NORDIC LIFEWATCH COOPERATION

FINAL REPORT

A joint initiative from Denmark, Iceland, Finland, Norway and Sweden

Financed by NordForsk and kindly contributed to by participating institutions

Coordinated by the Norwegian Institute for Nature Research

Hanssen, F.(ed)[1],Heggberget, T. (ed) [1], Bladt, J. [2], Endresen, D. [3], Forsius, M. [4], Gudmundsson, G. [5], Gärdenfors, U. [6], Heiðmarsson, S[5]., Kindvall, O[6]., Koch, W[7]., Koviula, K. [8], Laiho, E.L. [8], Laine, K. [9], Obst, M. [10], Skov, F. [2], Telenius, A. [11], Vallan[7], N., Wasowicz, P.[5] and Wremp, A.M[6].

1) Norwegian Institute for Nature Research (NINA), 2) Aarhus University, 3) Natural History Museum- University of Oslo, 4) Finnish Environment Institute (SYKE), 5)Icelandic Institute of Natural History, 6)Swedish Species Information Centre/LW Sweden, 7) Norwegian Biodiversity Information Centre, 8)Finnish Museum of Natural History-University of Helsinki, 9)University of Oulu, 10)University of Gothenburg and Natural History Museum- University of Stockholm

Introduction and goals

Presently large European research infrastructures are being developed under the auspices of the European Strategy Forum on Research Infrastructures (ESFRI). LifeWatch is one of several ESFRI projects which aim to establish eInfrastructures and databases in the field of biodiversity and ecosystem research.

The ultimate goal for the present Nordic project is that an enhanced cooperation between Nordic countries within biodiversity shall result in increased scientific collaboration and more effective policies and management strategies within this area.

The Nordic dimension in further development of core European infrastructures will be strengthened by establishing a Nordic cooperation. The possibilities for the Nordic countries to interact successfully with the European level will increase if the Nordic countries can act as a unity. Development of common eInfrastructures and analytical eTools will be cost-efficient by allowing construction costs to be shared among the participating countries. The use of available expertise will also be more efficient. A Nordic LifeWatch consortium will help to promote the Nordic countries in the development of the European LifeWatch construction phase.

The added value of a Nordic LifeWatch collaboration is expected to be:

- Increased scientific opportunities for Nordic researchers by increased availability of compatible biodiversity data.
- Reduced construction and operational costs for participating countries by establishing compatible eInfrastructure solutions.
- Improved interaction with the European level through Nordic collaboration.

These themes are outlined and discussed in the present report, where both possibilities and bottlenecks for future development of Nordic LifeWatch are outlined.

During the project period, two collaborative meetings have been arranged; a kick-off workshop in Stockholm, Sweden 7 – 9 November 2012, and a follow-up workshop in Akureyri, Iceland 6 – 9 May 2013. In addition, several Skype meetings and frequent email contacts among the participants have been performed during the one and a half year project period. In addition, the Nordic LifeWatch has been represented at several EU LifeWatch conferences and workshops.

The present report has been written jointly by the participants. The contribution from NordForsk covers in principal all direct costs, including travel expenses for all participants. The time spent for meetings and writing is covered by the participating institutions. In the role as project coordinator, NINA would like to take the opportunity to thank both NordForsk and the participants for their efforts to realize this project.

Trondheim March 2014
Tor G. Heggberget
Project leader

Summary

The main goal of the present report is to outline the possibilities for an enhanced cooperation between the Nordic countries within eScience and biodiversity. LifeWatch is one of several ESFRI projects which aim to establish eInfrastructures and databases in the field of biodiversity and ecosystem research.

Similarities between Nordic countries are extensive in relation to a number of biodiversity related issues. Most species in Nordic countries are common, and frequently the same challenges concerning biodiversity and ecosystem services are addressed in the different countries.

The present report has been developed by establishing a Nordic LifeWatch network with delegates from each of the Nordic countries. The report has been written jointly by the delegates, and the work was organized by establishing working groups with the following themes: strategic issues, technical development, legal framework and communication.

Written during two workshops, Skype meetings and emailing, the following main issues are discussed in the present report:

- Scientific needs for improved access to biodiversity data and advanced eScience research infrastructure in the Nordic countries.
- Future challenges and priorities facing the international biodiversity research community.
- Scientific potential of openly accessible biodiversity and environmental data for individual researchers and institutions.
- Spin-off effects of open access for the general public.
- Internationally standardized Nordic metadata inventory.
- Legal framework and challenges associated with environmental-, climate-, and biodiversity data sharing, communication, training and scientific needs.
- Finally, some strategic steps towards realizing a Nordic LifeWatch construction and operational phase are discussed.

Easy access to open data on biodiversity and the environment is crucial for many researchers and research institutions, as well as environmental administration. Easy access to data from different fields of science creates an environment for new scientific ideas to emerge. This potential of generating new, interdisciplinary approaches to pre-existing problems is one of the key features of open-access data platforms that unify diverse data sources. Interdisciplinary elements, access to data over larger gradients, compatible eSystems and eTools to handle large amounts of data are extremely important and, if further developed, represent significant steps towards analysis of biological effects of climate change, human impact and development of operational ecosystem service assessment techniques.

It is concluded that significant benefits regarding both scientific potential, technical developments and financial investments can be obtained by constructing a common Nordic LifeWatch eInfrastructure.

Several steps concerning organizing and funding of a future Nordic LifeWatch are discussed, and an action plan towards 2020 is suggested. To analyze the potential for future Nordic LifeWatch in detail, our main conclusion is to arrange a Nordic LifeWatch conference as soon as possible. This conference should involve Nordic research councils, scientists and relevant stakeholders. The national delegates from the participating countries in the Nordic LifeWatch project are prepared to present details from the report and developments so far as a basis for further development of Nordic LifeWatch.

The present work is financed by NordForsk and in-kind contributions from participating institutions.

Contents

1. Project background, aims and current relations to national and international strategies for development of research infrastructure

The Nordic LifeWatch initiative has evolved from the Nordic Research council (NordForsk)[1] and their need to review potential strategic, scientific and financial benefits of harmonizing large scale research infrastructures at the Nordic level. Several infrastructures in the domain of biodiversity and ecosystem research are under development within the European Strategy Forum on Research Infrastructures (ESFRI)[2] and the global Group on Earth Observations Biodiversity Observation Network (GEOBON)[3].

The European LifeWatch[4] infrastructure is one of several ESFRI initiatives. Finland, Sweden, Norway and Denmark contributed in the preparatory phase of the European LifeWatch (2008 to 2010). The European LifeWatch entered the construction phase in 2011 and will be one of the very first European Research Infrastructure Consortia (ERIC)[5] to be organized as a new legal entity developed for pan-European infrastructures, governed by the member states (Italy, Spain, Greece, the Netherlands, Romania and Hungary). So far none of the Nordic countries have joined the European LifeWatch ERIC as partners. Norway, Finland, Sweden and Denmark each signed a MOU with LifeWatch Europe at the construction phase start-up meeting in Amsterdam, January 2011.

LifeWatch aims to align with other parallel initiatives. GEOBON coordinates global activities relating to the Societal Benefit Area (SBA) on Biodiversity of the Global Earth Observation System of Systems (GEOSS)[6]. Some 100 governmental, intergovernmental and non-governmental organizations are collaborating through GEOBON to organize and improve terrestrial, freshwater and marine biodiversity observations globally and make their biodiversity data, information and forecasts more readily accessible to policy makers, managers, experts and other users. Moreover, GEOBON has been recognized by the Parties to the Convention on Biological Diversity. GEOBON is both a community of practice and a task force in the global GEO Work Plan[7]. This is a voluntary partnership that is guided by a steering committee comprising the key stakeholders, including DIVERSITAS[8], GBIF[9], IUCN[10], NASA[11], UNEP-WCMC[12] and others. The European Biodiversity Observation Network (EUBON)[13] will establish the European part of GEOBON. EUBON will build on existing components mainly from the

[1] http://www.nordforsk.org/no
[2] http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
[3] http://www.earthobservations.org/geobon.shtml
[4] http://www.lifewatch.eu
[5] http://www.lifewatch.eu/towards-eric
[6] http://www.earthobservations.org/gci_gci.shtml
[7] http://www.earthobservations.org/docshow.php?id=129
[8] http://www.diversitas-international.org/
[9] http://www.gbif.org
[10] http://www.iucn.org/
[11] http://www.nasa.gov/
[12] http://www.unep-wcmc.org/
[13] http://www.eubon.eu/

Global Biodiversity Information Facility (GBIF), LifeWatch and national biodiversity data centers. Several institutions from Norway, Sweden, Denmark and Finland are partners in EUBON.

The scope of this Nordic LifeWatch pilot project is to identify the scientific potential of a common Nordic infrastructure based on inventories of user needs, existing data repositories, and challenges/constraints related to data sharing in general. Based on these findings, strategies and a proposal will be developed for funding of a **Nordic LifeWatch construction phase in close collaboration with stakeholders** (Nordic research councils, Ministries, and scientific communities), the National LifeWatch consortia, the European LifeWatch, GBIF and EUBON.

The national LifeWatch projects in the Nordic countries are at very different status levels. The national status for each country is described in detail later in this report. The Nordic LifeWatch project aims to stimulate development of national LifeWatch initiatives (including Iceland and the Baltic countries), and also to act as a common interface between the national projects and the European LifeWatch.

The Nordic LifeWatch aims to develop a Service Oriented Architecture (SOA) which is compatible with the ESFRI LifeWatch, EUBON, AlterNet, LTER, GBIF and other contributing infrastructures/networks present at inter-Nordic and national level. This is highly achievable given the fact that many of the current Nordic LifeWatch-partners already are national coordinators in several of these initiatives. In addition, the Nordic and the Baltic countries already have made great international efforts to promote data sharing through the Global Biodiversity Information Facility (GBIF), the Baltic Biodiversity project[14], the Norwegian Biodiversity Information Centre (NBIC)[15] and the Swedish Species Information Centre[16]. The ultimate goal for the Nordic LifeWatch is full integration of the biodiversity research community, including citizen's science, through a common Nordic interdisciplinary e-Infrastructure. Biodiversity informatics is meant to facilitate and improve biodiversity research for the future.

The European Commission has defined open science as the main direction for proposals under the HORIZON 2020[17] funding initiative. This is also in accordance with the Nagoya protocol adopted at the 10th meeting of the Parties to the Convention on Biological Diversity (CBD) (Hardisty *et al.* 2013).

Making publicly funded research widely available is one of the top priorities for the European Commission. A proposal on Rules for participation and dissemination[18] in HORIZON 2020 was launched by the Commission in November 2011. The Director-General of research and innovation at the Commission, Robert-Jan Smits, said in an interview[19] with the *Times Higher Education* that "*open access, which typically involves making research papers freely available within months or a year of publication, 'will be the norm' for research funded through Horizon 2020. With our €80 billion we can make one hell of a difference*".

[14] http://balticdiversity.eu/
[15] http://www.artsdatabanken.no/
[16] http://www.slu.se/artdatabanken/
[17] http://ec.europa.eu/research/horizon2020/index_en.cfm
[18] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0810:FIN:en:PDF
[19] http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=419949&c=1

The practical implication of these rules towards open data access is however unclear. The current advancement of electronic publishing of literature has identified the need for online citation mechanisms securing future recognition and retrieval. Electronic publishing of data gives related but much more complex challenges. Best practices and international data citation standards are important measures in this context. Section 7.5.3 in this report focuses on this.

Nevertheless, it is clear that this HORIZON 2020 openness approach will challenge the Nordic biodiversity research community and how we design our future research infrastructures. To address this challenge, the Nordic LifeWatch will organize necessary efforts at both national and inter-Nordic levels. The national LifeWatch consortia will have to provide adequate infrastructures for data management and data flow, whilst the Nordic LifeWatch consortium will act as a regional building block in the European context providing exploration tools and virtual laboratories for interdisciplinary research of different thematic and inter-Nordic purposes.

The world governments missed their target to reduce the rate of biodiversity loss by 2010. One of the main reasons for this was few reliable indicators and the general shortage of available biodiversity information. Adopting the Strategic Plan for Biodiversity 2011 - 2020 (including the Aichi Biodiversity targets)[20], many governments now put greater emphasis on preservation, restoration and maintenance of ecosystems. Biodiversity informatics plays a fundamental role in achieving the Aichi Biodiversity targets and the Strategic Plan for Biodiversity 2011 to 2020. It will also be of fundamental importance for the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES)[21]. IPBES aims at strengthening the science-policy interface for biodiversity and ecosystem services in conservation and sustainable development. IPBES has also identified informatics and improved access to information as central key capacity building needs for the future. The Nordic LifeWatch partners will actively contribute in biodiversity informatics capacity building towards IPBES.

The Nordic LifeWatch will also actively contribute to, and align with, the Global Biodiversity Informatics Outlook (GBIO)[22] framework where important focal areas and action components are identified in order to coordinate future efforts and funding and to enable improved interaction of initiatives and projects. Figure 1 below illustrates the focal areas, action components and their current progress.

[20] https://www.cbd.int/sp/
[21] http://www.ipbes.net/
[22] http://www.biodiversityinformatics.org/

Figure 1: The GBIO Framework

A Nordic LifeWatch should interact with, and take advantages of, progressive international ICT-resources. In this context the technical progress of the European Data Infrastructure project (EUDAT) needs to be monitored carefully. EUDAT aims to develop a collaborative, generic infrastructure for scientific data management that can be used by a diversity of research communities and existing infrastructures. EUDAT now offers online e-Services that allow users to find, access, share, store, replicate and compute big research data. EUDAT is coordinated by the Finnish IT Center for Science (CSC) and has 25 partners (data centers, technology providers, research communities and funding agencies) from 13 countries (including Norway and Sweden).

2. Current scientific needs for improved access to biodiversity data and advanced eScience research infrastructure in the Nordic countries

We have not surveyed current scientific needs directly, but instead used available literature where major challenges and obstacles for data exchange are discussed as an integrated part of the current scientific needs. The present chapter will give a short overview of this.

For stakeholders and users in general, the storage of vast amounts of robust, secure, and easily accessible data is crucial. But despite the technical biodiversity informatics "state of the art", biodiversity science will always rely on data that is not fully available, linkable, discoverable and accessible as it should be. In addition, services and tools to process these data are not yet available for "plug and play". Finally, models for different parts of the biodiversity system are not yet linked across time, space and scales (Hardisty *et al.* 2013). Solving such issues will require much more openness and collaboration.

The "*eInfrastructure Scientific Opportunities Panel*" (appointed by the Norwegian Research Council) launched in 2010 "The scientific case for eInfrastructure in Norway"[23] (Gisler *et al*. 2010). The aim of this report was to assess the future growth of scientific needs for eInfrastructure in Norway, and to advise the Research Council of Norway and Norwegian decision makers in pointing out the opportunities for scientific progress that will be enabled by further sustained development of the national eInfrastructure. The report focuses on the eInfrastructure needs in several fields of scientific research in Norway. The report identifies current scientific needs that in general are common with the other Nordic countries.

Recent developments in bioinformatics and systems biology illustrates the increased usage and need of eInfrastructure. Bioinformatics is both data and computer intensive, with heavy use of statistics to build models, make inferences and perform validations. In future, nearly all biologists will use bioinformatics methods through well-adapted programs and databases, but may need expert support. Bioinformatics databases have grown rapidly, and represent a vital source of information for molecular biology and medicine. This has facilitated studies of biological processes and has required new ways of organizing, analyzing, and interpreting data. The agenda of bioinformatics research relies heavily on current technology in molecular biology and biotechnology. In return, biological research is strongly influenced by advances in bioinformatics.

The field of system biology focuses on interactions between biological systems, and reflects a paradigm shift in molecular and cellular biology. The aim is the modeling and discovery of emergent properties, properties of a system that arise from collective interactions and are understood using e.g. dynamical systems theory. Systems biology obtains and analyses complex data from multiple experimental sources. The enormous range of temporal and spatial scales involved in systems biology requires experimental and computational methods that span those scales competently. The

[23] http://www.forskningsradet.no/prognett-evita/Documents/1226485742881

methodology of systems biology includes the use of robots and automated sensors to aid in experimentation and data acquisition. Computational models, data mining techniques, and very large online databases put stringent demands on available eInfrastructure.

The great variety among the life sciences presents a challenge in building an efficient infrastructure for the diverse needs of biologists. Local facilities should be strengthened, and communication among different institutions must be enhanced. The educational challenges of multidisciplinarity, mutual understanding, and the cultural barriers between the different fields of sciences remain a problem. Bio-imaging must deal with visualizing and exploring large-scale, multidimensional data.

Finally, in many experimental situations the dimensionality of the data can be orders of magnitude larger than the number of available samples so that feature selection and model selection are difficult. The newer mathematics of data mining and machine learning, like compressed sensing and random matrix theory, are needed. Researchers and clinicians need to acquire the knowledge and tools to effectively use future eInfrastructure and successfully apply the lessons learned. The field of genomics has already exploded through second generation sequencing technology. This shifts focus towards enormous efforts in functional and comparative genomics generating huge amounts of data within this particular field. Completely new sets of tools for the analysis of genome data will be needed. It is also to be expected that genomics will have an impact within the entire life science field, and outside the life sciences in psychology, anthropology, archaeology, and humanistic sciences as well.

HORIZON 2020 demands more data sharing and collaboration in their focus on open science. Bridging the fragmented international biodiversity informatics landscape is feasible at the Nordic level. The relatively small size of the Nordic countries, combined with the general similarity and common understanding on a vast number of issues, implies that we should have a great potential to address future needs and challenges facing the Nordic biodiversity research community.

The needs for Nordic eScience infrastructure was recognised already in 2007 by the Nordic Council of Ministers when they launched the Nordic eScience strategy[24]. The strategy was followed up in an action plan by the Nordic eScience initiative eNORIA (2008)[25]. The action plan was commissioned in 2009. Actions in eScience infrastructure will be pursued with a view to open access and EU policies. Efforts will complement relevant priorities within the EU framework and the ESFRI Roadmap. The action plan is targeted at defined eScience needs in higher education, research and ICT Infrastructure. The organizational framework for implementing the action plan was the Nordic Council of Ministers, NordForsk, the national research councils and the national eScience infrastructure providers.

The eScience action plan was in 2009 developed into the Nordic eScience Globalization initiative (NeGI)[26] and the Nordic Infrastructure Collaboration (NeIC)[27] focusing on Infrastructure and data. These initiatives have a thematic focus on climate change and health conditions. They are now one of

[24] http://www.ndgf.org/ndgfweb/ndgf_strategy_workshop_2008_attchmt/Nordic%20eScience.pdf
[25] http://www.nordforsk.org/files/nordic-escience-action-plan
[26] http://www.nordforsk.org/no/programs/programmer/escience/copy_of_escience-globaliseringsinitiativ
[27] http://neic.nordforsk.org/

the main priorities for the Nordic Council of Ministers. NordForsk is coordinating both initiatives (2011-2015). Future collaboration between the Nordic LifeWatch and NeGI/NeIC would be very beneficial for the Nordic biodiversity research community.

2.1 General need for interoperable tools and services

The Nordic research community needs interoperable eScience tools and services in order to address new, complex and interdisciplinary knowledge challenges. The term eScience Infrastructures include computer networks, high performance computing and visualization systems, federated databases, and network-enabled research instrumentation. It also includes ICT GRIDS, the distributed technology that provides access to remote sources and enables collaboration among distributed virtual organizations. Such collaboration which often requires cross-disciplinary research teams, vast data collections, large-scale computing resources and high speed networks will revolutionize the way science is undertaken in the 21st century. Having such infrastructure established will facilitate more interdisciplinary research collaboration, innovation as well as increased commercial and academic competitive power.

2.2 Harmonization of priorities, strategies and policies

To realize such eInfrastructure developments there is a great need for harmonization of both national and Nordic priorities, strategic assessments and policies to support research fields where activities are high and where the investment need is high. This includes collaboration, coordination and work division on funding principles between research institutions, research councils and the Ministries.

There is also a great need for earmarked funding and targeted financial means such as tax based depreciation for infrastructure investments, business models for how to share costs and basis budget allowance for management and maintenance of infrastructures when cost levels are expected to be higher than projects or institutions can bear.

One essential question is how to ensure continuity in maintaining an infrastructure that is developed with project funding? Prior to construction phase infrastructure projects should consider to develop strategies for long-term funding. The funders on their side have to communicate their clear expectations to the institutions holding infrastructures. Hosting an eScience infrastructure demands good business plans, management plans and a sustainable maintenance and refinance economy.

Investments should not only be allocated to physical equipment. The biodiversity research community in the Nordic countries also needs greater investments in scientists that can manage and use the upcoming deluge of data and online eServices. To attract scientific software developers and database curators into academia there is a great need to define career paths that are competitive with the ICT industry. Establishing an eInfrastructure should also include a long-term commitment to maintaining and continuously improving the skills and competence of its human resource base. Research, development, support, training, community development, outreach, and education are all fundamental aspects, and lead to strength, flexibility and the capability to respond to new areas of science.

There is a lacking awareness in the biodiversity research community about the importance of data sharing and eInfrastructure. There is therefore a great need to target information campaigns aimed at data stakeholders and data users. Negative attitudes at personal or institutional levels have to be met with facts about challenges, opportunities and cost-benefits. This may leverage the willingness to participate in data mobilization.

The Nordic research community also has a great need for development of best practices and policies regarding data quality enhancement and data sharing. This can best be achieved in collaboration with GBIF and the national biodiversity information centers in Norway and Sweden, as they are offering guidelines, targeted capacity building programs, training, mentoring and tools for enhanced data sharing capacity.

2.3 Nordic collaboration on data resources

The Nordic Council of Ministers (NMC) published in 2012 the report "Reinforced Nordic collaboration on data resources"[28] (Sandberg 2012) where significant challenges and obstacles for data exchange are mapped. This report focus mainly "*on existing national databases and registers established for administrative purposes but also the question of newly-generated scientific data is handled*". The challenges are analysed from several perspectives. These perspectives are relevant in describing general scientific needs facing the whole Nordic research community. "*The broad scope of the report targets primarily policy makers involved in eScience development on national and/or Nordic level*". The Nordic Council of Ministers, NordForsk, the Finnish Ministry of Education and Culture and the Finnish IT Center for Science (CSC) were involved in the study.

2.3.1 The organizational perspective

From the NMC-report (Sandberg 2012) the following is cited: "*The main challenges found from an organizational perspective are the scattered resources, meaning both the high numbers of different databases/registers as well as the wide range of institutions involved in maintaining these data resources*". Another organizational challenge emphasised in the NMC-report is that decision makers are lacking a *research perspective*. Also, it appears to be *no common language* between researchers and decision makers.

"*For register-based research the difficulty to access personal data is a specific organizational bottleneck*". "*Researchers own hesitation towards data sharing* "is also a major challenge. Finally "*The lack of a Nordic perspective in national strategies and initiatives targeting research data resources also poses a challenge for reinforced Nordic cooperation, since there is a risk of developing policy directives which are not aligned on a Nordic level*". All these challenges highly underline that

[28] http://www.norden.org/no/publikationer/publikasjoner/2012-514

there is a huge scientific need for interoperable workflows and improved harmonization, communication and strategy development at the Nordic level.

2.3.2 The legal perspective

Sandberg (2012) writes: The "*main challenges found*" "*from a legal perspective*" are the "*different legislations regarding mainly access to personal data in the Nordic countries and between different authorities. There is also an insecurity regarding the revision of the EU Data protection legislation and how it will affect the national legislations and possibilities for register-based research*". "*[A] new EU legislation with requirements of revising national data protection laws (...) could give a necessary momentum for the Nordic countries to look into further harmonization of legislation on a Nordic level*". To advance open science and more scientific collaboration in the Nordic countries there are clearly a need for a harmonization of national legislation.

2.3.3 The financial perspective

Sandberg (2012) writes: "*The financial perspective highlights the benefits of cost-efficient research. A huge amount of public funds are spent on maintaining existing public registers and databases as well as gathering new data and it is highly cost efficient to use these resources to their full potential in research. A selection process needs to take place to determine which databases and registers are worth the costs of improving and making truly sustainable by integrating them into a reinforced Nordic database infrastructure. To promote newly-generated scientific data sharing funding incentives may be used. Data sharing could be promoted through a common Nordic system where research projects and researchers receive funding credits for sharing data. From a financial perspective it is important with pooling of national funds for Nordic research benefit; national funds should be set aside to ensure Nordic data resources collaboration. One solution could be to fund ESFRI projects supporting sharing of data resources, but to promote a Nordic perspective within these EU initiatives - a second layer of collaboration inside the bigger collaboration.*"

2.3.4 The ethical perspective

Sandberg (2012) writes: "*The ethical perspective mainly handles register-based research and Bio banks where the challenge is to use the advantage of the unique Nordic conditions for register-based research without violating the personal privacy and integrity. One important issue is to keep the societal trust and an open ethical debate to towards this type of research, not only on a national level but also on a Nordic level by for instance making room for more Nordic channels and platforms for public dialogue and involvement. The role of national Ethics Advisory Boards and Ethics committees in a reinforced Nordic cooperation also has to be clarified and the working methods need to be aligned.*"

2.3.5 The technical perspective

Sandberg (2012) writes: "*The technical perspective deals mainly with the challenge of interoperability and how to go from pilot project to large-scale solutions. The technical perspective also underlines the* *importance of political support to develop technical solutions and the monitoring and use best practices of technical solutions in international initiatives and projects.*"

3. Future challenges and priorities facing the international biodiversity research community

In their BMC Ecology white paper "*A decadal view of biodiversity informatics - challenges and priorities*"[29], Hardisty *et al.* (2013) reviewed the last decade of biodiversity informatics in Europe. This community consulting paper position the role of biodiversity informatics for the next decade, recommending necessary actions to link various infrastructures and to facilitate supportive understanding both to business and policy-makers. The recommendations they propose are meant as a background from which decisions can be made in making and evaluating proposals, allocating funds or directing work to build infrastructures. These recommendations will be highly valuable in the construction of a Nordic LifeWatch infrastructure.

History has proved that biodiversity informatics has an increasingly central role in enabling the research communities to address relevant scientific conservation and sustainability issues. During the last decade there have been great strides in establishing a unified framework for data sharing, where the fields of taxonomy and systematics have been perceived as the most prominent disciplines. To some extent this is inevitable since species names are the pivot around which biodiversity information is organized. However, to fully understand the ecological complexity related to nature conservation, land-use, environmental change, sustainability, food security and ecosystem services the scientific community need to invent a system approach that moves significantly beyond pure species observation and taxonomy.

Hardisty *et al.* (2013) claim that the grand challenge for biodiversity informatics is to develop an infrastructure that can address questions related to human impact and at the same time capture the variety, distinctiveness and complexity of the nature. Biodiversity processes are very complex and occur as inter-dependent processes over a breadth of scales. Capturing this complexity is beyond the capability of current information management and modelling methods. To have an impact on biodiversity conservation, sustainability and urgent social questions we need to consider all aspects of biodiversity in a holistic approach across time, space and scales. Captured data on biodiversity has to be analysed in their observational and temporal context. Creating and maintaining such infrastructures will require long-term commitment and funding from all involved stakeholders.

To enhance collaboration and reduce duplication of efforts Hardisty *et al.* (2013) recommend that:

1. **Open data have to be normal practice** and should embody the intelligent openness principles of being accessible (easy to find and use data), assessable (easy to assess data reliability), intelligible (easy to understand data) and usable (easy to reuse data for different purposes).

[29] http://www.biomedcentral.com/1472-6785/13/16

2. **Data encoding** should make analysis at different scales possible.
3. **Infrastructure projects have to devote significant resources to market the services they develop in order to attract significant numbers of users.** To target this effort, projects should release their services as early as possible with frequent updates according to user feedback.

To enhance usability deployment of existing technologies Hardisty *et al.* (2013) recommend that:

4. A **complete list of taxon names should be established with a statement of their interrelationships** (spelling variations, synonyms and so on). This is a much simpler challenge than building a list of valid taxon names.
5. **Every resource should have a Persistent identifier (PID) so that resources can be linked to each other**. Part of the PID should have a common syntactic structure such as a Digital Object Identifier (DOI). In this way any instance can be made available in a free-text search.
6. A **system of author identifiers** should be implemented. This combined with the PID will allow the computation of impact of any contributions and the provenance of any resource.
7. **Trusted third-party authentication measures should be implemented to allow users to work with multiple resources** without separately having to log into each one.

The foundational technologies referred above partly exist, but need to be integrated. To develop new structures by exploiting existing technologies Hardisty *et al.* (2013) recommend that:

8. A **classification repository** has to be constructed to allow automatic construction of taxonomies in order to bridge gaps in data coverage.
9. Develop **a single portal for currently accepted taxon names**.
10. **Tools and standards are required to make data linkable**. This can be done using the potential of vocabularies and ontologies for all biodiversity facets (taxonomy, environmental factors, and ecosystem functioning, services and data streams from DNA to the genome level).
11. **Mechanisms to evaluate data quality and fitness-for-use should be developed**.
12. **A next-generation infrastructure is required** to manage the ever-increasing amounts of observational data.

According to Hardisty *et al.* (2013) the decadal vision is to develop services that deliver data or information analysis with the use of interchange standards. Implementing this vision depends on a balance between top-down and bottom-up approaches by making appropriate funding decisions.

The top-down approach includes strategy development and actions at European level (encouraging community adoption of standards and scoping through targeted funding calls, workshops and meetings). The bottom-up approach includes initiatives from individuals with ideas for specific problem solving. Both approaches recognize individuals and groups as important key-players in order to bridge the many isolated islands of infrastructures.

EU has funded a number of projects to address the challenges of deploying eInfrastrastructure for biodiversity science. This includes Framework programs, the Networks of excellence (ALTERNet[30], LTER-Europe[31], EDIT[32], PESI[33], MarBEF[34], Mars[35], EuroMarine etc.) and others (4D4Life/i4Life[36], agINFRA[37], Aquamaps[38], iMarine[39], BioFresh[40], BioVel[41], ENVRI[42], EU BON[43], EU-BrazilOpenBio[44], Fauna Iberica[45], MicroB3[46], OpenUp![47], proiBiosphere[48] and Vibrant[49]).

All these initiatives share similar characteristics, but appear to be very different regarding architecture and technology. There are some overlaps but several dead-ends and often completely lack of mainstream industrial involvement.

This illustrates a lack of a common understanding on how to best deploy eInfrastructures and underline the importance of a community based consensus around a decadal vision. Community consensus has to be combined with scoped funding of projects interacting within a coherent programme. This can be achieved through a common roadmap (e.g. like ESFRI) focusing on relevant architectural approaches and necessary construction steps.

Future roadmap projects has to be aligned and coordinated in order to maximize the benefit from past, present and future investments. Hardisty *et al.* (2013) underline that this requires adequate coordination, dissemination, education and training capabilities within the HORIZON 2020 framework. It is very important that future bottom-up project proposals fits under the umbrella of the community`s decadal vision. They should leverage existing project results, avoid developing incompatible alternatives and being specific on how they can engage the biodiversity expert community is the key factor for advancing scientific knowledge. Important contributions from citizen scientists on species occurrence and distribution data are also highly valuable. The main challenge for the biodiversity informatics community is to develop a framework that addresses the requirements citizen science projects rise:

---

[30] http://www.alter-net.info/
[31] http://www.lter-europe.net/
[32] http://www.e-taxonomy.eu/
[33] http://www.eu-nomen.eu/portal/
[34] http://www.marbef.org/
[35] http://www.marsnetwork.org/
[36] http://www.4d4life.eu/
[37] http://aginfra.eu/
[38] http://www.aquamaps.org/main/home.php
[39] http://www.i-marine.eu/Pages/Home.aspx
[40] http://www2.freshwaterbiodiversity.eu/
[41] http://www.biovel.eu/
[42] http://envri.eu/
[43] http://www.eubon.eu/
[44] http://www.eubrazilopenbio.eu/Pages/Home.aspx
[45] http://www.fauna-iberica.mncn.csic.es/english/
[46] http://www.microb3.eu/
[47] http://open-up.eu/
[48] http://www.pro-ibiosphere.eu/
[49] http://vbrant.eu/

- The need to cover all steps in the development and implementation of such projects
- Perform automated validation and annotation of data from such projects
- Develop incentives to encourage participation in processing, analysis and use of data
- Develop virtual research and teaching environments for citizen scientists
- Improve systems for automated image recognition based on existing technologies
- Promote best practices with successful examples from nature conservation
- Ensure continued economic viability of services provided from citizen science

To further identify the future challenges and priorities Hardisty *et al.* (2013) elaborate on important issues related to the biodiversity informatics fundamental backbone, required steps and tool development. Last, but not least, they focus on the human interface of a modern eInfrastructure for biodiversity and ecosystem research. The following sections of chapter 3 are largely based on Hardisty *et al.* (2013).

3.1 The fundamental backbone

3.1.1 Organization of taxon names

Taxon names are the pivot around which biodiversity information is organized. While content about taxon names must be assembled by nomenclaturalists, taxonomists and managers of biodiversity information, there is an urgent need for development of vision driven architectural and engineering solutions. Cyber infrastructures like the Global Name Architecture (GNA)[50] focus on the Global Names Index (GNI)[51] and name-usage instances like the Global Names Usage Bank (GNUB)[52]. GNUB does not yet exist but will provide cross-linked semantic relationships at the nomenclatural level.

Names have traditionally been organized using Latin (where the first part of the name is shared by a group of similar organisms, and the second part differentiates between members of the group) or hierarchical classifications (genera, families, orders, classes and phyla). As science advance relationships will change. Hierarchies can instead be built from instances of name strings, however this is inefficient. A required solution for this includes a classification bank combined with a name list to produce a taxonomic hierarchy automatically for species groups that have not recently received taxonomic attention.

The Species 2000[53] and Integrated Taxonomic Information System (ITIS)[54] Catalogue of Life (CoL)[55] partnership provides a global taxonomic reference system. Names in CoL represent concepts, but there are no links to the concepts themselves, and therefore identification cannot be unequivocally

[50] http://www.globalnames.org/
[51] http://gni.globalnames.org/
[52] http://www.globalnames.org/GNUB
[53] http://www.sp2000.org/
[54] http://www.itis.gov/
[55] http://www.catalogueoflife.org/

verified. Other name classification systems such as the NCBI taxonomy[56] or the WoRMS systems[57] can also be used as organizational frameworks. However, none of these systems are interoperable.

The Global Name Architecture (GNA)[58] promotes the development of an infrastructure that is capable of linking available information about biological names. iPlant`s Taxonomic Name Resolution Service (TNRS 3.2)[59] corrects and standardize scientific plant names against particular taxonomies. ZooBank[60] is a new initiative to move the process by which new names become recognized in the digital age. However tools for alignment and cross mapping of taxonomies can only partially be automated because taxonomic knowledge is very hard to fully codify. i4Life[61] have developed a useful first draft cross-map, but in order to be authoritative, future solutions have to be able to link nomenclatures, taxonomic compendia, other classifications, literature sources and phylogenies in order to cover the whole spectrum of biodiversity complexity. Taxon names provide important access keys, but it is essential that they can be linked to descriptions, traits and habitats.

3.1.2 Can biodiversity studies be done without names?

Despite the fact that taxonomists have described approximately 1.5 million species so far, the biodiversity community still lacks a single authoritative list of names. The numbers of species left to be discovered are substantial and, given that current taxonomy is the product of the last 250 years of effort, it is not realistic to have a complete catalogue of life using the current accepted methods. New methods like DNA barcoding are highly effective in revealing much of the undescribed biodiversity. These techniques cannot be used to name new species, but they are valuable to assess ecosystem biodiversity. Ecological research will benefit greatly from such new approaches classifying and understanding genomic biodiversity based on functions, their evolution and distribution.

To advance biodiversity data beyond names, we need to integrate name indexed biodiversity information with information on functional biodiversity; various organizational levels of biodiversity (genes, organisms, ecosystems and landscapes); relationships between facets of biodiversity and ecosystem functioning and services; Physical environmental data and variables, and element fluxes through the environment.

3.1.3 To link resources we need persistent identifiers

To succeed with this endeavour we have to develop better methods for data linkage. This will require a system of universally persistent identifiers (PID). Such identifiers can be attached to resources of any kind (taxonomic concepts, genetic sequences, data and services). In addition to be stable and unique these identifiers has to conform to widely used syntactic definitions and free text search. The identifiers should be resolvable and archived with the resource in a sustainable manner.

---

[56] http://www.ncbi.nlm.nih.gov/taxonomy
[57] http://www.marinespecies.org/about.php
[58] http://www.globalnames.org/
[59] http://tnrs.iplantcollaborative.org/sources.html
[60] http://zoobank.org/
[61] http://www.i4life.eu/

The technology required for establishing PIDs such as the Digital Object Identifiers (DOI) is well-explored and will not present major challenges. They are now familiar in many publications. Initiatives such as DataCite[62] have reduced both the cost and complexity of using DOI. However, they are for some reason not yet commonly applied in the biodiversity informatics community.

The obstacles to implementation of PIDs in the biodiversity informatics community seem to be mainly of social character with community reluctance about changing current working practices. Being unique, the PIDs could be discovered using standard searching tools like Google, but resolvable and machine readable PIDs are required for establishing more elaborate linkage mechanisms. The main challenge for community-wide implementation of PIDs remains the introduction of simple and easy-to-use technologies for building linkages. Solutions for Linked Data show great promise in this respect.

Both networked and centralized services have their advantages and disadvantages. Although networked services are desirable in order to maintain consistency and to focus resources for maintenance, they are often unable to transfer big data at a usable speed of response.

Making local copy repositories are often a practical solution in order to perform effective computation of big data. Automated workflows require networked web services that handle such big data transactions. Developing these services is however technically challenging, and requires suitable search facilities that minimize the number of host-client interactions, and the bandwidth necessary to keep response times short.

Centralized services often assemble big data submitted by data creators in a common structure. These services have scale advantages in terms of economy and the ability to tune the system performance. They also have significant advantages when applied in large scale data generation systems. The disadvantage of centralized services is that they are difficult to change once they are established.

Another drawback of centralized cashed services, as experienced in the genomics community, is that it is time-consuming to transfer big data to where the computational resources for analysis and modelling are located. As a response to this, workflow strategies are presently being considered for how to transfer computation algorithms to the data instead of moving the data to the site where the data analysis is performed.

3.1.4 Professional and non-professional contributions

Engaging the biodiversity expert community is the key factor for advancing scientific knowledge. Important contributions from citizen scientists on species occurrence and distribution data are also highly valuable. The main challenge for the biodiversity informatics community is to develop a framework that addresses the requirements citizen science projects rise:

- The need to cover all steps in the development and implementation of such projects.
- Perform automated validation and annotation of data from such projects.

[62] http://www.datacite.org/

- Develop incentives to encourage participation in processing, analysis and use of data.
- Develop virtual research and teaching environments for citizen scientists.
- Improve systems for automated image recognition based on existing technologies.
- Promote best practices with successful examples from nature conservation.
- Ensure continued economic viability of services provided from citizen science.

3.1.5 User engagement

Standard applications or resources should not be imposed upon the community. It is a better strategy to let the users themselves decide what products fits to their requirements. Biodiversity informatics projects should invest significant resources into marketing of their products and services. They also should engage with real users and align the product with requirements from user feedback. Doing this early and often in a project will be wise.

3.1.6 User identification

The tradition of peer reviewed paper publishing and modern e-publishing have several consequences and typographical restrictions. First, the system of citations is important for individuals in their career development. Secondly, the cost of print-on-paper has driven data presentation to compact and often summarized formats. Third, the financial interests of the publishers constitute a barrier for open access.

Two important aspects of the citation mechanism are provenance and impact. Provenance is about reliability (e.g. who has generated the data). Impact is about where, how and for what purpose the data have been re-used. Impact is in this context a measure that can be used in the managerial assessment of an individual career.

Modern digital publication could effectively remove typographical restrictions. Today many digital publishers (e.g. Pensoft[63]) have introduced publication in parallel formats (paper, PDF, HTML and XML). The new publishing paradigm focuses on the development of new methods for consistent identification of both data contributors and data users. The identification can be transferable across social networking environments (e.g. Facebook, Google and Twitter). Approaches to this are currently being developed in ORCID[64], ResearcherID[65] and VIAF[66]. These consortiums design a common system of identification for scholarly authors and users involved in report compilation and assessments.

The transition to reusable data that can be associated and linked with identified individuals or group of individuals is central in the Open Science movement[67]. This has now been implemented by the US National Science Foundation as they require that applicants for funding to list their “products” rather

[63] http://www.pensoft.net/
[64] http://orcid.org/
[65] http://www.researcherid.com/
[66] http://www.oclc.org/en-europe/viaf.html
[67] http://en.wikipedia.org/wiki/Open_science

than “publications” (See NSF Grant proposal guide Chapter II - Proposal Preparation Instructions[68]). This means that a scientist's worth is not dependent solely on publications. Data sets, software and other non-traditional research products will count too. This illustrates that the value of public contribution is recognized beyond traditional paper publication.

Open access data and services provide user anonymity for some level of access. Some forms of interactions will nevertheless require the users to identify themselves (e.g. when posting comments, making corrections and downloading data). Social media tools (like Facebook, Google and Twitter) offer third party authentication mechanisms that can be used for access control. This gives two main advantages providing easier access and stronger security check. Anyway, some resources will always require a stronger form of authentication. However, access to public financed biodiversity data should always be unrestricted unless they have some kind of sensitive information content.

3.1.7 The importance of metadata

Metadata is the key to evaluate fitness for use of the data you consider to re-use. Registering metadata is often considered as an overhead in ordinary data capture. However, metadata is absolutely essential if generated data are to be discoverable and re-usable. Current metadata practices have to be improved and there is an urgent need for community agreement on metadata standards and the development of automated mechanisms to collect and append metadata (e.g. workflows that make use of standard services to create data-recording templates). In short-term perspective, the data producers will have to take on extra effort of metadata production. Developments of automated metadata harvesting tools are conceivable and will ease the work-burden of metadata production on the longer term.

3.1.8 Sustaining the physical infrastructure

Improved digital and online access to the physical infrastructure of databases, museum collections, experimental facilities, monitoring facilities and genomics facilities will generate much greater impact of appropriate biodiversity informatics tools than is currently physically practical.

3.2 Required steps

3.2.1 Data sharing

Currently there are many well-functioning data sharing practices in different fields of expertise. The PARSE.Insight project[69] and the Science Magazine[70] performed two surveys to understand how data are treated by scientists. From a total of 2902 multidisciplinary international responses (from both surveys) it can be deduced that data are not often shared openly. Across all disciplines it is clear that only 6-8 % of the researchers share their data in external archives. The most common environment

[68] http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp
[69] http://www.parse-insight.eu/
[70] http://www.sciencemag.org/

for storage, managing and re-use of data is the lab or the internal working environment, down to personal computers and portable storage carriers. Servers are mainly used as file servers behind a firewall with restricted access for defined groups of registered users. According to the Science Magazine most of the respondents (80,3 %) said that there are not sufficient funding available for data curation. Clearly relevant actions have to be addressed in order to ensure that data sharing will become a normal multidisciplinary international practice.

3.2.2 Vocabularies and ontologies

Shared and standardized vocabularies are the foundation for data sharing. Data integration and analysis will require a semantic consistency and a syntactic standardization. Terms are organized into vocabulary lists, vocabulary lists evolves into a thesaurus, and as formal relationships between terms are agreed, an ontology is established.

Biodiversity informatics can learn much from ontologies developed by other communities (e.g. Google, Medical informatics, Agriculture informatics etc.). Biodiversity community ontologies often tend to isolate communities rather than to enable a more open data sharing. We see an emerging demand for ontologies that span multiple communities and domains.

Building community standard terminologies and ontologies are familiar challenges in other informatics communities. Lessons learnt from human genetics and model organism functional genomics are described below. First, terminologies and ontologies need to be owned by the community. Their maintenance is an ongoing process which requires stable funding and community coordination and interaction. Secondly, tools biologists find intuitive have to be developed both for data coding and analysis, making the process efficient and effectively invisible. Third, ongoing terminology and syntax development need expert construction and are not just problems of computer science. Fourth, it is hard to communicate all changes in terminology lists to sites that consume such data. A central catalogue is needed to avoid this problem. Fifth, mapping of data coded by legacy terminologies and integration of data coded by different species-specific ontologies are problems that are already addressed by some communities.

Biodiversity data has a semantic interoperability potential. Realizing this potential and entering a new paradigm of open semantic approaches will require quite basic research and IT development. Transferring “legacy” data models into semantic-aware technologies is clearly desirable because these existing models are often accurate, comprehensive and represent a great deal of effort from the scientific community. There is however a need for a pragmatic strategy for mobilizing this knowledge. Such mobilization may also assist in broadening the user acceptance, which is often more problematic than the associated technical issues. Developing and applying vocabularies will be challenging and requires the existence of persistent identifiers to be effective. It will also require organization and cooperation, and in summary both goodwill and cash funding.

In its 2006, 2007 and 2008 Technical Roadmaps the Taxonomic Databases Working Group (TDWG)[71] identified community-supported vocabularies and ontologies expressing shared semantics of data as

[71] http://www.tdwg.org/

required components of standard architecture, in order to provide basic interoperability across open systems. Other required components are common exchange protocols and use of persistent identifiers for the data. The TDWG Darwin Core glossary of terms[72] is one of the most widely deployed biodiversity vocabularies. Both its management and relation to the TDWG Ontologies can be a model for other vocabularies.

A GBIF Task group highlighted in 2011 the need for GBIF to identify sustainable support mechanisms for shared vocabularies, and commissioned the White paper recommendations for the use of Knowledge Organisation Systems (KOS) by GBIF[73]. Here GBIF concluded that core technologies are available and well understood. But the uptake of the community is not ideal. The challenge is to develop and deploy tools within the overall biodiversity informatics infrastructure that make the implementation of knowledge organization system effectively invisible. The GBIF Integrated Publishing Toolkit (IPT)[74] is one example in this direction. TDWG convened in 2012 a task group, Vocabulary management task group (VoMaG), to provide guidelines for the governance of vocabularies for biodiversity information terminology[75]. The VoMaG guidelines were presented at the 2013 TDWG conference (Baskauf *et al*. 2013)[76]. The VoMaG report provides an evaluation of the TDWG Ontology and provides recommendations for actions in relation to existing TDWG standards such as the Darwin Core. Another VoMaG outcome is the TDWG Terms Wiki[77], based on the Semantic MediaWiki software, and offered as an open platform for the collaborative development of terminology for the biodiversity information standards.

3.2.3 Data integration

The foremost challenge in biodiversity informatics is to make the multitude of participating international systems interoperable in order to support discovery and access to those data. Common exchange technologies (e.g. like XML and JSON) allow the syntactic exchange of data.

But such exchange technologies also need to understand the semantics of delivered data in order to process it meaningfully. If the data do not share a common reference model then the exchange will require some processing. For instance is this the problem with the Darwin Core Standard which is dedicated to occurrence. The unit of information in Darwin Core is a physical specimen or observational record. This is of limited value in the context of e.g. genomics that may contain information about environmental function without mentioning a taxonomic entity, or information about communities of taxa. It is therefore crucial that future efforts in this area take account of the ongoing work in major global initiatives like GEOBON, GBIF and the Genomics Standards Consortium as well as the novel approaches of eco-informatics.

3.2.4 Data reliability

[72] http://rs.tdwg.org/dwc/
[73] http://www.gbif.org/resources/2576

[74] http://code.google.com/p/gbif-providertoolkit/
[75] http://community.gbif.org/pg/groups/21382/vocabulary-management/
[76] http://www.gbif.org/resources/2246
[77] http://terms.tdwg.org/wiki/Main_Page

Information on fitness for use about how data is captured has to be described and explained. The challenge for biodiversity informatics in this context is to provide appropriate tools for data cleaning and automated procedures for data consistency checking. User feedback is the best way to validate data. A system is required for data publishers to display comments from identifiable users, providing feedback mechanisms or essentially an open peer review of data.

3.2.5 Physical infrastructure

Biodiversity information resources have so many different requirements, so there is no need for bespoke ICT technologies. A continued use of a wide variety of platforms and approaches is to be expected.

The biodiversity research domain will however place heavy capacity demands on the computing infrastructure. This requires robustness, stability and persistence. Over the hardware infrastructure lays the spatial data information infrastructure. All these features are associated with key institutions with long-term funding.

The leveraging of information from other domains will be increasingly important in the future (e.g. digital literature resources, image, environmental and climatic information databases). Specifying and supporting services from these domains built for long-term funding of biodiversity science use cases is required from the biodiversity community.

The social infrastructure is a major challenge. The technological capability is present but we need to increase its uptake by the community. For this we have to strengthen the socially connected network of experts spanning the ICT and biodiversity communities.

3.3 New tools

3.3.1 Digitization of legacy collections

The worldwide biological collections contain an estimated amount of 1.2 to 3 billion specimens (Duckworth *et al*. 1993, Ariño 2010). Only 10 percent of these specimens have been captured in databases, and much less is captured as digital images. This means that 90 percent of the collections are currently unavailable for use through the Internet. Making this available manually will be very labour intensive. Using mass- digitizing techniques and partly automated workflows through imaging techniques should however make the task feasible. Once digitized, the images have to be georeferenced, this can be computer assisted or with the help of gazetteer services.

The task can also be crowd-sourced by using volunteers, or outsourced to private companies. Distributed digitization infrastructures may become essential parts of most natural history collections. A major challenge is that collections grow much faster than they are digitized. As private collections also must be digitized by their owners, this requires easy to use and inexpensive tools that can be deployed at large scale. To effectively deliver this infrastructure service, digitization requires prioritization and its own funding channels.

3.3.2 Generating targeted and reliable data

Biodiversity data infrastructure needs more advanced informatics support, not only mainstream ICT, but also the ability to deal with the specifics of biodiversity feature and data. This includes informatics to support observations, detect events, species identification, data transfer, storage, filtering, and other kinds of data processing. New data gathering tools, new observatories and sensor networks have to be designed, created and tested. There should be automatic processes allowing for feedback from data interpretation back to the observation or detection at site. This combination of techniques and biodiversity informatics tools is expected to herald a revolution in biodiversity research, resolving much of the current fragmented data coverage and knowledge. This technological evolution can be accelerated through public-private partnerships.

3.3.3 The role of mobile devices

Developments in mobile platforms (smart phones and tablet PC`s with GPS) offer numerous opportunities for innovation in data collection and user information services. Combined with relevant Apps they can deliver reference products like identification keys and generate image-vouchered, location-tagged observations uploaded in a central database. Performing science where participants have varied level of expertise requires new techniques for data harvesting, processing, cleansing and validation.

3.3.4 Accessing data

Today most biodiversity data are searchable through Google and GBIF. These facilities are often designed for use by humans rather than for automated data retrieval, and may have in-built limitations or constraints. Users may therefore need to use more specialized resources in addition to GBIF (e.g. Users can use GBIF for retrieval by species name but have to use Catalogue of Life in order to get alternative names for species-based searching).

It is often not possible to refine keyword- type search in large volumes of searchable data. Finding the “needle in the haystack” is therefore difficult especially in the absence of widely- used vocabularies. Contextualizing information means establishing relations between data elements in order to improve search functionality. This is possible, but currently slow and difficult. Implementing persistent identifiers will make the construction of metadata portals much easier. A mature search mechanism that contextualizes information rather than simply indexing it would be far more powerful. Several techniques exist, and some are under development, making use of visualization methods to detect patterns and issues in data collections. These techniques could be useful for quality and fitness-for-use assessment in large datasets such as the LTER-Europe data index or the GBIF index and taxonomic nomenclatures. Data publishers need to go further in assisting the users to find data matching their requirements using persistent identifiers (PID) and knowledge organization system (KOS).

3.3.5 Data extraction from publications

Important information is often embedded in publications as text blocks or tables in a way that inhibits automated reuse. Semantic technologies (such as data mining) offer potential for liberating such data, but have not yet demonstrated the necessary flexibility or speed for broad uptake by the descriptive disciplines of taxonomy and ecology. Sometimes it is not possible to extract information because the style of writing demands that the reader know the context of the text in order to understand it.

New tools are therefore needed for data extraction from publications. These tools must use the vocabularies, ontologies and knowledge organization systems (KOS) to establish context between data elements, then to extract and assemble those elements into a format suitable for the user`s purposes.

Copyright held by commercial publishers remains a serious obstacle to recover the non-copyright factual data. Some older non-copyright publications are currently being digitized but serious issues remain. Errors with the Optical Character Recognition (OCR)[78] process means that search will return incomplete results. This results in an additional level of difficulty over and above the problems with extracting information from electronically published literature.

3.3.6 Data aggregation

Several data-aggregating initiatives have emerged over the last two decades for various areas of biodiversity informatics. Some of them were done on a project basis, while others were embedded in national structures making them more reliable in the long-term. These initiatives have important beneficial side-effects for biodiversity informatics in term of enhanced data availability, standardization and duplication (backup). Common aggregation problems related to hidden duplication, proper attribution, harvesting and storage remain. Each aggregator has tried to solve this on their own by developing their own data provider network and internal infrastructure. Increasingly they are recognizing the need to streamline and to avoid duplication of effort. To facilitate integration of these aggregators it is recommended to develop a catalogue of large-scale resources with associated metadata and a concept map.

3.3.7 Data quality

Non- interoperable data sources/domains are not the only obstacle to analyse complex patterns of biodiversity. Accessible data often have high quality with respect to reliability but low quality with respect to consistency. Data aggregated today have been collected for different purposes and on different spatio-temporal scales leaving significant gaps in the assembled data, and seriously hampering the analysis of complex patterns with data from different domains. Gaps can be filled by developing more comprehensive biodiversity observatory networks (BONs) and associated eTools to support the collection, aggregation and discovery of data from these observatories. There is also a fundamental need to reconsider previous efforts in terms of re-examining and reformulating existing

[78] http://en.wikipedia.org/wiki/Optical_character_recognition

data to make it more homogenous, to remove non-biodiversity factors and to make it suited for the kinds of analysis we foresee for the future.

3.3.8 Virtual research environments

Virtual research environments (VREs) or Virtual laboratories are online systems helping researchers to carry out their work. They include environments both to publish data (e.g. Scratchpads[79]) and to execute operations on data (myExperiment[80]) or both (AquaMaps[81] and iMarine[82]). VREs also include facilities to support collaboration between individuals. The challenge will be to build integrative and flexible eScience environments using standardized building blocks and workflows with access to data from various sources.

For a successful uptake of VREs they must generate immediate benefits for their users. For non-frequent users they must be simple and easy to use. For developers they must provide a pool of services and other resources that can be linked simply (e.g. BioVel[83]). VREs must perform functions that people find useful. VREs should also act as social networking applications having a central role in making some of the available technology better and more usable.

3.3.9 Future usage of data

Biodiversity eInfrastructures should provide more than generation, transfer, storage, long term archival and processing of biodiversity data. They should also support functionality for data analysis, prediction modelling and decision support. This will require user friendly VRSs; predictive multiscale models; feedback mechanisms to prompt new data; integrated interaction between data parameters, models and visualized results; and new approaches for decision support when model result in various scenarios.

3.4 The human interface

Experience from recent years of bioinformatics show that advancements will benefit from a balance between top-down and bottom-up approaches. The European Network for Biodiversity Information (ENBI)[84] concluded in 2005 that a modular infrastructure could provide both the infrastructure and the sustainability to overcome the partial and *ad hoc* solutions from the last twenty years of bioinformatic developments. This led to the design of the ESFRI LifeWatch Infrastructure[85].

History from the last decade show that this approach has not led to the expected development of trusted repositories or stable funding, and therefore have failed in generating user confidence. All

[79] http://scratchpads.eu/

[80] http://www.myexperiment.org/
[81] http://www.aquamaps.org/main/home.php
[82] http://www.i-marine.eu/Pages/Home.aspx
[83] http://www.biovel.eu/
[84] http://www.enbi.info/forums/enbi/index.php
[85] http://www.enbi.info/forums/enbi/index.php

biodiversity informatics projects share the same problem, namely how to keep the service running after the funding period. If users do not have confidence that an environment will last on a long-term they will not invest time and effort in contributing to those environments. To build substantial user confidence the service managers should invest much more than they have before in marketing to create new social norms in the biodiversity research communities.

Traditional citing systems measure impact and provenance related to a publication unit, rather than to code from a software library or data from a repository. New systems are required to generate comparable measures from open science resources. Data creators or software code contributors would benefit from knowing the number of people/projects using their data (impact). In addition, users would benefit from knowing who created the code or the data they use (provenance).

The ability to credit the data creator or code author is the primary basis for trust in the quality of the data or a service. This requires the resolving of several challenges. First, we need a robust system of attribution in a distributed network, using persistent identifiers and author identities. Secondly, licensing is poorly understood in the community both by producers and consumers. For data, the flavours of the Creative Commons license involving “non-commercial” clauses make risk-averse consumers careful in using the data even when the original intention of the producer was free use. For software code, the terms of open source licensing and free-use are similarly subtle. In both cases, there is a widespread failure to understand the distinction between licensing and copyright. Third, copyright often creates a barrier to data use and reuse, although in the academic world there are no instances of case-law that has been identified for this. Guiding principles are based on commercial publishing case law predicated on financial loss. The open wider community is pushing hard to clarify this situation.

Academic career progression is enormously influenced by citation metrics as a proxy for impact. This keeps the biodiversity informatics community tied to the paper publication model. Working to develop user products often gain nothing to the individual career progression.

People often hesitate to share data freely. They need to be given new tools that facilitate controlled data sharing. New metrics need to be defined that measure how often a data set is used and where conclusions on those appear. This will require the use of persistent identifiers and a system to identify contributors. This is the single largest problem we face in persuading people to share their data.

4. What is the scientific potential of openly accessible biodiversity and environmental data for individual researchers and institutions?

Easy access to open data on biodiversity and the environment is crucial for many researchers and research institutions, as well as environmental administration. Contemporary studies in biodiversity need to employ interdisciplinary approaches that require data input from different, often distant fields of science. Because the potential data users are not necessarily from same field of science as the originator of the data, it has become necessary to create mechanisms to acknowledge the effort of making well documented data openly available. This can support a continuously public validation and peer review of scientific data. Data made openly available are accessible for review allowing corrections, thereby increasing data quality. Openly available data give the researcher increased visibility that may lead to increased scientific accreditation. Increased recognition from decision makers and the public in general are other benefits.

Tools to support harvesting and gathering of data from multiple sources easily accessible for scientists will make the research more efficient. Easy access to data from different fields of science creates an environment for new scientific ideas to emerge. This potential of generating new, interdisciplinary approaches to pre-existing problems is one of the key features of open-access data platform that unify diverse data sources.

Open biodiversity and environmental data can be used in a wide range of research projects addressing the key scientific questions relating to the research field. Possible benefitting research fields are: Invasive alien species, climate change, conservation, food and farming, human health, and species distributions.

Unification of different data sources creates a potential to connect contemporary data with complementary information from the past (held in museums) and allows researchers to build a collection of observations repeated over time. In other words, when merging data from multiple sources, informatics becomes a vital part of research methodology. Its role is to ensure the interoperability and ”fitness-for-use” of the merged data before it can be analysed.

Biodiversity informatics is a relatively new discipline. Whereas the term “bioinformatics” is often used synonymously with the computerized handling of data in the specialized area of molecular biology, biodiversity informatics is the application of informatics techniques to biodiversity information for improved management, presentation, discovery, exploration and analysis. Without biodiversity informatics and its results, creating, developing and maintaining a research infrastructure becomes useless, because it can’t answer the needs of multidisciplinary (and/or interdisciplinary) research.

Biodiversity informatics can also reveal new hypotheses with its own methods. Data-intensive science organizes large volumes of data from multiple sources and fields and then analyses them using techniques tailored to the discovery of complex patterns in high-dimensional data through visualizations, simulations, and various types of model building. Through interpreting and analysing these models, truly novel and surprising patterns that are “born from the data” can be discovered. These patterns provide valuable insight for concrete hypotheses about the underlying ecological processes that created the observed data. Data-intensive science allows scientists to analyse bigger and more complex systems efficiently, and complements more traditional scientific processes of hypothesis generation and experimental testing to refine our understanding of the natural world.

Citizen science, the involvement of volunteers in research, has increased the scale of ecological field studies with continent-wide, centralized monitoring efforts and, more rarely, tapping of volunteers to conduct large, coordinated, field experiments. The unique benefit for the field of ecology lies in understanding processes occurring at broad geographic scales and on private lands, which are impossible to sample extensively with traditional field research models. Citizen science produces large, longitudinal data sets, whose potential for error and bias is poorly understood. Because it does not usually aim to uncover mechanisms underlying ecological patterns, citizen science is best viewed as complementary to more localized, hypothesis-driven research. In the process of addressing the impacts of current, global "experiments" altering habitat and climate, large-scale citizen science has led to new, quantitative approaches to emerging questions about the distribution and abundance of organisms across space and time.

Informatics should guide development of research infrastructures to better understand and document the "fitness-for-use" of different data types and also both collecting and managing of the data to enable the reuse of data no matter how and why it was collected.

5. Spin-off effects of open access for the general public

5.1 Research based policy (through public awareness)

"*Sustainable governance of our biological resources demands reliable scientific knowledge to be accessible and applicable to the needs of society. A system that facilitates open access to taxonomic data is essential because it will allow a sustainable provision of high quality data to partners and users, including eScience infrastructure projects as well as global initiatives on biodiversity informatics*" (http://www.eubon.eu/). In order to create, implement and evaluate sustainable policies affecting biodiversity on any scale, reliable, up-to-date research results need to be openly available.

5.2 Efficiency in management sector, publicly funded

Opening up government data not only enables the private sector to create services based on publicly shared data, it can also contribute towards improved public services. Increased reuse and redistribution of data also results in less duplication of work and increased efficiency.

Sharing of government data will provide the basis for new and better digital services, whether developed by the private sector or by the public sector itself. In either case, the result is increased and improved services for you and me from the public and private sector.

Data that are both openly available and easy to find are cost effective, as the need for investing resources into collecting the same data multiple times is removed. This frees resources to develop better services or invest in new data collection efforts, guided towards areas where it is needed the most.

5.3 Commercial interest and opportunities

Re-users need open data as a basis for creating new applications and services that rely on information provided by the public sector. These innovations can be sold in a continually growing market for apps and digital services. Open data represent raw materials that can stimulate innovation and new services of benefit to both the individual and society through more diversified services.

We are currently witnessing a technological revolution in the areas of analysis, utilisation and processing of data. The value of data is currently equated with capital and labour (The Economist) and can potentially contribute to an annual increase in value creation estimated at some 140 billion euros annually for the entire European Union (Vikery 2011). There are many examples of ideas that can contribute to the development of valuable services, but that have not yet been realized due to the lack of access to data.

5.4 Educational values of open data access

Open data can be used and exposed to general public with educational purposes on the Internet. There are many initiatives and ongoing projects, where information is made available together with other related resources like multimedia and occurrence data. These sites utilize standard interfaces for showing information and data from multiple sources, but also offer tools to allow for mobilizing and linking of new biodiversity data to be publicly available. Importing and linking data from multiple biodiversity resources and allow data to be re-used with attribution is very important in this context . Such global biodiversity initiatives is for example the Encyclopedia of Life, which is expanding to become a global community of collaborators and contributors serving the general public, enthusiastic amateurs, educators, students and professional scientists from around the world.

5.5 Public awareness (citizen science)

People have been observing and collecting specimens from the natural world for centuries - minerals, plants, fungi and animals. Originally the purpose was to understand weather and phenology for agricultural purposes. Early citizen science was mainly a tradition of collecting or hunting specimen. Today, there are an estimated two billion specimens housed in natural history museums around the world! These biological collections document where species and populations exist now and where they existed decades and centuries before, so they hold irreplaceable information necessary for uncovering the patterns of changes in species distributions and ecosystem composition over time. Scientists use such data and information in order to address the key environmental issues we are facing right now, such as the impacts of climate change and how diseases affect wildlife and humans.

5.6 Democratization

Open government data improves access to public processes and thereby helps keep public sector activities in check. We gain better insight into the bases for decisions and prioritization, improved knowledge upon which to assess our political administrators, and broader insight into the political process. This can strengthen the public confidence in public administration and the political system.

6. An internationally standardized Nordic metadata inventory

Metadata is important to guide the researchers in finding data, describe data quality for the "fitness-of–use", and use existing standards for vocabularies and protocols. Metadata is required to describe the scope and purpose of the study and make explicit the limitations of the data sets.

6.1. Metadata standards

Shared standards for metadata descriptions provide a common understanding of the meaning of data collected in data sets and will describe the purpose and methods followed when the data was collected. Numerous data standards for metadata have been developed, including Dublin Core as one very general example. A useful principle could be to follow common metadata standards for each respective data type. ISO 19115 provides a widely used standard for describing geospatial information and a common base for many of the datasets that are relevant for the LifeWatch community. With the growth of the semantic web we can now represent meaningful and interoperable metadata using a mix-and-match of terminology from many different ontologies (e.g. Baskauf *et al*. 2013).

The open geospatial consortium (OGC)[86] provides a framework of standards to describe geospatial information. Of particular interest for the LifeWatch community is the catalog service for the web (CSW), web map service (WMS), web feature service (WFS), web coverage service (WCS) and web processing service (WPS). The GeoNetwork opensource[87] project provides an interesting open source metadata catalog with implementation of the CSW standard that is of particular relevance for building a network of interoperable LifeWatch services.

GBIF recommends following a subset of descriptors[88] from the Ecological Metadata Language (EML)[89] when publishing species based biodiversity information as Darwin Core archives (GBIF 2011). The elements of the EML derived GBIF metadata profile are grouped into categories such as: Geographic coverage, taxonomic coverage, temporal coverage, project description, sampling methodology and

[86] http://www.opengeospatial.org/

[87] http://geonetwork-opensource.org/

[88] http://rs.gbif.org/schema/eml-gbif-profile/

[89] http://knb.ecoinformatics.org/software/eml/

protocols, the associated institutions, collections and people, and intellectual property rights and licensing.

6.2. Finding an appropriate level of ambition for metadata harvesting

A portal provided from LifeWatch (national or Nordic level) need a catalogue of environmental layers. Linking the availability of environmental layers in the portal and allowing users of the portal to describe and provide EML metadata for the relevant environmental layers they want to use could be a mechanism contribution to the population of the metadata catalogue. Gaining access to use the data sets they describe, could provide the incentive for them to do the work of describing them.

For a LifeWatch linking/analysis portal to access environmental data sets, they need to be available on an OGC compliant web service such as WFS. If the required OGC web service endpoint cannot be established and hosted long term from the data owner or a partner institute, we could foresee a mechanism for uploading data sets other than occurrence data where a LifeWatch structure could host the data to provide the OGC web service endpoint, required by the data linking/analysis portal to access the data set. For observation data such solutions already exists (e.g. The Norwegian Artsobservasjoner.no and GBIF) that should be used.

6.3. The Nordic template for metadata harvesting

The Nordic template for metadata harvesting was developed in accordance with ISO 19115[90] and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[91]. The template is therefore compatible with the European INSPIRE[92] and the American GCMD[93]. All mandatory fields are marked with red text - in Appendix 1.

6.4. Nordic biodiversity data inventory list

The Nordic countries are all members of GBIF and share a substantial part of the occurrence data made available through the GBIF data portal. In addition to the occurrence data shared through the GBIF infrastructure, the Nordic countries also hold other ecological data sets relevant to LifeWatch. This chapter provides an overview for some of these data sources.

Status GBIF (2013-11-29)	Data sets	Occurrences
Denmark[94]	46	9 326 214

[90] http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020
[91] http://www.openarchives.org/OAI/openarchivesprotocol.html
[92] http://inspire.jrc.ec.europa.eu/
[93] http://gcmd.nasa.gov/Aboutus/index.html
[94] http://www.gbif.org/country/DK/publishing (visited 29 November 2013)

Finland[95]	57	14 666 474
Iceland[96]	4	458 705
Norway[97]	85	12 697 671
Sweden[98]	48	43 382 912

6.4.1. Norway

See the Norwegian metadata inventory list from http://talos.nodc.no/infrastruktur/listmetadata.html

LifeWatch Norway performed during the pilot project a temporary mapping exercise on existing biodiversity data in Norway. The mapping exercise used technology that was developed by the Norwegian Marine Data Centre (NMDC)[99] for metadata harvesting. ,

,

The metadata mapping was performed in two ways:

- Biodiversity metadata already captured by Artsdatabanken (the Norwegian Biodiversity Information Centre).
- Direct consultation with 103 institutions and experts asking them to register metadata.

This metadata mapping exercise resulted in a list of 281 unique data series from about 31 Norwegian Institutions and Universities. Previously registered metadata from long time-series in different environmental monitoring campaigns (the Norwegian Research Council, 2004[100]) are not included in this mapping exercise, but should absolutely be registered in a Norwegian LifeWatch Construction phase. Figure 2 shows the distribution registered metadata by ISO-category:

[95] http://www.gbif.org/country/FI/publishing (visited 29 November 2013)
[96] http://www.gbif.org/country/IS/publishing (visited 29 November 2013)
[97] http://www.gbif.org/country/NO/publishing (visited 29 November 2013)
[98] http://www.gbif.org/country/SE/publishing (visited 29 November 2013)
[99] http://www.imr.no/forskning/faggrupper/norsk_marint_datasenter_nmd/en
[100] http://www.forskningsradet.no/no/Nyheter/Ny_oversikt_over_marine_dataserier/1236685434600?lang=no

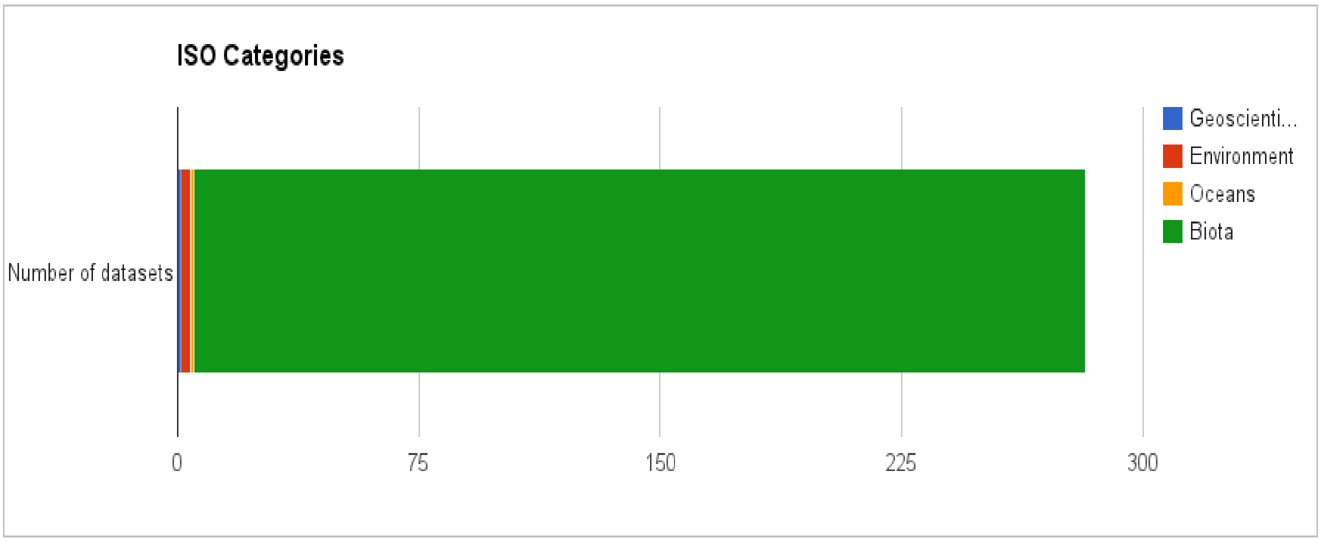

Figure 2: Data sets distributed by ISO-category.

6.4.2. Sweden

In Sweden there are two main initiatives that provide metadata on data sets and web services relevant for LifeWatch, i.e., Environment Climate Data Sweden (ECDS)[101] and the Geodata Portal (Sweden Bit by Bit)[102]. The data sets in these to catalogs mainly consist of environmental data but also species observations. The scope differs little between the two initiatives. ECDS is focusing on research data and will probably to a greater extent than Geodata, handle rather small and scattered data sets, which are at best available in non-standardized file formats for manual download. The Geodata Portal has much focus on data sets relevant for INSPIRE.

6.4.3. Denmark

There is no recent comprehensive metadata inventory of biological data from Denmark. DanBIF (national Danish GBIF-node) provides 46 occurrence dataset with more than 9 million species observations, but these only represent a minor part of the biodiversity data stored in databases by Danish authorities, universities, natural history museums, NGO-s and private data repositories from terrestrial, freshwater and marine environments. DanBIF is currently in contact with several new data providers and expect to make significant amounts of biodiversity data available to the international community via GBIF through 2014 and onwards.

6.4.4. Finland

[101] http://www.smhi.se/ecds
[102] http://www.geodata.se/

Finnish organizations currently share 61 datasets with 14,7 million records via GBIF. Even if the numbers are quite impressive, the Finnish Biodiversity data is still stored in a very separate way and much remains invisible to data sharing networks like GBIF, LTER or LifeWatch. More than 20 organizations functioning under three different ministries as well as private sector including NGO`s have been identified to both collect and use biodiversity data. There are also many separate networks that have surveyed data sources from their viewpoint, but only a few metadata repositories exist.

The Finnish Museum of Natural History surveyed occurrence data in collections by capturing metadata from Finnish museums in 2009. This metadata is maintained and accessible at: http://collections.luomus.fi/

6.4.5. Iceland

The largest datasets in Iceland concerning occurrences of different species are kept at the Icelandic Institute of Natural History and several of them are shared via GBIF. More datasets will be ready for sharing soon. At the IINH big datasets concerning ecological data are being compiled and digitized and those datasets will be available for LifeWatch during 2014.

Several other institutions keep biodiversity data in Iceland as the Marine Research Institute (www.hafro.is) and the Institute of Freshwater Fisheries (www.veidimal.is) among others.

There is no recent metadata inventory concerning biological data kept in Icelandic institutions. An inventory was conducted during the BioCASE-project in 2002. That inventory did, however, not result in full coverage of the biodiversity data present at Icelandic institutions.

7. Legal framework and challenges associated with environmental, climate, and biodiversity data sharing

Presently, we describe the legal framework and the current challenges associated with environmental and biodiversity data sharing. This document will give an overview of the national specifics as well as common international principles, laws and guidelines governing environmental and biodiversity data sharing.

“Open data” is the idea that data should be freely available for all to use and publish as they please, without restrictions from intellectual property rights, patents and other control mechanisms. The goals for the open data movement are similar to those of other open movements such as open source, open content and open access. The philosophy behind open data has long been established, but the concept of open data in itself has been recently popularized through the growth of the Internet and the World Wide Web.

Substantial amounts of publicly financed environmental and biodiversity data are currently either not accessible or difficult to access for society. If these resources were to be made freely available in standardized formats, the added value will greatly surpass the costs involved with the investment. In addition it will contribute to a more open public sector with increased cooperation.

In order to implement such data sharing practices, it is necessary to look at current principles, laws and guidelines that govern data sharing. With some exceptions the main rule is that publicly funded data should always be available for the society. Various technical, organizational, financial and legal factors can however be obstacles in various actors' ability and will to share data. These barriers are especially pronounced in the relationship between public and private institutions, where economical terms are exceedingly different. It is also important that efforts on making more data openly available are in accordance with society's general-, and research's specific requirements regarding quality controlled relevant datasets.

International principles, laws and guidelines that particularly govern data sharing are the OECD Principles and Guidelines for Access to Research Data from Public Funding, the European Directive on the Re-use of Public Sector Information (EU Directive 2003/98/EF), and INSPIRE (Directive 2007/2/EC of the European Parliament and of the Council of 15 May 2007 establishing an Infrastructure for Spatial Information in the European Community).

In 2004, all OECD member states signed a statement that publicly funded data should be made publicly available. Following a request and an intense discussion with data producing institutions in the member states, the OECD published their Principles and Guidelines for Access to Research Data from Public Funding in 2007 as a "soft law" recommendation.

7.1. The OECD principles

The OECD principles are guidelines aimed towards assuring open, sustainable and cost-effective access to publicly funded data. The principles define infrastructural requirements regarding interoperability, flexibility, quality, and documentation (metadata), as well as mechanisms for training and rewarding. The OECD principles point out the need for data sharing rules, and recommend that these are to be in line with guidelines for national research programmes and adapted to requirements regarding national security, intellectual property, competitive information, citation, copyright, sharing, limits to use, financial agreements, ethical rules, license agreements, legal agreements, and sustainable archiving. The OECD principles specify that intellectual property rights of private data producers should not necessarily limit open data sharing in publicly financed projects. In cooperative projects, the OECD principles advise to formalize responsibilities and rules for data sharing at the earliest possible stage. The OECD principles point out the need for data sharing to be implemented in accordance with applicable ethical standards and values in the scientific communities. The contractor can be given exclusive access to the resulting data for own scientific publishing for a limited time, but this should be explicitly agreed upon by the funder and researcher at the onset of the project. The OECD emphasize that the public sector should assure sufficient financing for the contractors data management. Research funding should not finance data management. Any costs involved in acquiring access to data should not surpass the actual costs involved in the delivery of the data.

7.2. The EU-directive on the re-use of public sector information

The EU approved the "European directive on the re-use of public sector information"[103] in 2003 in order to make use of the great potential for re-use of data from the public sector. The directive strengthens member states' obligations to share data from the public sector through minimum requirements regarding access to publicly financed data for viewing and re-use.

7.3. INSPIRE

INSPIRE (Directive 2007/2/EC of the European Parliament and of the Council of 15 May 2007 establishing an Infrastructure for Spatial Information in the European Community[104]) aims to establish a shared international infrastructure for spatial information. The directive commands public sectors to share spatial information electronically and to enable the public to search, view and download the content.

7.4. Implementation of international laws, principles and guidelines in national legislation governing data sharing

7.4.1. Norway

The OECD Principles and Guidelines, the European Directive on the Re-use of Public Sector Information (EU Directive 2003/98/EG), and INSPIRE are all implemented and supplemented by the Freedom of information act (Offentlighetsloven), the Spatial Data Act (Geodataloven), the Infrastructure for Spatial Information Act (Geodataforskriften), the Act Relating to the Right to Environmental Information and Public Participation in Decision-making Processes Relating to the Environment (Miljøinformasjonsloven), and the Nature Inspectorate Act (Naturoppsynsloven).

The Spatial Data Act (Geodataloven) obliges the public management sector and others with management tasks decreed by law to share publicly funded data and to offer free public services for

[103] http://ec.europa.eu/digital-agenda/en/european-legislation-reuse-public-sector-information
[104] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32007L0002:EN:NOT

searching, displaying, transforming and downloading spatial data. Except the Freedom of information act (Offentlighetsloven) and the Act Relating to the Right to Environmental Information and Public Participation in Decision-making Processes Relating to the Environment (Miljøinformasjonsloven), public and private research institutions are not decreed by law to share data unless it is regulated by contract or if the institution performs management tasks decreed by law.

7.4.2. Sweden

The INSPIRE directive has been implemented into Swedish law. The Act and Ordinance on spatial information (Lagen och förordningen om geografisk miljöinformation) regulates the Swedish implementation of INSPIRE. The EU Directive on Public Service Information (PSI-direktivet) (2003/98/EG) is implemented in Swedish Law through "Lagen om vidareutnyttjande av handlingar från den offentliga förvaltningen", also called the PSI law (re-use of Public Sector Information).

7.4.3. Denmark

The Danish Act on Infrastructure for Spatial Information (GI-loven) incorporates the regulations, principles and associated guidelines of the INSPIRE Directive into Danish law and the Act on the re-use of public sector information (Lov om videreanvendelse af den offentlige sektors informationer) implements the European Directive on the Re-use of Public Sector Information (EU Directive 2003/98/EG).

The Act Relating to the Right to Environmental Information and Public Participation in Decision-making Processes Relating to the Environment (Miljøoplysningsloven) and Order on Active Dissemination of Environmental Information (Bekendtgørelse om aktiv formidling af miljøoplysninger), the Danish Public Administration Act and the Public Records Act implements the Aarhus Convention.

The present status of the legislation in Denmark seems to be very similar to the Norwegian as public and private research institutions are not decreed by law to share data unless the above-mentioned laws decreed it or it is regulated by contracts.

7.4.4. Finland

As a guideline for implementing international and national laws and principles, the Finnish Partnership for Research on Natural Resources (LYNET Consortium) worked to form a data policy statement for environmental and biodiversity data collecting, management and sharing. The aim was to ensure more open access to research data, which is an essential prerequisite for effective public-funded research, the implementation of economic innovation chains and fruitful international cooperation. It also serves as a suitable outline for other institutions data policies, when collecting, producing or working with similar kinds of data.

The policy on data was formed in order to:

- Give data openly available as soon as possible after the first use,
- Give data to a wider and more efficient use,
- Increase the co-operation and collaboration between producers and re-users of data.

This policy covers data acquired, assembled or created through statistics, survey, research and monitoring.

The LYNET institution’s publicly funded, and acquired data sets are openly available unless otherwise restricted by law or regulation. Because of commercial or other specific reasons some exceptions might occur.

Data sets are “free-of–charge” licensed and will be online accessible for individuals, public and the private sector. Copyrights belong to the institution responsible for the specific data set.

The personnel of institutions applying this data policy have a right for the 2 years first use period (embargo time) of data sets. After that data sets will be released for open access.

7.4.5. Iceland

Legal basis for open access to publicly funded data were established in 2010 and later with Act on public information (2012 nr. 140) and Act on open access to environmental information (2006 nr. 23). The legislation is adapted to the European Directive of Re-use of Public Sector Information although it has not been actively implemented at all institutions. There is, however, rarely any legal problem with accessing data that has been collected by governmental institutes. However, poor metadata might cause a problem in finding relevant data. Lack of institutional resources may also be an obstacle if data needs restructuring to be made accessible.

7.5. Common Nordic barriers towards making publicly funded research data openly accessible

7.5.1. Technology, standards and financial framework

Infrastructure, capacity, competence and lack of funding represent barriers for making public funded data openly accessible. Harmonizing technology and databases towards interoperable data protocols for open access is in itself a practical issue, though highly influenced by the institutional priorities, in-house ICT-competence and financial capacity.

Many institutions would like to share data but lack the ability to prioritize it, and in the longer run, to realize it. Making data available is a very important effort, but to ensure that this work does not depress the institutional research activities, authorities should target specific capacity building programmes towards the institutions. This is especially an issue in Norway where many research institutes are organized as foundations outside the governmental sector.

7.5.2. Institutional culture and individual researcher attitudes

Conflicting informal agendas both within and between research institutions will always influence the actual data sharing ability. These agendas represent as such potential barriers for open access to publicly funded data. Institutions should therefore be obligated to develop strict data policy strategies in order to prevent such barriers from evolving.

Metadata standards may be too comprehensive and complicated to use, resulting in a very time-consuming metadata mapping. This problem was apparent in the Norwegian LifeWatch metadata mapping based on the international ISO 19115/139 metadata standard (compatible with both INSPIRE and GCMD), resulting in a surprisingly low response rate from some data owners. This problem has also been reported from similar international projects (Schmidt-Kloiber *et al*. 2013). The reluctance to report metadata could have several explanations.

Metadata reporting can be both time- and work consuming. When such resources are limited, the individual scientist may fear that this work has to be done at the expense of doing real science. One solution to avoid this researcher's dilemma could be to allocate sufficient resources for mandatory metadata reporting in project contracts. Beniston *et al*. (2012) suggest developing an EU-directive in the form of a best practice guideline for data management, data sharing and open access. Another, non-financial means of compensating for metadata reporting burden could be to urge scientists to use social scientific networks such as Mendeley[105] or ResearchGate[106] to showcase their reported metadata.

Some researchers fear that sharing metadata with the general public implies losing the intellectual properties to the data itself. It is therefore very important to underline that publishing metadata does not automatically imply making data freely available. Information about intellectual property, criteria for use, and contractual arrangements need to be specified in the metadata.

When it comes to licensing and accreditation, lack of knowledge can be an obstacle for researchers to share under an open license. While the function of an open license is to make data more widely reusable while maintaining ownership and ensuring due to accreditation, it is often felt as 'giving something away' rather than sharing. This holds especially true for a license allowing for commercial use, where many may feel that it allows others to profit financially from work that one has shared freely. In reality, it is all but impossible to earn money from work under an open license, as the license and owner have to be clearly stated. It does however allow third parties to invest in the usage of the data, adding to its value to such an extent that it is economically viable as a new product. It is important to communicate these aspects well, to safeguard both the influx of data as its usability under the appropriate terms.

---

[105] http://www.mendeley.com/
[106] http://www.researchgate.net/

Some data owners fear that open access could lead to misuse and/or misinterpretation by other scientists. These data owners also fear that they will not be credited if other scientists use their data. Again these worries could be met by a thorough metadata provenance scheme. Uncertainty and lack of knowledge on the consequences of open access often lead to scepticism among scientists and data owners. As a response to this the Norwegian Agency for Public Management and eGovernment has developed an Open data handbook addressing legal, social and scientific aspects of open data sharing. This handbook only addresses these issues at a general level, and extended documentation for scientific issues is therefore desirable.

7.5.3. The need for academic accreditation of open data access

Academic accreditation is important for any researcher's scientific career. Currently, open data sharing does not directly honour researchers the same way as published scientific articles in high ranking journals.

Electronic publishing of scientific literature calls for mechanisms of online reference citations that ensure future recognition and retrieval. Electronic publishing of data represents similar but more complex challenges. Validated data citation standards and best practices have to be implemented to meet these challenges. Several institutions, countries and disciplines have been working on this task for some time. Political and technical approaches have been introduced by the American National Information Standards Organisation (NISO)[107].

Others have been working on the development of Persistent Identifiers (PID) which are live and easily maintainable identifiers referring to digital objects, data files (such as documents, pictures or software installation files) or to physical entities (such as collection specimens). PIDs are not like hyperlinks, as they will remain valid after being moved between clients and organizations. Hyperlinks (HTTP URLs) provide a location service, while PIDs provide a naming service for the entity. Several standards have reached a mature level of development:

- Uniform Resource Name (URN)
- Persistent URL (PURL)
- Digital Object Identifier (DOI)
- National Bibliography Numbers (NBNS)
- Archival Resource Key (ARK)
- Open URL
- Universally Unique Identifier (UUID)

The Digital Object Identifier (DOI)[108] became an ISO-standard (26324)[109] in May 2012, driven by the international non-profit DOI foundation (established in 1998).

---

[107] http://www.niso.org/home/
[108] http://www.doi.org/
[109] http://www.iso.org/iso/catalogue_detail?csnumber=43506

Delegates from several CODATA-committees (Committee for Data for Science and Technology)[110], the ICSTI (International Council for Scientific and Technical Information)[111] and others have established an international task group on data citation standards and practices[112] to investigate relevant issues, coordinate activities and to suggest common practices and standards for the scientific community. http://www.nature.com/press_releases/scientificdata.html

Other important groups are the DataCite initiative[113] and the group composed of SCOR (Scientific Committee on Oceanic Research)[114], IODE (International Oceanographic Data and Information Exchange)[115], and MBLWHOI (Marine Biological Laboratory/Woods Hole Oceanographic Institution Library)[116], meeting annually to discuss data storage, interdisciplinary data publication, and interaction with scientific publishers. In 2012 the group made an online bibliography[117] of institutions focusing on data citation and referencing practices.

In 2012 GBIF[118] and PenSoft Publishers[119] pioneered a workflow[120] between the GBIF Integrated Publishing Toolkit (IPT)[121] and Pensoft journals such as PhytoKeys[122], ZooKeys[123], BioRisk[124], NeoBiota[125] and Nature Conservation[126] to automatically export metadata into the form of a data paper manuscript, based on the Ecological Metadata Language (EML)[127]. Data papers are scholarly journal publication whose primary purpose is to describe a dataset or a group of datasets, rather than to report a research investigation. As such, it contains facts about data, not hypotheses and arguments in support of the data, as found in a conventional research article. Pensoft has recently established a new journal Biodiversity Data Journal[128] dedicated to publishing data papers (Smith *et al*., 2013).

---

110 http://www.codata.org/
111 http://www.icsti.org/
112 http://www.codata.org/taskgroups/TGdatacitation/index.html
113 http://www.datacite.org/
114 http://www.scor-int.org/
115 http://www.iode.org/
116 http://www.mblwhoilibrary.org/
117 http://www.codata.org/taskgroups/TGdatacitation/Bilbliography_Links.html
118 http://www.gbif.org/
119 http://www.pensoft.net/
120 http://www.gbif.org/communications/news-and-events/showsingle/article/new-incentive-for-biodiversity-data-publishing
121 http://www.gbif.org/communications/news-and-events/showsingle/article/new-incentive-for-biodiversity-data-publishing
122 http://www.pensoft.net/journals/phytokeys
123 http://www.pensoft.net/journals/zookeys/
124 http://www.pensoft.net/journals/biorisk/
125 http://www.pensoft.net/journals/neobiota
126 http://www.pensoft.net/journals/natureconservation
127 http://knb.ecoinformatics.org/software/eml/
128 http://biodiversitydatajournal.com/

The Nature Publishing Group will launch in May 2014 a new platform named Scientific Data[129], for open-access, online-only publication of the descriptions of scientific valuable data sets. Scientific data provides formal peer-review for scientific data sets and a solution for citation of data sets in a similar manner as for citation of other scientific works. Scientific Data will build further on more general data repositories such as Dryad and Figshare where the actual data sets to be described in Scientific Data will be uploaded and can thus be accessed. The Dryad data repository is also open for researchers to share scientific data sets underlying scientific publications accepted by peer-review journals. Figshare provides an even more flexible platform for researchers to share their scientific results in any file format and in a scholarly citable, sharable and discoverable manner. Also GBIF (Global Biodiversity Information Facility) will start collaborating with Scientific Data, the collaboration will develop methods for authoring submissions to the new journal using the standard formats for metadata recommended for sharing data through the GBIF network.

7.5.4. Data management, strategies and contractual arrangements

Making publicly funded data accessible from structured databases is an institutional responsibility that has to be implemented in long-term strategies. Internationally, several scientific funding sources have started to demand data management plans for project funding. Such data management plans should act as a guideline checklist to ensure that data are managed according to all relevant aspects of data management.

The international Data Curation Center (DCC) in the UK has developed a data management plan template[130]. DCC has also developed an online solution for developing and maintaining such a data management plan. Examples on data management plans are also available from the Australian National Data Service (ANDS)[131].

When it comes to actual strategies regarding development and implementation of data management plans there is substantial variation among Scandinavian research institutions. GBIF has developed the "Best practice guide for Data Discovery and Publishing Strategy and Action Plans"[132]. Strategies and plans have to outline all aspects related to responsibilities, intellectual property rights, utility value and synergies associated with data sharing.

It is very important that these issues are formalized in contractual arrangements at the start of an inventory project or a research project. The source of funding is responsible for consistently embedding such practices in contracts, and to see to their implementation by the contractor. The contractor is then legally responsible for the collection, maintenance, documentation and sharing of data.

Factual data, also in the form of prosaic text, is not copyrightable, and thus does not require a license to allow its reuse. There are a number of reasons to apply a license nevertheless. First and foremost,

129 http://www.nature.com/scientificdata/
130 http://www.dcc.ac.uk/resources/data-management-plans
131 http://ands.org.au/datamanagement/index.html
132 http://www.gbif.org/orc/?doc_id=2755

a sufficiently open license ensures the usability of a dataset by covering the anomalous copyrightable data within it, without requiring a check for such data. In addition, a license setting requirements to attribution is psychologically important as it seemingly guarantees data sharers proper scientific accreditation, even though scientific accreditation of data sources is not governed by copyright.

The most common options for scientific data are either licensing under a license requiring attribution, such as Creative Commons Attribution (CC BY), or a waiver such as Creative Commons Zero (CC0) where the dataset is released into the public domain, thus freeing it from possible copyright restrictions. In the latest version of CC BY (4.0) accreditation is flexible enough to remove the need for anyone re-using such data to determine how to properly attribute these, which would have renders large, compound datasets practically non-reusable. Scientific practice requires scientific attribution, regardless of enforcement by a license, so the more open CC0 guarantees the usability of the data without annulling proper scientific accreditation of the source.

In practice, however, data owners are more reluctant to share even non-copyrightable data without a license demanding attribution. This may lead to data not becoming available, whereas it would have under a license that, where legally applicable at all, does not extend beyond common scientific practice. It can thus be more pragmatic to allow a CC BY license with an added predefined agreement in which attribution is specified in a way that does not hinder use within or outside the scope of LifeWatch.

It is essential for reuse of large, compound datasets that both content and licensing are machine readable: Directly readable and processable by a computer so that it can readily be shared between IT systems.

7.6. Discussion

In general, open data sharing will immediately showcase data, methods and data producers to the public. Collectively, the management and research sector collect a broad range of data in their work. Making these data freely available contributes to an increase in business development, value creation, innovation, research, efficiency, transparency and democratic control within society. Furthermore, sharing data allows for quality control by multiple external parties.

7.6.1. Norwegian experiences

The pilot project LifeWatch Norway has (as have earlier investigations) brought to light a major need for users to have easier access to publicly financed environmental- and biodiversity data across institutional- and national borders. Experiences with the institutions cooperation with sharing data through the Norwegian Biodiversity Information Centre[133] and GBIF[134] have been positive, but only partially meet current user demands.

[133] http://www.artsdatabanken.no

[134] http://www.gbif.no

The Norwegian LifeWatch user survey shows that there is a large demand for access to more types of observational-, collection- and time series data that remains unmet. There is also a great demand for environmental- and climate data that is only partially accessible from the respective owners and the national geographical infrastructure "Norway Digital"[135].

Recent years have brought the society an increasing access to publicly, sector based map data through a multitude of web-based map services and distribution channels where data can be downloaded. Several of our respondents express a need for better cooperation between the distribution channels.

Improved access to data in itself is not sufficient to meet all the requirements of research and management. Many respondents have expressed a need for differentiation of functionality, such as data aggregation, distribution map generation, statistical analyses, and possibilities to link observational data to environmental data and different map layers for further GIS analyses, GIS modelling, and numerical modelling.

The Norwegian LifeWatch Consortium recommends several actions for the realization of a national infrastructure for sharing publicly funded environmental and biodiversity data. These include the development of contract templates, strategies and action plans for improved data management, capacity building, development of data mobilization tools, and an evaluation of current terms and conditions towards the research institute sector in Norway.

On both national and international levels there are many parallel initiatives on data sharing. In Norway we will seek to align the efforts (and the associated resources) of LifeWatch, EU BON, GBIF, The Norwegian Biodiversity Information Centre (NBIC) and the Norwegian Marine Data Centre (NMDC).

To realize its goal of increased effort in public innovation, the Research Council of Norway has started to develop strategies aiming to improve the flow of knowledge between research, industry and the educational sector. A national LifeWatch eInfrastructure for environmental- and biodiversity data will contribute to increased quality, innovation and efficiency within these three sectors.

At this stage it is hard to specify how a future national LifeWatch eInfrastructure should be organized. The specific needs for such an eInfrastructure should be scoped to satisfy the needs identified in this pre-project. Furthermore it is very important that this eInfrastructure will be sufficiently aligned with existing parallel initiatives on national, Nordic and international levels.

[135] http://norgedigitalt.no/

The Norwegian LifeWatch Consortium currently works on including other relevant ministries and raising funds from the Research Council of Norway for a large scale Norwegian LifeWatch infrastructure.

7.6.2. Swedish experiences

The Swedish LifeWatch consortium was formally established in 2011 when a contract was signed by six national parties. The project has been successful in following the construction plan and building a national eInfrastructure for biodiversity data. Many of the central components are already in place and in 2014 the infrastructure will reach its full capacity according the plan for the first construction phase.

The major part of the funding for the Swedish LifeWatch project comes from the Swedish Research Council (SRC) but all consortium partners contribute with co-funding. The first contract period with the SRC spans from 2010 to 2014 and an application for a second funding period (2015 to 2019) will be submitted in March 2014. Although there are no guarantees for future funding, the SRC has signalled that their investments in research infrastructures are considered long-term engagements.

Swedish LifeWatch integrates databases containing species observations from citizen science, research, monitoring, inventories, and museum collections, in easily accessible formats. The infrastructure also provides access to a large range of environmental data, through web map services, which can be directly used in analyses and models. A coherent Analysis portal provides tools for handling, visualizing and analysing data.

The Swedish LifeWatch strategic plan covers the construction phase 2010 to 2014 and is updated on a yearly basis. The plan presents an overview of goals in terms of deliverables and a prioritization for reaching the goals. The main goal is constructing a strong eInfrastructure for biodiversity research, where distributed databases holding data on biodiversity and explanatory variables will be connected through web services, using agreed standards and a common taxonomy. Other strategic goals are to provide easily accessible resources, achieved by user support, strategic communication, and education. The strategic plan has also included show cases and pilot studies to test the infrastructure and demonstrate its potential to stakeholders and funders.

Another important objective for the Swedish LifeWatch has been to establish a dialog with future and potential users to meet the actual demands of the research society. This has mainly been achieved through user meetings/workshops arranged at different universities, reference groups and web surveys.

In September 2012, a web survey was sent out to researchers at all Swedish universities holding research and education within biology and/or environmental sciences. The survey also reached conservation biologists, mainly at regional county administrations (länsstyrelser) and consultants at SMEs within relevant fields. In total, the survey was sent out to ca. 2500 people and 312 answers were submitted. Of the respondents, two thirds were working with research and one third with nature conservation and management.

Results from the survey indicated that although 89% of the respondents use biodiversity data to a large extent in their daily work, 93% would need better access to biodiversity data. Occurrence data for plants, birds and insects were most requested, but there was a clear need for occurrence data for all major groups of organisms as well as other environmental data like habitat, ecological features, species traits, climate data, land-use, topography, hydrology, etc. A user-friendly interface, no charges and good access to support were other important desired features.

In the survey, researchers also expressed a strong interest for sharing data for open access through the infrastructure, under condition that the data can be non-public or shared only within the research group while awaiting publication.

Experiences from user meetings and workshops have been very useful when planning the functionality and priorities of the Analysis Portal. One important outcome of the user meetings has been that the development of specialized analytic tools should be less prioritized, as many researchers prefer that the infrastructure focuses on data accessibility and possibilities to search, filter, visualize and download data, while they generally prefer to do advanced analyses with already established and well-known programs and tools.

7.6.3. Danish experiences

There is yet no official Danish LifeWatch project in Denmark, but plans are afoot to establish a national consortium in spring 2014 with participation of major universities, museums, NGO's and the authorities. During the past years, however, public access to biodiversity data has been discussed widely and a number of portals now offer biodiversity data. These efforts are still largely uncoordinated and they provide raw data and rarely tools for data aggregation or analysis. No full survey has been made, but experience suggests that a common portal for storage, retrieval and analysis of biodiversity data for research, decision making and for the public is in great need. A Danish LifeWatch node could fill this gap.

In Denmark there are several sources for sharing of environmental and biodiversity data. One of the largest and most dynamic datasets is maintained by the Ministry of the Environment. The Ministry is responsible for The Nationwide Monitoring and Assessment Programme for the Aquatic and Terrestrial Environments (NOVANA) that entered into force on 1 January 2004. This program is designed to meet Denmark's obligations in relation to EU legislation, international conventions and national legislation. Furthermore, the program is expected to contribute to the development of the scientific basis and to develop various tools to be used in regional or national action plans. All data collected in this program are publicly available on an internet-portal: http://naturdata.miljoeportal.dk. Most museums and many NGO's that collect and store data have made their data accessible through DanBIF (the Danish GBIF node - danbif.dk).

7.6.4. Finnish experiences

The biggest impediment in forming a national infrastructure for sharing of publicly funded environmental and biodiversity data has been the fact, that the responsibility falls under three different ministries (Ministry of Environment, Ministry of Forestry and Agriculture and Ministry of

Education and Culture). The Ministry of Finances has recognized that the opening of publicly funded data would greatly enhance the efficiency and effect of many governing processes in public sector as well as it would offer a boost to small and medium sized companies, but intersectional funding tools are still lacking from the funding programs. Secondly, biodiversity-informatics as a methodological research field has not been recognized as an essential component in developing the existing research infrastructures further. So far, very limited research infrastructure funding has been allocated for LifeWatch activities in Finland (100 000 € for coordination activities 2013 to 2014). Furthermore, a LifeWatch project proposal was not accepted to the second evaluation round in the updating process of the national research infrastructure roadmap in spring 2013.

7.6.5. Icelandic experiences

Currently there is no LifeWatch project in Iceland. The need for easier access to publicly financed environmental- and biodiversity data across institutional- and national borders, is though generally recognized, and Iceland has towards that end participated in cooperation with sharing some data through GBIF. Several types of observational-, collection- and time series data are still not as openly accessible as possible, since these data are distributed in various institutions on different formats. Presently these data also have limited functionality, regarding linking of data, GIS mapping, statistical analyses, etc. The most feasible way to implement LifeWatch in Iceland would be to establish a close cooperation with the other Nordic countries, i.e. by building on already existing infrastructure in these countries for sharing of publicly funded environmental and biodiversity data.

8. Communication and training

8.1. Communication - the key to success

Communication is a key to success for the establishment of a Nordic LifeWatch infrastructure. Internal communication is needed for a successful construction and cooperation, and external communication in order to raise awareness about the project amongst the target groups and key stakeholders, and to make the project and its services well known and well used. In addition, an efficient communication and dialogue with potential users will help us to meet the actual needs of the research community and the users.

We strongly recommend that financial and personnel resources are allocated already at the onset of a Nordic LifeWatch construction phase to guarantee an efficient and well planned communication.

Ambitious communication plans that could serve as inspiration and models have been formulated as work package deliverables by related projects like LifeWatch Europe in its preparatory phase (LifeWatch communication plan) and EU BON (D8.2 Dissemination and Communication Strategy and Implementation Plan).

8.1.1. Communication objectives

The communication objectives and target groups must be thoroughly defined and identified. This includes formulation of communication objectives. Some suggested communication objectives for external communication are:

- Raise awareness about the Nordic LifeWatch Infrastructure and its services amongst the target groups and key stakeholders.
- Encourage potential users to become active users.
- Demonstrate the added value of a Nordic LifeWatch Infrastructure.
- Learn from users to improve the services provides.
- Communicate important findings and results achieved using the infrastructure.

Some suggested communication objectives for internal communication are:

- Effective communication within and between project working groups.
- Effective working groups achieving the goals set up for the working groups, within the deadline.
- Transfer of knowledge and experience between the project partners.

8.1.2. Target groups

Identifying the relevant target groups will be another important task when planning the strategic communication. Some of the most important external target groups would likely be:

- The biodiversity informatics community and other relevant related initiatives (LifeWatch Europe, national LifeWatch projects, GBIF, BioVeL, Vibrant, EU BON, etc.)
- The scientific community – researchers, postgraduate and graduate students within relevant disciplines
- Major data providers
- Policy and decision makers
- The environmental management sector

8.2. Education and training

Another key to success is education and training activities to train researchers and PhD students in using the infrastructure and demonstrate its potential. Cooperation with graduate and undergraduate education at universities, workshops and research schools are recommended actions. Sufficient financial and personnel resources should be allocated for education and training activities.

9. How should the identified scientific needs be targeted technologically with relevant services and tools?

According to Hardisty et al. (2013) the decadal vision in biodiversity informatics is to develop services that deliver data or information analyses with the use of interchange standards. Realising this vision into operational practice depends on a balance between top-down and bottom-up approaches, in addition to appropriate funding decisions.

The national and Nordic LifeWatch developments will have a special focus on how user-needs can be met with state-of-the-art interoperability and operational exchange protocols implemented at a national node level. These nodes can be coordinated from a Nordic LifeWatch Service Centre that may be established as a virtual organization relying on existing national ICT infrastructures, by creating technical interfaces for data sharing as well as best practices to guarantee data quality and consistency.

The national Swedish LifeWatch Infrastructure are the most advanced among the Nordic countries. Sweden plan to make all major national biodiversity databases interoperable, and thereby accessible through standardized web services. The Swedish LifeWatch Analysis portal provides a single access point to all data connected to the infrastructure, together with a range of analytical services. There are some data exchange protocols already in use internally in participating countries. In Sweden several protocols have been applied. Species observation data are harvested from local providers with the use of SOAP, WFS, ODBC and the Baltic Diversity Protocol. The Swedish LifeWatch Analysis Portal, as well as several other applications within the Swedish LifeWatch, mainly utilizes SOAP-services.[136] However, as the services specified by the OGC standard were suggested to be the main type in the European LifeWatch (Hernandez-Ernst *et al*. 2010), the Analysis portal has been constructed to integrate both WFS and WMS mainly for handling environmental data. As different protocols are useful for very different purposes, Swedish LifeWatch has adopted an approach where Species Observation data is provided by several protocols. The main Swedish LifeWatch node for Species Observations are two SOAP-services (Swedish Species Observation Service and Analysis Service). This node can handle all the records that are provided including the non-public ones. The open data that is free for public use is also provided with the OGC protocols WMS and WFS in addition to the export via GBIF protocols.

The Nordic LifeWatch Consortium will pay special attention to the rapidly arising amount of services and tools provided by the European Data Infrastructure (EUDAT). The European LifeWatch and LTER-Europe represents the biodiversity and ecosystem community in EUDAT, and cooperates to design infrastructure services that can better meet the community requirements. The involvement of LifeWatch and LTER-Europe is providing a use case with large and complex data sets from various origins. The biodiversity and ecosystem community is looking for a data infrastructure to better manage its data (which is generated using different systems in various locations that span the globe).

[136] http://www.svenskalifewatch.se/en/data-management/

It is expected that EUDAT will explore a range of data infrastructure services, such as automated assignment of identifiers and metadata, authentication when users (including machines, such as sensors) are operating from different interfaces (even in extreme field situations), support for data discovery, and mechanisms for fast data access, retrieval and transfer. It is apparent that the EUDAT advances will provide a tremendous source of vertical synergies from the European levels to the Nordic and national levels in this context.

At a Nordic level the Nordic eInfrastructure collaboration (NEIC) will facilitate development of e-Infrastructure services in areas of joint Nordic interest. New, innovative services can be developed and deployed by bringing together the competencies of all the Nordic e-Infrastructure providers, and challenging this pool of highly skilled IT experts with the needs of the Nordic user communities. This is precisely what the NEIC aims to do in key areas of common Nordic interest. By combining the expertise already found in the Nordic region, NEIC can provide leverage for more efficient and responsive e-Infrastructure, thereby avoiding costly duplication of work in each individual country. An important focus of the NEIC in the coming years will be on innovation in new e-Infrastructure solutions, in addition to selected services such as the Nordic Tier-1 service for the global collaboration that links grid infrastructures and computer centres worldwide.

Most partners in the Nordic LifeWatch Consortium are active in networks like GBIF, LTER, EUBON, BIOVEL and VIBRANT. This is of great importance when it comes to coordination of efforts in developing services and tools. In addition it will be of great importance to actively contribute to, and align with, the results of the Global Biodiversity Informatics Outlook (GBIO). Hardisty et al. (2013) recommends necessary actions for the next decade to link infrastructures, and to facilitate supportive understanding from business and policy-makers. These recommendations will be highly valuable in the development of a Nordic LifeWatch infrastructure and should be thoroughly followed up by the Nordic LifeWatch consortium.

10. Strategic steps towards realizing a Nordic LifeWatch construction and operational phase

10.1 Status of LifeWatch in Nordic countries

Country	Member of Nordic LifeWatch Consortium	State of the art of national LifeWatch	Preparatory funding	Construction funding	Operational funding
Norway	Yes	National Consortium established	500 K NOK.	New call 2014 Norwegian Research Council	
Sweden	Yes	Has a national consortium, which is		45 000 KSEK (36 000 from SRC + 9	Application 2014

		in a constructional phase (2010-2014), aims for entering the operational phase from 2015		000 from participating partners)	
Iceland	Yes	Not yet started, will be housed by the Icelandic Museum of Natural History	No		
Denmark	Yes	Not yet started	No		
Finland	Yes	National consortium established	100 k€ (2013-14)	Proposal made to national research infrastructure call	-

10.1.1. Finland

An integrated LifeWatch/LTER proposal was accepted to the national research infrastructure roadmap in 2009. The Academy of Finland has funded a national LifeWatch coordination project in 2013-2014 (100 000 €) to set up a "national node". The Finnish Environment Institute (SYKE) is in charge of the project. The Academy also started a process to update the RI roadmap in 2013, and a consortium submitted an updated LifeWatch proposal to this process. However, this proposal was not accepted to the 2nd evaluation stage.

The Academy also arranged a call for research infrastructure projects in spring 2013, and two projects to support national LifeWatch construction were submitted: i) Finnish Biodiversity Information Facility; ii) Finnish LTER network. Pilot study funding was granted autumn 2013.

10.1.2. Norway

The national LifeWatch pre-project, financed by the NRC, was finished in January 2013. The report enlightens Norwegian challenges in an international context. The pilot project has mapped user needs and metadata, and considered the most relevant principles, laws and guidelines for sharing of public funded environmental and biodiversity data. Further on, main challenges, barriers, terms and conditions for sharing of public funded research data are considered by the Norwegian LifeWatch Consortium (coordinated by NINA in close cooperation with the Norwegian Institute on Marine Research (IMR), The Norwegian Biodiversity Information Centre (NBIC), the Norwegian Institute for Freshwater Research (NIVA) and the Natural History Museum at the University of Oslo (which holds the Norwegian node in the Global Biodiversity Information Facility).

Nordic interaction and strategy development will be very important in future ESFRI-projects. The aim of this pilot project has been to investigate the possibilities and challenges related to the realization of a Norwegian eInfrastructure for free sharing of biodiversity and environmental data across institutional and national borders, The Norwegian LifeWatch pilot project is tightly integrated with

LifeWatch Europe through the NINA participation in the European LifeWatch Preparatory Phase (2008-2010), and through the NINA observatory seat in the LifeWatch Board meetings. A strong Nordic cooperation within LifeWatch Europe could have a long-term strategic importance for Nordic research institutions. Also, the Nordic LifeWatch level could represent a common interface between the different national levels and the European level.

The national LifeWatch consortium is planned to be expanded, and will explore the possibilities for future funding through the “infrastructure program” in Norwegian Research Council. The next call will be in 2014, and preparations for the call will continue, also in interaction with the Nordic LifeWatch consortium.

10.1.3. Sweden

The Swedish Research Council funded the preparatory and developmental phase of a Swedish LifeWatch (SLW) for the period 2010-2014. In 2011, a SLW consortium was established through the signing of a mutual agreement between the six parts:

The Swedish Species Information Centre (ArtDatabanken) at the Swedish University of Agricultural Sciences (being responsible for management and coordination of SLW), Department of Biology at Lund University (LU), Department of Zoology at Gothenburg University (GU), Department of Ecology and Environmental Sciences at Umeå University (UmU), The Swedish Museum of Natural History (NRM) including the Swedish GBIF, and the Division of Environment and Security at the Swedish Meteorological and Hydrological Institute (SMHI).

During the period 2011-2013 much of the core infrastructure has been developed, including an authorization- and authentication service, a taxonomic backbone (Dyntaxa), web services providing species observations from different data sources and web map services providing both species observations and environmental data. Currently a focal point, the so called Analysis portal https://www.analysisportal.se, is being developed. Here data sources can be discovered through catalogue services, selected, filtered (at taxonomic, spatial, temporal, ecological, legislative levels), aggregated, analysed, visualized and downloaded for further analyses.

During March 2014 funding for operation and further development of SLW in 2015-2019 will be applied for.

10.1.4. Denmark

In Denmark, the government has since 2007 allocated in excess of DKK 800 million for investments in large-scale research infrastructures of national importance. In March 2010, six academic panels were appointed to help building up a solid basis of knowledge concerning the needs of research for national and international research infrastructures. The panels were tasked with mapping and prioritizing needs over the coming years for research infrastructures within their respective fields, and this resulted in a Danish Roadmap for Research Infrastructures, in which a total of 19 promising proposals for concrete projects or initiatives have been identified. One of the panels, 'Energy, climate and environmental sciences', suggested that DKK 30 million were allocated to a research infrastructure for biodiversity and environmental data. This infrastructure is envisaged as the Danish contribution to LifeWatch and as a framework for the Danish node of GBIF; which is served by a Danish secretariat. Danish membership of LifeWatch-ERIC for the short term is dependent on developments in the international infrastructure project, national interest and availability of other funding which could be supplied in part by the Danish research councils and partly from private foundations.

The final decision for a grant award is expected in 2014 or 2015.

10.1.5. Iceland

LifeWatch project has not started yet in Iceland but the preparatory phase has been organized by the Icelandic Institute of Natural History which serves as a GBIF-node for Iceland and keeps big parts of the biodiversity data of Iceland. The recently established Icelandic Museum of Natural History has accepted the challenge of leading the LifeWatch project in Iceland with support from the IINH.

Funding of an Icelandic LifeWatch project is still uncertain but the project fits the governmental policy of making biodiversity data freely and openly accessible and might therefore gain governmental funding earmarked to obtain certain level of data accessibility.

10.2 Interface with related initiatives at National, Nordic and EU levels.

Table 2: An overview of related LifeWatch initiatives at national, Nordic and EU levels

	Explanation of the project	Norway	Sweden	Iceland	Denmark	Finland
LifeWatch EU	http://www.lifewatch.eu/web/guest/home	X	X			X
LifeWatch Belgium	National LifeWatch (http://www.lifewatch.be/		X			
BioVeL (MO)			X			X
EMBRC	http://www.embrc.eu/		X			
EU BON (AT)	European part of GEO BON	X	X		X	X
GEO (MF)	Effort to build a Global Earth	X	X	X	X	X

	Observation System					
GBIF (AT)	GBIF is an international initiative for open access to biodiversity data.	X	X	X	X	X
PSEI	http://www.eu---nomen.eu/pesi/		X			
EUDAT	http://www.eudat.eu/		X		X	X
DINA (AT)	Internationally acknowledged and implemented collection management system		X		X	
NMDC	A Norwegian 5yr project to increase access to marine data	X				
CBMP (CAFF) SH)	Harmonizing and enhancing Arctic monitoring efforts, thereby improving the ability to detect and understand significant trends, see: http://www.caff.is/about-the-cbmp	X	X	X		
ViBRANT	http://vbrant.eu/		X			

LTER-Europe	http://www.lter-europe.net/	(X)	X			X

EU LifeWatch is a European research infrastructure in development. The first services to users are planned for 2013. Users may benefit from integrated access to a variety of data, analytical and modeling tools as served by a variety of collaborating initiatives. Another service is offered with data and tools in selected workflows for specific scientific communities. In addition, LifeWatch will provide opportunities to construct personalized 'virtual labs', also allowing entering new data and analytical tools. New data will be shared with the data facilities cooperating with LifeWatch.

LifeWatch Belgium is currently developing the LifeWatch taxonomic backbone, habitats database with a facility to identify disturbance and restoration of habitats. Summary reporting on ecosystem dynamics. Service mechanism for delivering land cover and land use mapping through remote sensing.

BioVeL - Biodiversity Virtual e-Laboratory is a virtual eLaboratory that supports research on biodiversity issues using large amounts of data from cross-disciplinary sources. BioVeL offers the possibility to use computerized "workflows" (series of data analysis steps) to process data, be that from one's own research and/or from existing sources.

EMBRC - European Marine Biological Resource Centre brings together 12 leading marine stations and EMBL. These institutes study marine organisms (microbes, plants, animals) with the latest technologies to study our seas. Cooperation with LifeWatch considered with respect to: (a) Collaboratively building a European Marine Biodiversity Observatory (also EU BON), (b) extending the biodiversity eLaboratories into the biomedical domain, and (c) joint stakeholder and knowledge transfer program.

EU BON (European Biodiversity Observation Network) is the European part of GEO BON that coordinates activities relating to the Societal Benefit Area (SBA) on Biodiversity of the Global Earth Observation System of Systems (GEOSS). The Nordic contribution provides a biodiversity data gap analysis, is responsible for developing tools for integration, interoperability, data analysis and presentation, and serves the implementation of GEO BON strategies and solutions at European and global levels.

GEO (Group on Earth Observations) is coordinating efforts to build a Global Earth Observation System of Systems, or GEOSS. GEO was launched in response to calls for action by the 2002 World Summit on Sustainable Development and by the G8 (Group of Eight) leading industrialized countries. These high-level meetings recognized that international collaboration is essential for exploiting the growing potential of Earth observations to support decision making in an increasingly complex and environmentally stressed world. GEO is a voluntary partnership of governments and international organizations. GEO is constructing GEOSS on the basis of a 10-Year Implementation Plan[137] for the period 2005 to 2015. The Plan defines a vision statement for GEOSS, its purpose and scope, expected benefits, and the nine "Societal Benefit Areas" of disasters[138], health[139], energy[140], climate[141], water[142], weather[143], ecosystems[144], agriculture[145] and biodiversity[146].

---

[137] http://www.earthobservations.org/documents/10-Year%20Implementation%20Plan.pdf
[138] http://www.earthobservations.org/geoss_di.shtml

GBIF (Global Biodiversity Information Facility) promotes and facilitates the mobilization, access, discovery and use of information about the occurrence of organisms over time and across the planet. National GBIF nodes exist in all the Nordic countries offering bi-directional data flow to and from LifeWatch.

PESI (Pan European Species directories Infrastructure) provides standardized taxonomic information by integrating Europe's authoritative species name registers, nomenclatures and the associated expert network.

EUDAT (European Data Infrastructure) work towards a collaborative data infrastructure for Europe with a global vision.

DINA (Digital Information System for Natural History Collections) is a natural science collection management system supported by a number of stakeholders (mainly natural history museums in Sweden, Estonia, Denmark, Germany, France and Canada) that enables efficient collection data flow. The project management is run by the Swedish Museum of Natural History.

NMDC (Norwegian Marine Data Centre) has as its ultimate goal to provide seamless access to marine data and greatly increase the efficiency of marine science in Norway and facilitate the generation of high quality research. This is a 5-year project started in 2012 and funded by the NRC and participating institutions. The overall budget is about 80 million NOK.

CBMP (**Circumpolar Biodiversity Monitoring Program ()** is an international network of scientists, government agencies, indigenous organizations and conservation groups working together to harmonize and integrate efforts to monitor the Arctic's living resources. It is initiated by CAFF - Conservation of Arctic Flora and Fauna - and focuses on three main ecosystems, marine, freshwater and terrestrial.

ViBRANT (Virtual Biodiversity Research and Access Network for Taxonomy).

LTER-EUROPE (European Long-Term Ecosystem Research Network) is a network of national networks and the European contribution to the global International Long Term Ecological Research (ILTER). The national networks consists of LTER sites and LTSER platforms. There are currently 21 national networks in Europe, but more national networks are emerging.

---

[139] http://www.earthobservations.org/geoss_he.shtml
[140] http://www.earthobservations.org/geoss_en.shtml
[141] http://www.earthobservations.org/geoss_cl.shtml
[142] http://www.earthobservations.org/geoss_wa.shtml
[143] http://www.earthobservations.org/geoss_we.shtml
[144] http://www.earthobservations.org/geoss_ec.shtml
[145] http://www.earthobservations.org/geoss_ag.shtml
[146] http://www.earthobservations.org/geoss_bi.shtml

10.3 Stimulating further LifeWatch initiatives in Nordic countries

10.3.1. National level

LifeWatch initiatives are at different levels in the Nordic countries, cf. Table 2. For future development of National LifeWatch initiatives towards operational levels, the national research councils and, possibly, the national environmental authorities will be the most relevant funding agencies. Their policies are different in the different countries. However, there seems to be an increased understanding that research infrastructures (including eInfrastructures) will be an important part of research funding also in Nordic countries. Several Nordic governments have an official policy of supporting and encouraging IT developments. Through Nordic collaboration, national initiatives will be more effective by learning from each other and by exchanging experiences and eSystems. The potential of developing compatible systems for exchange of data is promising, but the lack of a common Nordic source of funding will restrict the possibilities. We therefore, recommend Nordic authorities to explore possible arenas for common funding of future Nordic LifeWatch activities.

10.3.2. Nordic level

The starting point is that LifeWatch activities presently are on a very unequal level in the Nordic countries. This is mainly due to funding decisions taken by the national research councils (Table 2, above). Sweden, where substantial funding has been available, is clearly the most advanced Nordic country regarding the technical infrastructure development. The other Nordic countries should take advantage of the Swedish experience and implement existing structures (where feasible) in a coordinated exercise. New infrastructures and tools can also be developed by coordinated planning and funding.

The proposed common visions for the Nordic LifeWatch development by 2020 are:

- Each Nordic country should be at an operational level regarding the technical LifeWatch structure.
- The technical capacity to harvest data from the national data sources for joint Nordic activities is operational.
- Some common data analysis environments (virtual laboratories) have been developed.
- Permanent working groups have been established for the different activities (technical development, scientific goals, communication etc.).
- The potential of the Nordic LifeWatch structure has been proved by producing some selected products of scientific and policy relevance.
- The Nordic LifeWatch infrastructure is a recognized regional part of the European LifeWatch structure and has contributed significantly to the European project.
- Stakeholder involvement is obviously a cornerstone in a successful implementation of LifeWatch. The ongoing debate and interaction between stakeholders and LifeWatch project groups is important to keep the focus on the most important research questions at any time, and to make LifeWatch relevant in the support of decision makers. The potential group of

relevant stakeholders is very diverse, making effective and dedicated communication a key focus point for the LifeWatch project group.

- As a part of the design and construction phase, a working group of communication experts will develop a dedicated communication plan to ensure efficient transfer of information between the LifeWatch project team and relevant stakeholders. This involves thorough analyses of stakeholder interests.Based on experience from other projects (e.g. GBIF) we expect workshops and training seminars to be one of the important ways to stimulate interaction and to facilitate the transfer of export knowledge about LifeWatch tools and opportunities to groups of stakeholders.

As a follow-up of the present report, we suggest that NordForsk consider hosting a Nordic LifeWatch conference and workshop, where the national Research Councils and national delegates from the participating countries in the Nordic LifeWatch project present details from the report and discuss further development of Nordic LifeWatch.

LifeWatch aims at providing the basis of science-based management of biodiversity and living natural resources and to catalyse sharing of knowledge. Integration of monitoring data from ground surveys, sensors and remote sensing with ecological, geographical and taxonomical expertise will provide high quality biodiversity data of animals, plants, microorganisms, fungi as well as habitat types and ecosystems. This will help in habitat prioritization, vulnerability analyses, risk assessment and public involvement in practical nature conservation. Furthermore, it will be possible to apply grid-based technology to perform advanced data-mining and pattern recognition. In the end it is the aim to create a knowledge-based platform for decision making which covers the total biological hierarchy from genes to ecosystems. Such a system will make it possible to distinguish natural development tendencies and dynamics from the changes caused by human impact of nature, and to observe the warning signals before large irreversible changes happen.

In detail, Nordic LifeWatch will make use of existing tools that help reducing the taxonomic impediment: To fill in the gaps in taxonomic knowledge and to more efficiently than hitherto use taxonomic expertise in focused investigations. The naming of species, species names organization and species identification is facilitated, and the implementation of biodiversity studies based upon mere sequence data is enabled. Coordinated Nordic LifeWatch activities will rest upon the development, implementation and use of persistent and universally unique identifiers, Common vocabularies and ontologies are crucial for centralized and networked services to facilitate both citation mechanisms, data validation and data re-use. Such actions will strengthen the engagement of both professional and non-professional actors. By making use of new techniques for digitizing collections, automated data collection and data mining, and by creating virtual research environments Nordic LifeWatch will contribute to more efficient, targeted and reliable investigations and predictions at wider geographical scale than ever before, and at shorter notice. Feedback mechanisms to prompt new data generation will result in integrated and more fine-tuned experiments and better decision support when model outcomes result in various scenarios.

10.3.3. Technical working plan based on existing national and international efforts

Based on the goals to be achieved at the Nordic level by 2020, and on the mapping of what resources already exist in the respective countries, a gap analysis should be performed. This will result in a list of services that must be developed or accustomed to meet needs at national levels, such as web services for species observations, web map services for other data, feature access services, catalogue services for metadata, authorization service, authentication service, and taxonomic backbone. These all require that common ontologies and other standards have been agreed on.

The plan will also include a charting of what has already been developed in other parts of the world, in particular so when it comes to applications for analysing and visualizing data and results, including workflows and other means for creating virtual laboratories.

All components identified to be essential to the Nordic LifeWatch network will be described technically, the cost for development or adaptation estimated, and a time-plan for their implementation conducted.

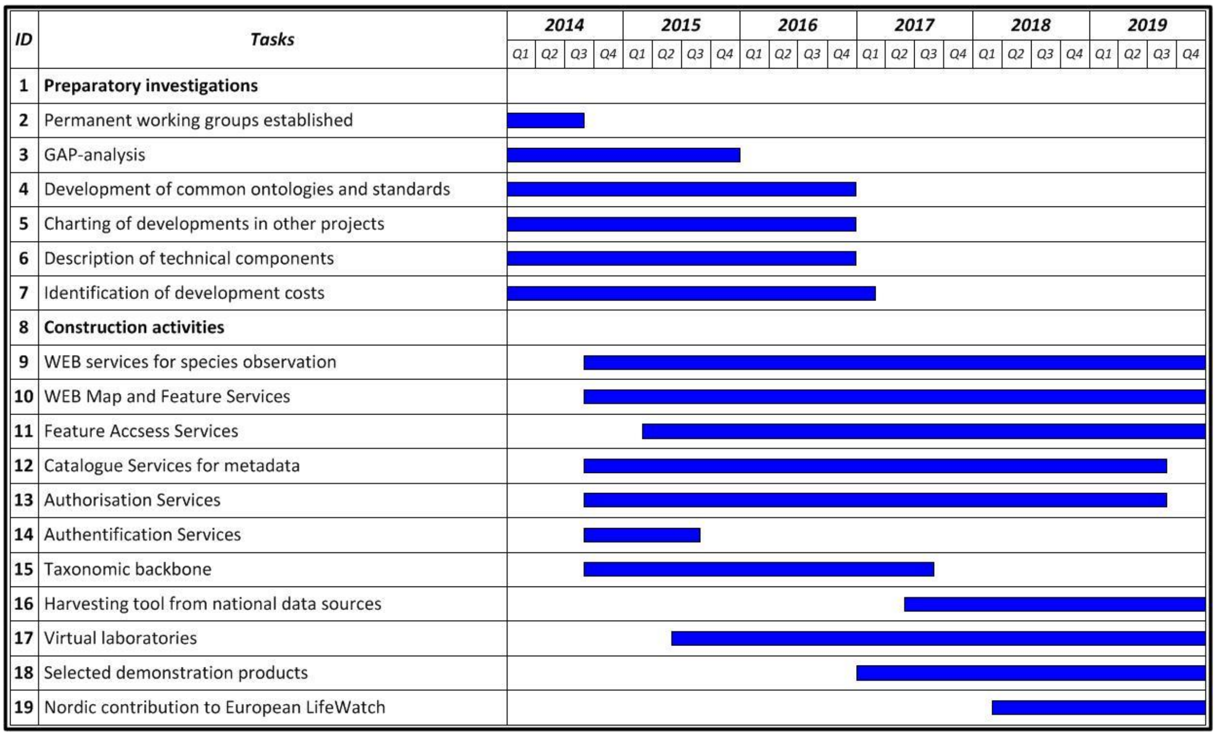

Figure 3: Activity plan for establishing a permanent Nordic LifeWatch by 2020.

It is obvious that the construction of a Nordic LifeWatch infrastructure would need a coordinated funding strategy by national research councils (and other national funding agencies), and Nordic funding instruments. The funding of the national structures would be the responsibility of national funders, whereas the construction of common data analysis environments (virtual laboratories) and Nordic demonstration products would be the main responsibility on the Nordic level.

10.3.4. Strategy development

Resources spent on biodiversity infrastructure projects must generate new insights and possibilities that are clearly communicated and indisputable. Admittedly, all infrastructure projects with different financial sources but partly overlapping aims are at risk of repeating each other at some stage. Therefore, it must be clearly stated what is to be achieved, by what means, technical standards, time schedule and philosophy that the goals will be fulfilled. Collaboration within the community is crucial, and re-use of established techniques and material is recommended.

The LifeWatch architecture represents 3 main levels; national, Nordic and international (European) level. The present Nordic LifeWatch project has shown the potential and interest to develop LifeWatch networks, both on national and Nordic levels. In some countries, national networks are well developed, while in other Nordic countries networks have to be further developed to include the most relevant institutions holding important biodiversity databases. At the international level, Nordic countries have a number of biologically similar challenges and opportunities. Baltic nations have similar opportunities, and are relevant for future LifeWatch collaboration. Therefore, one should consider exploring the possibilities to include additional Baltic countries for future extended Nordic LifeWatch collaboration.

A good coordination and collaboration with the EU LifeWatch will be essential in the future. The communication with the EU level will be strengthened if a Nordic LifeWatch portal is established.

The Nordic LifeWatch project has made significant progress in documenting the present situation and outlining the main areas for further work. There are significant benefits regarding both technical developments and savings regarding financial investments in continuing close collaboration and establishing a common Nordic LifeWatch infrastructure. A core funding is necessary to continue development of the Nordic network after the present project is ended. The Nordic and national funding agencies should therefore discuss funding of the national LifeWatch infrastructures and the common Nordic parts, and the potential for sharing costs between the different activities. The participating delegates in the Nordic LifeWatch network are prepared to continue their work towards an operating stage of the Nordic LifeWatch, given that appropriate funding exists. This process will involve strategic issues, technical development, communication and organization. Also, the developments of the European LifeWatch and the potential for common Nordic action in this respect should be paid attention in the near future. The Nordic LifeWatch network should also follow the activities of key international activities within relevant areas (for instance, IPBES, GEOSS etc.), thus identifying the potential for joint Nordic action towards these.

To outline the potential and funding for future Nordic LifeWatch in detail, our main conclusion is to arrange a Nordic LifeWatch conference as soon as possible. This conference should involve Nordic research councils, scientists and relevant stakeholders. The national delegates from the participating countries in the present Nordic LifeWatch project are prepared to present details from the report and developments so far, as a basis for further development of Nordic LifeWatch.

11. References

Ariño, A. (2010). Approaches to estimating the universe of natural history collections data. Biodiversity Informatics 7(2): 81–92.Baskauf, S., Ó Tuama, É., Endresen, D., and Hagedorn, G. (eds.) (2013). Report of the TDWG Vocabulary Management Task Group (VoMaG). 25 pp. Available online at http://www.gbif.org/orc/?doc_id=5422, verified 21 January 2014.

Beniston, M., Stoffel, M., Harding, R., Kernan, M., Ludwig, R., Moors, E., Samuels, P., and Tockner, K., (2012). Obstacles to data access for research related to climate and water: implica- tions for science and EU policy-making. Environmental Science & Policy 17: 41–48.

DIFI (2013). Del og skap verdier - Veileder i tilgjengeliggjøring av offentlige data. Direktoratet for forvaltning og IKT (DIFI). Available from http://data.norge.no/document/del-og-skap-verdier-veileder-i-tilgjengeliggj%C3%B8ring-av-offentlige-data, verified 21 January 2014.

Duckworth, W.D., Genoways, H.H., Ros, C.L. (1993). Preserving natural science collections: chronicle of our environmental heritage. Washington, D.C. iii+140 pp.

GBIF (2011). GBIF Metadata Profile, Reference Guide, Feb 2011, (contributed by Ó Tuama, É., Braak, K., Copenhagen: Global Biodiversity Information Facility, 19 pp. Accessible at http://www.gbif.org/resources/2559, verified 21 January 2014.

Gisler, G., Celledoni, E., Helgaker, T.U., Iversen, T., Jakobsen, K.S., Jones, C., Lipniacka, A., Lundervold, A., Olsen, N.R.B., De Smedt, K., Koster, J., and Høst, G. (2010). The scientific case for eInfrastructure in Norway. The Norwegian Research Council, Oslo. ISBN 978-82-12-02832-6.

Hardisty, A. and Roberts, D. (2013). A decadal view of biodiversity informatics: challenges and priorities. BMC-Ecology, DOI 10.1186/1472-6785-13-16

Hernandez-Ernst, V., Poigné, P., Giddy, J., and Hardisty, A. (2010). LifeWatch Technical Construction Plan – Reference Model. Fraunhofer IAIS. Available at http://www.eubon.eu/getatt.php?filename=LW-RMV0.5_4310.pdf, verified 21 January 2010.

OKFN (2012). Open data handbook. Open Knowledge Foundation (OKFN). Available at http://opendatahandbook.org/, verified 21 January 2014.

Sandberg, M. (2012). Reinforced Nordic collaboration on data resources: Challenges from six perspectives. Nordic Council of Ministers, Copenhagen. TemaNord 2012:514, ISBN 978-92-893-2342-0. Available at http://www.norden.org/no/publikationer/publikasjoner/2012-514, verified 21 January 2014

Schmidt-Kloiber, A., Moe, S.J., Dudley, B.J., Strackbein, J., and Vogl, R. (2013). The WISER metadatabase: the key to more than 100 ecological datasets from European rivers, lakes and coastal waters. Hydrobiologia. doi:10.1007/s10750- 012-1295-6.

Smith, V., Georgiev, T., Stoev, P., Biserkov, J., Miller, J., Livermore, L., Baker, E., Mietchen, D., Couvreur, T.L.P., Mueller, G., Dikow, T., Helgen, K.M., Frank, J., Agosti, D., Roberts, D., and Penev, L. (2013). Beyond dead trees: integrating the scientific process in the Biodiversity Data Journal. Biodiversity Data Journal 1: e995. DOI: 10.3897/BDJ.1.e995

Townsend Peterson, A., Knapp, S., Guralnick, R., and Soberon, J. (2010). Perspective: The big questions for biodiversity informatics. Systematics and biodiversity 8(2):159-168. DOI:10.1080/14772001003739369

Vickery, G. (2011). Review of recent studies on PSI re-use and related market development. Paris, 41 pages. Available at http://www.umic.pt/images/stories/publicacoes6/psi_final_version_formatted-1.pdf, verified 21 January 2014.