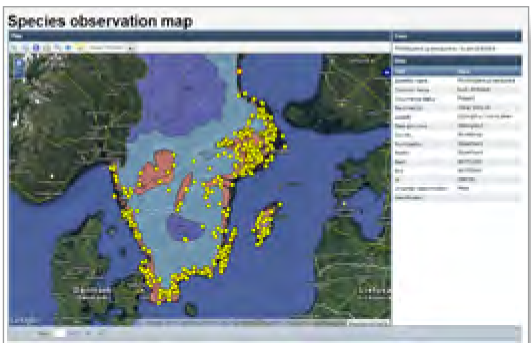
Species Richness Grid Map

The Analysis Portal makes it possible to calculate various types of grid-based spatial statistics. Here is an example of a species richness grid map where the number of species per grid cell is shown in different shades of blue. In this case, all public observations currently available from the Swedish LifeWatch national web service for species observation are used. Similar calculations are possible for any group of organisms or user-defined set of taxa in order to explore biodiversity patterns.



Time Series Histogram Species Observation Map

The variation of species observations over time can be explored in the Analysis Portal. This figure shows temporal variation of all species observations currently available in the Swedish LifeWatch infrastructure. Sampling efforts must be taken into consideration when analysing observation data statistically. A tool for analysing time series based on an abundance index, which provides controls for sampling effort, will soon also be available in the portal.



Species Observation Map

The Analysis Portal makes it possible to co-analyse species observations and environmental data. This screen shot shows observations of the bush cricket *Pholidoptera brachyptera* (yellow dots) and the average time for the first nights of frost in autumn. The frost occurs later in the red zone than in the bluish zones. In this case, species observation data is provided by Artportalen. The environmental data is provided by means of an OGC Map Service (WFS) at the Swedish Meteorological and Hydrological Institute (SMHI).

3 questions for Matthias Obst, scientific coordinator for BioVeL:

What is the main advantage of an e-Infrastructure?

– We currently use e-Infrastructures all the time, both at work and privately, to store and share documents, music and pictures, collaborate online or exchange specific information within a network. Most sectors of the economy, such as the music industry, commerce and the media, have quickly adapted to the opportunities offered by digital enterprise, but biodiversity research is still far from leveraging its digital assets effectively. As a result, biodiversity datasets are still difficult to find, reuse and integrate. Swedish LifeWatch is the first national infrastructure offering a 'streaming service' for biodiversity data. The future of biodiversity and conservation research as a scientific discipline operating truly globally is dependent on the rise of similar e-Infrastructures around the world, facilitating seamless integration of biodiversity datasets and analysis tasks across repositories, biotas and countries.

How can Swedish LifeWatch and BioVeL create synergy effects?

– Swedish LifeWatch has many system components, each with a particular focus. The core mission of Swedish LifeWatch services is to provide quality-controlled data from national repositories, while the BioVeL infrastructure allows Swedish researchers to access data from outside the country as well. One of the most popular ways of using the BioVeL portal involves extraction of environmental and ecological values by overlaying species distribution records with global data layers on factors such as chlorophyll, habitat type, salinity and tidal amplitudes, for example.

What major challenges do biodiversity e-Infrastructures face?

– e-Infrastructures are not visible in the same way as physical research infrastructures, such as laboratories, field stations or research vessels. This can create problems as regards long-term support and finance. Researchers usually take e-Infrastructures for granted once they are in place, but they would not expect to pay for them; especially when data should be free and open to everyone in the future. Many international biodiversity e-Infrastructures are struggling to establish a long-term sustainability model, but the European LifeWatch initiative may provide a solution when forming the European Research Infrastructure Consortium (ERIC).

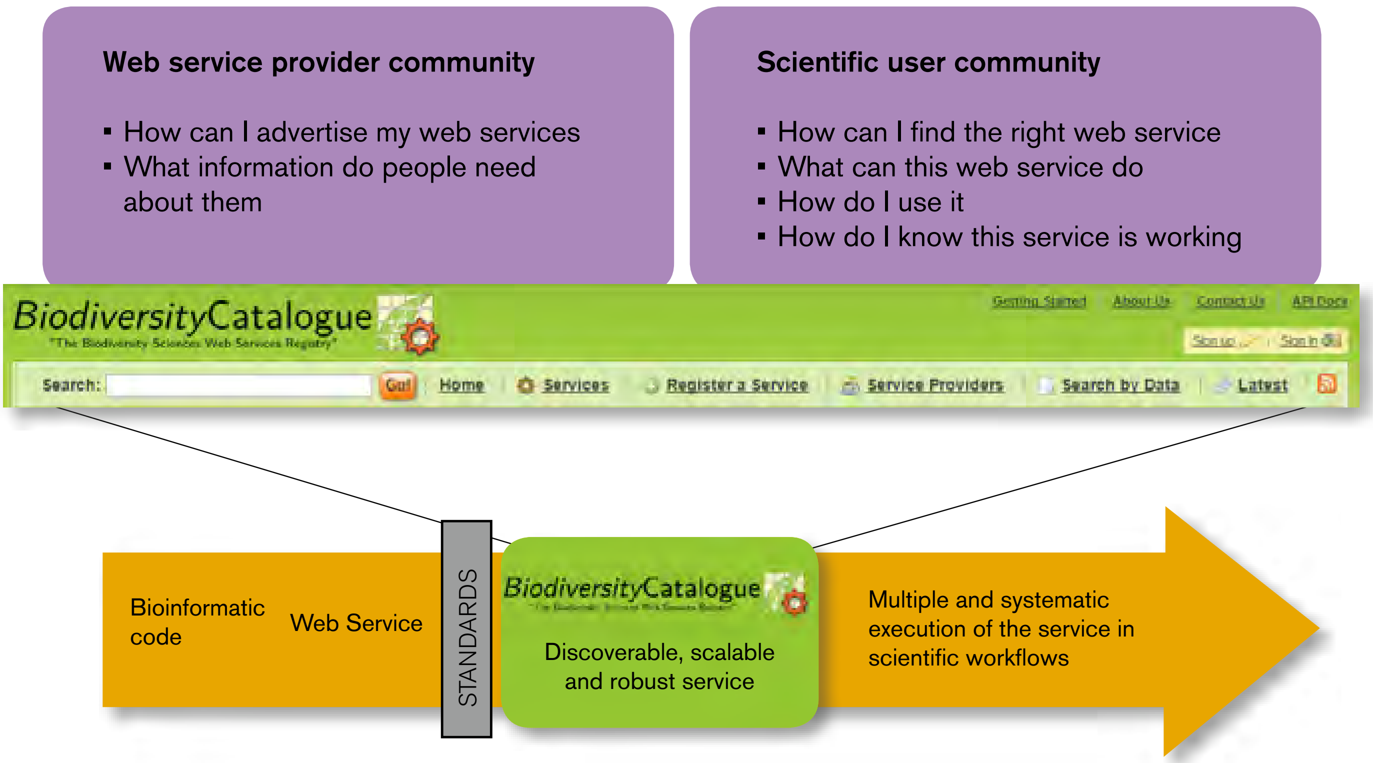
BioVeL – advanced workflows for biodiversity analysis

BioVeL is a virtual e-laboratory for biodiversity science and ecology data access, analysis and modelling. The platform offers functions for accessing and analysing data through curated web services and for performing complex in silico analysis through exposure of R programs in workflows. The infrastructure was developed as part of the European Union's Seventh Framework Programme and now includes a portal (http://portal.biovel.eu/), a web services catalogue (www.biodiversitycatalogue.org/) and a documentation wiki (https://wiki.biovel.eu/).

The BioVeL catalogue currently features more than 70 biodiversity-specific web services. This is a modest number compared to the 1187 web services offered by similar infrastructures in the Life Sciences domain (www.biocatalogue.org/). However, the community of biodiversity informaticians and scientists is constantly growing and the BioVeL platform is actively supporting this development. Scientists are using the e-laboratory for on-line collaborations in particular as the system permits sharing, repetition and documentation of complex analyses that draw data from many distributed sources throughout the world. The infrastructure also permits seamless links between intense analyses and high-performance computing infrastructures.

BioVeL and Swedish LifeWatch

BioVeL is currently supported by a number of European institutions and infrastructure programmes dedicated to sharing, sustaining and developing open access biodiversity research tools. The BioVeL virtual e-laboratory is currently supported by the Swedish LifeWatch infrastructure and in turn offers international data services and products to Swedish scientists.



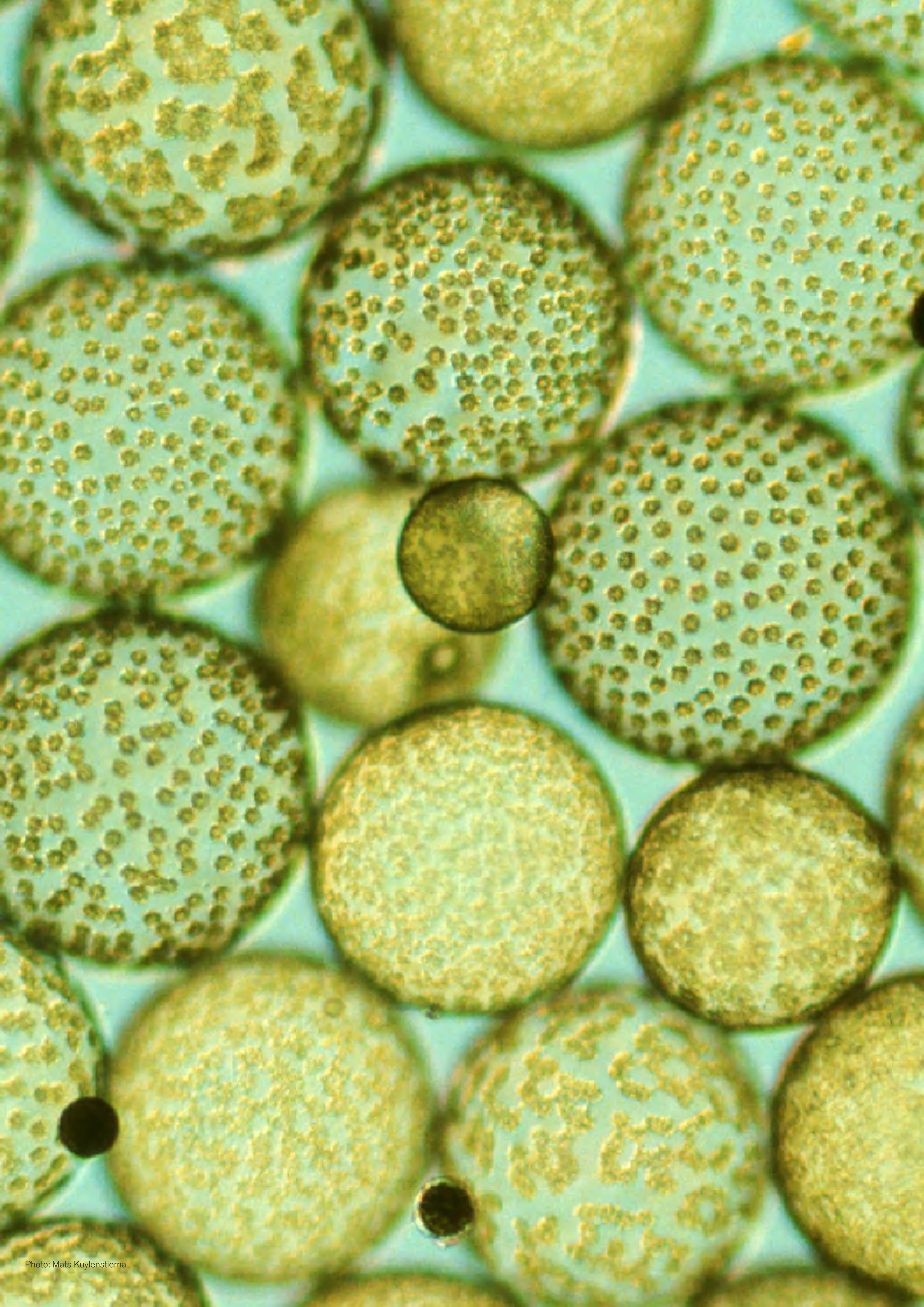
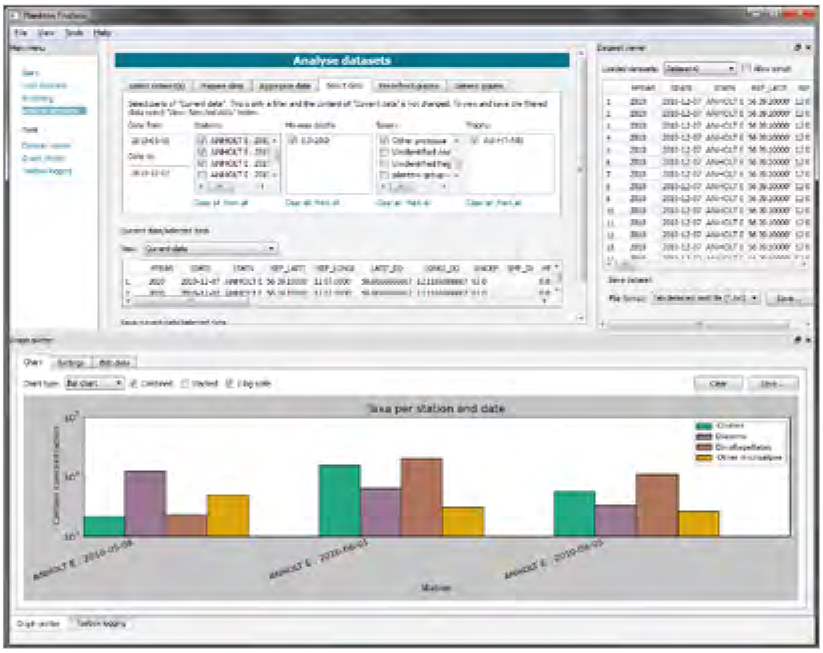


Photo: Mats Kuylenstierna

Plankton Toolbox – open source software making it easier to work with plankton data



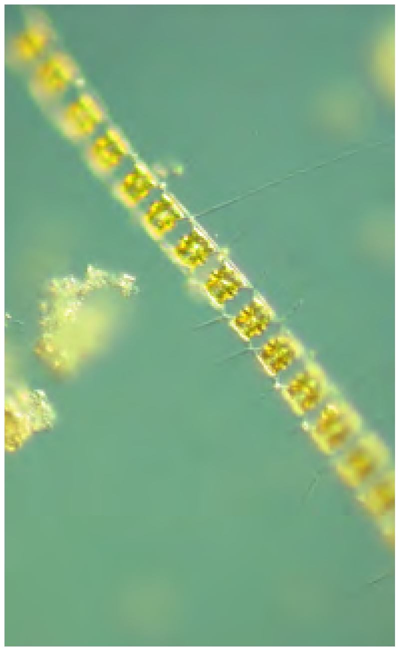


Photo: Ann-Turi Skjevik

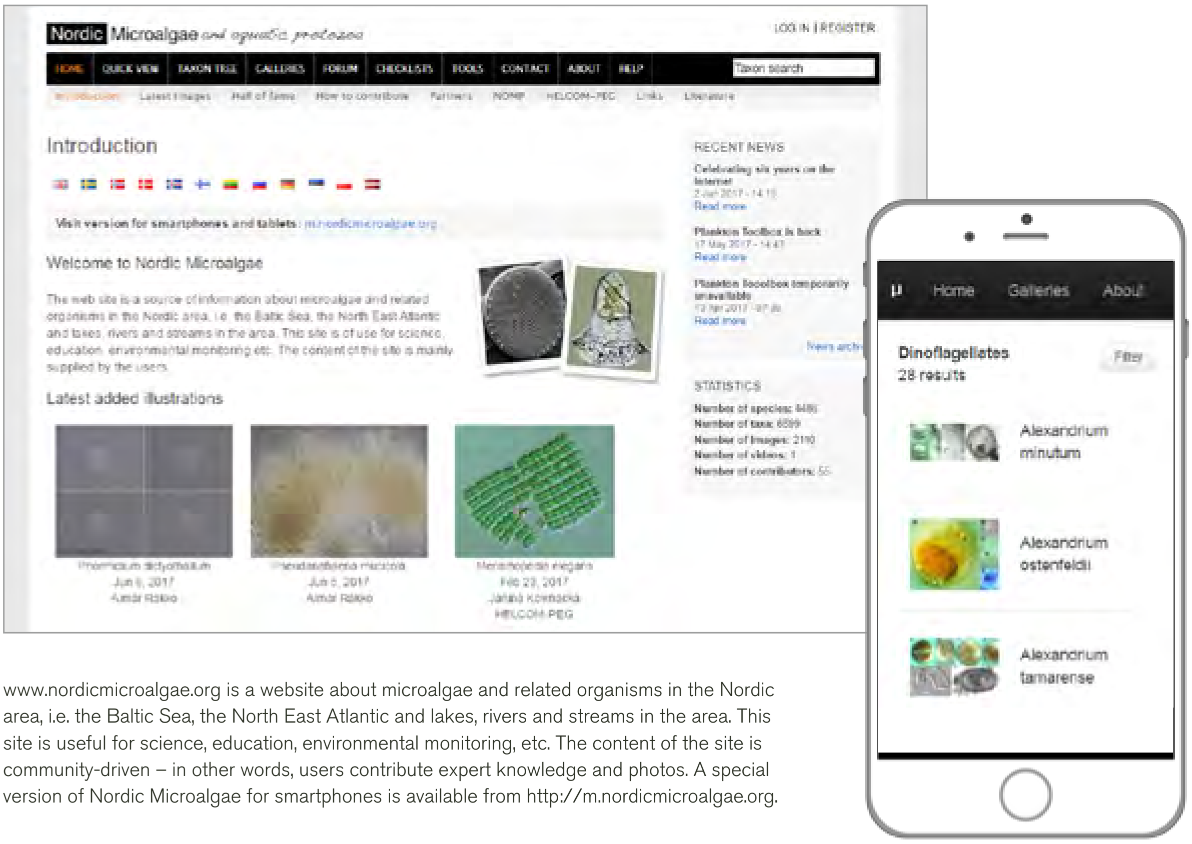
The Plankton Toolbox is open source software designed to make it easier to work with phyto- and zooplankton data. The software is freely available at http://nordicmicroalgae.org/tools.

Phytoplankton form the base of the marine food web. Modern database systems make it possible to access large datasets on the occurrence of phytoplankton and zooplankton, e.g. at the Swedish Oceanographic Data Centre, http://sharkweb.smhi.se. This provides insights into geographical distribution of species, occurrence of harmful algae, bloom development, etc. Datasets on plankton biodiversity and abundance, including harmful algal bloom species, may be time-consuming to work with. As part of Swedish LifeWatch we have developed the Plankton Toolbox, an open source, standalone application, to facilitate the process. The software is mainly aimed at working with data from water samples analysed microscopically, but it is applicable for other data as well.

The workflow includes importing data, a quality control step, ways to filter out data for specific purposes, functions for aggregating data, plotting tools and several tools for statistical analyses. Various data import and export formats are supported. For advanced statistical analyses, exporting to other software such as R is useful.

Another feature of the Plankton Toolbox is the counting module. This is used by microscopists/planktonic specialists when analysing samples using a microscope. The system is based on quality-controlled species lists available at http://nordicmicroalgae.org, which are cross-checked with AlgaeBase and the World Register of Marine Species. Users are free to work with their own checklists if they so wish. Cell volume and species carbon content lists are taken from the HELCOM Phytoplankton Expert Group for the Baltic Sea and the Nordic Marine Phytoplankton Group (NOMP) for the North Atlantic. Users can also define their own lists.

The software was developed using Python and runs on personal computers using a Windows or MacOS operating system. The software is freely available at http://nordicmicroalgae.org/tools.



www.nordicmicroalgae.org is a website about microalgae and related organisms in the Nordic area, i.e. the Baltic Sea, the North East Atlantic and lakes, rivers and streams in the area. This site is useful for science, education, environmental monitoring, etc. The content of the site is community-driven – in other words, users contribute expert knowledge and photos. A special version of Nordic Microalgae for smartphones is available from http://m.nordicmicroalgae.org.

Nordic Microalgae – community-driven expert platform and taxonomic updates

A collaboration between Nordic microalgae specialists has been developed as part of Swedish LifeWatch. As part of this project, the Nordic Microalgae website (http://nordicmicroalgae.org) was set up and a first version was released in 2011. The website includes taxonomy, images, traits and other information on microalgae. Nordic Microalgae includes 4486 species and 2109 images at present (May 2017).

Information on factors such as size, cell volume, carbon content and trophic type (phototrophic, mixotrophic and heterotrophic) are provided by Nordic microalgae researchers. There are two main networks: the Helsinki Convention Phytoplankton Expert Group (HELCOM-PEG) and the Nordic Marine Phytoplankton Group (NOMP). Both groups meet annually. There are also networks for freshwater microalgae.

LifeWatch has contributed to several workshops on microalgae, harmful algae and the use of Nordic Microalgae and the Plankton Toolbox.

Nordic Microalgae also includes information on harmful microalgae. This information comes from the Intergovernmental Oceanographic Commission (IOC) Intergovernmental Panel on Harmful Algal Blooms (IPHAB), which maintains a list of harmful (toxin-producing) algae.

Taxonomy of microalgae

In research is critical to use correct and consistent names of organisms. Changes in the taxonomy of unicellular eukaryotic organisms such as many phytoplankton and microzooplankton may be difficult for the non-specialist to follow. Relatively new gene-based data is resulting in new insight into the phylogeny and taxonomy of organisms. The global AlgaeBase is used to keep track of changes in taxonomy and nomenclature relating to algae. The World Register of Marine Species includes information from AlgaeBase, with the aim of providing a consistent database for all marine organisms. The Swedish Species Information Centre (ArtDatabanken) and the Norwegian Biodiversity Information Centre (Artsdatabanken) update update Dyntaxa and Artsnavnebase using information from AlgaeBase which has been checked by Nordic microalgae specialists.

DINA – Digital Information System for Natural History Data

DINA is a web-based collection management system developed by the Swedish Museum of Natural History in cooperation with partners and consortium members outside Sweden (currently Canada, Germany, Denmark and the UK) and includes a number of components for efficient management of the database, integration of sequence databases and tools for management of loans and inventory data and displaying species information (see https://dina-web.net/). It also includes a reference library-based DNA key. In Sweden, DINA will be applied over the next few years to all major Swedish natural history collections and also supply data to the Analysis Portal.

iStockphoto

rAquaMaps

Using resources from both Swedish LifeWatch and EU-BON, the Swedish Museum of Natural History has continued its development of an AquaMaps modelling tool in R which was initiated under the Swedish LifeWatch project. AquaMaps (http://aquamaps.org) is a visual modelling tool, combining niche modelling and expert knowledge. One concern with AquaMaps has been the relatively slow computation time and limited compatibility with open source spatial software, both factors relating to the database format. Our solution is rAquaMaps, a standalone R application based on the published AquaMaps algorithm and associated data sources. rAquaMaps reduces computation time significantly and is built using an open framework, promoting long-term sustainability. The rAquaMaps package provides advanced users with enormous modelling flexibility and will be made available both as a web service and via an easy-to-use web interface. The first beta version was released in 2015.

Research Data Storage

A considerable volume of data generated by researchers tends to remain on researchers' own computers and will eventually be lost. Nowadays, journals and research councils are increasingly requiring data generated by scientific studies to be stored centrally and made readily accessible. This is often considered by some researchers to be a strenuous, time-consuming task.

To reduce the burden, Swedish LifeWatch has initiated and funded Artportalen functionality enabling researchers to create additional custom data fields for their own studies. Thus there are now three configurable data field types available that can be used in any combination; project-specific value lists (containing parameters that have certain allowable values, or a yes/no choice), additional numeric values and plain text fields. This development has provided Artportalen with enhanced data storage flexibility and the ability to meet the often complex and unique requirements of research projects or inventories. Moreover, uploaded data can be tagged with a time embargo for public access so that the researcher can publish the information before releasing the data. This gives researchers a database system for their studies that is almost ready to use, but all the data entered is stored automatically and can be made publicly accessible.

iStockphoto

Core Web Services

At the core of the Swedish LifeWatch infrastructure is a network of web services for data harvesting, taxon handling, analysis and user administration. With a service-oriented architecture, resources can be easily reused and combined in many types of applications.

SONJA LEIDENBERGER, SLU AND UNIVERSITY OF GOTHENBURG

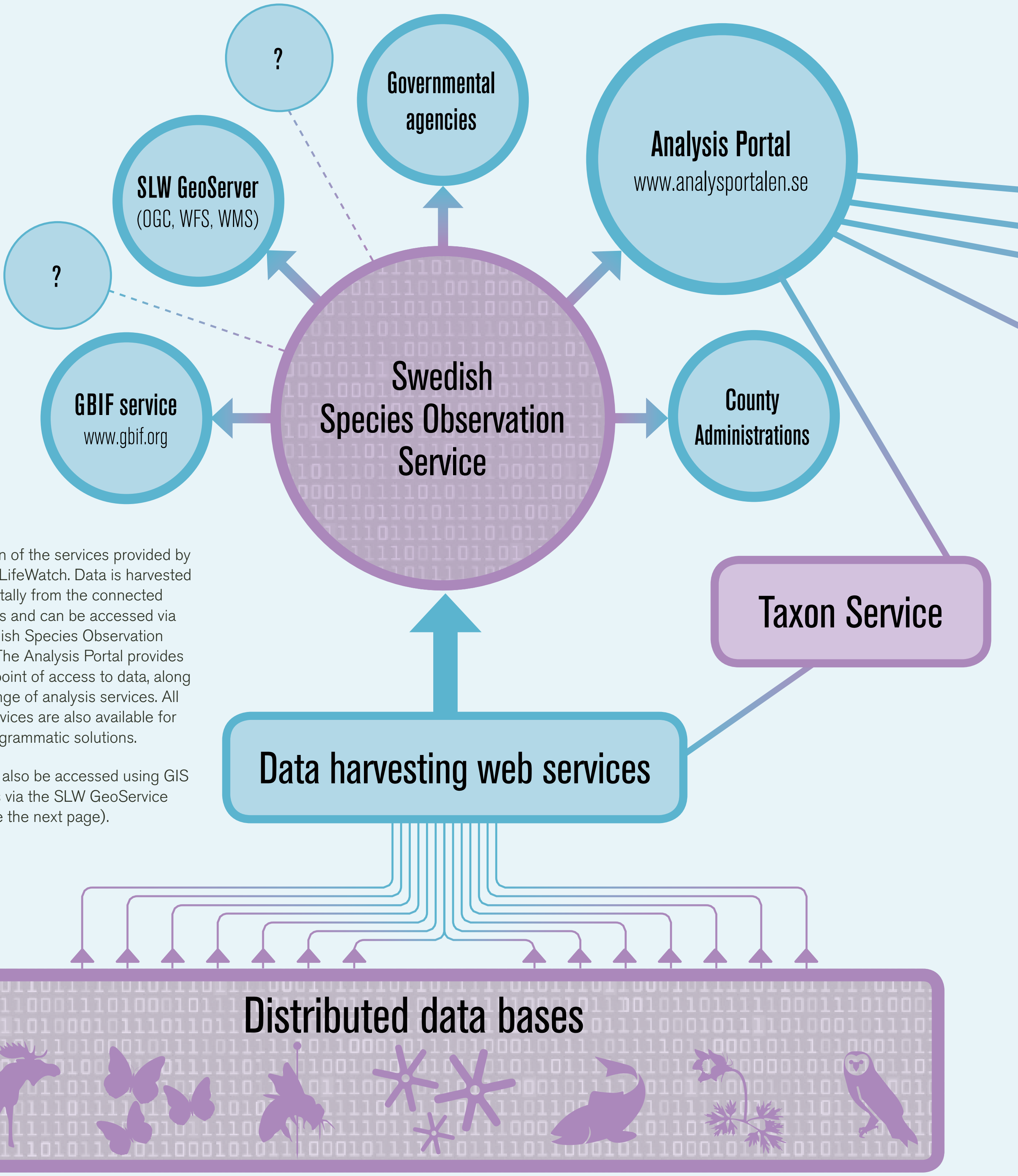


Illustration of the services provided by Swedish LifeWatch. Data is harvested incrementally from the connected databases and can be accessed via the Swedish Species Observation Service. The Analysis Portal provides an easy point of access to data, along with a range of analysis services. All these services are also available for other programmatic solutions.

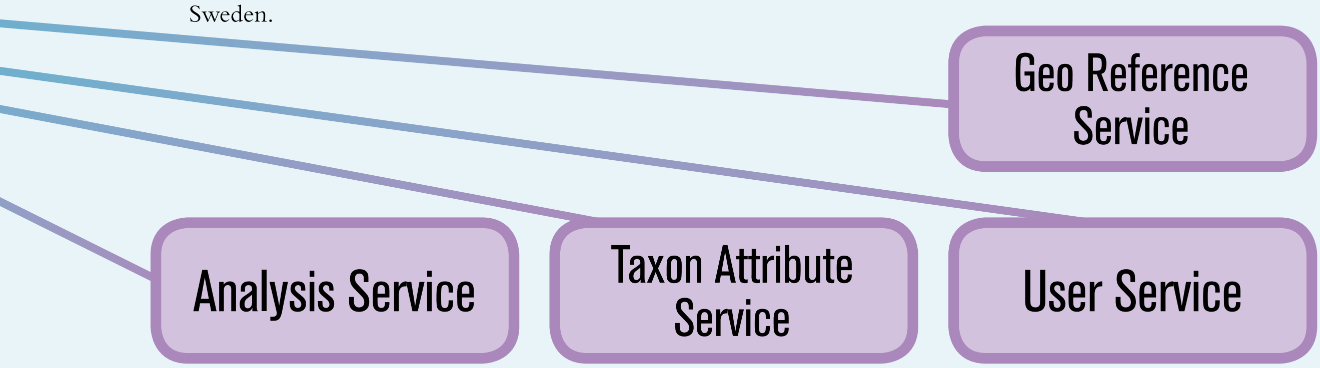
Data can also be accessed using GIS programs via the SLW GeoService WFS (see the next page).

Illustration: Katarina Nyberg

The basic concept of the LifeWatch infrastructure involves handling both data retrieval and analytical processing by means of web services. All core end-user applications are constructed on the basis of web services, making it possible to integrate the same services in an infinite number of alternative applications in the future.

Swedish Species Observation Service

This service constitutes the main focal point of Swedish biodiversity data in terms of species observations and occurrence data. It provides a couple of methods that can be used to retrieve species observations originating from several data sources such as Darwin Core records. It also has methods that support incremental harvesting of observation data. One of these methods is used by GBIF to harvest public occurrence data from Sweden.



User Service

This is the main service used by the UserAdmin web application, which is the main Swedish LifeWatch administration tool for authorisation. The service supports generic construction of application-specific authority objects and user roles which can be used to regulate which users are allowed to retrieve or edit data of various kinds. The service supports data handling regulation based on taxonomic, temporal and spatial specifications. It also supports delegation of authorisation in a rather simple yet effective way, making it possible – for example – for certain members of the Swedish County Administrations to authorise their own trusted employees to handle sensitive data within their geographical regions. The User Service is used by all other core services for authentication and to check user roles and authorisations.

Taxon Service

This is the main service used by the Dyntaxa web application (www.dyntaxa.se) to search and display taxonomic information on practically all multicellular taxa occurring naturally in Sweden. This service also provides all methods utilised by Dyntaxa for all content editing in the Swedish Taxonomic Database. The Taxon Service provides the infrastructure with globally unique identifiers for all the taxonomic concepts handled by Swedish LifeWatch core web services. The Dyntaxa taxon ID, which is the name of the taxon concept ID in Dyntaxa, provides the main link between the different data types essential for biodiversity in Sweden, such as species observations, taxon names and hierarchies, taxon traits and other attributes, as well as information on legislation and Red List status.

Taxon Attribute Service

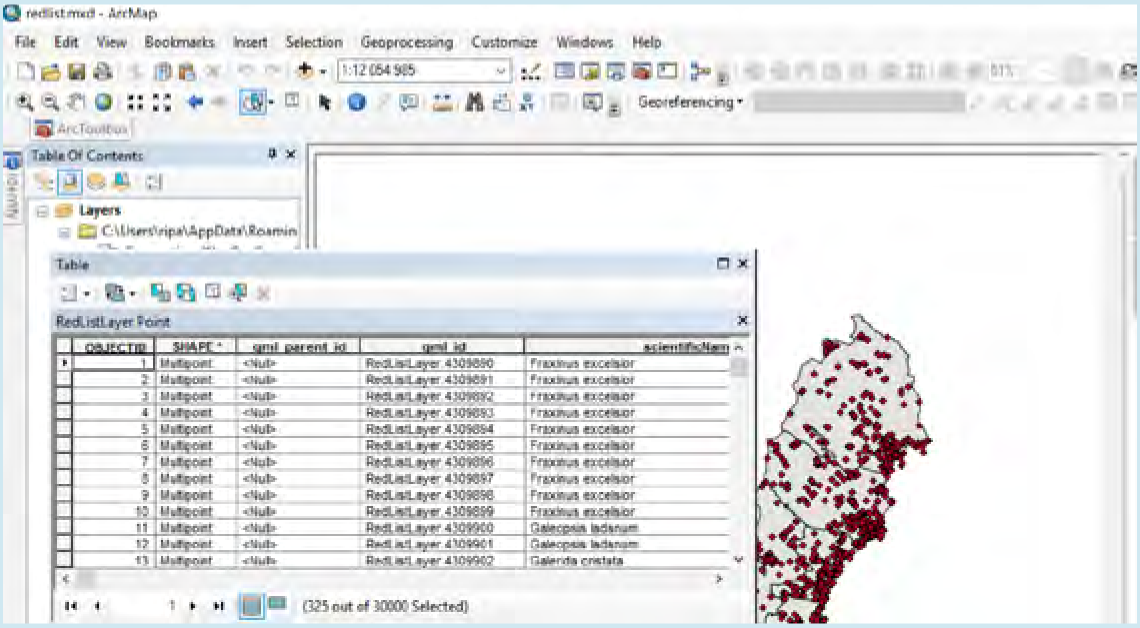
This service has the potential to handle all taxon generalisations, including habitat and substrate preferences and usage, interspecific interactions, life history traits, threats, Red List classification and legislation. It handles more than 2000 factors that are evaluated in relation to the Swedish taxa provided by the Taxon Service. Within the scope of LifeWatch, the service is used mainly for retrieving taxon lists determined by different factors or combinations of factors and taxonomic hierarchies.

Analysis Service

Unlike the other services, this service does not handle any particular type of data. Instead, it is dedicated to all sorts of data processing or data retrieval tasks that involve transformation of data types from their basic representation to something else. Most functions are related to species observations in one way or another, and many of these functions are controlled using the same search criteria (WebSpeciesObservationSearchCriteria) as those used when retrieving species observations or Darwin Core records from the Swedish Species Observation Service. Instead of providing records per se, the Analysis Service supplies summary statistics aggregated taxonomically for species observations. The Analysis Service also includes a number of processing methods that use data from OGC WFS. These methods can be used for calculating grid statistics based on the features in a specified data layer.

Geo Reference Service

Information on existing Swedish regions of different kinds, such as counties and municipalities.



Data access in ArcGIS and QGIS

For experienced GIS users, a web feature service is a convenient way of accessing data directly in ArcGIS or QGIS. Swedish LifeWatch occurrence data is provided in a geoserver service.

MARIA BARRET RIPA, ARTDATABANKEN SLU

Community and municipal planners, foresters and other end-users are often experienced in the use of GIS tools such as ArcGIS or QGIS. These tools provide them with data, maps, analysis models and other elements that they can combine with data such as species observations from Swedish Life-Watch. A Web Feature Service is a standardised way of accessing data and adding it to existing systems where end-users create their own specific analyses, maps or reports.

The WFS Standard provides an interface allowing requests for geographical features via the web using platform-independent calls. A geographical feature is more just than a map image as it consists of primary data, including the coordinates and attributes linked with, permitting spatial analysis by the end-user. GeoServer is a product used commonly to implement the WFS standard.

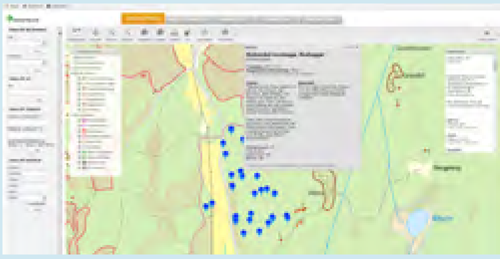
The GeoServer instance used in Swedish Life-Watch publishes a dataset called All Swedish Occurrences, which includes all public occurrences data harvested by the project in the Darwin Core format so that users can access it. GeoServer supports access to metadata, data and a method for listing all accessible datasets.

Two processed datasets has been published in addition to the complete dataset mentioned above.

These are *OccurrencesAndTaxaCountPer10KmGridCell* and *OcurrencesCountPer10KmGridCellAndTaxon* where data has been aggregated to a 10 km grid in two different ways. This is also a dataset where observations of all red-listed species have been extracted to a specific dataset called RedListLayer. The data is maintained by a program that harvests the Swedish LifeWatch database on a nightly basis and recreates the other three datasets each week. Metadata is published to the Swedish metadata node Geodata.se.

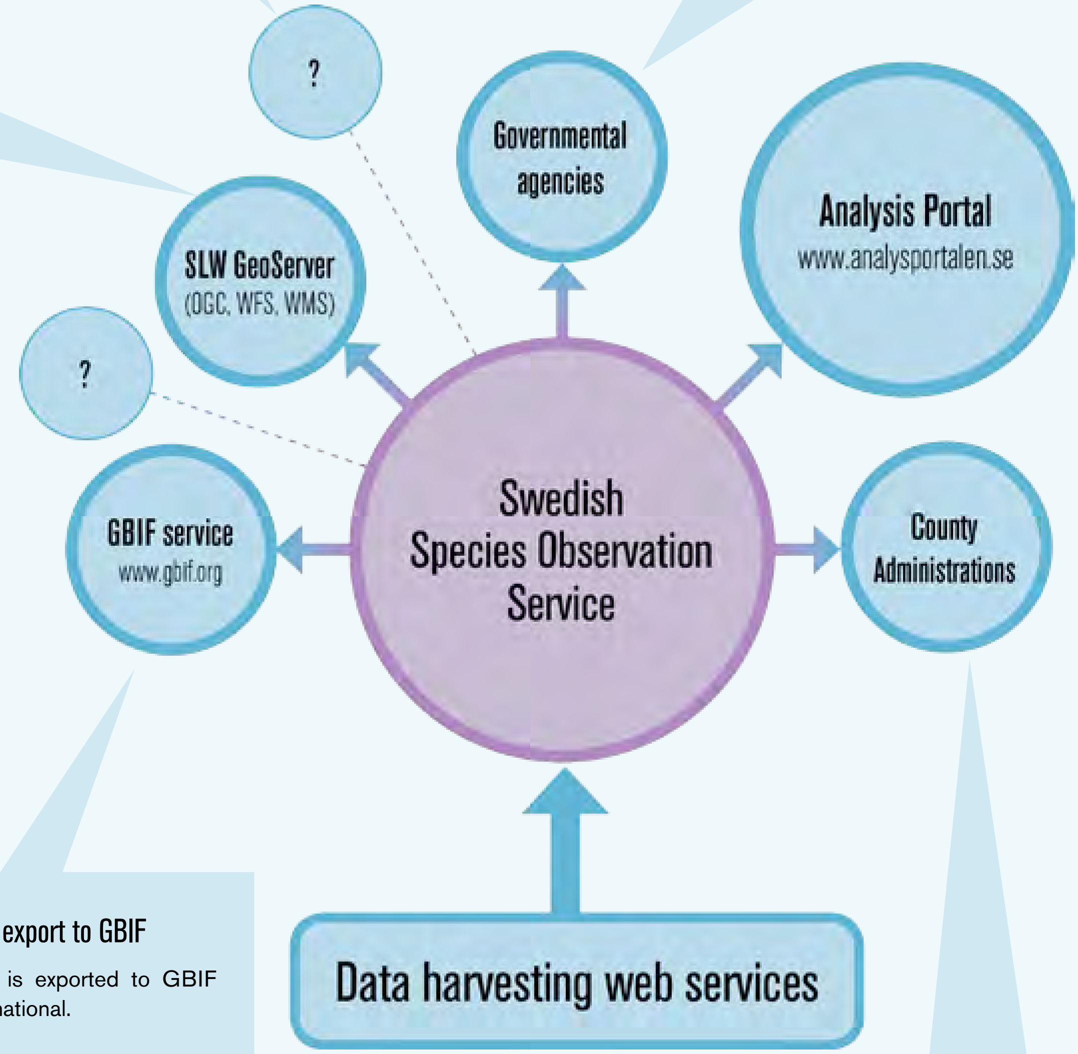
New applications

As the SLW infrastructure is modular, units communicating and remaining accessible via web services (APIs), data and tools can be used and combined in many contexts besides the Analysis Portal. Thus researchers (or enterprises, NGOs, governmental organisations, etc.) can have direct access to any part of the infrastructure for development of their own applications.



Governmental agencies

Landowners connected to the Swedish Forest Agency's web where they can see which species are known from their properties and also get tailored information about red-listed species and advices how to manage their forests with respect to these.



Data export to GBIF

Data is exported to GBIF international.

County Administrations

SLW data is integrated into the GIS handling systems of all 21 Swedish county administrative boards and many municipalities, enabling faster and better planning and decisions.

LifeWatch – a European vision for the 21st century

The LifeWatch vision was formalised in January 2011, when representatives of five European countries signed the first Memorandum of Understanding. The idea was to construct a European research e-Infrastructure project for biodiversity science and ecosystems research (www.lifewatch.eu).

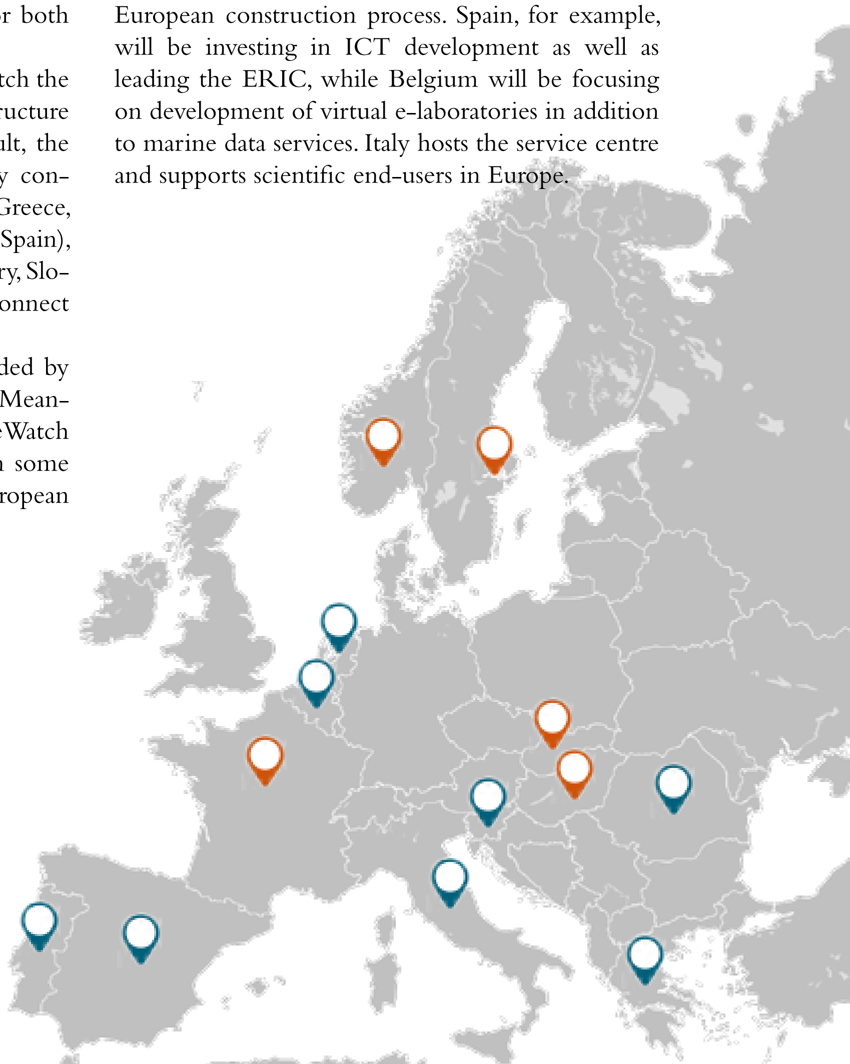
The concept for a European e-Infrastructure for biodiversity and ecosystem research came into being in January 2011, when representatives from Hungary, Italy, the Netherlands, Romania and Spain signed the first Memorandum of Understanding. The idea was to construct a distributed architecture that links existing information systems in Europe, making important biodiversity and ecosystem data available to European scientists. Important elements in a distributed infrastructure of this kind are 1) information and communication technology (ICT) services that host and connect data and tools, 2) virtual e-laboratories for scientists to create and share their data, analyses and models, and 3) a thorough capacity building and support programme for both scientists and developers.

The European Commission granted LifeWatch the official status of a European Research Infrastructure Consortium (ERIC) in early 2017. As a result, the infrastructure is now a legal European entity consisting of seven member countries (Belgium, Greece, the Netherlands, Italy, Romania, Portugal and Spain), while some countries (Finland, France, Hungary, Slovakia, Slovenia and Sweden) have decided to connect to the ERIC as observers initially.

The inauguration of the ERIC was preceded by Italy setting up a service centre back in 2012. Meanwhile, activities and development of the LifeWatch concept have taken place at national level in some countries as well. Sweden was the first European country to build a standalone national LifeWatch e-Infrastructure for biodiversity data. All Swedish LifeWatch services and applications have been designed to become part of the European infrastructure in the future. Belgium in particular has focused on marine biodiversity, but also sensor data, a taxonomic backbone for LifeWatch as a whole and certain analysis and visualisation tools. Greece is also focusing on marine considerations and provides several services and virtual laboratory analyses. Portugal has started working with certain functions. Other member countries, as well as Norway, are also planning to set up national LifeWatch nodes.

Many countries will have a thematic focus in the European construction process. Spain, for example, will be investing in ICT development as well as leading the ERIC, while Belgium will be focusing on development of virtual e-laboratories in addition to marine data services. Italy hosts the service centre and supports scientific end-users in Europe.

LifeWatch



Members of LifeWatch ERIC

Observing or collaborating countries

GEO BON (Group on Earth Observations)

The Group on Earth Observations Biodiversity Observation Network (GEO BON) is an initiative aimed at improving the availability of biodiversity change data to decision-makers and scientists in support of policy. GEO BON initiates and coordinates efforts to design and implement interoperable national and regional biodiversity monitoring programmes. GEO BON supports the sharing and dissemination of information and technology for biodiversity observations through its global network of organisations and experts. GEO BON also supports the application of the most recent scientific knowledge to advance collection, integration and interpretation of biodiversity observations. GEO BON focuses on developing a network of observation systems that supplies enhanced and harmonised biodiversity information to facilitate better decision-making from local to global levels. GBIF-Sweden and the Swedish Museum of Natural History are Sweden's GEO BON representatives.

EU-BON

The Swedish Museum of Natural History has also contributed to the EU BON project, which has created an EU-funded European infrastructure aimed at developing platforms for interoperability between GEO BON and GEOSS (the Global Earth Observations System of Systems). In March 2017, the final EU BON meeting served as a platform for the presentation of key outputs from the FP7 EU-funded project EU BON "Building the European Biodiversity Observation Network". EU BON represents a joint effort of 31 partners from 15 European countries, Israel, the Philippines, Brazil and more than 30 associated partners. The project has worked on establishment and adoption of new data standards, development of tools, integration of advanced data analysis techniques and development of new approaches and strategies for future biodiversity monitoring and assessment. One major outcome has been the launch of the European Biodiversity Portal (http://biodiversity.eubon.eu), which offers access to biodiversity observations and ecological data, statistics and analyses of changes over time, along with tools for sharing or discovering data and products generated by scientific and analytical processes.

EU BON

Photo: Diliff via Wikimedia Commons,

Biodiversity Informatics Horizons 2013

The international conference BIH2013 (Biodiversity Informatics Horizons) was held in Rome by LifeWatch and 17 other related initiatives in September 2013. The aim of the conference was to structure the biodiversity informatics community and prepare for the release of funding calls for Horizon 2020 in a spirit of cooperation rather than competition.

180 participants, including more than 40 speakers, spent four days reviewing challenging areas and promising technologies in biodiversity informatics, pathways to sustainable implementation and changes to the community culture. One important outcome of the conference was the definition of common goals for future collaborations under the LifeWatch umbrella.

Marine LifeWatch working groups

There is a strong community of marine scientists and developers within the European LifeWatch initiative. These marine working groups work in close cooperation with other ESFRI programmes such as the European Marine Biological Resource Centre to further develop and integrate existing information systems for marine biodiversity research. Examples are the World Register of Marine Species, the European Ocean Biogeographic Information System, the European Marine Observation and Data Network, the Marine Regions standard list and the World Register of Introduced Marine Species. The European Marine Data Centre in Ostend coordinates the working groups, and Sweden plays a dominant role in this process by contributing the Swedish Ocean Archive (http://sharkdata.se), the Nordic Microalgae information system (http://nordicmicroalgae.org) and development of a marine analysis portal.

NeIC Biodiversity

Nordic-Baltic collaboration on e-Einfrastructure for Biodiversity Informatics

The informal Nordic LifeWatch Network and the Nordic eInfrastructure Collaboration (NeIC) initiated a collaboration in 2014. NeIC facilitates the development and operation of high-quality e-Infrastructure solutions in areas of joint Nordic interest.

FRANK HANSEN, NINA (NORWEGIAN INSTITUTE FOR NATURE RESEARCH)

There has been fragmented development of biodiversity and ecosystem e-Infrastructures in the Nordic-Baltic countries over the last 16 years. Given the different priorities and funding regimes, the relevant countries are currently at different levels of maturity in terms of mandatory support, implementation and technological development.

The informal Nordic LifeWatch Network and the Nordic e-Infrastructure Collaboration (NeIC) initiated a collaboration in 2014. NeIC facilitates the development and operation of high-quality e-Infrastructure solutions in areas of joint Nordic interest. Building on the findings in an unpublished Nordic LifeWatch report (2014) and a subsequent Nordic-Baltic survey (2015), NeIC and the Nordic LifeWatch Network organised a Nordic Workshop on e-Infrastructures for Environmental Research in Oslo in 2015.

NeIC and a team of Swedish, Norwegian, Finnish, Icelandic, Danish and Estonian research institutions initiated a NeIC Biodiversity Informatics project (2017-2020) based on identified needs. This project is closely related to the NeIC strategy focus areas. The overarching aim of the project is to facilitate and intensify collaboration on e-Infrastructure matters for biodiversity informatics in the Nordic-Baltic region, and so the project will establish strong links with national and regional activities. This will help to improve coordination of common efforts, reduce redundant investments and efforts and boost future regional scientific achievements and global competitiveness. In turn, this will also provide educational support for e-Science and strengthen the Nordic-Baltic science policy interface.

The project will focus on i) coordination and alignment of RI competence (WP1), ii) knowledge sharing and competence development (WP2), and iii) exploring technologies for compilation, analysis and visualisation of biodiversity data (WP3). The major value added by the project is summarised below:

- A better understanding of national strategies, priorities, competence levels and gaps in order to address relevant measures for closing the gaps and aligning national RI strategies
- A Nordic-Baltic Virtual Support Centre pilot for capacity building in biodiversity and ecosystem RI
- Workshop on mechanisms for integration of national taxonomic backbones
- A common knowledge platform on current training activities and documentation for data storage, data publication and data sharing.
- Guidelines on how to publish taxonomic information as Linked Open Data on the web
- Guidelines on how to compile, analyse and visualise standardised biodiversity data (observation and event data) already shared via GBIF and similar open data infrastructures
- Guidelines for the development of access policies and long-term support of services
- An overview of existing visualisation services and technologies, and the potential for making these interoperable
- Facilitated workshops on collaboration and shared development. The aim of these workshops is to develop a common understanding of the status, needs and feasibility of shared interoperable Nordic-Baltic portals and to explore practical solutions (in terms of organisation and sharing resources)

Contact person: Dr. Thomas Röblitz (NeIC)

Nordic pre-study funded by NordForsk

In 2012, NordForsk provided funding for a pre-study aimed at investigating the opportunities for establishing a joint Nordic LifeWatch. A Nordic LifeWatch collaboration was formally initiated during a start-up meeting held in Stockholm in November 2012.

Representatives from Sweden, Norway, Denmark, Finland and Iceland gathered for a second meeting and workshop in Akureyri, northern Iceland on 6-8 May 2013 to formulate a joint Nordic LifeWatch vision and work on the final report. The scope of the Nordic LifeWatch pre-study was to identify the scientific potential of a common Nordic infrastructure based on inventories of user needs, existing data repositories and challenges and constraints related to data sharing in general. Based on these findings, a proposal was developed for funding and strategies relating to a Nordic LifeWatch construction phase in close collaboration with stakeholders (research councils, ministries and scientific communities), national LifeWatch consortia, the European LifeWatch Service Centre and relevant parallel initiatives.

A final report from the pre-study was published in 2014:

Hanssen F (ed), Heggberget T (ed), Bladt J, Endresen D, Forsius M, Gudmundsson G, Gärdenfors U, Heiðmarsson S, Kindvall O, Koch W, Koviula K, Laiho EL, Laine K, Obst M, Skov F, Telenius A, Valland N, Wasowicz P, Wremp AM (2014) Nordic LifeWatch Cooperation. Final report. 68 pp.

Baltic-Nordic collaboration

Marine organisms do not perceive national boundaries. Collaboration in respect of marine biodiversity started early on, and there is ongoing collaboration in the Nordic countries and the Baltic States with regard to marine biodiversity. This collaboration has been developed further as part of the LifeWatch project. The HELCOM Phytoplankton Expert Group uses the Nordic Microalgae website for documentation of organisms by means of photography. The website is also used to distribute information about the traits of Baltic phytoplankton. The Nordic Marine Phytoplankton Group, which covers the Kattegat, Skagerrak and North Sea area as well as the Norwegian Sea, the Barents Sea and the Greenland Sea, also uses Nordic Microalgae and the Plankton Toolbox software described elsewhere in this report. There is also collaboration in respect of freshwater microalgae, such as NORBAF – the Nordic-Baltic Network for Benthic Algae in Freshwater.

Representatives from Sweden, Norway, Denmark, Finland and Iceland met for a workshop in Akureyri, Iceland 6-8 May 2013


Photo: Louise Tenn


Photo: Mikael Axelsson

3 questions for Anders Telenius, GBIF-Sweden node manager at the Swedish Museum of National History

What tools and services have been developed by the Swedish Museum of Natural History as part of Swedish LifeWatch?

– The services developed at the Swedish Museum of Natural History based on funding from Swedish LifeWatch include components for the DINA collection management system, as well as a modernised underlying structure that enhances the functionality of AquaMaps, which is an existing tool for visualisation and modelling of the distribution of marine organisms. Many of the data sources found in the Analysis Portal were originally presented by GBIF-Sweden as well, which is hosted by the museum.

LifeWatch involves more than just tools and services – networking and outreach, to name just two examples. In what way does the Swedish Museum of Natural History contribute in this regard?

– It is crucial for a modern national dynamic biodiversity infrastructure to successfully stay abreast of international efforts aimed at creating standards and systems for user interfaces and any underlying systems management tools. The Swedish Museum of Natural History is contributing to the Swedish LifeWatch infrastructure by building a variety of elements that are more or less clearly linked to one another, and that puts them in perspective when considering ongoing development by our colleagues with similar agendas in other countries.

The Swedish Museum of Natural History interacts in a global biodiversity informatics context. Which contacts are particularly important as regards your collaboration with Swedish LifeWatch?

– As well as participating in the practical efforts of the international organisation for standardisation of biodiversity informatics tools and processes, Biodiversity Information Standards/TDWG provides one of the most important elements in this context; the activities of GBIF-Sweden, the Swedish element of the international GBIF (Global Biodiversity Information Facility).

– GBIF-Sweden also represents Sweden in GEO BON (Group on Earth Observations – Biodiversity Observation Network). System developers at the museum have also contributed to the EU BON project, by means of operations funded by Swedish LifeWatch, thereby helping to create an EU-funded European infrastructure aimed at developing platforms for interoperability between GEO BON and GEOSS (Global Earth Observations System of Systems). This includes creating and enhancing tools for assessment of biodiversity data, analysis, visualisation and publication of information on biological diversity, linking biological and other environmental data at national and regional level and interacting with citizens, businesses and governments.

GBIF
www.gbif.org

GBIF-Sweden is the national node of Global Biodiversity Information Facility (GBIF), hosted by the Swedish Museum of Natural History.

Photo: PhilAugustavo / iStockPhoto

Marine invasive species

Swedish LifeWatch data and BioVeL workflows have been used in two studies to establish a risk assessment and monitoring tool for invasive species spreading with the ballast water of trading ships.

SONJA LEIDENBERGER & MATTHIAS OBST, UNIVERSITY OF GOTHENBURG

Biological invasions have dramatically increased with the development of global trading, causing the homogenisation of communities and the decline of biodiversity. Ballast water exchange from shipping and introductions from aquaculture are the two main vectors of invasive species in the marine environment, rendering eradication very difficult. Modelling approaches are invaluable for predicting the impact of potentially invasive species before they establish themselves as breeding populations, thereby allowing preventive measures to be put into place.

Predicting habitat suitability with models built from distributed data resources

Scientists at the University of Gothenburg have developed a set of automated analysis pipelines for data mobilisation, ecological niche modelling and statistical analysis which can be reused to predict marine invasion paths in relation to shipping (Leidenberger et al., 2015a) for any species and any region in the world. These data and modelling services are available through Swedish LifeWatch and the BioVeL infrastructure. In this scientific study, habitat suitability in the Baltic Sea and the Northeast Atlantic was analysed for a 'black species list' of 18 marine invasive species in Northern Europe, divided into four ecological groups: zoobenthos, phytobenthos, zooplankton and phytoplankton. A taxonomic data service was used to mobilise more than 22,000 occurrence records from public databases and integrate these with observations from reference works. Suitable habitats were modelled using ecological niche modelling algorithms, and species distributions were subsequently analysed statistically.

The study found several potential risk zones (hotspots) for invasive species in the Skagerrak and Kattegat region, a transitional area for invasive species entering the Baltic Sea. Likewise, a number of 'cold spots' with a low risk of invasive species spread were found in the Bothnian Bay. The model-based results have been compared to traditional risk assessment methods based on salinity matching in order to assess the risk of spread along a sample shipping route (Gothenburg to St. Petersburg). The comparison study highlights the potential of e-science approaches when providing scalable tools for rapid integration of biodiversity data and producing predictive models that improve the prevention and management of marine invasions.

Similarly, a report submitted to the European Space Agency in 2013 used Swedish LifeWatch data together with BioVeL services to establish a semi-automated risk assessment and monitoring tool for the spread of invasive species with the ballast water of trading ships (Stelzer et al., 2013). In particular, this study explores the potential of combining satellite (Earth Observation) data with biological data for prediction of biological invasions caused by certain shipping routes. The results highlight the application potential of e-science approaches for effective integration of satellite environmental data with biological data. This report contributes to the implementation of the International Convention for the Control and Management of Ships' Ballast Water and Sediments (BMWC 2004), developed by the International Maritime Organisation (IMO). According to this legislation, all ships will have to treat their ballast water from 2018, but the regional convention programmes OSPAR and HELCOM suggested originally that there exceptions may be permitted for certain ships operating on short routes or in enclosed seas. However, such decisions need to be based on biological risk assessments of shipping routes, and this is where the data and analysis services offered via the LifeWatch e-Infrastructures may be very valuable.

Photo: Ron Offermans / Wikimedia commons

A common shipping route in Northern Europe was analyzed for the spread of invasive species through ballast water. Here, the potential of ecological niche modelling was tested as a component in marine non-indigenous species risk assessment in general. One of the species included in the study was Chinese Mitten Crab (*Eriocheir sinensis*).

Modelling distribution patterns of marine food chains under climate change

A recent scientific study used Swedish LifeWatch data and BioVeL services to compare the range shift among species in different levels of a Baltic food web. The results suggest that climate-induced changes will increase the grazing pressure to isolated populations of macroalgae in the Baltic Sea and thereby elevate the extinction risk for some of these species.

SONJA LEIDENBERGER & MATTHIAS OBST, UNIVERSITY OF GOTHENBURG

A recent scientific study (Leidenberger et al. 2015b) used Swedish LifeWatch data and BioVeL services to compare the range shift among species in different levels of a Baltic food web. The results suggest that climate-induced changes will increase the grazing pressure to isolated populations of macroalgae in the Baltic Sea and thereby elevate the extinction risk for some of these species.

The Baltic Sea is one of the world's largest semi-enclosed brackish water bodies characterized by many special features, including endemic species that may be particularly threatened by climate change.

Baltic habitats are often species-poor brackish environments. Hence the biological communities tend to be less complex and often consist of only a few key species. This makes Baltic communities interesting study systems for system-level biological responses to climate change. In this study researchers for the University of Gothenburg In this study, researchers from the University of Gothenburg, have together with international colleagues analyzed potential climate driven changes in distribution patterns of a benthic community that is connected in a food chain across three trophic levels. The studied community included macroalgae (*Fucus vesiculosus, F. radicans*), grazing crustaceans (*Idotea balthica, I. chelipes, I. granulosa*), and predatory fish (*Gasterosteus aculeatus*).

The authors examined possible impact on the food web from climate change scenarios in the Baltic Sea. For this purpose, two open-source analytical services were developed as a part of the BioVeL infrastructure: one for ecological niche modelling and another for raster layer comparison to compute the extent and intensity of change in species' potential distributions.

The results of the analysis suggest that habitat suitability for the dominant crustacean grazers in the Baltic Sea are largely determined by temperature and ice cover rather than by salinity. In addition, the predictions for the year 2050 show a northern/north-eastern shift in the Baltic Sea for all species included in the study, while the distribution ranges for some species (*Idotea granulosa*, *Gasterosteus aculeatus*) become patchier. Furthermore, in each trophic level species were identified who gain and loose suitable habitat in the future, a process that will distort the current balance in the food web. Finally, climate-induced changes will increase the grazing pressure to one of the isolated macroalgae populations in the northern part of the Baltic Sea and thereby elevate the extinction risk for this species.

Main conclusions

For the Baltic Sea, climate-induced changes resulted in a gain of suitable habitats for *F. vesiculosus*, *I. chelipes* and *I. balthica*, whereas lower habitat suitability was predicted for *I. granulosa*, *F. radicans* and *G. aculeatus*. The predicted north-eastern shift of *I. balthica* and *I. chelipes* into the distribution area of *F. radicans* in the Baltic Sea may result in increased grazing pressure. Such additional threats to isolated Baltic populations can lead to a higher extinction risk for the species, especially as climate changes are likely to be very rapid.

Sonja Leidenberger

is a researcher at the Department of Marine Sciences at the University of Gothenburg. After her PhD in marine ecology, she has worked with biodiversity informatics for BioVeL and the Swedish LifeWatch project. She contributed as project secretary (2015) and system manager for the Analysis Portal (2016). She has had an active role in teaching biodiversity informatics at university and to the public.

Vector-borne infections
and cost effective surveillance systems

This project has tested the basis of a risk-based surveillance system for vector-borne diseases using BioVeL workflows by combining ecological niche modelling for mapping the distribution of mosquitos with pathogen introduction and transmission models.

GUNNAR ANDERSSON[1], ANDERS LINDSTRÖM[1] & MATTHIAS OBST[2]

NATIONAL VETERINARY INSTITUTE (1) AND UNIVERSITY OF GOTHENBURG (2)

Mosquito-borne infections are some of the most important new and emerging diseases worldwide and in Europe. Diseases such as West Nile fever, chikungunya, dengue fever, the Usutu virus and the Sindbis virus have appeared in Europe for the first time or re-emerged over the past few decades. Globalisation and climate change provide pathogens and vectors with opportunities to colonise new areas. The need for mosquito surveillance on a European level has been highlighted in reports from European Centre for Disease Prevention and Control, ECDC, for example. Swedish mosquito fauna include at least eleven species that act as vectors for zoonotic diseases in other countries. Mass appearance of nuisance floodwater mosquitoes can seriously affect quality of life and cost huge amounts of money every year.

Active surveillance

Active surveillance of vectors and pathogens is expensive, as this requires large numbers of traps and repeated sampling over large areas in order to gain a good overview of the current situation. There is a high risk of missing an outbreak in an initial phase as these things begin with a few cases in isolated areas. Focused surveillance efforts in areas of

We can predict areas and periods in which potential disease transmission can occur by modelling climatic and environmental parameters specific to each pathogen and vector. We can increase the chances of detecting early-phase outbreaks by focusing surveillance efforts on areas and times when disease transmission is possible. This makes for more cost-effective surveillance and increases the chances of successful control.

Aedes cinereus. Photo Anders Lindström.

high risk due to vector presence and environmental parameters suitable for pathogen replication could dramatically increase the sensitivity and cost-effectiveness of surveillance.

The transmission of mosquito-borne diseases is largely determined by climatic and environmental parameters acting on both the vector and the pathogen. The likelihood of an outbreak can be estimated by models describing the biological processes relevant to disease transmission. We can predict areas and periods in which potential disease transmission can occur by modelling climatic and environmental parameters specific to each pathogen and vector. We can increase the chances of detecting early-phase outbreaks by focusing surveillance efforts on areas and times when disease transmission is possible. This makes for more cost-effective surveillance and increases the chances of successful control.

Mass appearance of mosquitos that are not disease vectors may still have a major impact on society and tourism in affected areas, and improving the ability to predict the appearance and abundance of major nuisance mosquitos will be helpful for local government and tourism and facilitate the planning of mosquito control.

Modelling vector borne disease (VICE)

Models indicating the risk of introduction and transmission of vector-borne diseases have been developed at DTU Vet (the National Veterinary Institute in Denmark) and CVI (the Central Veterinary Institute in the Netherlands) as part of the EMIDA-VICE project "Vector-borne infections and cost-effective surveillance". The transmission models use vector abundance and temperature input, whereas introduction models also need information on trade routes and winds, for example. The limiting factor for applying such models in Sweden is the availability of data for the spatial and temporal distribution of vector-competent mosquitos.

Mosquito distribution in Sweden

We used crowdsourcing for the Swedish part of the EMIDA-VICE project to gather mosquito distribution data throughout Sweden as part of a project entitled *Myggjakten*, the Mosquito Hunt. While providing a unique mosquito distribution dataset on a broader scale, this is actually just a snapshot. The abundance of vector-competent and nuisance mosquitos in any one area may differ dramatically from year to year, and ideally decision support would be based on real-time collection of mosquitos through a network of observation stations. This is not possible in practice. Firstly, the systematic collection and analysis of mosquitos at a large number of locations is not feasible from an economic standpoint; and secondly, the observation would represent only small points, and interpolation between points would be no trivial matter as mosquitoes are not evenly distributed in a region but associated with particular ecological niches and habitats, and some species appear following certain weather events such as heavy rainfall.

Pilot work with ecological niche modelling

Pilot studies using ecological niche modelling were conducted using Myggjakten data as part of the EMIDA-VICE project ending in 2015. Environmental data was extracted using an ecological niche modelling workflow available via the Biodiversity Virtual e-Laboratory (BioVeL), which is part of the Swedish LifeWatch infrastructure.

This workflow provides access to data on precipitation and temperature from (BIO1 – BIO19) WorldClim (http://www.worldclim.org), topsoil data from the Harmonised World Soil Database (HWSD) and Corine Land Cover (http://www.eea.europa.eu/publications/COR0-landcover).

The variables were used as-is for topsoil and WorldClim layers. The predictor variables for land

BioVeL workflows for ecological niche modelling

BioVeL workflows give researchers access to environmental and ecological data relevant to study correlations between species occurrences and certain environmental conditions. The data repositories accessed by the workflows offer a wide range of local and global data layers for both terrestrial and marine habitats, including variables relating to climate, soil, land use, hydrology, marine water quality, geography and tides.

BioVeL workflows have been used in a large number of scientific studies, including investigations of vector-borne infections (this article), marine invasions (Leidenberger et al., 2015a, Laugen et al., 2015), climate impact on marine food webs (Leidenberg et al., 2015b) and identification of coastal protection areas for endangered species (Vestbo et al., 2017). A detailed description of the workflows is provided by Hardisty et al., 2016.

WWW.BIOVEL.EU

cover were specifically translated for analysis of suitable habitats for mosquitoes. For this purpose, we recalculated the proportion of the land type of interest within a radius of 300 or 3000 metres from the Corine Land Cover trap site. Ecological niche modelling was executed on the BioVel portal, treating data as presence-only data. The data was also modelled in R using RandomForests and generalised linear models, treating data as presence-absence.

The pilot study indicated that ENM could be used to predict the distribution of some mosquito species, but also identified several caveats. Predictive performance was tested for the 21 most abundant species and was found to be low to moderate with ROC AUC, ranging from <0.5 to 0.69 for *Aedes rusticus*. Furthermore, some variables were identified, using Monte Carlo feature selection and linear regression, as significant for predictive performance. However, the results also showed that feature selection may result in severe overfitting, and that it is crucial to employ validation schemes that test for this.

Lessons learned and current work

The results of the pilot study suggest that Corine variables may not be ideal for modelling mosquitos. One reason for this may be that many classes of land cover will correspond to habitats that are very different. For instance, the class "coniferous forests" will include pine forests on dry eskers as well as damp "blueberry-spruce" forests. Most mosquitos breed in pools of water that are generally too small to be viewed as water bodies in the Corine Land Cover inventory. Although some topsoil variables such as gravel and clay content may correlate indirectly with soil drainage, we hypothesise that the lack of hydrological predictors is a weakness in the currently tested models.

Several topsoil variables were among the most significant factors, but we noticed that some selected features, such as carbonate soils and salinity, showed strong spatial autocorrelation, raising the suspicion that their appearance may be a result of confounding effects.

This work continues in the project "Risk-based vector surveillance from an authorities perspective", which is financed by the Swedish Civil Contingencies Agency (MSB). The aim of this project is to construct more accurate models of mosquito distribution in Sweden. More representative data will be obtained by exchanging data with neighbouring countries. Data from Sweden, Denmark and Germany has been complied in Darwin Code format to date. Layers representing hydrological variables are being constructed as part of a collaboration between BioVel and the Swedish Meteorological and Hydrological Institute (SMHI).

Gunnar Andersson

Gunnar Andersson is a researcher at the National Veterinary Institute (SVA). He comes from a background in experimental biology, including mycology and virology, at Uppsala University. He later shifted his focus towards bioinformatics and biostatistics. His research interests focus on decision-making using uncertain information.

Recent projects include sampling for hazardous substances in animal feed, time of death estimation with forensic entomology and surveillance of vector-borne diseases.

Anders Lindström

Anders Lindström is a researcher at the National Veterinary Institute (SVA). He is an entomologist who has worked in medical and veterinary entomology, as well as forensic entomology. His research interests include the distribution and abundance of Swedish mosquitoes and surveillance of vectors.

The role of Swedish LifeWatch in biodiversity citizen science

Swedish LifeWatch has an important part to play in facilitating the integration of citizen data in biodiversity research by improving scientists' awareness of and accessibility to citizen biodiversity data. Studies described here show that the use of a suite of modelling approaches allows citizen data to be applied reliably to land management conservation decision-making, demonstrating that citizen data made widely available through Swedish LifeWatch can also be a valuable forecasting resource.

MARI JÖNSSON, TORD SNÄLL & LOUISE MAIR, ARTDATABANKEN SLU

In the field of citizen science (CS), large data volumes are collected and classified by volunteer contributors as part of projects initiated by scientists. The largest current focus of CS is on research in the fields of biology, conservation and ecology, and citizen scientists' primarily collect and classify data (Kullenberg & Kasperowski, 2016). In many countries, national web service-oriented e-Infrastructures for open biodiversity data provide the foundation for this development. Swedish LifeWatch and the Swedish Species Observation System (Artportalen) are examples of such web services.

Although Swedish LifeWatch and associated services are being used increasingly by scientists and practitioners in biodiversity conservation, the potential for scientific output is still in its infancy. An international review showed that of 400 CS projects studied, only 20% had resulted in scientific output in terms of scientific publications (Kullenberg & Kasperowski, 2016). For biodiversity CS projects specifically, it has been shown that only 12% of 388 projects reviewed provided data for scientific publications (Theobald et al., 2015). Although this low utilisation of citizen data in published research is likely to have been underestimated as a result of published articles often failing to provide sufficient acknowledgement of citizen origins of data, this clearly represents a missed opportunity for both science and society.

Swedish LifeWatch has an important part to play in facilitating the integration of citizen data in biodiversity research by improving scientists' awareness of and accessibility to citizen biodiversity data. It has been shown that the likelihood of scientific publication in the field of biodiversity research increases in line with the spatial and temporal extent of citizen data, by having data openly accessible online, and by means of taxonomic rigour (Theobald et al., 2015). The web services provided by the Swedish Species Information Centre (SSIC, ArtDatabanken) cater for these important functions, and this has resulted in an increasing number of CS projects and scientific publications on a broad selection of organism groups and ecosystems. The Centre has also been actively carrying out CS biodiversity and conservation research for several years.

Examples of citizen science biodiversity research outputs from SSIC's web services

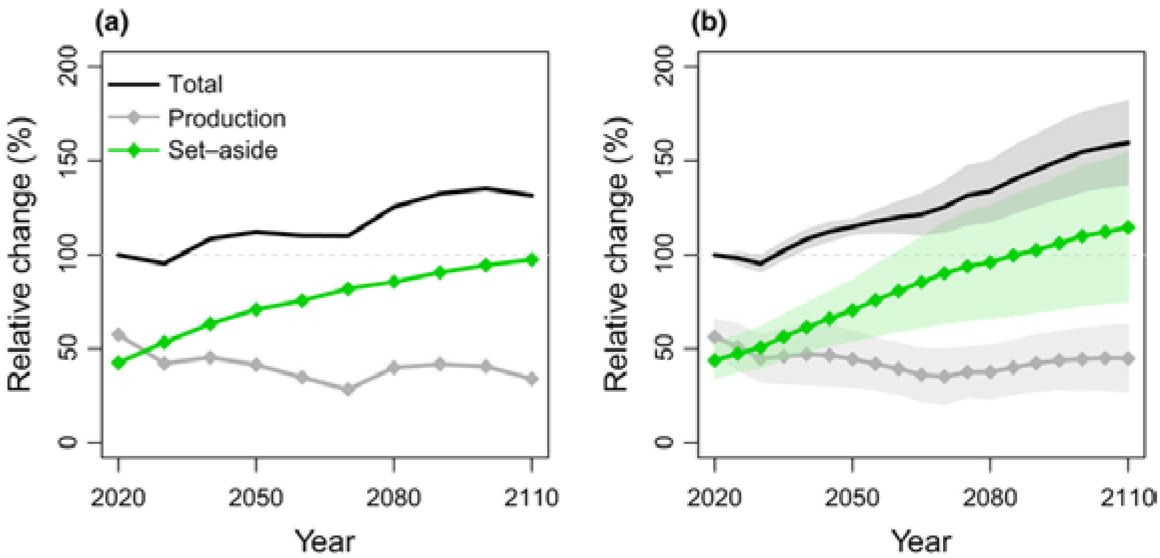
A group of researchers at the Swedish Species Information Centre, headed by Professor Tord Snäll, have evaluated citizen science data for forecasting the response of the old-forest indicator fungus *Phellinus ferrugineofuscus* to national forest management (Mair et al., 2017) and climate change (Mair et al. in press). These results show that the use of a suite of modelling approaches allows citizen data to be applied reliably to land management conservation decision-making, demonstrating that citizen data made widely available through Swedish LifeWatch can also be a valuable forecasting resource.

The results also highlight the importance of filtering and managing citizen data for reducing detection and sampling bias, and emphasise the importance of recording both presences and absences of species with as much spatial precision as possible. The research group is currently extending its research to citizen data for multiple species of wood fungi and forest birds (see the Siberian jay example) in order to estimate both current and predicted future distributions and habitat suitabilities.

Citizen data has also been used for habitat suitability modelling and risk assessment of non-indigenous species in marine environments (Leidenberger et al., 2015a), including the invasion of the Pacific oyster (*Crassostrea gigas*) in Scandinavian coastal waters (Laugen et al., 2015).

Other marine examples include current distributions and predicted future changes in distributions of organisms in marine food webs in the Baltic region, based partly on citizen data in the Swedish Species Observation System uploaded to the supranational Global Biodiversity Information Facility (Leidenberger et al., 2015b).

The research output from terrestrial environments spans a broad range of organisms and ecosys-



Forecasts of relative change in *Phellinus ferrugineofuscus* habitat suitability in response to projected forest management over the coming century from (a) a coloniation–extinction model based on systematically collected data by professionals and (b) averaged projections and standard deviations from the models based on citizen science data collected by non-professionals. Relative change is presented for a total of all forest types, and for production and set aside high conservation value forests separately. Reproduced from Mair et al. (2017) in Ecology and Evolution, Volume 7, Issue 1, pages 368-378, 20 Dec, DOI: 10.1002/ece3.2601.


Photo: Michael Krikorev

Phellinus ferrugineofuscus is a well-known and easily recognisable wood-living fungus that has been used extensively as an indicator of high conservation value coniferous forests. It occurs in large parts of the Norway spruce (*Picea abies*) distribution range in Sweden, where it is relatively common in the north and rarer in the south. The species population is estimated to have decreased by at least 15% over the past 30 years, putting it in the Near Threatened Red List category.

tems. Examples include the distributional patterns of plant species richness (Cousins et al., 2015), range expansions in resident butterflies (Audusseau et al., 2017), past distributions of rare insects (Burman et al., 2016), seasonal migration of a moth to high altitudes (Chapman et al., 2012), predicted landscape occurrence of an invasive muskrat (Ecke et al., 2014), abundance of red-listed species along roads and road verges (Helldin et al., 2015), the spread of potentially invasive trees by a non-native bird seed disperser during climate change (Hof, 2015), future climate-induced species distributions of mammals in the (sub)Arctic (Hof et al., 2012), long-term changes in spring arrival of migratory birds (Kullberg et al., 2015), flower abundance and vegetation height as predictors for nectar-feeding insect occurrence in Swedish semi-natural grasslands (Milberg et al., 2016), and effects of drought in Africa on delays to the arrival of European songbirds (Tøttrup et al., 2012).

For further examples of scientific output, please see www.slu.se/en/site/swedish-lifewatch/published/scientific-publications/.

Mari Jönsson

Mari Jönsson is a researcher at ArtDatabanken at SLU and works mainly with forest ecology and conservation. Much of her research concerns the distribution and dynamics of forest-dwelling cryptogams. She is also part of a team of scientists headed by Professor Tord Snäll who are working on research in terrestrial environments based on citizen science data stored in Artportalen.

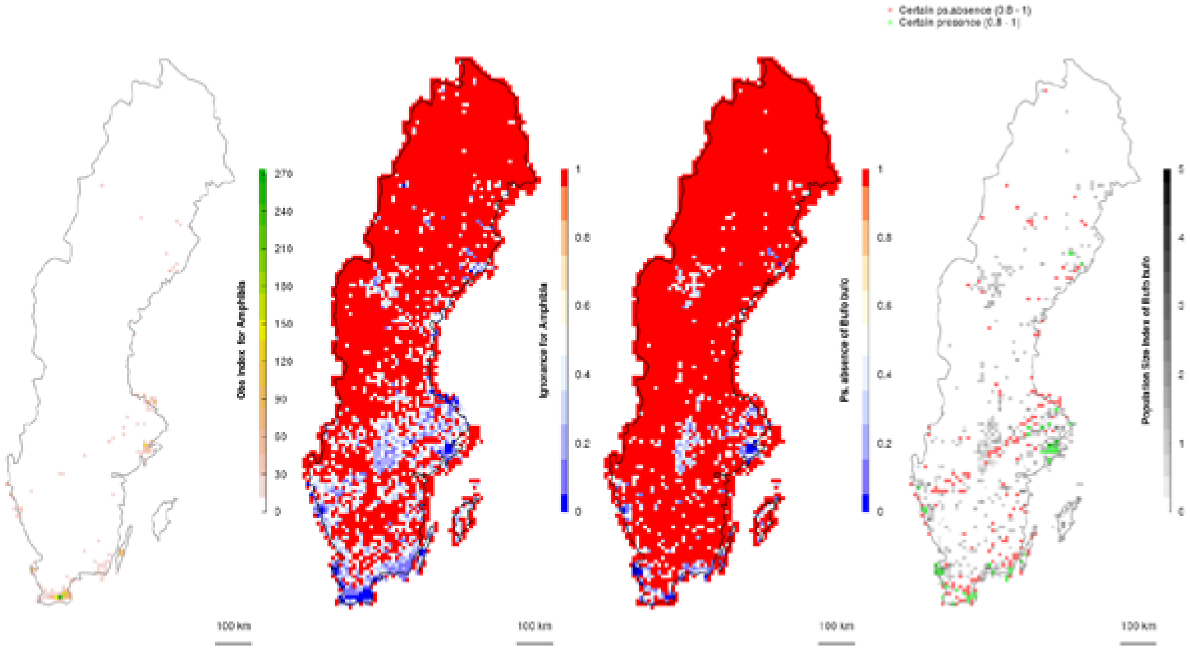
Ignorance scores for primary biodiversity data

How can we evaluate the reliability of non-systematically collected biodiversity data (e.g. museum collections and citizen science data)? An algorithm to identify gaps and spatial and temporal bias can offer a quick visual quality report that can be implemented in web-based tools or Application Programming Interfaces (APIs).

ALEJANDRO RUETE, SLU AND GREENSWAY AB

Primary biodiversity data (observations of species) stored in biodiversity databases such as GBIF or Artportalen offers a constantly growing source of data, extensive both temporally and spatially, that is available for a wide range of uses. However, as it often includes non-systematically collected data (e.g. museum collections and citizen science data) it has limitations which include sampling bias in favour of recorder distribution, lack of survey effort assessment and lack of coverage of species distribution. These limitations are not always explored, but any technical assessment or scientific research should include an evaluation of the uncertainty of its source data and researchers should acknowledge this information in their analyses. Ignorance maps are an easy way to explore the quality of the data visually and filter out unreliable results.

I was commissioned by Swedish LifeWatch to develop an algorithm intended to help curators and users of primary biodiversity data to identify gaps, assess spatial and temporal bias inherent in the data and evaluate the relative gain in knowledge as a result of new observations. The potential of this tool lies in the simplicity of its algorithm and the few assumptions required, giving users the freedom to tailor analyses to their specific needs. Any infrastructure for biodiversity information can implement this approach to provide a quick visual quality report implemented in web-based tools or Application Programming Interfaces (APIs). However, it can also be used offline in the researcher's preferred analysis environment. Quantifying the sampling effort of the observation allows users to incorporate uncertainty into analyses of species richness and distributions, and to identify areas where unreliable results are expected.



Number of observations and ignorance maps for the class Amphibia in Sweden for the period 2000–2014. The third map shows the ignorance scores for the common species *Bufo bufo*, while the fourth map shows the likelihood of true presence (green) and pseudo-absences (red) over the species' observation frequency (black). Ignorance scores were produced using the half-ignorance algorithm (O = 1). The black contour shows a 10 km buffer around Sweden's land area. Grid resolution is 10 x10 km.

The algorithm behind the ignorance scores is designed for comparison of bias and gaps in primary biodiversity data across taxonomic groups, time and space. That is to say, it can be applied to any species groups, but can also be applied at species level; it can be used to aggregate or dissect bias over time; it compares information content per pixel and can be summarised over irregular polygons. The simple visualisation and comparison of the ignorance scores across these dimensions can help data providers to set data mobilisation priorities for areas of particular interest, help observers to identify undersampled areas to be targeted on their next campaign or excursion, and help data end-users to evaluate the fitness of the data for the intended use.

How it works

This approach represents the availability of data on a scale of 0 to 1 (0 being theoretical absolute knowledge and 1 being absolute ignorance). This rationale is based on the assumption that species groups share similar bias. Observations are reported by people with varied field skills and accuracy. However, observers are assumed to be fond of or specialists on one or more taxonomic groups (such as families, orders or even classes), rather than individual species. As it is likely that an entire group of species observed using similar methods will share similar bias, it is appropriate to use species groups as a surrogate for sampling effort. Hence it is straightforward to assume that a lack of reports of any species from the species group – birds at a particular location, for instance – is most likely due to a lack of ornithologists in that specific location, rather than the total absence of birds. The inverse logic also holds true. That is to say, the larger the number of observations of species from the species group in any one location, the more likely it is for a lack of reports of a particular species to reflect a true absence.

Alejandro Ruete

Alejandro Ruete is an analyst at Greensway AB and an external researcher at the Department of Ecology at SLU. He is a quantitative ecologist whose aim is to develop mathematical and statistical tools for nature conservancy, and to translate findings into knowledge applicable to conservation biology.

In October 2016, Alejandro Ruete won the prestigious GBIF Ebbe Nielsen Challenge for his approach to measuring and comparing spatial and temporal gaps in biodiversity data.

Photo: Lennart Hansson / MostPhotos

Siberian jay distribution in Sweden

– testing citizen science data from Artportalen

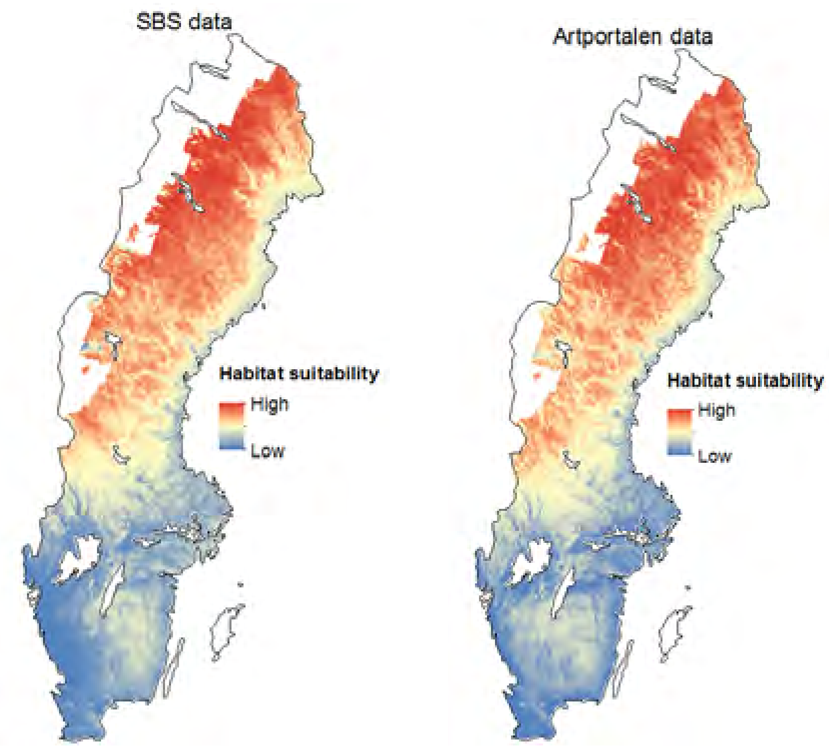
Citizen science data has the potential to offer a significant resource to help address conservation problems in a rapidly changing world. It is important, therefore, to understand whether inferences from such data can be reliable, and if so which modelling methods produce the best results.

UTE BRADTER, LOUISE MAIR, MARI JÖNSSON, ALEX SINGER & TORD SNÄLL, ARTDATABANKEN SLU

Efficient conservation depends on knowing where species occur. Systematic monitoring programmes can provide this information, but they do not exist in all countries, even for popular groups such as birds. Many such programmes may also provide little data on rarer species, species with localised habitats (e.g. water bodies) or species that are active outside the main survey times (e.g. at night). Birdwatchers, on the other hand, are often particularly interested in rarer species, and many meticulously note and upload such observations to databases such as ebird or Artportalen.

Such databases can quickly gather an impressive number of bird observations. Yet unlike systematic monitoring programmes, they lack a systematic sampling design. Systematic monitoring programmes are carefully designed to provide representative and comparable data, with survey instructions specifying where, when and how to carry out surveys. There is no such design for Artportalen data, and inferences drawn from such data have been criticised as there can be a degree of bias. However, along with an increase in the popularity of citizen science, data analysis methods continue to be developed and adapted so as to account for such biases more effectively.

Citizen science data has the potential to offer a significant resource to help address conservation problems in a rapidly



Distribution of Siberian jay, derived from Swedish Bird Survey (SBS) and Artportalen data. White areas in the north-west of Sweden are not covered due to a lack of data on forest conditions. The maps were produced using SLU Forest Maps, SMHI climate data, Lantmäteriet topographic data and data from Statistics Sweden.

changing world. It is important, therefore, to understand whether inferences from such data can be reliable, and if so which methods produce the best results. Hence we asked the question: can analysis of the distribution of a species using Artportalen data produce similar results to an analysis of systematically collected data from the Swedish Bird Survey (SBS)? And if so, which of several methods performs best? We selected the Siberian jay (*Perisoreus infaustus*) as the species for our study, and observations were obtained from the Analysis Portal.

The methods we selected rely on different input data. Some require only information on locations where the species has been seen, plus information about the environmental conditions at these and other locations. Others also require information on where the species has not been seen, information currently not recorded in Artportalen; only species observations are recorded. Implicitly, however, species observations include information on non-detection if paired with information about reporters. If a reporter is able to identify a species of interest by sight and sound and will always report this species, locations visited with no record of the presence of the species become non-detection locations. We approached some of Artportalen's most active forest bird reporters to obtain information on their reporting behaviour and bird identification skills, and subsequently used data from 38 individuals whose Artportalen entries gave us information about non-detection of the Siberian jay.

It is worth noting that the combined entries from these 38 reporters exceeded 2 million records for the period 2000–2013! This demonstrates the potential of citizen science to collect large data volumes and the dedication of individuals to birdwatching and ensuring that their observations are preserved.

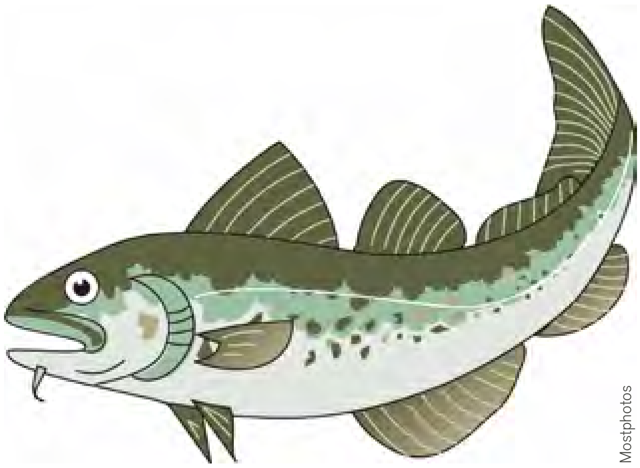
We have not tested all methods fully as yet, but some preliminary results are noted here. After analysis, including identifying key environmental variables, we found that all methods suggested relationships that we expected to see, based on species ecology. For example, all methods suggested that the chances of finding Siberian jay increased in line with the availability of mature forest. However, some methods were less able to cope with non-systematically collected data and also suggested relationships contrary to expectations and contrary to those found in the SBS model.

We produced a map of the predicted distribution of Siberian jay from each method. Across Sweden overall, all maps showed a similar distribution pattern to the map compiled from SBS data (Fig. 1). On a more local scale, however, differences between the methods became apparent, particularly in areas with higher predicted habitat suitability for Siberian jay. Methods that identified relationships consistent with expectations and similar to the SBS model demonstrated more of a correlation with the map pattern compiled from SBS data.

Therefore, with application of a suitable method and, at least, data with similar characteristics as for Siberian jay, Artportalen data can produce species distribution maps that are comparable to maps compiled from SBS data and lead to similar conclusions.

Ute Bradter

is a post-doc at ArtDatabanken (Swedish Species Information Centre) at SLU and works mainly with bird populations and remote sensing in various landscapes.


Mostphotos

Kattegat case study

Substantial datasets exist for the Kattegat marine area, and the aim of this case study by SMHI and the University of Gothenburg was to test the Swedish LifeWatch systems and demonstrate data flow opportunities and bottlenecks.

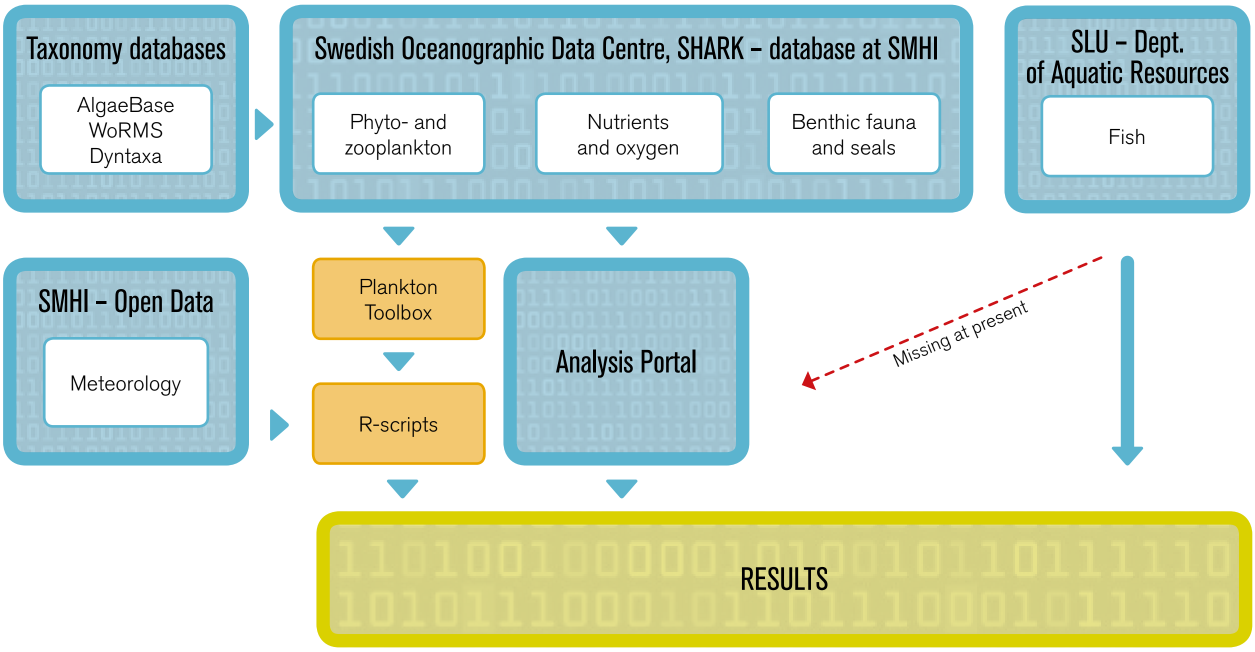
BENGT KARLSON, SWEDISH METEROLOGICAL AND HYDROLOGICAL INSTITUTE (SMHI)

In the Swedish LifeWatch project, several different types of observation data relating to biodiversity and environmental parameters are connected through a system of databases. Data can be combined, analysed and visualised in the Analysis Portal for Biodiversity Data. Case studies have been carried out in order to test the systems and demonstrate data flow. The Swedish LifeWatch marine group identified the Kattegat as a suitable area for a case study and the development of and pressures on the cod

Data type	Data repository	Start of time series	Comments
Phytoplankton		1989 (some data from 1971)	2-4 locations
Zooplankton		1996 and 2007	2 locations
Benthic fauna		1994 (one from 1993)	~8 locations
Harbour seals		1988	Sampling effort has varied
Nutrients, oxygen		1982 (some data from 1965)	1-8 locations
Fish	SLU-Aqua	1979 (winter) 1991 (autumn)	International Bottom Trawl Survey (IBTS)
Meteorology	SMHI (Open Data)	1961	Hallands Väderö


Photo: Bengt Karslon

In the 1980s, algal bloom was noted in Laholm Bay and reported on widely by the media. This bloom caused acid deficiency in the Bothnian Sea, resulting in the deaths of mussels and other aquatic animals. The picture shows the dinoflagellate *Tripos muelleri* (former name *Ceratium tripos*).



population (*Gadhus morhua*) as a suitable topic. The aim of this study was to test part of the Swedish LifeWatch system and to address one or more scientific issues.

Taxonomy databases are necessary to allow historical data to be used consistently. The World register of marine Species (WoRMS) is the most commonly used database for marine organisms. AlgaeBase supplies information to WoRMS. Marine biological data and marine climate data, i.e. physical and chemical oceanographic data, was supplied by the Swedish Meteorological and Hydrological Institute (SMHI) via the National Oceanographic Data Centre's SHARK database. The data is sourced from long-term national and regional marine monitoring programmes and short-term scientific investigations. Data on fish was available from the Department of Aquatic Resources (SLU-Aqua) at the Swedish University of Agricultural Sciences. Meteorological data was downloaded from SMHI Open Data.

Overview of available data

In the study, the Kattegat was defined as specified by the International Council for the Exploration of the Sea (ICES) for reporting fish data. The northern limit was set at a line between the northern tip of Jutland (Grenen, north of Skagen) and Tistlarna, south-west of Gothenburg. The line between Gilbjerg Head and Kullen in Sweden forms the southern limit, i.e. the boundary towards the Sound (Öresund).

Conclusions

There is a substantial dataset for the Kattegat. Long-term biodiversity data includes fish, harbour seals, phytoplankton, zooplankton, benthic flora and fauna. Other available data includes temperature, salinity, chlorophyll, nutrients, oxygen and meteorological data. Data on direct terrestrial input, e.g. river flow, was not within the scope of the case study. Scientists planning to use the data will need to combine the various datasets. The systems provided by Swedish LifeWatch are a great help, but scientists will need a good knowledge of data processing and statistics. Free software such as R and Python are commonly used. The Plankton Toolbox, produced as part of SLW, is useful for working with plankton data. One missing link in the data flow involves marine fish data, which is currently unavailable via Swedish LifeWatch. Meteorological data has recently become freely available via SMHI Open Data.

Bengt Karlson

Bengt Karlson is a researcher at the Department of Research & Development, Oceanography, SMHI, in Gothenburg, Sweden. His current scientific interests include marine phytoplankton diversity and the dynamics of harmful algal blooms, as well as the effects of climate change on the marine ecosystem. Bengt has been promoting free access to marine data for a long time.

Selected publications

Ruete, A., Pärt, T., Berg, Å., Knape, J. 2017. Exploiting opportunistic observations to estimate changes in seasonal site use: an example with wetland birds. Ecology and Evolution DOI:10.1002/ece3.3100.

Mair, L., P. J. Harrison, M. Jönsson, S. Löbel, J. Nordén, J. Siitonen, T. Lämås, A. Lundström, and T. Snäll. 2017. Evaluating citizen science data for forecasting species responses to national forest management. Ecology and Evolution 7:368–378.

Mair L, Harrison PJ, Räty M, Bärring L, Strandberg G & Snäll. 2017. Forest management could counteract distribution retractions forced by climate change. Ecological Applications, in press.

Hardesty et al. (2016). BioVeL: a virtual laboratory for data analysis and modelling in biodiversity science and ecology. BMC Ecology 16:49. doi: 10.1186/s12898-016-0103-y.

Leidenberger S, Käck M, Karlsson B, Kindvall O (2016) The Analysis Portal and the Swedish LifeWatch e-infrastructure for biodiversity research. Biodiversity Data Journal 4: e7644. doi: 10.3897/BDJ.4.e7644.

Mair, L. & Rute, A. (2016): Explaining spatial variation in the recording effort of citizen science data across multiple taxa. PLOS ONE DOI:10.1371/journal.pone.0147796.

Cousins S, Auffret A, Lindgren J, Tränk L (2015). Regional-scale land-cover change during the 20th century and its consequences for biodiversity. AMBIO 44: 17-27.

Helldin J-O, Wissman J, Lennartsson T (2015). Abundance of red-listed species in infrastructure habitats – ”responsibility species” as a priority-setting tool for transportation agencies' conservation action. Nature Conservation 11: 143-158. Doi: 10.3897/natureconservation.11.4433.

Laugen AT, Hollander J, Obst M, Strand A (2015). The Pacific Oyster (Crassostrea gigas) invasion in Scandinavian coastal waters in a changing climate: impact on local ecosystem services. In: Biological Invasions in Aquatic and Terrestrial Systems: Biogeography, Ecological Impacts, Predictions, and Management. De Gruyter, Warsaw. DOI: http://www.degruyter.com/view/product/458036.

Leidenberger S, Obst M, Kulawik R, Stelzer K, Heyer K, Hardisty A, Bourlat SJ (2015a). Evaluating the potential of ecological niche modelling as a component in invasive species risk assessments. Marine Pollution Bulletin 97:470–487. DOI: http://dx.doi.org/10.1016/j.marpolbul.2015.04.033.

Leidenberger S, De Giovanni R, Kulawik R, e t al. (2015b). Mapping present and future predicted distribution patterns for a mesograzer guild in the Baltic Sea. Journal of Biogeography, 42 (2):241–254. http://onlinelibrary.wiley.com/doi/10.1111/jbi.12395/abstract.

Ruete A (2015). Simple algorithms to display ignorance maps for raw distributional data accessed from species observation databases: SLW implementation. Biodiversity Data Journal 3: e5361.

Strand M, Panova M (2015). Size of genera – biology or taxonomy? Zoologica Scripta, 44 (1):106-116. http://onlinelibrary.wiley.com/doi/10.1111/zsc.12087/full

Gärdenfors U, Jönsson M, Obst M, Wremp AM, Kindvall O, Nilsson, J (2014). Swedish LifeWatch – a biodiversity infrastructure integrating and reusing data from citizen science, monitoring and research. Human Computation 1(2): 147163. http://hcjournal.org/ojs/index.php?journal=jhc&page=article&op=view&path[]=21

González Talaván A, Mathew C, Obst M, Paymal E (2014). Data Refinement Using the BioVeL Portal. 67 pp. Copenhagen: Global Biodiversity Information Facility Available online at http://www.gbif.org/orc/?doc_id=5983

Mathew C, Guentsch A, Obst M, Vicario S, Haines R, Williams AR, de Jong Y, Goble C (2014) A semiautomated workflow for biodiversity data retrieval, cleaning, and quality control Biodiversity Data Journal 2: e4221. http://biodiversitydatajournal.com/articles.php?id=4221

Snäll T, Forslund P, Jeppsson T, Lindhe A, O'Hara RB (2014) Evaluating temporal variation in Citizen Science Data against temporal variation in the environment. Ecography 37(3): 293-300. http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0587.2011.00544.x/abstract

Holl S, Garijo D, Belhajjam K, Zimmermann O, Giovanni RD, Obst M, Goble C (2013) On specifying and sharing scientific workflow optimization results using research objects. Proceedings of the 8th Workshop on Workflows in Support of Large-Scale Science. ACM: 2013: 28–37. http://dl.acm.org/citation.cfm?doid=2534248.2534251.

Klein, T., Langner, J., Frankenberg, B., Svensson, J., Broman, B., Bennet, C. & Langborg, T. (2013): ECDS – A Swedish research infrastructure for the open sharing of environment and climate data. Data Science Journal, 12 (March): 1–9.

Reports

Kindvall, O., in collaboration with Roscher, S., Bailly-Maître, J. and Šípková-Gaudillat., 2015. Dyntaxa taxon concept administration and how to handle information related to taxa.TC/BD report to the EEA.

Hanssen F (ed), Heggberget T (ed), Bladt J, Endresen D, Forsius M, Gudmundsson G, Gärdenfors U, Heiðmarsson S, Kindvall O, Koch W, Koviula K, Laiho EL, Laine K, Obst M, Skov F, Telenius A, Valland N, Wasowicz P, Wremp AM (2014) Nordic LifeWatch Cooperation. Final report. 68 pp.

Dettki, H., Ericsson, G., Giles, T. & Norrsken-Ericsson, M. (2012): Wireless Remote Animal Monitoring (WRAM) – A new international database e-infrastructure for telemetry sensor data from fish and wildlife. pp. 247-256. In: Proceedings Etc 2012: Convention for Telemetry, Test Instrumentation and Telecontrol (Eds. The European Society of Telemetry). Books on Demand, pp. 292, ISBN: 978-3-7322-5646-4.

Organisation

Steering Committe

Ulla Mörtberg (chair)
KTH Royal Institute of Technology

Anna-Lena Axelsson
Swedish University of Agricultural Sciences

Stefan Claesson
Swedish Natural History Museum

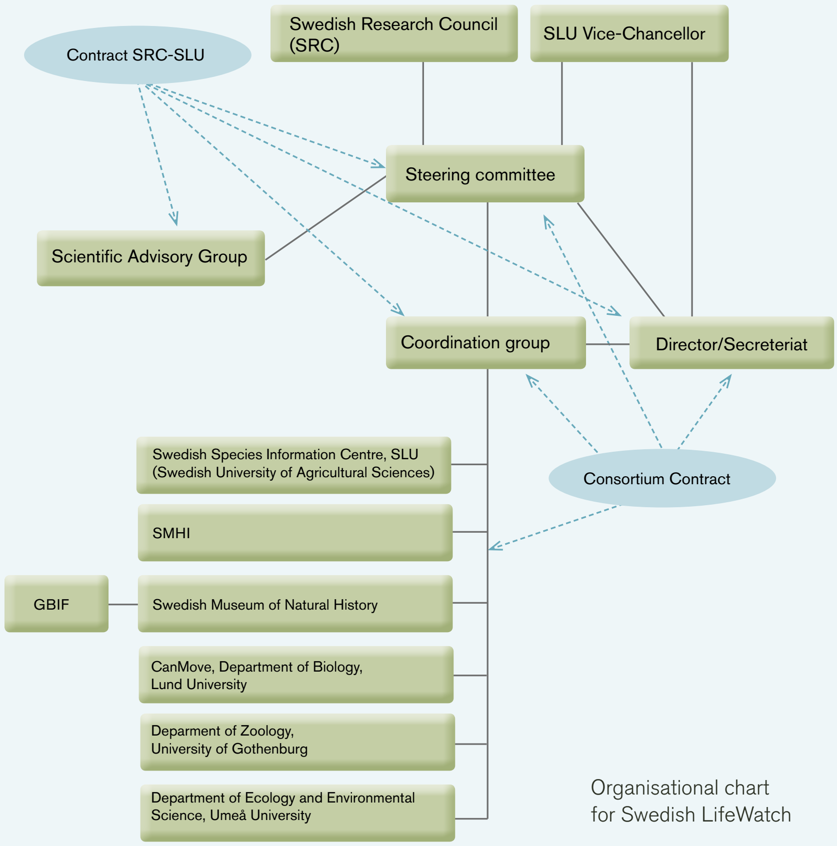
Maria Sjö (2011–2013)
Swedish Environmental Protection Agency

Henrik Svedäng
Swedish Institute for Marine Research

Uno Wennergren (2011–2013)
Linköping University

Lars Westerberg (from 2014)
Linköping University

Hanna Östergård (from 2014)
Swedish Environmental Protection Agency



Organisational chart
for Swedish LifeWatch

Scientific Advisory Board

Walter G. Berendsohn
Professor and Director of the Department of Biodiversity Informatics and Laboratories, Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Germany

Lena Gustafsson
Professor in Conservation Biology at the Department of Ecology, Swedish University of Agricultural Sciences, Sweden

Anna Gårdmark
Research Scientist at the Institute of Coastal Research, Department of Aquatic Resources, Swedish University of Agricultural Sciences, Sweden

Jannicke Moe
Research Scientist, Norwegian Institute for Water Research (NIVA), Norway

Cynthia Parr
Director of Species Pages Group of the Encyclopedia of Life, National Museum of Natural History, Washington D.C., U.S.

Hannu Saarenmaa
Project Manager, Digitarium – Digitisation Centre of the Finnish Museum of Natural History and the University of Eastern Finland, Finland

Niclas Jonzén † 2015
Associate Professor, Department of Biology, Theoretical Population Ecology and Evolution Group, Lund University, Sweden

Coordination Group

Representatives from:

- ArtDatabanken, SLU
- CanMove (Centre for Animal Movement Research), Lund University
- Department of Aquatic Resources, SLU
- SMHI (Swedish Meterological and Hydrological Institute)
- Swedish Museum of Natural History
- UC-WRAM, SLU
- Umeå University
- University of Gothenburg

Chair: Matthias Obst, University of Gothenburg

References

Audusseau, H., Le Vaillant, M., Janz, N., Nylin, S., Karlsson, B., Schmucki, R. (2017) Species range expansion constrains the ecological niches of resident butterflies. Journal of Biogeography, 44: 28–38.

Basset A, Los W (2012) Biodiversity e-Science: LifeWatch, the European infrastructure on biodiversity and ecosystem research. Plat Biosystems 146(4): 780–782

Burman J., Westerberg L., Ostrow S., Ryrholm N., Bergman K.O., Winde I., Nyabuga F.N., Larsson M.C., Milberg P. (2016) Revealing hidden species distribution with pheromones: the case of Synanthedon vespiformis (Lepidoptera: Sesiidae) in Sweden. Journal of Insect Conservation. 20: 11–21.

Burin G, Kissling DW, Guimaraes PR, et al (2016) Omnivory in birds is a macroevolutionary sink. Nat Commun 7:11250 doi: 10.1038/ncomms11250.

Chapman J.W., Bell J.R., Burgin L.E., Reynolds D.R., Pettersson L.B., Hill J.K., Bonsall M.B.,Thomas J.A (2012) Seasonal migration to high latitudes results in major reproductive benefits in an insect. Proc Natl Acad Sci USA, 109: 14924-14929.

Cousins S, Auffret A, Lindgren J, Tränk L (2015) Regional-scale land-cover change during the 20th century and its consequences for biodiversity. AMBIO 44: 17–27.

Cressey D (2015) Tropical paradise inspires virtual ecology lab. Nature. 2015;517:255–6.

Davies N, et al (2016) Simulating social-ecological systems: the Island Digital Ecosystem Avatars (IDEA) consortium. GigaScience 5:14.

Ecke F., Henry A., Danell K. (2014) Landscape-based Prediction of the occurrence of the invasive muskrat (Ondatra zibethicus). Annales Zoologici Fennici, 51: 325–334.

Evans MR, Bithell M, Cornell SJ, et al (2013) Predictive systems ecology. Proc Biol Sci. 2013;280:20131452. doi:10.1098/ rspb.2013.1452.

Gärdenfors U, Jönsson M, Obst M, et al (2014) Swedish LifeWatch – a biodiversity infrastructure integrating and reusing data from citizen science, monitoring and research. Human Computation 1(2): 147–161.

Hanssen FO, Heggberget TG, Bladt J, et al (2014) Nordic LifeWatch cooperation, final report: A joint initiative from Denmark, Iceland, Finland, Norway and Sweden. NordForsk, 68 pp. http://hdl.handle.net/11250/2395084

Hardisty et al (2016) BioVeL: a virtual laboratory for data analysis and modelling in biodiversity science and ecology. BMC ecology 16 (1), 49. DOI: 10.1186/s12898-016-0103-y

Helldin J-O, Wissman J, Lennartsson T (2015) Abundance of red-listed species in infrastructure habitats – "responsibility species" as a priority-setting tool for transportation agencies' conservation action. Nature Conservation 11: 143–158.

Hof A.F. (2015) Alien species in a warming climate: a case study of the nutcracker and stone pines. Biological Invasions, 17: 1533–1543.

Hof A.R., Jansson R. & Nilsson C 2012 Future climate change will favour non-specialist mammals in the (sub) Arctics. PLoS ONE 7(12): e52574. doi:10.1371/journal.pone.0052574

Kullberg C., Fransson T., Hedlund J., Jonzén N., Langvall O., Nilsson J., Bolmgren K. (2015) Change in spring arrival of migratory birds under an era of climate change, Swedish data from the last 140 years. Ambio, 44: 69–77.

Kullenberg C, Kasperowski D (2016) What is Citizen Science? – A scientometric Meta-Analysis. PLos ONE 11(1): e0147152.

Kindvall, O., in collaboration with Roscher, S., Bailly-Maître, J. and Šípková-Gaudillat, Z., 2015. Dyntaxa taxon concept administration and how to handle information related to taxa. TC/BD report to the EEA.

Kissling et al (In review) Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biological Reviews.

Koureas D, Arvanitidis C, Belbin L, et al (2016) Community engagement: The 'last mile' challenge for European research e-infrastructures. Research Ideas and Outcomes 2: e9933 doi: 10.3897/rio.2.e9933

Laugen AT, et al (2015) The Pacific Oyster (*Crassostrea gigas*) invasion in Scandinavian coastal waters in a changing climate: impact on local ecosystem services. In: Canning-Clode J, editor. Biological invasions in aquatic and terrestrial systems: biogeography, ecological impacts, predictions, and management. Berlin: De Gruyter Open, p. 230–52.

Leidenberger S, Obst M, Kulawik R, Stelzer K, Heyer K, Hardisty A, Bourlat SJ (2015a) Evaluating the potential of ecological niche modelling as a component in non-indigenous species risk assessments. Marine Pollution Bulletin. 97: 470-487. doi:10.1016/j.marpolbul.2015.04.033.

Leidenberger S, De Giovanni R, Kulawik R, e t al. (2015b) Mapping present and future predicted distribution patterns for a mesograzer guild in the Baltic Sea. Journal of Biogeography, 42: 241-254. doi:10.1111/jbi.12395

Leidenberger S, Käck M, Karlsson B, Kindvall O (2016) The Analysis Portal and the Swedish LifeWatch e-infrastructure for biodiversity research. Biodiversity Data Journal 4: e7644. doi: 10.3897/BDJ.4.e7644.

Mair L, Harrison PJ, Jönsson M, Löbel L, Nordén J, Siitonen J, Lämås T, Lundström A, Snäll T (2017) Evaluating citizen science data for forecasting species responses to national forest management. Ecology and Evolution, 7: 368–378.

Mair L, Harrison PJ, Räty M, Bärring L, Strandberg G & Snäll (in press) Forest management could counteract distribution retractions forced by climate change. Ecological Applications.

Milberg P., Bergman K.O., Cronvall E., Eriksson Å.I., Glimskär A., Islamovic A., Jonason D., Löfqvist Z., Westerberg L. (2016) Flower abundance and vegetation height as predictors for nectar-feeding insect occurrence in Swedish semi-natural grasslands. Agriculture, Ecosystems & Environment, 230: 47–54.

Miraldo A, et al(2016) An Anthropocene map of genetic diversity. Science 353 (6307), 1532-1535. doi: 10.1126/science.aaf4381

Pereira HM, Ferrier S, Walters M, et al (2013) Essential Biodiversity Variables. Science 339, 277–278

Purves D, Scharlemann J, Harfoot M, Newbold T, Tittensor DP, Hutton J, et al. Ecosystems: time to model all life on Earth. Nature. 2013;493:295–7. doi:10.1038/493295a.

Rockström J, et al (2009) A safe operating space for humanity. Nature 461(7263): 472–475.

Seebens H, Schwartz N, Schupp PJ et al (2016) Predicting the spread of marine species introduced by global shipping. Proc Natl Acad Scie USA 113(20): 5646–5651. www.pnas.org/cgi/doi/10.1073/pnas.1524427113

Stelzer K, Heyer K, Bourlat S, Obst M (2013) Application of Niche Modeling and Earth Observation for the risk assessment and monitoring of invasive species in the Baltic Sea. Report MarCoast II - Marine and Coastal Environmental Information Services, Ballast Water Option, pp 57.

Theobald EJ, Ettinger AK, Burgess HK, DeBey LB, Schmidt NR, Froehlich HE m.fl. (2015) Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation, 18: 236–244.

Tøttrup A.P., Klaassen R.H.G, Kristensen M. W., Strandberg R., Vardanis Y., Lindström Å., Rahbek C. Alerstam T., Thorup K. (2012) Drought in Africa Caused Delayed Arrival of European Songbirds. Science 338, 1307.

Vilhena DA, Antonelli A (2015) A network approach for identifying and delimiting biogeographical regions. Nat Commun 6: 6848

Vestbo S, et al (In review) Present and future distributions of horseshoe crabs determine potential areas for conservation. J Biogeogr.

Economy

	2010	2011	2012	2013	2014	2015	2016	Total
Allocated funds	9 000 000	9 000 000	9 000 000	4 524 085	4 475 915	4 499 996	1 800 000	42 299 996
Co-financing	2 550 000	2 420 263	2 818 119	3 399 793	1 945 921	2 047 666	8 203 498	23 385 260
Other revenue	28 652	5 348	0	297 776	-242 454	6 429	0	95 751
Total revenues	**11 578 652**	**11 425 611**	**11 818 119**	**8 221 654**	**6 179 382**	**6 554 091**	**10 003 498**	**65 781 007**
Staff costs	1 645 296	4 919 800	5 887 798	5 890 842	3 875 080	5 382 912	4 263 696	31 865 424
Consultant costs	0	1 297 888	1 995 773	2 813 178	1 763 007	912 236	2 858 259	11 640 341
Other costs	279 153	401 545	305 475	393 656	823 604	1 115 774	499 697	3 818 904
Depriciation	0	140 993	241 702	241 702	117 482	416 621	15 038	1 173 538
Premises	110 859	273 185	333 447	304 622	148 104	238 715	127 693	1 536 625
Overhead costs	776 721	2 622 252	3 030 023	2 514 303	1 467 970	2 483 916	2 337 075	15 232 260
Total costs	**2 812 029**	**9 655 663**	**11 794 288**	**12 158 303**	**8 195 247**	**10 550 174**	**10 101 458**	**65 267 092**

The overarching mission of Swedish LifeWatch (SLW) is to make all Swedish biodiversity data openly available in standardised formats through interoperable web services, and to develop tools and virtual laboratories for advanced biodiversity and ecosystem analysis.

The Swedish LifeWatch consortium was founded in 2010 as a national collaboration between four universities, including the Swedish University of Agricultural Sciences (SLU), the University of Gothenburg, Lund University and Umeå University, the Swedish Meteorological and Hydrological Institute (SMHI) and the Swedish Museum of Natural History, including the Swedish GBIF node. The project is coordinated by the Swedish Species Information Centre (ArtDatabanken) at SLU and financed by the Swedish Research Council (VR).

This report summarises the achievements of Swedish LifeWatch; data content, tools and services, international collaboration, pilot studies and practical applications from the start-up in 2010 until 2016.

www.slu.se/lifewatch

SLU SMHI Naturhistoriska riksmuseet GBIF Vetenskapsrådet

UNIVERSITY OF GOTHENBURG

LUND UNIVERSITY
Centre for Animal Movement Research

UMEÅ UNIVERSITET

NATUR VÅRDS VERKET