KLAIPĖDA UNIVERSITY
COASTAL RESEARCH AND PLANNING INSTITUTE

ALEKSAS NARŠČIUS

DEVELOPMENT AND APPLICATION OF THE INFORMATION SYSTEMS ON AQUATIC NON-INDIGENOUS SPECIES

Doctoral dissertation

Biomedical sciences, ecology and environmental sciences (03B), hydrobiology, marine biology, aquatic ecology, limnology (B260)

Klaipėda, 2013

Dissertation prepared 2008–2012 at Klaipėda University, Coastal Research and Planning Institute.

Supervisor:

prof. habil. dr. Sergej Olenin (Klaipėda University; Biomedical Sciences, Ecology and Environmental Sciences – 03B).

TABLE OF CONTENTS

1. INTRODUCTION

Relevance of the problem

A broad spectrum of human activities may lead to introductions of non-indigenous species (NIS) into marine, brackish and freshwater areas. Some species are introduced deliberately for aquaculture, stocking or habitat management, while the great majority of other organisms are unintentionally transported, for example in ship ballast water or on the hulls of vessels (Carlton and Geller, 1993; Gollasch et al., 2002). Climate alterations and global trade are likely to result in further range extensions of taxonomically diverse invaders (Occhipinti-Ambrogi and Galil, 2010; Olenin and Minchin, 2011). Since the first generalization of the biological invasion phenomena in the fundamental study by Elton (1958), NIS expansions and their impacts have become an area of increasing interest for scientists, managers, policy makers and members of the public. Biological invasions (bioinvasions) are recognized as a worldwide threat to native biodiversity, ecosystem functioning, economies and human health (IUCN, 2000). It was indicated that the impacts of invasive alien species on ecosystems and society are increasing among all taxonomic groups, and in all environments causing at least € 12.5 billion damage each year in the European Union (Shine et al., 2008; NEOBIOTA, 2012). This, in turn, has led to global concern and management advice by international organizations (IUCN, 2000; IMO, 2004; Shine et al., 2008).

Scientific and managerial attention to bioinvasion problem results in growing number of electronic resources on NIS. Currently there are more than 250 websites on NIS worldwide (GISIN, 2008). The geographical scope of these information resources varies from global (e.g. GISD, 2012), pan-European (EASIN, 2013) to regional (e.g. Baltic Sea Alien Species Database, 2013; CIESM Atlas of Exotic Species in the Mediterranean, 2013) and national (e.g. Mastitsky et al., 2013). NIS databases are increasingly being used for many aspects of bioinvasion research, e.g. to aid the compilation of NIS lists for specific areas (e.g. Zaiko et al., 2007; Occhipinti-Ambrogi et al., 2011), to rank the most impacting NIS (e.g. Olenina et al., 2010; Savini et al., 2010; Zaiko et al., 2011), to define the major pathways and vectors responsible for NIS introductions (e.g. Savini et al., 2008; Minchin et al., 2009), to analyze species traits and ecological preferences (e.g. Paavola et al., 2005; Strayer, 2010), to assess the risks posed by alien species on economies and ecosystem functioning (e.g. Baker et al., 2005; Campbell et al., 2008).

In general, it is widely acknowledged that scientifically validated, updated and continuously maintained databases are the most reliable source for integrated information on NIS, their population dynamics, ecology and means of control (Genovesi, 2001; Olenin et al., 2011). In order to be effective, information must be placed within the proper context and organized in a manner that is both logical and standardized (Simpson et al., 2006).

In this study an information system consisting of the database on Aquatic Non-Indigenous Species (AquaNIS) and equipped with a special tool for Biological invasion impact assessment (BINPAS) is developed. It is designed to assemble, store and disseminate comprehensive data on NIS as well as to provide meaningful information for solving research problems and evaluating progress towards bioinvasion management.

Aim and objectives of the study

The aim of this study was to develop information systems on aquatic non-indigenous species, which provide additional exploratory value to multiple accumulated data, and apply them for analyzing patterns of biological invasions. The following tasks were raised for this work:

1. To identify and formalize essential features of biological invasion process in aquatic environment for construct data-driven structure of NIS information systems.

2. To develop a NIS information system to store and disseminate knowledge on NIS taxonomy and biological traits, recipient and source regions, pathways and vectors, and other relevant documented data.

3. To develop a computerized application to assess the level of bioinvasion impacts by accumulating data on abundance and distribution range of NIS and their impacts on communities, habitats and ecosystem functioning according to biopollution (BPL) index method.

4. To analyze patterns of NIS distributions in European regional seas, rate of new introductions and prioritize invasion pathways using the developed NIS information system.

5. To generate classification rules allowing optimization of BPL assessments using machine learning algorithms.

Novelty of the study

In this study a new information system on aquatic NIS was developed to integrate multiple data on NIS taxonomy, biological traits, environmental characteristics, bioinvasion impacts, introduction events, pathways and vectors, and other relevant geographical and ecological data. The system is integrated with a specialized computerized application, which facilitates usage of biopollution assessment method, accumulates knowledge on bioinvasion impacts on native communities, habitats and ecosystem functioning, and makes possible cross-taxon and interregional comparison of bioinvasion effects. The developed system differs substantially from the existing NIS information sources in its structure, functionality, maintenance principles and output potential for end-users. It provides an added heuristic value to multiple data, allowing its comprehensive analyses in different aspects, such as identification of most invasive species, important introduction pathways, predominant taxonomic groups of NIS, etc.

Scientific and practical significance of the results

The application of AquaNIS and BINPAS showed that these information products can be used for regional and global level analysis and overview of biological invasion problems. The systems allow identifying most common introduction pathways and vectors; accumulating and storing information on abundance and distribution range of various NIS as well as their impacts on communities, habitats and ecosystem functioning; comparisons between different species, ecosystems and time periods. The data search and extraction tools allow getting data sources in a format suitable for further analysis with specialized statistical packages, and provide meaningful information for solving research problems and evaluating progress towards bioinvasion management goals. The completeness of datasets for solving various research questions can be analyzed thus helping to determine the level of knowledge accumulated in the information systems. The structure of AquaNIS and BINPAS is flexible enough to accommodate specialized modules to assess NIS impacts on Water framework directive ecological quality parameters, Marine Strategy Framework Directive qualitative descriptors and socio-economic impacts.

Defensive statements

1. The rate of aquatic NIS introductions is increasing, both at pan-European and regional scales.

2. Aquatic NIS are represented by a broad spectrum of free-living and parasitic multicellular and unicellular organisms. In species poor systems like the Baltic Sea the relative number of cryptogenic species is lower than in fully saline species rich marine regions.

3. Specific environmental conditions and pathways are important factors shaping NIS compositions in the regional European seas.

4. The homogeneity of the Baltic Sea region in terms of NIS compositions is increasing during recent decades.

5. The developed information systems can serve as multipurpose tool useful for research and practical for management, providing a flexible platform for bioinvasion data storage, extraction and analysis.

6. Machine learning algorithms can be used for optimizing the application of the Biopollution index method.

Scientific approval

The results of this study were presented at 10 international and 6 regional conferences and seminars:

- Scientific conference of young researchers: research innovations fundamentals. Klaipėda, Lithuania, April 2009.
- Regional conference "Marine and coastal research". Nida, Lithuania, April 2009.
- International Seminar on introduced aquatic species: "Introduced marine species: what should we study now and why?" Bergen, Norway, October 2009.
- 6th international conference on biological invasions "NEOBIOTA: Biological Invasions in a Changing World – from Science to Management". Copenhagen, Denmark, September 2010.
- Regional conference "Marine and coastal research". Palanga, Lithuania, April 2010.
- 5th International Student Conference: "Biodiversity and functioning of aquatic ecosystems in the Baltic Sea region". Palanga, Lithuania, October 2010.
- Regional conference "Marine and coastal research". Palanga, Lithuania, April 2011.
- Project VECTORS workshop. Rome, Italy, May 2011.

- 2nd World Conference on Biological Invasions and Ecosystem Functioning “BIOLIEF”. Mar del Plata, Argentina, November 2011.
- Scientific seminar of young researchers. Klaipėda, Lithuania, February 2012.
- 7th international conference on biological invasions “NEOBIOTA: Halting Biological Invasions in Europe – from Data to Decisions”. Pontevedra, Spain, September 2012.
- 51th International Symposium “ECSA: Research and management of transitional waters”. Klaipėda, Lithuania, September 2012.
- 6th international student conference “Aquatic environmental research”. Palanga, Lithuania, October 2012.
- Project VECTORS Regional Baltic workshop. Copenhagen, Denmark, November 2012.
- Project VECTORS Annual Meeting – Conference. Tarragona, Spain, March 2013.
- 7th national conference “Marine and coastal research”. Klaipėda, Lithuania, April 2013.

The material of this dissertation was presented in three original publications, published in peer-reviewed scientific journals (#1-3) and in two additional articles in a specialized bioinvasion information bulletin (#4) and applied computer science proceedings (#5).

1. Olenin S., **Narščius A.**, Minchin D., David M., Galil B., Gollasch S., Marchini A., Occhipinti-Ambrogi A., Ojaveer H., Zaiko A., 2013. Making non-indigenous species information systems practical for management and useful for research: an aquatic perspective. *Submitted.*
2. **Narščius A.**, Olenin S., Zaiko A., Minchin D., 2012. Biological invasion impact assessment system: From idea to implementation. Ecological Informatics 7, pp. 46-51.

3. Zaiko A., Lehtiniemi M., **Narščius A.**, Olenin S., 2011. Assessment of bioinvasion impacts on a regional scale: a comparative approach. Biological Invasions, 13, 8, pp. 1739-1765.

4. Olenin S., **Narščius A.**, 2010. "Beaufort Scale" for bioinvasion impacts. Aliens: The Invasive Species Bulletin. Newsletter of the IUCN/SSC Invasive Species Specialist Group. ISSN 1173-5988, 29, pp. 52-54.

5. Baziukaitė D., **Narščius A.**, 2010. The modified Renyi-ClipX conceptual clustering algorithm. Applied computer science, Institute for Environment, Engineering, Economics and Applied Mathematics. ISSN: 1792-4863, ISBN: 978-960-474-225-7, pp. 396-400.

Structure of the Thesis

The dissertation includes *seven* chapters: Introduction, Literature review, Materials and methods, Results, Discussion, Conclusions, and References. The material is presented in 134 pages, 39 figures and 19 tables. The dissertation refers to 182 literature sources. Appendix contains supplementary information on the most spread NIS. Dissertation is written in English with extended summary in Lithuanian and English language.

Abbreviations used in the study

ADR	Abundance and Distribution Range
ANOSIM	Analysis of Similarity
AP	Assessment Period
AquaNIS	Information system on Aquatic non-Indigenous species
AU	Assessment Unit
AWE	Agile Web Engineering
BINLIT	Lithuanian State Science and Studies foundation project "Biological invasions in Lithuanian ecosystems under the climate change: causes impacts and projections", agreement number: C-04/2008/2

BINPAS	Biological Invasion Impact / Biopollution Assessment System
BPL	Biopollution level
CBD	Convention on Biological Diversity
CR	Cryptogenic species
CSS	Cascading Style Sheets
CSV	Comma Separated Values
DAISIE	Delivering Alien Invasive Species Inventories for Europe
DEVOTES	The EU FW7 project DEVOTES "Development of innovative tools for understanding marine biodiversity and assessing good environmental status", project number: 308392
EASIN	European Alien Species Information Network
EC	European Commission
ECCHM	European Biodiversity Clearing House Mechanism
EEA	European Environment Agency
EU	European Union
GISD	Global Invasive Species Database
GISIN	Global Information System for Invasive Species
HELCOM	HELsinki COMmission
HTML	Hyper Text Markup Language
IABIN	Inter-American Biodiversity Information Network
IAS	Invasive Alien Species
ICES	International Council for the Exploration of the Sea
IMO	International Maritime Organization
IOC	Impacts On Communities
IOC	Intergovernmental Oceanographic Commission
IOE	Impacts On Ecosystem functioning
IOH	Impacts On Habitats
LME	Large Marine Ecosystem
MEECE	The European FP7 project "Marine Ecosystem Evolution in a Changing Environment", project number: 212085

MSFD	Marine Strategy Framework Directive
MVC	Model-View-Controller
NIS	Non-indigenous Species
NISBase	International Nonindigenous Species Database Network
NOAA	National Oceanic and Atmospheric Administration
NOBANIS	North European and Baltic Network on Invasive Alien Species
PHP	Hypertext Preprocessor
VECTORS	The EU FW7 project "Vectors of Change in Oceans and Seas Marine Life, Impact on Economic Sectors", project number: 266445
WFD	Water Framework Directive
WGBOSV	Working Group on Ballast and Other Ship Vectors
WGITMO	Working Group on Introductions and Transfers of Marine Organisms
WoRMS	World Register of Marine Species
XML	Extensible Markup Language

Acknowledgements

Firstly, I would like to thank my supervisor Sergej Olenin for special believe, encouragement, motivation, comprehension, positive attitude, patience, help at hard moments, support, useful ideas, and care during the whole PhD theses preparation time. I am especially thankful to Anastasija Zaiko, Greta Srėbalienė and Dan Minchin for comfort atmosphere at Biological invasions working group, activity, offered opportunities, support, kind words, valuable criticism, help and provided knowledge.

Also I am very grateful to Informatics Department at Klaipeda University for provided background and experience during bachelor and master studies, as well as a motivation to take a new challenge. I am fully grateful to Coastal Research and Planning Institute at Klaipeda University for provided trust, infrastructure and opportunity to work with active, friendship, kindness and supporting team!

For datasets and active cooperation I am very grateful to my colleagues: Anastasija Zaiko, Anna Occhipinti, Bella Galil, Dan Minchin, Greta Srėbalienė, Henn Ojaveer, Maiju Lehtiniemi, Sergej Olenin, Stephan Gollasch, Romualda Chuševe.

Also I would like to say a special thanks to Marija Kataržytė, Raimonda Kybrancienė, Rasa and Julius Morkūnas, Daniel Depellegrin for providing extra time, comments, support, friendship; Darius Daunys and Vitalij Denisov for fair criticism and valuable comments; Vitalija Gasiūnaitė and Ingrida Bagdanavičiūtė for nice drawings and maps; Andrius Narbutas for the server management.

I largely appreciate my close friends Jurgita Neverdauskaitė for interest, support and warmth, Kęstutis Bružas for advices on programming and fisheries issues, Martynas Astrauskas for the superior friendship, support and concern.

Finally, my deepest gratitude goes to my mama, dad, sister, brother, Asta and Danielė for their patience and support.

Financial support for this study was provided by:

- Coastal Research and Planning Institute, Klaipėda University (KU CORPI).
- Lithuanian State Science and Studies foundation project BINLIT “Biological invasions in Lithuanian ecosystems under the climate change: causes impacts and projections” (agreement number: C-04/2008/2).
- The European FP7 project MEECE “Marine Ecosystem Evolution in a Changing Environment” (project number: 212085).
- The EU FW7 project VECTORS “Vectors of Change in Oceans and Seas Marine Life, Impact on Economic Sectors” (project number: 266445).
- The EU FW7 project DEVOTES “Development of innovative tools for understanding marine biodiversity and assessing good environmental status” (project number: 308392).

2. LITERATURE REVIEW

2.1. Essential features of a bioinvasion process

Since the first generalization of biological invasion phenomena by Elton (1958), non-native species expansions and their impacts have become an area of concern for scientists, managers, policy makers and members of the public, and this in turn has led to global advice and, or, management by international organizations (Carlton, 1989; Williamson, 1996; Leppäkoski et al., 2002; Alimov and Bogutskaya, 2004; Rilov and Crooks, 2009; Executive Order, 1999; IUCN, 2000; IMO, 2004; Shine et al., 2008; Olenin et al., 2010). These efforts have resulted in strategies to prevent or reduce the risks from biological invasions (Wittenberg and Cock, 2001) with several reviews of invasions ranging from taxonomic studies to impacts, and how alien species are spreading (Olenin et al., 2010).

Biological invasions (bioinvasions) are recognized as a worldwide threat to native biodiversity, ecosystem functioning, economies and human health. Although "biological introductions" and "biological invasions" are related, they are not synonymous: an invasion may be caused by either natural or human activity, while an introduction is always human mediated, either intentionally or unintentionally (Olenin et al., 2010). In this Section definitions concerning biological invasions are considered. Also invasion process as such is overviewed in relation to pathways and vectors.

2.1.1. Non-indigenous, cryptogenic and invasive alien species

Terms and concepts crucial to understanding ecology have often been criticized for their tautological, ambiguous or nonoperational nature (Peters, 1991; Colautti and MacIsaac, 2004). Invasion ecology has enjoyed a rapid ascension in the public domain, owing in part to the extensive use of adjectives like "invasive", "alien", "noxious" and "exotic" (Chew and Laubichler, 2003). A number of definitions exist both in scientific literature (e.g. Leppäkoski et al., 2002; Occhipinti-

Ambrogi and Galil, 2004; Carlton, 2009; Pyšek et al., 2009; Olenin et al., 2010) and within legislative/administrative documents (e.g. Office of Technology Assessment, 1993; Executive Order, 1999; EC, 2008). As a result, currently there are several widely used definitions of alien and invasive species (Table 1).

Table 1. Definitions of alien and invasive species. Applied from Heger et al. (2013).

Term / Reference	Definition
Alien species (Davis and Thompson, 2000)	A species introduced outside its natural past or present distribution.
Invasive alien species (CBD, 2002)	An alien species whose introduction and/or spread threatens biological diversity.
Alien species (Wilson et al., 2009)	A species that has shown extra-range dispersal owing directly or indirectly to human activity.
Invasive species (Wilson et al., 2009)	An introduced species that has sustained self-reproducing populations and can produce reproductive offspring at considerable distances from parent plants.
Invasive alien species (IUCN, 2011)	Invasive alien species are animals, plants or other organisms introduced by man into places out of their natural range of distribution, where they become established and disperse, generating a negative impact on the local ecosystem and species.
Alien species (Heger et al., 2013)	Any species that occurs at a location beyond its area of origin; the occurrence of the species in the new area must have been prevented in the past by a dispersal barrier, not by unsuitable conditions. It does not matter whether the passing of the major geographical barrier was aided by humans or not.
Invasive species (Heger et al., 2013)	Species spreading in the new area, i.e. they are colonizing sites beyond the area of the founder population(s).

In some cases the true origin of a species remains obscure either because of insufficient taxonomic knowledge, due to a lack of records from the time they became introduced, or for other reasons. These are termed as cryptogenic species or cryptogens (Carlton, 1996, 2009). Many of the species now categorized as cryptogenic have been previously treated as being native. Also, some species treated previously as being alien have appeared in cryptogenic species lists (Carlton, 2009).

Subconscious associations with preconceived terms, particularly emotive ones, can also lead to divergent interpretations and a confusion of concepts and theory (Simberloff, 2003; Colautti and MacIsaac, 2004). So, globally approved terminology is necessary. In this study, the terminology proposed by the EC Joint Research Center / International Council for Exploration of the Seas Task Group 2 “Non-indigenous species” developing Good Environmental Status Descriptor for the EU Marine Strategy Framework Directive is used (Olenin et al., 2010 and references therein) (Text Box 1).

Tex Box 1. Definitions for non-indigenous, invasive alien and cryptogenic species

Non-indigenous species (NIS) are species (synonyms: alien, exotic, nonnative, allochthonous and introduced), subspecies or lower taxa introduced outside of their natural range and outside of their natural dispersal potential. This includes any part, gamete or propagule of such species that might survive and subsequently reproduce. Natural shifts in distribution ranges do not qualify a species as a NIS.

Invasive alien species (IAS) are a subset of established NIS, which have spread, are spreading or have demonstrated their potential to spread elsewhere, and have an adverse effect on biological diversity, ecosystem functioning, socio-economic values and/or human health in invaded regions.

Cryptogenic species (CR) are species, which cannot be reliably demonstrated as being either introduced alien or native.

2.1.2. Invasion process, introduction pathways and vectors

The invasion process has been summarized in several recent overviews (Colautti and MacIsaac, 2004; Occhipinti-Ambrogi and Galil, 2004; Leung et al., 2008; Walther et al., 2009; Pyšek and Richardson, 2010; Olenin et al., 2010, 2011). A typical invasion sequence is followed by five key steps (Figure 1).

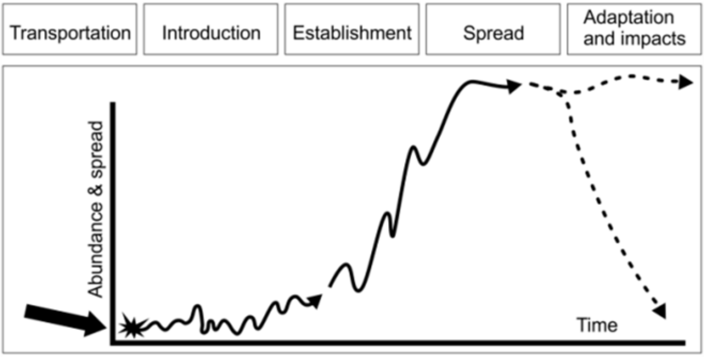


Figure 1. A typical invasion sequence (based on Reise et al. 2006; Olenin et al. 2011). Explanation in text.

Each step is characterized by different status of a NIS population:

1. Survival and transportation. Individuals of the target species are collected and transported from their native geographical range to new locations, where they do not occur naturally.

2. Success in introduction in new area. The target species are introduced into the new location, where they are alien species.

3. Establishment in new areas. Individuals become established at the point of introduction.

4. Spread in new areas. The established population subsequently grows and spreads to other locations.

5. Ecological and economic impacts. The invaders become a nuisance and cause ecological and economic impacts.

Different management scenarios may be applied for each step. However, it is assumed generally, that the risk of new biological invasions can be effectively reduced by precautionary measures, while control or eradication of existing NIS is more challenging (e.g. Olenin et al., 2010).

A NIS arrival in a new location directly from its native region is called a *primary introduction*, while its subsequent spread from the founding site is considered to be a *secondary introduction*. This spread may occur through a combination of natural dispersal and human-associated transport mechanisms, i.e. pathways and vectors, such as water dispersal, transfer by wind, or spreading by animals (Olenin et al., 2010).

A *pathway* is the route an alien species takes to enter or spread through a non-native ecosystem. Species become spread by a wide range of pathways either deliberately or inadvertently (Carlton, 2001; Minchin, 2001; Minchin et al., 2009). Each pathway may have a number of vectors that are involved in a species transmission.

A *vector* is a transfer mechanism, the physical means by which species are transported from one geographic region to another (Carlton, 2001; Verling et al., 2005), e.g. ballast water or ship's hull. Several vectors within a pathway may be involved in a transmission; also, the role of vectors may change over time and may differ regionally (Galil et al., 2009; Minchin et al., 2009). In many cases the operating vector remains unknown (Olenin et al., 2010). As an introduced species expands its new range, other opportunities may arise by additional vectors to increase its range. On occasion, the arrival of a NIS may arise with additional vectors acting in relay to convey an alien elsewhere (Figure 2).

It is very important to gather data not only on pathways and vectors, but also on biological traits of NIS as they can be potential predictors indicating whether or not they will be impacting (Leung and Dudgeon, 2008). Biological parameters are the basis for the ecological status assessment (Olenin et al., 2010), and they vary among different stages of invasion and are likely taxonomic specific.

Figure 2. Interaction of various biological introduction vectors in estuarine and coastal areas: 1. arrival of NIS with shipping; 2. range expansion through the canal systems; 3. transfer of fouling organisms on small craft and to marina sites from sea and overland transport of boats; 4. stocking of organisms to provide leisure pursuits or for fishery management; 5. releases from aquaria or from cold water ponds; 6. releases of organisms intended as live food; 7. releases by anglers or from their equipment; 8. aquaculture escapees; 9. discharges of wastes following processing; 10. movements associated with fishing gear or of discards. Reproduced from Minchin et al. (2006); drawing: Vitalija Gasiūnaitė.

2.1.3. Understanding impacts of invasive alien species

Impacts of invasive alien species (bioinvasion impacts) at levels that adversely alter environmental quality by effects on an individual organism, a population, community, habitat or an ecosystem are defined as *biological pollution* (or *biopollution*) (Elliott, 2003; Olenin et al., 2007, 2010). It is biopollution that causes major concern for researchers, environmental managers and nature conservationists. However, the term "biological pollution" has been used in different meanings. Recently this term became widely accepted due to the increasing number of studies regarding the presence and dispersal of invasive species. Elliott (2003) suggested considering such non-indigenous and invasive species and their geographical spread as "biological pollution" and "biological pollutants", and Arbačiauskas et al. (2008) used the term "biocontamination".

The nature of the impact varies greatly from the obvious effects on industries to the loss of species in a particular area or alteration of communities, habitat changes and, in some cases, ecosystem functioning (Simon and Townsend, 2003; Reise et al., 2006; Wallentinus and Nyberg, 2007; Nentwig et al., 2009; Olenin et al., 2007, 2010). Ecological impacts are altering biological, chemical and physical properties of ecosystems, resulting from single prey-predator interactions between non-native and native species to massive shifts in ecosystem functioning at all levels of biological organization from genes to communities and biotopes (Olenin et al., 2010). Virtually all established NIS have at least some impact on the environment in the area where they dwell, feed and occupy a certain territory. Unlike terrestrial systems where bioinvasions have caused significant damage to economic interests, the majority of demonstrable marine bioinvasion impacts appear to be primarily on native biodiversity and ecosystem health with few direct impacts on economic values (Olenin et al., 2010).

In a recent review Olenin and Minchin (2012) summarized the needs of managers in data on NIS impacts for:

1. Development of early warning systems based on regular sampling. For some species early information may enable eradication.
2. Assessing, on the basis of risk, expected arrivals of targeted species that cause known harmful effects.
3. Cost-benefit analysis of different management options.
4. Environmental status assessments taking into account the marine bioinvasion effects that are only possible where regular monitoring is in place.

The lack of impacts in most cases is assumed by the absence of evidences (Carlton, 2002). Neither, very few online sources contain substantiated data concerning NIS impacts, though legislators and environmental managers are mainly interested in NIS populations that have significant impacts on the environment, quality of life, economy and/or human health (Olenin et al., 2011). However, the European experts Task Group (Olenin et al., 2010) recommended that the assessment of NIS impacts generally should begin at the local scale or in areas of special interest (e.g. marine protected area, aquaculture site). By this approach species need to be classified according to the magnitude of their effects on environment. Although there is growing recognition of a need to include NIS into overall environmental status assessment (e.g. Cardoso and Free, 2008; Orendt et al., 2009), but formal classification of biological invasion impacts has been challenging due to the nature and diversity of impacts, and lack of an agreed terminology.

In order to address this need, a standardized method to assess the magnitude of the bioinvasion impacts, the “Biopollution Level” (BPL), was devised by Olenin et al. (2007). According to it the BPL calculation is based on abundance and distribution range of the NIS under consideration and the magnitude of its bioinvasion impact. The assessment should be performed for a defined aquatic area (e.g. a marina, a sand bank, a coastal lagoon, an estuary or a whole sea region) for a certain period of time. Also other initiatives exist to assess the environmental impacts, e.g. Kenis et al. (2012) derived the assessment protocols for alien plants and for other pests.

To manage invasion risks, it is necessary to register all known NIS in any region. Additionally some classification indicators may be used, e.g. distribution and abundance of selected alien species; rate of increase in aquaculture-related introduced species in the marine environment in European Seas (EEA, 2003); total number of invasive species as a percentage of relevant groups (McGeoch et al., 2006). Indicators are needed because managers are mainly interested in those species that result in some effect on human health, quality of life, economic impact or have a significant impact on the environment. Species biological characteristics (e.g. life forms, reproductive frequency, type and toxicity), traits in recipient region (e.g. migration pattern, reproductive duration), population and species statuses, and other relevant data are necessary for solving problems of biological invasions. Application of information systems on NIS to manage, share huge datasets of different origin is needed for both understanding and managing of the bioinvasion problems.

2.2. Sharing biological invasion data

In order to be effective, information must be placed within the proper context, organized in a logical and standardized manner, and as it was stated, scientifically validated, updated and continuously maintained databases are the most reliable source for information on NIS (Genovesi, 2001; Simpson et al., 2006). However, gathering data into databases has a number of challenges, e.g. involvement of participants, sufficient financial support, different data scales, varying temporal coverage and resolution, various formats and types.

NIS databases, in principle, do not differ from biodiversity information systems, which are also emerging in recent years (Costello et al., 2008, Costello, 2009). In both cases, properly collected and maintained data are fundamental for research, policy making and management. However, a lot of issues occur as bioinvasion is the multidisciplinary studies, involving not only taxonomy, genetics, physiology, but also data on impacts, habitats, geographical origins, etc. In this Section the data sharing needs and problems are reviewed in both biodiversity and NIS related fields in order to outline the roadmap towards possible technological and organizational solutions for an ideal information system.

2.2.1. Data sharing needs and problems

Data sharing is the practice of making ones data available to others, reusing it for subsequent analyses openly and publicly, or at least within a closed community of researchers (Enke et al., 2012). Long term data storage is only reasonable if the content is easy accessible and reused. In biodiversity research many international initiatives providing infrastructure for data sharing exist, but not all data (e.g. locations of endangered species, not published material) can be shared openly, and there is still a lack of comparable data (Bertzky and Stoll-Kleemann, 2009; Costello, 2009). Data sharing is ensured by databases and other supporting technologies, while records are provided by repositories, e.g. contributors. Most of limitations in data sharing arise from technical and organizational ground.

The technical limitations include differences in data structure, lack of data extraction services for both data sharing and further analysis with sophisticated software. A major obstacle to data access can be institutional barriers, where data are not centralized, but are stored in various formats with little compatibility (Beniston et al., 2012). Even if data are accessible, the lack of proper data documentation and dissemination after the termination of the project impedes reuse of the data (Refsgaard et al., 2007). Another reason may be that there often is no contact person responsible for managing requests.

The organizational issues are mainly related to uncertainties regarding the intellectual property rights and human attitude. For example, one reason why the data from former projects are not used more extensively can be that scientists, who produce data, may be unwilling to share them, due to strong traditions, competition for funding and other circumstances (Costello, 2009). However, usually it happens at the end of a project, at retirement or other external pressure, but not immediately (Enke et al., 2012). The main concern regarding data sharing is “loss of control”, the lack of time to get the data in the standardized form, not unified politics for data citation and rules of reuse. Sometimes researchers avoid not being acknowledged for sharing data, and are not optimistic to deposit data if no long term funding, maintenance and data availability is provided by projects or organizations. There is a direct relationship between data sharing and reuse: if users agree to share data they are also expecting to reuse data, and bring research into a broader context (Enke et al., 2012). However, the motivation to share data may exceed such fears.

The main drivers for data sharing are the availability of large datasets for comprehensive analyses, transparency of results, networking with other researchers, and researcher's visibility in the community (Constable et al., 2010; Olenin et al., 2013). In some cases requirement to publish data is requested from funding agencies. However, the storage and reuse of complex data are becoming increasingly important in attempts to cope with the impacts of environmental change. In order to answer important questions, enable better analysis, researchers nowadays depend on the accessibility and the reuse of combined datasets.

2.2.2. Publically available NIS information sources

Recent efforts to standardize unprecedented amount of NIS data have resulted in a great variety of open-access online resources: more than 250 websites are listed from 2004 (GISIN, 2008). These online sources collect and disseminate data for researchers, policy makers and members of the public. Stored information is often retrieved from a variety of resources, such as online databases, peer-reviewed and grey literature, published and unpublished national, regional projects or institutional datasets. Obtained knowledge from these online data sources may be integrated into policies, whose successful implementation also depends on the quality and availability of the alien species records (Vandekerkhove and Cardoso, 2011).

In past years there have been a number of initiatives to integrate country specific information into national, regional or even global databases. Although Europe for long time has poor coverage of NIS information, recently it has transformed itself to one leading in the world (Hulme et al., 2009a). This is attributed to the parallel developments in invasive species inventories undertaken over the last decade through several major initiatives. For example:

- Baltic Sea Alien Species database, which is available online since 1997 (Olenin et al., 2002; Baltic Sea Alien Species Database, 2013). The database provides a qualified reference system on alien species for the Baltic Sea area, their biology, vectors of introduction, spread, impacts on environment and economy. The system encourages the exchange of data among different geographical regions and thereby serves a node in the Global Information System for Invasive Species (GISIN, 2008).
- North European and Baltic Network on Invasive Alien Species, a regional portal supplying information on alien species in northern and central Europe, was launched in 2000 (NOBANIS, 2013). The database is currently delivering data through a network of cooperating competent authorities of 20 countries.
- Delivering Alien Invasive Species Inventories for Europe (DAISIE, 2013) was initiated to deliver a pan-European inventory of invasive alien species. DAISIE collates data for all 27 European Union member states (and separately for their significant island regions), and for other European states (Hulme et al., 2009b).
- Global Invasive Species Database (GISD, 2013) aims to increase awareness about invasive alien species,to facilitate effective prevention and management activities. GISD focuses on invasive alien species that threaten native biodiversity and natural ecosystems and covers all taxonomic groups from microorganisms to animals and plants in all ecosystems.
- International Nonindigenous Species Database Network, available online since 2003, is a distributed database providing information concerning NIS (NISBase, 2013). At least five databases on aquatic (marine) non-indigenous species may be searched simultaneously through this network.

- European Alien Species Information Network (EASIN, 2013) was launched in late 2012. The system aims at improving the access to data and information on alien species in Europe. EASIN facilitates the exploration of alien species information in Europe from distributed resources through a network of interoperable web services, following internationally recognized standards and protocols.

While there are many studies involving compilations of NIS, the associated information principally relates to their likely vectors and geographical regions they have invaded (e.g. IABIN, 2013). Most of them provide certain information to end-user; however, there are no online tools to perform any meaningful analysis. Some databases are organized to share data with external resources, and are distributed within cross-searchable information networks. However, quantity and quality of stored and disseminated data are typically affected by taxonomical biases, poor geographical representation and a lack of confidence in data quality (Olenin et al., 2011). Most NIS databases are based on summarized secondary data, which is of limited use for managers and policy makers (Olenin et al., 2013).

No one of the above mentioned and other online sources address the problem of biological pollution as such – in some cases impacts are considered, but this presentation doesn't clearly inform managers, why public and/or shipping/ports industry money should be spent to prevent new introductions, and doesn't indicate feasibility of control measures (Olenin et al., 2011; Vandekerkhove and Cardoso, 2011). This happens as bioinvasion studies lack being more analytical and predictive because of a variety of uncertainties, such as information on vectors involved in the transportation process (Minchin, 2007), numbers of established and spreading NIS, propagule pressure (Johnston et al., 2009), biological traits of invaders (Karatayev et al., 2009) and their environmental tolerance limits (Olyarnik et al., 2009), functional role (Crooks, 2009) and the impacts on environment, economy and human health (Olenin et al., 2007).

This variety of study fields results that some existing NIS databases store just anecdotal information on impacts (Narščius et al., 2012). Moreover, in most cases there is no structure to accommodate data on impacts (e.g. on environment, public health, public well-being and quality of life, uses of marine areas and economy) in a standardized way.

As a result, it is difficult to undertake comparisons on the magnitude of the impacts arising from the same species in different regions or between different species as information on biological invasion impacts is seldom quantified and usually descriptive. More advanced information systems on NIS are needed to support bioinvasion management, including standard descriptive methods to classify the degree of impacts in order to prioritize those species that warrant particular attention.

2.2.3. Towards an ideal NIS information system

Many biodiversity databases have emerged in recent years in order to support integrated research and effective data reuse. However, some specialized databases are mainly known by close community of users (Enke et al., 2012). The reason that makes NIS data gathering and sharing complicated is that different datasets are developed independently (e.g. different technological solutions, formats, lifetimes), and it is challenge task to accumulate large volumes of complex and heterogeneous biological data (Soberón and Peterson, 2009; Pullin and Salafsky, 2010; Bach et al., 2011).

To increase the amount of shared content, researchers should collect data already in a database compatible format to save time spent on editing and converting data into the accepted format (Van House, 2003). Furthermore, databases could provide guidelines of comprehensive usage, availability for user to work safely on datasets, combine and present different datasets, apply simple analysis tools with optional release of results for public access. Published data cannot be modified, but if some changes have to be made later, versioning of records is needed. The availability to link existing databases and provide the integrated search, include additional functionality (e.g. to check species taxonomic information) is necessary.

In all cases the quality of published data is very important aspect, and often potential users are worried about data available from unfamiliar sources (Van House, 2003). An annotation facility and additional information (i.e. comments, references) would improve the value of database over time, and the probability of data being reused. A more sustainable way for motivating an activity to store data into databases could be support from journals as an obligatory criteria required for publications (GenBank, 2009). Sustainable repositories and long term funding for continuous data storage are also needed (Bastow and Leonelli, 2010).

To ensure smooth and coherent biodiversity data exchange and dissemination, accepted standards are required. Although biodiversity data are very heterogeneous (Bowker, 2000), enormous efforts have been invested in developing such standards (Meng, 2005). This heterogeneity may be classified into several classes of standards: syntactic-semantic, technical, data model and structural (Halevy, 2005).

Syntactic heterogeneity is the problem of using multiple syntactic descriptions for the same value, e.g. comma or point for floating numbers. It is quite easy to handle when a standardized vocabulary is used, making the data in different systems directly comparable (Meng, 2005). Semantic heterogeneity refers to the differences in meaning, interpretation and usage of data due to homonyms and synonyms. To overcome these problems semantic standards (e.g. taxonomic and geographic names) are compiled: European Cooperative Programme for Plant Genetic Resources (ECPGR, 1980), National Biological Information Infrastructure (Frame et al., 2002), the global change master directory keyword list (Olsen et al., 2007), the world register of marine species (WoRMS, Appeltans et al., 2010), etc. Also more general and broadly applicable metadata standards have been established: Ecological Metadata Language (EML, Fegraus et al., 2005), Access to Biological Data Collections (ABCD), Dublin Core (Weibel, 2005), Darwin Core (Wieczorek et al., 2012), geographic information standards ISO19xxx, etc.

Technical heterogeneity refers to the problem of accessing and exchanging data as different systems require specific data formats, export and import procedures. Technical standards have been developed to facilitate the proper exchange of data and interoperability for databases: Herbarium Information Standards and Protocols for Interchange of Data, a biological collection access service for Europe (Berendsohn et al., 2002), Distributed Generic Information Retrieval (Blum et al., 2001), etc. Also global standards, such as Z39.50 (Lynch, 1991), may be used. Exchange protocols employ commonly available data standards like comma separated values (CSV), extensible markup language (XML).

Data model heterogeneity is known as the problem that systems use a different database model to store data. Most of them are based on relational database management systems: PostgreSQL, Microsoft Access, MySQL (MySQL A.B, 1997), Oracle, etc. Other technologies like the genuine Resource Description Framework data model have been developed to store ontologies.

Structural heterogeneity addresses the problem that information can be represented in multiple ways for a given data model. To prevent these kinds of problems, structural standards, like Minimum information about an environmental sequence (Yilmaz et al., 2011), are used for describing the underlying database schema. Many of structural standards may not receive acceptance because they impose restrictions on scientific freedom (Halevy, 2005).

Researchers, data providers, managers, developers and other potential data users have different technical and contextual perspectives, requirements on data sharing and access (Simpson et al., 2006). These requirements may be gathered and generalized using standardized software engineering methods. Suggestions based on literature review (Lane and Edwards, 2007; Malcolm and Walter, 2007; White, 2007; Ruusalepp, 2008; Kuipers and Hoeven, 2009; Bach et al., 2011; Tenopir et al., 2011; Enke et al., 2012; Michener et al., 2012) for the best-practice solutions are summarized as follows: easy and user-friendly access to the information, retrieving it by browsing or searching; information on data origin and version control; clear definitions for attributes and their values; technical support and specialized courses at usage; interactive tools for data analysis; feedback from users; usage of common data standards; data quality, sustainability and the protection of intellectual property assurance; various services to ensure that the functionality of the system is of sufficient value to users and the system is extensible by an addition; linkage of digital data from disparate sources to answer complex questions.

3. MATERIAL AND METHODS

3.1. Development of the information systems

There is a difference between standard software and web application development. In web content a hyper-media paradigm is used, web applications perform similarly on each platform, and they are developed for unlimited number of users, which behavior sometimes is unpredictable (Barry and Lang, 2001). However, web technology is changing faster than software, and they are needed to be maintained more often than standard software applications.

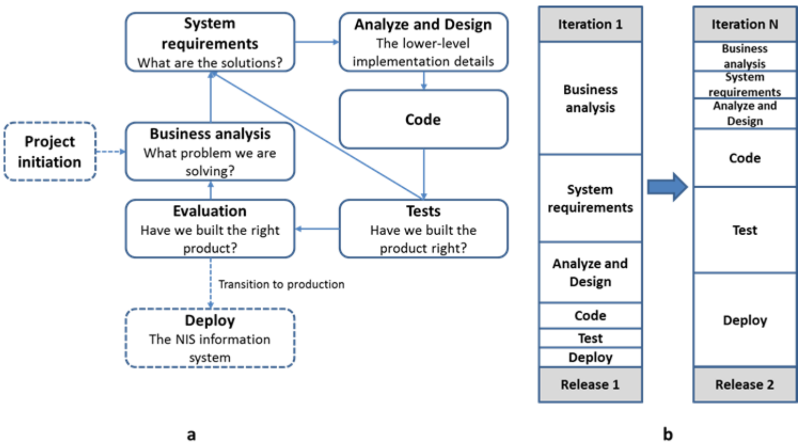


Figure 3. The iterative AWE life-cycle process (McDonald and Welland, 2005), which was used to develop the NIS information systems (a). Each iteration results in decreasing number of detected problems, increasing amount of code and test procedures (b).

In this study the Agile Web Engineering process life-cycle (AWE; McDonald and Welland, 2001, 2005) was applied according to the Rational Unified Process (Krutchen, 2003). AWE is a continuous process of discovery, invention and implementation, where iterations force to drive a project in a predictable and repeatable way (Figure 3).

Application of AWE life-cycle resulted in active development, multidisciplinary team, clearly identified needs and requirements, etc.

The Project initiation phase of BINPAS took place in 2008, and AquaNIS in 2011. The former system was developed in a framework of two projects: MEECE and BINLIT, the latter is the outcome of the ongoing projects VECTORS and DEVOTES. Specialized workshops were organized to determine the concepts of both systems, overview experiences of project participants and set up working groups. The work groups consisted of experts with different academic and professional background working in the field of biological invasions. The geographical distribution of experts covered six countries: Lithuania, Estonia, Ireland, Italy, Israel, and Germany. Several team meetings and discussions were arranged during collaborative workshops at 2008-2013 in Lithuania (2008, 2012), Italy (2011), Denmark (2012) and Spain (2012 and 2013).

The Business analysis phase involved identifying and prioritizing the objectives of AquaNIS and BINPAS. The process was based on data-driven principle applied to the analysis of data describing essential features of the bioinvasion process (see Section 2.1 for details). The purpose of this phase was to understand the problems to be addressed in the information systems (e.g. steps in BPL method application, data exchange with external databases).

The System requirements phase detailed concrete criteria for solving the objectives, which were identified in the business analysis phase. At this stage the requirements to be addressed in the next iteration were selected (e.g. to get the WoRMS species list daily).

After the requirements were identified, the Analyze and Design phases involved reasoning about possible solutions and determining the high-level implementation details, and Coding – the implementation activities. The Testing phase measured the deliverables against the requirements, while in the Evaluation phase – against the business objectives. The Deployment involved activities required to move the deliverables from a development to end-user environment.

The implementation of the NIS information systems was based on the Model-View-Controller (MVC) principle (Deacon, 1995), which separates three classes:

- The Model manages the behavior and data of the application domain, responds to requests for information about its state, and responds instructions to change the current state. It contains all code that relates to a database and other data structures.
- The View manages all user interface elements, such as CSS, Flash, HTML, JavaScript code, PHP and others.
- The Controller interprets the mouse and keyboard inputs from the user, informing the Model and/or the View to change as appropriate. In general, it gathers Models and Views together.

An open source web application framework CodeIgniter (EllisLab, 2002), which is loosely based on the MVC development pattern, was involved for use in AquaNIS and BINPAS. CodeIgniter is often noted for its speed when compared to other PHP frameworks. AquaNIS was constructed of these main controllers:

- *Aquanis*, which manages web site first screen, news, basic information, login scenarios, etc.
- *Contributors and Users* controllers are developed to administrate user rights, groups, assignments, access, login data, etc.
- *Editorial* controller monitors performed actions to records and ensures performance of the editorial board.
- *Extract* controller is developed to share data with other databases and to extract in various files.
- *Impact* controller manages data on species impacts and specific impacts in recipient regions.
- *Manage* controller is responsible for data management.
- *Open* controller performs availability to make suggestions/changes on records by not registered users.

- *Resources* controller allows uploading files.
- *Search* controller is developed for data search, comparison and metadata management.
- *Species* and *Introductions* controllers organize data on species and introduction events for public domain.

BINPAS contains most of these controllers, where specific ones (*Assessment units* and *Species assessments*), managing data on assessment units, periods and NIS impacts, are included.

Technical implementations of the NIS information systems mainly were realized using HTML, MySQL and PHP. Additionally JavaScript code and CSS were involved.

3.2. Data gathering

Initial datasets for BINPAS and AquaNIS were gathered during European projects MEECE and VECTORS, and regional project BINLIT. Currently the system integrates data from other national, regional, pan-European research projects and non-project resources on NIS: EU Concerted Action "Testing Monitoring Systems for Risk Assessment of Harmful Introductions by Ships to European Waters" (1997-1999) (Rosenthal et al., 2000); EU FP6 Integrated Project ALARM "Assessing Large-scale environmental risks with tested methods" (2004-2009) (Settele et al., 2010); EU FP6 project DAISIE "Delivering Alien Species Inventory for Europe" (2005-2008) (DAISIE, 2008); EU FP6 project IMPASSE "Environmental impacts of invasive alien species in aquaculture" (2006-2008) (Savini et al., 2010); European Census of Marine Life (2009-2010); Baltic Sea Alien Species Database (1997-2012) (Olenin et al., 2002); ICES WGITMO; ICES WGB WGBOSV; "Allochthonous Species Group" of the Italian Society of Marine Biology (SIBM).

3.3. Species taxonomy standardization

A unified taxonomic list is required to share and integrate biological data according to organism names via the internet. To standardize organism names several steps were applied – direct mapping of names and authorship from obtained lists to names in the master list, and after that assigning standard species codes. World Register of Marine Species (WoRMS; Appeltans et al., 2010) was chosen as a species master list. It combines verified and confirmed data from local collections, regional registers of marine species, various international initiatives, a permanent search of literature, the internet, sample collections and other sources. A list of marine species is stored in the database called Aphia as a part of Marine Biodiversity and Ecosystem Functioning EU Network of Excellence (Costello et al., 2008), which contains valid species names, full taxonomy, synonyms and vernacular names, statuses, biogeographic data, geographical distribution and extra information of literature data. Each record has a unique identifier AphiaID, which allows linking a species to the master list.

Primary mapping of species scientific names provided for AquaNIS and BINPAS was problematic due to homonyms when authorship was not included, use of common names in various languages, ranks instead of scientific names, different synonyms for same taxon, spelling and authorship variations, authorship year differences, etc. After unification of species names was completed, standard codes were obtained. This allowed accessing common and valid names, authorships, ranks, different spellings and languages, additional metadata, synonym taxa, linkage to parent record/taxonomic hierarchy, etc. from the master list (WoRMS). Additionally, if the validity status for a name changes in the master list, it also is changed in AquaNIS and BINPAS. If a species position is not found in the master list, a negative value to the standard code field is assigned as suggested by Branton et al. (2007). After that these new names with additional metadata are transferred to the managers of the master list with the request to consider inclusion into the master list. If the name is accepted and appended to the master list, then the record code is assigned the new non-negative code value for this name.

3.4. Data sharing

Darwin Core (Wieczorek et al., 2012) was selected as a suitable and flexible standard to facilitate finding, sharing and management of information from AquaNIS and BINPAS with various databases. This standard is related to Dublin core (Weibel, 2005) and is closely concerned with biodiversity object. It defines sets of terms from various groups resulting in many different variants (TDWG, 2010). The idea is to promote use of the accepted terms in every appropriate context, and to leave the implementation details to specific applications. The terms are organized into nine categories, six of which cover broad aspects (Event, Location, Geographical context, Occurrence, Taxon and Identification) of the biodiversity domain, and remaining ones cover relationships to other resources, measurements and generic information about records.

Although not all AquaNIS and BINPAS attributes are accepted terms (e.g. Habitat modifying ability potential), but these may be registered as candidates to be reviewed and included. While data can be shared in a variety of encoding schemes (CSV, XML, JavaScript Object Notation, etc.), AquaNIS and BINPAS are organized to share datasets using XML schemes. For example, the list of registered species with taxonomy is stored in textual file "species.csv":

```
<?xml version="1.0"?>
<archive xmlns="http://rs.tdwg.org/dwc/text/">
<core encoding="UTF-8" linesTerminatedBy="\r\n"
fieldsTerminatedBy="," fieldsEnclosedBy="&quot;"
ignoreHeaderLines="1"
rowType="http://rs.tdwg.org/dwc/terms/Taxon">
        <files>
                <location>species.csv</location>
        </files>
```

<id index="0"/>
<field index="1"
term="http://rs.tdwg.org/dwc/terms/scientificName"/>
<field index="2"
term="http://rs.tdwg.org/dwc/terms/scientificNameAuthorsh
ip"/>
<field index="3"
term="http://rs.tdwg.org/dwc/terms/genus"/>
<field index="4"
term="http://rs.tdwg.org/dwc/terms/family"/>
<field index="5"
term="http://rs.tdwg.org/dwc/terms/order"/>
<field index="6"
term="http://rs.tdwg.org/dwc/terms/class"/>
<field index="7"
term="http://rs.tdwg.org/dwc/terms/phylum"/>
</core>
</archive>

3.5. Biopollution index

In order to address the need to measure, report and verify the impacts of NIS, a standardized method to assess the magnitude of the bioinvasion impacts, the "Biopollution Level" (BPL; Olenin et al., 2007), was chosen as an object for BINPAS. This method utilized the general ecological concepts, such as "key stone species" (Payton et al., 2002), "functional groups" (Pearson, 2001) as well as descriptive accounts on invasive alien species impacts (Carlton, 2002; Grosholz, 2002; Payton et al., 2002; Reise et al., 2006; Simon and Townsend, 2003; Wallentinus and Nyberg, 2007). The BPL method is based on a classification of the abundance and distribution range of alien species and numerically expresses the magnitude of their impacts on communities, habitats and ecosystem functioning aggregated in a BPL index.

To determine the biopollution level, the abundance and distribution ranges of each alien species are assessed separately. Then the magnitude of impacts of each on community, habitat and ecosystem functioning is evaluated. The overall biopollution level is made-up of a combination of classifying the abundance and distribution ranges (ADR) and of the impacts of aliens on communities (C), invaded habitat (H) and ecosystem functioning (E).

Abundance of the aliens is ranked either as *low*, *moderate* or *high*. Distribution in the index is scored as *one locality*, *several localities*, *many localities* or *all localities*. The combination of the abundance and distribution scores gives five classes of ADR (Table 2).

Table 2. ADR classes and how they are determined.

ADR class	Description of ADR class
A	A species occurs in low numbers in one or several localities.
B	A species occurs in low numbers in many localities or in moderate numbers in one or several localities or in high numbers in one locality.
C	A species occurs in low numbers in all localities, or in moderate numbers in many localities, or in high numbers in several localities.
D	A species occurs in moderate numbers in all localities, or in high numbers in many localities.
E	A species occurs in high numbers in all localities.

After ADR is determined, it is then related to the magnitude of bioinvasion impacts scored at five levels (BPL) ranging from 0 to 4: *no impact* (0), *weak impact* (1), *moderate impact* (2), *strong impact* (3) and *massive impact* (4). The three categories of impacts relate to communities (ranging from C0 to C4), habitats (H0 to H4) and ecosystem functioning (E0 to E4) have been considered (Table 3). The particular description of the method and all the classifications are given in Olenin et al. (2007).

Table 3. Classification of alien species impacts on communities (C), habitats (H) and ecosystem functioning (E): 0 (no measurable impact), 1 (weak), 2 (moderate), 3 (strong), 4 (massive). If impact is not known, value "N" is identified. Modified from Olenin et al. (2007).

Code	Description
CN	Impact on communities is unknown.
C0	No displacement of native species, although NIS may be present. Ranking of native species according to quantitative parameters in the community remains unchanged. Type-specific communities are present.
C1	Local displacement of native species, but no extinction. Change in ranking of native species, but dominant species remain the same. Type-specific communities are present.
C2	Large scale displacement of native species causes decline in abundance and reduction of their distribution range within the assessment unit; and/or type-specific communities are changed noticeably due to shifts in community dominant species.
C3	Population extinctions within the ecosystem. Former community dominant species still present but their relative abundance is severely reduced; alien species are dominant. Loss of type-specific community within an ecological group.
C4	Population extinction of native keystone species. Extinction of type-specific communities occurs within more than one ecological group.
HN	Impact on habitats is unknown.
H0	No habitat alteration.
H1	Alteration of a habitat(s), but no reduction of spatial extent of a habitat(s).
H2	Alteration and reduction of spatial extent of a habitat(s).
H3	Alteration of a key habitat, severe reduction of spatial extent of habitat(s); loss of habitat(s) within a small area of the assessment unit.
H4	Loss of habitats in most or the entire assessment unit, loss of a key habitat.
EN	Impact on ecosystem function(s) is unknown.
E0	No measurable effect.
E1	Measurable, but weak changes with no loss or addition of new ecosystem function(s).
E2	Moderate modification of ecosystem performance and/or addition of a new, or reduction of existing, functional group(s) in part of the assessment unit.
E3	Severe shifts in ecosystem functioning in part of the assessment unit. Reorganisation of the food web as a result of addition or reduction of functional groups within trophic levels.
E4	Extreme, ecosystem-wide shift in the food web and/or loss of the role of a functional group(s).

3.6. Invasive alien species indicators

In 2002 through the Convention on Biological Diversity (CBD, 2010) the world leaders committed to achieve a reduction in the rate of biodiversity loss by 2010. To report a progress toward “2010 target” 31 indicators were selected according to such criteria as state (Wild Bird Index, Red List Index, Water Quality Index, etc.), pressure (Climatic Impact Indicator, Nitrogen deposition rate, Number of alien species in Europe, etc.), response (Extent of Protected Areas, Area of forest under sustainable management, etc.) and benefits (IUCN Red List Index for species used for food and medicine, etc.).

The primary reason for using IAS indicators is to specify and monitor the status of alien species invasion (CBD, 2010). To assess progress toward reduction of the threat from IAS, information is needed on the number and status of alien species, and on actions underway to reduce the number of IAS (McGeoch et al., 2006). As an effective tool, ecological indicators are widely used in ecosystem monitoring, assessment, management, and play an increasingly important role.

The “cumulative number of alien species in Europe since 1900th”. This is the first developed regional indicator of trends in NIS for Europe (ECCHM, 2005). High rates of increase in this number suggest that introduction pathways are numerous or wide open (Carlton and Ruiz, 2005), whereas a decline or stabilization in this number suggests that control and management actions are effective. Comparison of relative numbers of IAS across nations also provides information on the global extent of the problem (Andow, 2005). It is believed that the regions with elevated numbers of NIS are at greater risk of exposure to human mediated vectors of introductions and hence to future invasions (Olenin et al., 2010).

The number of documented introduction pathways. A measure of the number of introduction pathways and vectors covered by operational management plans to prevent or minimize the introduction of propagules will thus provide information on actions being taken to reduce the future potential status of IAS (Carlton and Ruiz, 2005). Although 5 major introduction pathways are recognized, such as air, sea, road, postal, and intraboundary translocation (Ruiz et al., 2000; Carlton and Ruiz, 2005), AquaNIS includes 11 pathways (see Section 4.1.2.1). Theoretically, management plans and policy actions for IAS introduction pathways should aim to achieve a reduction the number and size of introduction pathways.

3.7. Statistical and data mining methods

Taxonomic Distinctiveness index. Taxonomic Distinctness index (Clarke and Warwick, 1998; Euler, 1999) is defined as the average path length between any two different species chosen at random along the taxonomic tree drawn using the Linnaean classification (species name, genus, family, order, class, phylum, etc.). Here the higher values reflect the higher diversity of samples. Average taxonomic distinctness is a measure of average degree to which species in a composition are related to each other. Advantage of this index is that making use of the average taxonomic distinctness and variation in taxonomic distinctness index, biodiversity between healthy, moderately degraded and heavily degraded habitats are compared using the 95% histogram or 95% funnel and ellipse plot (Leonard et al., 2005). In this study, the taxonomic distinctiveness index was applied to analyze patterns of the Baltic Sea region homogeneity using the statistical package Primer v6.

Classification trees. Classification trees (Maindonald and Braun, 2007) are used to explore the relationship between a single response nominal variable and two or more nominal explanatory variables. Tree models deal better with non-linearity and interaction between explanatory variables than regression, generalized linear and additive models. Using too many explanatory variables results in a model that over fits the data, but only a few explanatory variables can lead to a poor model fit. Using a small tree might result in a poor data fit

(Baziukaitė and Narščius, 2010). The tree algorithm applies a cross-validation that gives a number of replicate values for the prediction error. In this study classification trees were used to identify bio-geographical regions of the Baltic Sea region using the statistical package Primer v6.

Classifiers and clusterers by Weka. Weka is a collection of machine learning algorithms for data mining tasks (Frank et al., 2004). These algorithms can be applied directly to a dataset or called from Java code. It is very important as they are going to be integrated into the server, where AquaNIS and BINPAS are being hosted. Weka contains tools for data pre-processing, classification, regression, clustering, association rules and visualization.

Below are listed classifiers from Weka (version 3.6.9) environment (Witten et al., 2011), which were applied on data stored in BINPAS to find patterns for the further BPL method revision:

- *Bayes Network Classifier* learns Bayesian nets under the assumptions that nominal attributes and no missing values are used. Search is done using *K2* or the *TAN* algorithm or more sophisticated methods based on hill-climbing, simulated annealing, tabu search, and genetic algorithms.
- *Decorate* builds ensembles of diverse classifiers by using specially constructed artificial training examples. Larger ensembles usually produce more accurate models but have greater training time and model complexity.
- *FT* builds functional trees for classification with linear functions at the leaves and, optionally, at interior nodes. It expands the choice of attributes to split on at interior nodes by creating synthetic attributes that hold the class probabilities predicted by that node's logistic regression model.
- *Logistic Regression* is an alternative implementation for building and using a multinomial logistic regression model with a

ridge estimator to guard against over fitting by penalizing large coefficients.

- *Multilayer Perceptron* is a classifier using neural networks with an input layer, number of hidden layers and an output layer. Input and output layers are necessary, but the number of hidden layers may vary.
- *NNGE* is a nearest-neighbor method for generating rules using non-nested generalized exemplars.
- *Radial basis function network* implements a Gaussian radial basis function network, deriving the centers and widths of hidden units using *k*-means and combining the outputs obtained from the hidden layer using logistic regression if the class is nominal and linear regression if it is numeric.
- *Random Committee* builds an ensemble of base classifiers and averages their predictions. Each one is based on the same data but uses a different random-number seed.
- *Random forest trees* construct random forests by bagging ensembles of random trees. Bagging generates a diverse ensemble of classifiers by introducing randomness into the learning algorithm's input.
- *Rotation Forest* combines the random subspace and bagging approaches with principal components feature generation to construct an ensemble of decision trees. In every iteration the input attributes are randomly divided into *k* disjoint subsets and after that principal components analysis is applied.

Additionally, these implementations of clusterers were used for the same datasets:

- *CLOPE* implements a fast clustering technique for market basket–type data. It uses a cluster-quality heuristic based on histograms.
- *Cobweb* implements both the Cobweb algorithm for nominal attributes. The ordering and priority of the merging and splitting operators differ between the original Cobweb and Classit papers.

- *DBScan* uses the Euclidean distance metric to determine which instances belong together in a cluster. It automatically determines the number of clusters, finds arbitrarily shaped clusters, and incorporates a notion of outlier.
- *DTNB* is hybrid classifier combining decision tables and Naïve Bayes. The predictions produced by the two methods are combined into an overall prediction using Bayes' rule.
- *EM* is a general method of finding the maximum-likelihood estimate of the parameters of an underlying distribution from a given dataset when the data is incomplete or has missing values.
- *Farthest First* clusters using the farthest first traversal algorithm. It is a fast, simple, approximate clusterer modeled on *k*-means.
- *LMT* builds logistic model trees. When fitting the logistic regression functions at a node using the *LogitBoost* algorithm, it uses cross-validation to determine number of iterations to run.
- *NB Tree* builds a decision tree with Naïve Bayes classifiers at the leaves
- *Prism* is a greedy algorithm that finds a minimum spanning tree for a connected weighted undirected graph.

4. RESULTS

4.1. AquaNIS: Information system on Aquatic Non-Indigenous Species

4.1.1. Concept

AquaNIS is an online information system on the aquatic non-indigenous species (NIS), and species which might be considered as NIS, i.e. cryptogenic species. This system is a product of the project Vectors of Change in Oceans and Seas Marine Life, Impact on Economic Sectors (VECTORS) funded during the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement Number [266445]. The system is accessible by internet at http://corpi.ku.lt/databases/aquanis. To cover needs of policy makers and researchers, and to ensure ethic data usage, these principles have been followed in the development and use of AquaNIS:

- AquaNIS stores and disseminates data on NIS introduction histories, recipient regions, taxonomy, biological traits, impacts, pathways of introduction, and other relevant documented data.
- AquaNIS seeks to ensure long-term maintenance and reliability of the database by continuous update and scientific validation of its data.
- AquaNIS users consist of registered (the chief editor, editors and contributors) and public access users (end-users).
- The chief editor is managing the database team and organizing data validation procedure.
- A voluntary wishing to contribute data to AquaNIS may freely subscribe to join the group of contributors.
- Only data supported with referenced documentation may be entered into the database by contributors.

- Contributed data is included into the database after validation by an editor, which is responsible for management of corresponding recipient region and/or NIS taxonomic group.
- Each database contributor is free to distribute (within and without AquaNIS) his own data.
- Data is not to be released for public access, either within of beyond AquaNIS, without the approval by the according contributor.
- Data for public access is freely shared for end-users.
- Registered users may access all information stored in the "Species" (NIS taxonomy, biological traits, native origin, etc.), "Impacts" and "Geography" data blocks.
- Data stored in the "Introduction event" block is available only to those corresponding contributors who submit data for countries which belong to a relevant Large Marine Ecosystem (LME) or LME sub-region.
- Acknowledgements for the usage of the AquaNIS system, referring to the version, and date of usage, should be made in the following way:

  AquaNIS. YEAR. Information system on Aquatic Non-Indigenous species AquaNIS, version X.X. Developed by: S. Olenin, A. Narščius, D. Minchin, A. Zaiko, B. Galil, S. Gollasch, A. Occhipinti-Ambrogi and H. Ojaveer. Accessed at www.corpi.ku.lt/databases/aquanis on YEAR-MM-DD.

4.1.2. Structure

AquaNIS is based on a flexible, easily extendible structure, where new data blocks and functional modules may be included as necessary. Presently (March, 2013) all data are organized in four interrelated blocks: "Introduction event", "Species", "Geography" and "Impacts" (Figure 4). However, "Impacts" data block is composite, containing information, which is divided within "Species" and "Introduction event" units. Such interrelated structure ensures that records entered in one data block (e.g. "Geography") can be reused by others without repeatedly entering them.

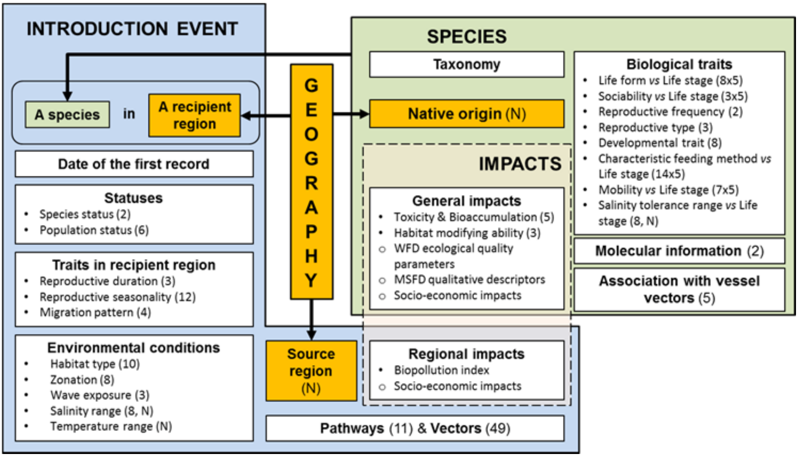


Figure 4. Data organization of AquaNIS: two main ("Introduction event" and "Species"), one supporting ("Geography") and one composite ("Impacts") data blocks and attributes within them. Numbers in brackets indicate how many predefined values are covered by each attribute. "N" means that the number of values may vary; "x" indicates a matrix of possible choices. Attributes indicated by open bullet points are under development.

Data within blocks are grouped according to thematically named attributes (Development traits, Pathways and vectors, etc. – hereafter in text names of data attributes are underlined). For most of them initial values (*brooding*, *vessels*, etc.) are included, likewise there is an availability to add additional ones. Verbal descriptions (definitions) are prepared for almost all values. To ensure data quality, references and comments are provided for every attribute.

4.1.2.1. "Introduction event" data block

The basic data entry in AquaNIS is an introduction event record, indicating a NIS introduction into a recipient region (Figure 5). There the recipient region is a country or country area within a Large Marine Ecosystem (LME, Table Box 2) sub-region.

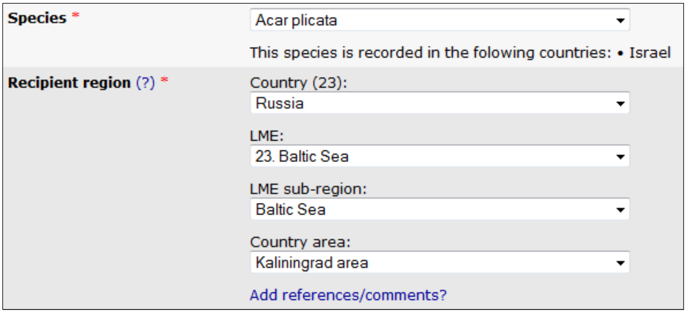

Species *
Acar plicata
This species is recorded in the folowing countries: • Israel
Recipient region (?) *
Country (23):
Russia
LME:
23. Baltic Sea
LME sub-region:
Baltic Sea
Country area:
Kaliningrad area
Add references/comments?

Figure 5. Print screen from AquaNIS showing the extraction of the introduction event registration form. After a species name is selected, additional information about its occurrences in other regions is displayed. A recipient region is constructed by selecting a Country, LME, LME sub-region and Country area.

Text Box 2. Large Marine Ecosystems of the World with AquaNIS additions

Large Marine Ecosystems (LMEs) are extensive areas of ocean space of 200,000 km² or more, characterized by distinct hydrographic regimes, submarine topography, productivity, and trophically dependent populations, adjacent to the continents in coastal waters where primary productivity is generally higher than in open ocean areas (sensu Sherman and Duda, 1999). The map of LMEs and all accompanying information is publically available at the NOAA (2012) website. In AquaNIS several additional LME-like regions (e.g. Caspian Sea, Macaronesia) were added to ensure better geographical coverage.

Occurrence of a NIS in the same recipient region can be registered only once. This information is supported by "Species" and "Geography" data blocks (see Section 4.1.2.2 and 4.1.2.3).

For each introduction event record the date of the first record (according to year, decade or century) indicates, when a species presence was firstly noticed in a concrete recipient region. This date should be supported by documented references. Additionally, the level of certainty (*High*, *Medium* or *Low*) is automatically indicated (Table 4). The *High* level of certainty means that the exact date is known; *Medium* – a period falls inside a decade. In other cases *Low* is indicated.

Table 4. Available combinations of attribute "Date of the first record".

Date to be specified	Date From	Date To	Level of certainty
Exact 1985	1985	1985	High
6th decade of 19th century	1954	1958	Medium
18th century	1701	1800	Low
Before 1700		1700	Low
After 2001	2001		Low

How a single NIS was introduced to a recipient region is stored under pathways and vectors. In AquaNIS each pathway (except *Suez Canal*) has a number of vectors. For example, the pathway *Vessels* includes such vectors as *Ballast tank sediments, Ballast water*, *Sea chests*. There more than one vector within a pathway may be involved in a transfer of a species. A drop down menu provides a choice of 11 pathways (aquarium trade, culture activities, leisure activities, live food trade, management, natural spread from neighbouring countries, other canals, research and education, Suez Canal, vessels and wild fisheries) linked to 49 vectors. Following an entry, the selected level of certainty (Table 5) identifies further data input procedure. If a level of certainty is selected as "*Direct evidence*", a pathway and at least one direct vector is required to be chosen. For a level of certainty "*Very likely*" likely or possible vectors are selected, and "*Possible*" – several known pathways and vectors may be checked.

Table 5. Explanation of levels of certainty for Pathways and Vectors.

Level of certainty	Criteria
Direct evidence	The species was actually found associated with the specific vector(s) of a pathway at the time of introduction to a particular locality within a country/country region.
Very likely	The species appears for the first time in a locality where a single pathway/vector(s) is known to operate and where there is no other explanation that can be argued for its presence except by this likely pathway/vector(s).
Possible	The species cannot be convincingly ascribed to a single pathway, but is known to be introduced by this pathway(s) elsewhere.
Unknown	Invasion of a given alien species cannot be clearly explained.

The source region of a NIS is a concrete recipient region a species was introduced from. Depending on the information available it may be assigned to a particular locality (e.g. port, port vicinity), country, LME, LME sub-region, Ocean region or Ocean (Figure 6). This information is important for identifying a NIS second spreads. All known source regions can be registered.

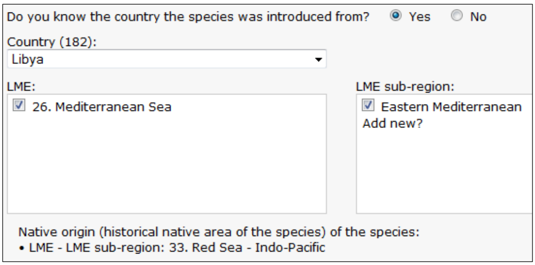


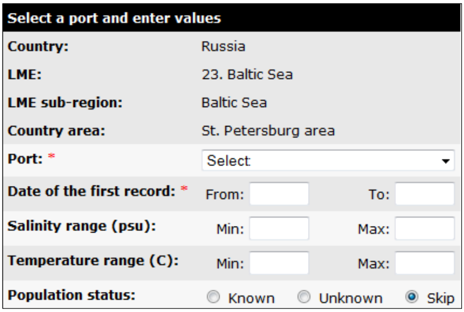
Figure 6. Print screen from AquaNIS showing a selection of a NIS source region. The source region is often confused with the native origin due to secondary spread of a NIS from a neighboring country or LME acting as a source. If a species native region is known, it is displayed near.

Ecological zones occupied by a NIS throughout its life cycle are stored by the attribute called zonation. The values are grouped into *benthic*, *pelagic* or *benthic-pelagic* zones. According to a selected group, one or more predefined values are suggested to be selected. The substratum type should be provided, if benthic ecological zone is occupied.

If salinity range of a recipient region, where a NIS is able to live and survive, is known, it should be filled in the attribute salinity range. For the exact salinity range minimum and maximum values are entered (in *psu*). In other case, the Venice system, which classifies marine waters according to salinity, may be applied by selecting proper categories. Accordingly, temperature range may be used to indicate the minimum and maximum annual temperature range values (in °*C*).

If a NIS habitat type, where it is able to live, produce and survive in a recipient region, is known, this information should be provided. For example, it can be *estuary* (river mouth, transition zone between riverine and marine environments, subject to influences from both), *lagoon* (shallow, enclosed water body separated from the sea by barrier islands, narrow spit or reefs), *offshore* (areas located at least 50 nautical miles from the shore), etc.

If a habitat type is selected as *port* (a location on a coast or shore containing one or more harbours where ships can dock and transfer people or cargo to or from land) or *port vicinity* (the area near a port where ballast water operations may occur, including areas where vessels may conduct ballast water discharge or uptake operations when approaching a port or leaving it), all known ports or port vicinities with additional information can be identified (Figure 7). Following the introduction the population status may change over time, the NIS may spread to other localities (e.g. ports, port vicinities) within the recipient region and appear at different levels of abundance. In this case recent information on a population status within ports or port vicinities should be updated.



Select a port and enter values				
Country:	Russia			
LME:	23. Baltic Sea			
LME sub-region:	Baltic Sea			
Country area:	St. Petersburg area			
Port: *	Select			
Date of the first record: *	From:		To:	
Salinity range (psu):	Min:		Max:	
Temperature range (C):	Min:		Max:	
Population status:	Known	Unknown	Skip	

Figure 7. Print screen from AquaNIS showing a NIS registration in a recipient port or port vicinity. A date of a NIS detection is required to be entered. Also salinity and temperature ranges may be entered for this period if they differ from defined in an introduction event record.

Reproductive duration indicates how long it takes a NIS to reproduce from one offspring to a new one: *long* (breeds in one or more discrete periods, each longer than three months), *medium* (breeds in one or more discrete periods, each longer than a week and less than three months) or *short* (breeds in one or more discrete periods within a week). Months, when a NIS reproduces in the invaded site, are stored by reproductive seasonality attribute. Moreover, types of a NIS movements between alternative habitats (e.g. *life-time* if a NIS produces one time migration between different habitats during its life cycle) in recipient regions are covered by migration pattern attribute.

Species status refers to either a species being NIS or cryptogenic. This option allows to separate introduction events for further data analysis.

Information on the population status is classified according to three levels of certainty: *Low*, *Moderate* and *High* (Table 6). A species population status may change over time after an introduction event.

Table 6. Certainty levels of a NIS population status in a recipient region.

Level of certainty	Explanations of available choices
Low	*Established* – a species is known to form a reproducing population in a wild. *Not established* – there is no evidence of a species' reproducing population in a wild. *Unknown* – there is no reliable information on population status of a species.
Moderate	*Extinct/no recent record* – there are old records where a species was recorded but have not been seen in the same region since. *Rare/single record* – there are only casual observations or a single record of a species' presence available. *Common* – a species with successfully reproducing populations in an open ecosystem, which are unlikely to be eliminated by man or natural causes. Not dominating native communities. *Abundant* – a species with successfully reproducing populations in an open ecosystem, which are unlikely to be eliminated by man or natural causes. Locally dominating native communities. *Very abundant* – a species with successfully reproducing populations in an open ecosystem, which are unlikely to be eliminated by man or natural causes. Largely dominating native communities. *Outbreak* – a species undergoing pulse-like, short-term (days to few months) exponential population growth during which it has an adverse effect on one or more of the following: biological diversity, ecosystem functioning, socio-economic values and human health.
High	Abundance and distribution range classes as required for the biopollution index assessment (sensu Olenin et al., 2007; for details see: Narščius et al., 2012).

4.1.2.2. "Species" data block

"Species" data block provides general information for each species, including its taxonomy, biological traits, native origin and other relevant information (Figure 4). Some stored attributes are reused by introduction events, so it is important to fill as much data as known.

Taxonomy is based on the updated accounts in major global organism-specific databases such as WoRMS (Appeltans et al., 2010). To simplify a new species record registration, only species genus should be selected and name with additional information (authority, synonyms, sub-species level, references and comments) entered (Figure 8).

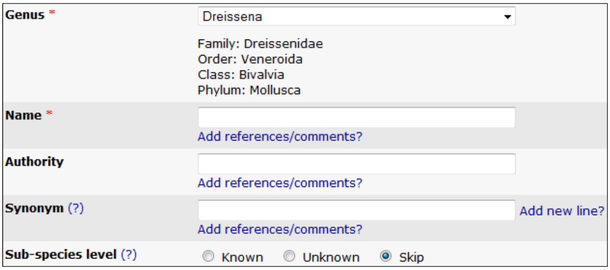


Figure 8. Print screen from AquaNIS showing an extraction of a species taxonomy registration. When a species genus is selected, its taxonomy (family, order, class and phylum) is displayed. Commonly used valid synonyms of a species (but not all of them) can be entered.

Native origin refers to a region, where a species originates from. It can be indicated according to its bio-geographical range at different scales from ocean to a LME/country (see Section 2.3). For example, native origin of black striped mussel *Mytilopsis sallei* is identified at a country-LME level: Guatemala (LME 12. Caribbean Sea) and USA (LME 5. Gulf of Mexico). In principle, it is presented in the same way as a source region in the "Introduction event" data block.

The special data entrance mechanism is provided for those attributes, which may change during the life history cycle (characteristic feeding method, life form, mobility and sociability). For them checkable matrices are generated, where available values are displayed in rows while life stages (*adult, juvenile, larvae, eggs* and *resting* stages) in columns. To simplify data entrance, theoretically not possible combinations are excluded for each attribute according to a species phylum by the Species trait helper tool.

Other attributes, reproductive frequency (species reproductive intensity which depends on reproductive system, e.g. *iteroparous*, *semelparous*), reproductive type (*asexual*, *self-fertilization*, or *sexual*) and developmental trait (different species life stages during the development, e.g. *brooding*, *spawning*), have checkable or selectable fields with predefined values (Figure 9).

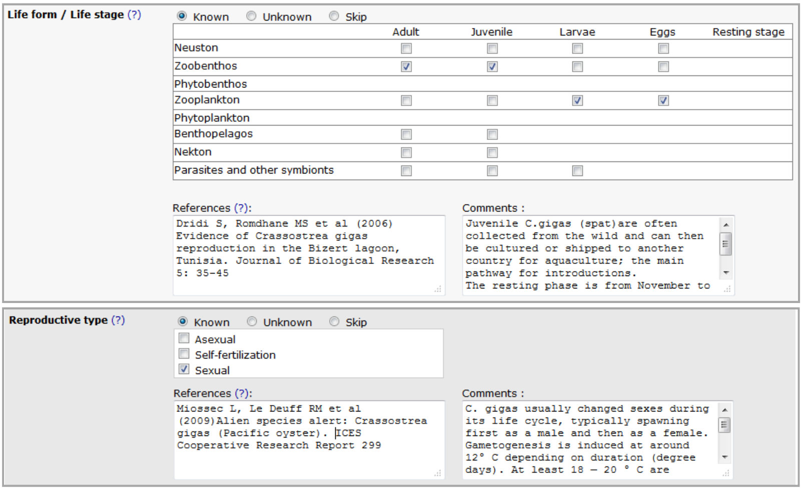


Figure 9. Print screen from AquaNIS showing examples of entry formats: Life form - Life stage matrix indicating life forms within the life cycle and option to specify reproductive type for the Pacific oyster *Crassostrea gigas*.

Salinity tolerance range describes a species ability to tolerate water salinity range in native environments. It may be indicated either by choosing the predefined Venice system zones (from limnetic < 0.5psu to hypersaline > 40psu) or by entering known minimum to maximum salinity range values (according to Figure 7).

Molecular information indicates whether there are genetic markers available for a given NIS. Such information may aid in identification of the source areas for those NIS occupying different geographical regions. If molecular information is known, references are required to be provided.

Association with vessel vectors provides verified and documented records of a species transmission by *anchor and anchor chains* (organisms found on anchors, anchor chain or within attached sediments, including anchor chain lockers), *ballast water* (water with its suspended matter taken on board a ship to control trim, list, draught, stability or stresses of the ship), *tank sediments* (matter settled out of ballast water within a ship), etc. from any world region. This is in contrast to “Introduction event” data block, where pathways and vectors have to be related to a given recipient region.

4.1.2.3. “Geography” data block

Supporting “Geography” data block is designed to maintain information on a recipient and source regions (“Introduction event” data block), and native origin (“Species” data block) of a species with the hierarchy of region scales ranging from oceans to LME’s (or country if known), recipient and source regions that have a finer tuned scale of sub-regions of LME’s, countries and localities (Figure 10).

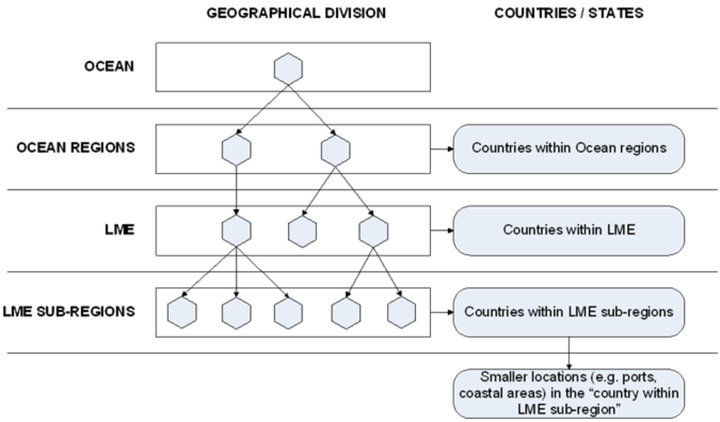


Figure 10. The principal scheme of the "Geography" data block. The levels Oceans, Ocean regions, LMEs, LME sub-regions are organized in hierarchical way. Countries and other smaller locations can be delivered from each level.

Additional sea regions, which are not covered by the LME system (originally there are 64 LMEs), are included to better complete geographical coverage of marine and coastal regions: 30A. Agulhas Current (tropical), 30B. Agulhas Current (temperate), A1. Macaronesia, A2. Caspian Sea, A3. Polynesia, A4. Melanesia and A5. Micronesia. The list of countries is adopted from the UN Population Division's quinquennial estimates and projections (WSSD, 2002). Almost all countries are linked to relevant LMEs or LME sub-regions. Here LME sub-regions are relatively large, geographically well-defined sea areas within an LME. This provides combinations "country + LME" or "country + LME sub-region", for different coasts and for a country that borders different seas, e.g. "Germany within the LME 23 Baltic Sea", "Italy within the Adriatic Sea, a sub-region of LME 26 Mediterranean Sea". The arrangement enables entering and retrieving information at different geographic scales defined by user. Large bio-geographical provinces may be combined using several LMEs, for example: Indo-Pacific, which includes the LMEs of the Tropical part of the Indian Ocean and West Pacific (LME 30A, 31, 32, 33, 34, 45; 35, 36, 37, 38, 39, 40); Ponto-Caspian region (as LME 62. Black Sea + A2. Caspian Sea).

The most important LMEs (Figure 11), covered by records in AquaNIS, are listed below:

- North Sea (LME 22).

- Baltic Sea (LME 23), where geographic limitation is defined according HELCOM Baltic definition. LME sub-regions are selected according to Baltic Sea Alien Species Database.

- Celtic-Biscay Shelf (LME 24) containing the area of the North-East Atlantic Ocean including the Celtic Sea, the English Channel and the French coast of the Bay of Biscay. The southern and western boundaries are delimited by the continental shelf, which drops away sharply. The eastern boundary is the borderline to the North Sea: the border between Belgium and the Netherlands to Dover. Three countries, Ireland, UK, and France border this LME. LME 24 sub-regions are Celtic Sea, English Channel and Biscay Gulf.

- Iberian coast (LME 25) including a continental shelf region of the NE Atlantic Ocean lying between approximately 36° N (Gulf of Cadiz) and 44° N (Cantabrian Sea) and bordered by Spain (Atlantic coast and the coast of the Bay of Biscay) and Portugal (Atlantic coast).

- Mediterranean Sea (LME 26) is limited in west by Gibraltar Strait, and in north east by Dardanelles Strait (i.e. Marmara and Black Seas excluded). It is divided into Western, Eastern Mediterranean Sea and Adriatic Sea.

“Geography” data block includes a list of ports. For each port exact location is provided – a LME, LME sub-region, country and country area. Annual minimum and maximum temperature and salinity values also can be entered. All this additional information may be reused in “Introduction event” data block by the attribute <u>habitat type</u>.

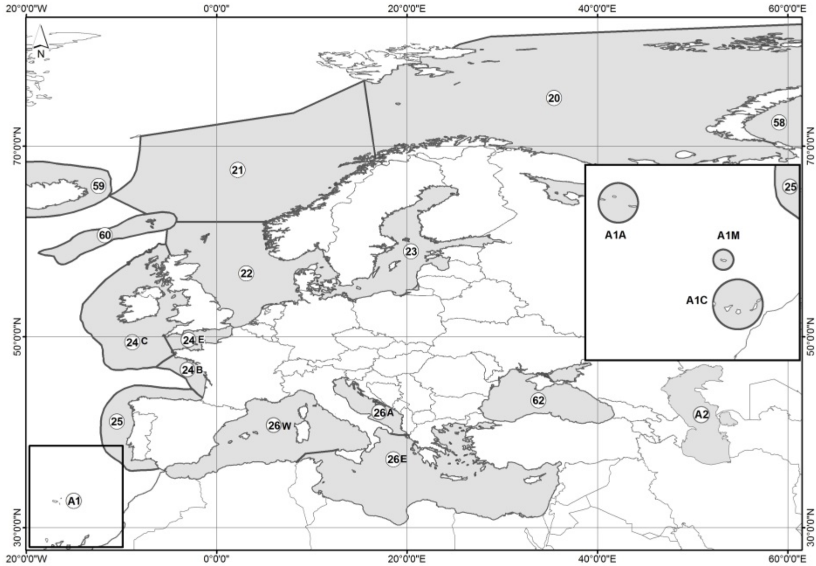


Figure 11. European and neighboring regions, where introduction events are recorded in AquaNIS. Numbers in open circles indicate Large Marine Ecosystems: 20 – Barents Sea; 21 – Norwegian Sea; 22 – North Sea; 23 – Baltic Sea; 24 – Celtic-Biscay Shelf with sub-regions (24C – Celtic seas, 24E – English Channel, 24B – Biscay); 25 – Iberian Coastal; 26 – Mediterranean Sea with sub-regions (26W – Western, 26A – Adriatic Sea, 26E – Eastern); 59. Iceland Shelf; 60 – Faroe Plateau; 62 – Black Sea. Additional LME-like regions: A1 – Macaronesia with sub-regions (A1A – Azores, A1M – Madeira, A1C – Canary Islands); A2 – Caspian Sea. Drawing: I. Bagdanavičiūtė.

To facilitate selecting a species native and source regions, the Geography Help tool was developed. This tool is useful if the same bio-geographical regions are often used (e.g. indicating West Pacific regions). It may be applied at three confidence levels: Country-LME-LME sub-region; LME-LME sub-region; Ocean-Ocean Region (Figure 12).

Figure 12. Print screen from AquaNIS showing the Geography Help tool. Every saved option has its own verbal name, e.g. "Baltic states". When adding a species native or source regions and selecting this name, the relevant regions are automatically checked as they were defined.

4.1.2.4. "Impacts" data block

"Impacts" data block is composite and includes information at two levels related to species impacts. Here one contains the global level (various environmental and socio-economic effects documented in the peer-reviewed literature from any world location) and another is based on region-specific knowledge for the introduction event based (related to "Introduction event" data block records). The region specific impacts may be more precisely incorporated into further data management, while others can be analysed as potential occasions. "Impacts" data block is especially important for bioinvasion management and risk assessments.

Global knowledge on impacts

The attribute toxicity refers to the known ability and specifically of a species to produce a poison. A species can be identified as being *poisonous* (capable of producing poison that gains entry to another organism body via the gastrointestinal tract, the respiratory tract, or via absorption through intact body layers), *venomous* (capable of producing poison, usually injected through another organism intact skin by bite or sting) or *not relevant* (neither poisonous nor venomous). These values are evaluated for different life stages using matrix form according to species phylum. Bioaccumulation defines accumulation of *natural toxins* (e.g. phytotoxins) and/or *anthropogenic chemical compounds* (e.g. pharmaceuticals, heavy metals, pesticides, dioxins) in tissues. The habitat modifying ability potential is to be entered, where a given NIS is known to change the environment via its own physical structures (*autogenic ecosystem engineers*), modify the environment by causing physical state changes in biotic and abiotic materials that, directly or indirectly, modulate the availability of resources to other species (*allogenic ecosystem engineers*), or is a keystone species (*sensu* Jones et al., 1994). These attributes (bioaccumulation, habitat modifying ability potential and toxicity) are valuated in the "Species" data block.

Two new data attributes are designed to support the decision process by managers and researchers measuring progress towards the implementation of the following two EU directives: EU Water Framework Directive (WFD; European Commission, 2001) and EU Marine Strategy Framework Directive (MSFD; European Union, 2008). Accumulated impacts on WFD ecological quality parameters will show how a NIS may change elements of biological *phytoplankton*, *phytobenthos*, *fish*, *macrofauna*, etc.), physico-chemical (*transparency/turbidity*, *temperature*, *salinity*, *oxygen conditions*, *pH*, *nutrients*, *specific pollutants*, etc.) and/or hydromorphological (*shore zones, substrate conditions, water flow, chanel patterns, sediment transport*, etc.) quality categories. For each element impact type (e.g. replacement of native species) is assessed, and description of mechanism with references is entered. All this data are arranged at a level of countries or smallest locations within them (Figure 13).

Altered WQE	Type of impact	Description of mechanism	Country	Reference	
Fish	Changes the structure of fish population	Decrease in food availability for native fish species (Mullus surmuletus)	France	Longepierre S.(2005).	
Fish	Changes the abundance of fish population	Changes in food availability for native species (Mullus surmuletus)	France	Levi F. (2004)	
Phytobenthos	Replacement of native species	Caulerpa taxifolia is outcompeting the native seagrass (Cymodocea nodosa)	Italy	Wallentinus, 2006	

Figure 13. Print screen from AquaNIS showing extraction of accumulated impacts on WFD ecological quality parameters: invasive alga *Caulerpa taxifolia* case.

Impacts on MSFD qualitative descriptors will show how NIS may alter ecosystem parameters used to evaluate the environmental status set by the MSFD qualitative descriptors (Olenin et al., in prep.). The socio-economic impacts sub-component will store data on NIS impacts on human uses within the aquatic environment, i.e. aquaculture activities, shipping, tourism including recreational boating, fisheries and also impacts on human health. The last two attributes are currently being developed.

Introduction event-specific knowledge on impacts

Environmental impact assessments in any specified locality affected by a NIS may be carried out using the biopollution index (BPL) approach (Olenin et al., 2007; Olenina et al., 2010; Zaiko et al., 2011). Assessments of BPL are guided by the Bioinvasion Impact / Biopollution assessment System, freely available online (Narščius et al., 2012) and which is linked to AquaNIS. In addition, recipient region specific socio-economic impacts for each introduction event will be included.

4.1.3. Functionality

The primary datasets were obtained in the spreadsheet files with multiple values in columns. The structure of these files didn't match the prepared new in AquaNIS. To accommodate the provided data the special algorithms for validation and importation were created. These algorithms detected a number of inaccuracies: problems with species taxonomy (e.g. not fully provided taxonomy, different taxonomy levels for the same species), bio-geographical regions (e.g. not unified geographical system, varying resolution) and missing introduction dates. Only checked and updated data were included.

Online user interface was developed for constant data management according to AquaNIS scheme (Figure 4). All selectable or checkable values of attributes are controlled by content management system, where they can be modified or added. Some attributes are organized in a hierarchical way (taxonomy, pathways-vectors and bio-geography data). For example, full taxonomic information (phylum – class – order – family) is automatically found after a species genus is selected. Introduction event records can be added only for NIS from “Species” data block. All these entries throughout AquaNIS are supported by explanations of terms and guidance, which are displayed near each attribute by popup if needed (Figure 14).

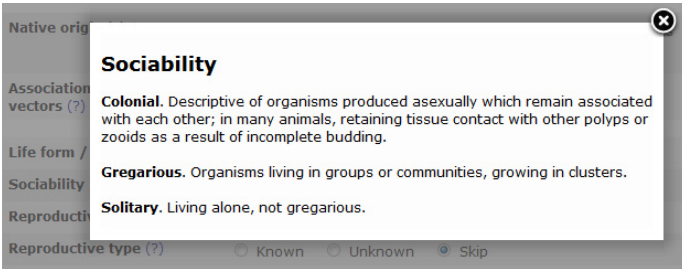


Figure 14. Print screen from AquaNIS showing an output of a definition. A definition is displayed after a question mark near attribute is clicked.

Drop down, checkable forms, which are designed to reduce data entrance errors, facilitate data entry. Free text fields are used to store references and comments. To keep data on update progress and ensure roll-back option, all entered changes are stored in a separate “track-changes” table. Version control is organized for all data blocks to identify modified attributes and contributors responsible for changes (Figure 15).

	Old value	New value
Species name	Reteporella jermanensis	Reteporella jermanensis
Synonym		Reteroporella jermanensis
	References:	References: Galil B (2007) Seeing Red: Alien species along the Mediterranean coast of Israel. Aquatic Invasions 4: 281-312
Contributor	Anna Occhipinti	Anna Occhipinti
Update date	2012-05-28	2012-06-14

Figure 15. Print screen from AquaNIS showing an example of version control output. Changes are highlighted.

Stored data are displayed by lists, where only basic information is included (e.g. species name, recipient region, date of the first record, contributor). Here pagination is enabled and alphabetic filters may be used. Detailed information is displayed after clicking a row in the list.

For specific needs the Advanced Search tool was prepared. The search function retrieves records satisfying the combination of various attributes from different categories (Table 7). These values can be selected from automatically prepared lists or some may be entered by wording (e.g. part of species name or its synonym).

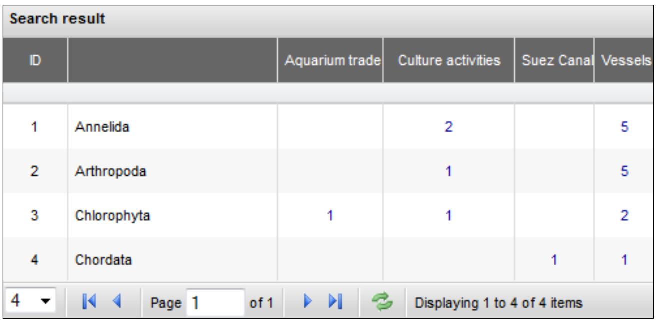
Table 7. Searchable categories and attributes within data blocks.

"Species" data block	**"Introduction event" data block**
Taxonomy Native origin Biological traits • Bioaccumulation association • Characteristic feeding method • Developmental trait • Habitat modifying ability potential • Life form • Mobility • Reproductive frequency • Reproductive type • Sociability • Toxicity • Unicellular/Multicellular? BPL Impacts • Abundance and distribution • Biopollution level • Impact on communities • Impact on ecosystems • Impact on habitats Other • Association with vessel vector • Molecular information	Recipient region Source region Pathway/Vector Traits of recipient habitat • Habitat type • Wave exposure Species status, traits • Environmental position • Migration pattern • Population status • Reproductive duration • Reproductive seasonality • Species status • Substratum • Zonation Date of the first record

Search queries can be executed with multiple criteria, which are combined using AND/OR logical statements. There AND means that the system should search for records satisfying all selected criteria in each record. Accordingly, statement OR includes records, where at least one search criteria is met in every record. Within some attributes (e.g. taxonomy, recipient region, species status) only OR is allowed to be used. Different categories are connected using logical statement AND. For example, if there is a need to get a list of introduction events of metazoan species, which have arrived by ballast waters or ballast tank sediments in the Western Mediterranean part of Spain, the search query should be like this:

[(Unicellular/Multicellular?) *Multicellular species*] AND
[(Vectors) *Ballast tank sediments* OR *Ballast water*] AND
[(Recipient region) *Western Mediterranean* AND *Spain*].

Executed search queries provide lists of found introduction events and corresponding species. Additionally the results may be presented as a table (matrix), where rows and columns are optionally constructed (Figure 16). For specific analysis some options may be applied, i.e. the selection of counting type (cumulative or standard), the inclusion of only the first NIS registration.


Search result

ID		Aquarium trade	Culture activities	Suez Canal	Vessels
1	Annelida		2		5
2	Arthropoda		1		5
3	Chlorophyta	1	1		2
4	Chordata			1	1

Figure 16. Print screen from AquaNIS showing a matrix of search results. Any values of attributes can be in rows and columns. The numbers indicate the match for corresponding row and column. Lists of introduction events or species are displayed after the number is clicked.

All lists and matrixes can be exported and downloaded. This opportunity greatly facilitates the dimension of the database for specific needs, e.g. further analysis with specialized statistical software.

To compare two or more search results, the Comparison of Search Results tool was developed. It detects the same species occurring in all results, and species which are found in one result but not in others (Figure 17). Also there is possibility combine several search results into groups and compare these groups. Output of this tool may be analysed using matrixes as demonstrated above.

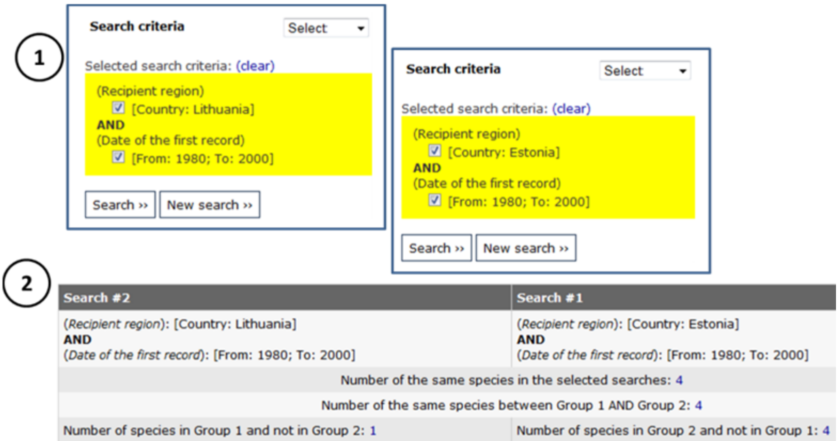


Figure 17. Print screen from AquaNIS showing the Comparison of Search Results tool. Step 1: Make two or more searches; Step 2: Compare results.

4.2. BINPAS: Biological Invasion Impact / Biopollution Assessment System

4.2.1. Concept

To develop BINPAS (Biological Invasion Impact / Biopollution Assessment System), a system aimed at translation of existing data on miscellaneous invasive species impacts into uniform biopollution measurement units, the theoretical BPL method (Olenin et al., 2007) background was used. The system is accessible by internet at http://corpi.ku.lt/databases/binpas. The purposes of the system are:

- To provide a user-friendly system to calculate biopollution level (BPL).
- To accumulate and store information on abundance and distribution range of various alien species in different geographical domains as well as their impacts on communities, habitats and ecosystem functioning.
- To enable comparisons between different species, ecosystems and time periods.

Registration to BINPAS is open for all willing to contribute with their data. On login to BINPAS a registered user can create a new assessment unit account for a certain assessment period, then to complete the assessment for an alien species estimating its ADR and ranking the impacts for either aquatic or terrestrial environment. It is important to stress that BINPAS is not producing new data; rather it is converting the existing data on alien species impacts into uniform Biological pollution level units.

4.2.2. Structure

The structure of BINPAS consists of three main interrelated data blocks (“Assessment unit”, “Assessment period” and “Species BPL account”), which constitute the framework of the BINPAS scheme, and allow the entry of the necessary data for computation of BPL (Figure 18). Data block “Assessment unit” contains data about assessment place (name, type of environment, temperature range, location, etc.). “Assessment period” stores not only assessment period, but also salinity, temperature ranges and trophic statuses if they differ from provided ones in “Assessment unit” data block for exact period. “Species BPL account” gathers data on species impacts, calculated values of abundance and distribution range (ADR) and biopollution level.

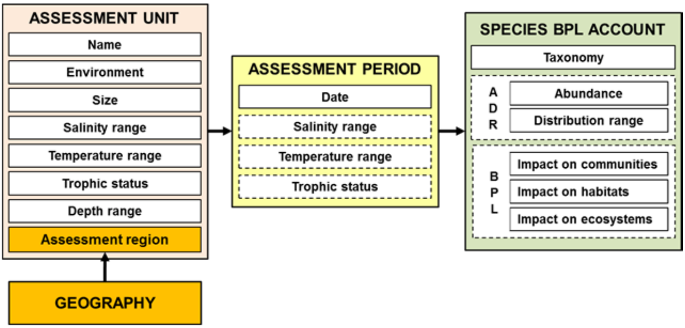


Figure 18. Data organization of BINPAS: “Assessment unit”, “Assessment period” and “Species BPL account”. The supporting blocks are excluded.

Supporting data blocks are organized to store additional data for BPL assessment procedure: “Trophic status”, “Lakes”, “Rivers”, “Large Marine Ecosystems”, “Geography location markers”, “Editorial board”, etc.

4.2.3. Functionality

The Biopollution level assessment procedure is made up of several consecutive steps, which should be performed for each assessment individually (Figure 19). This data entrance procedure considers assessment units, periods and species accounts, which are newly or additionally assessed. The sequence of steps may vary according to the purpose of a study.

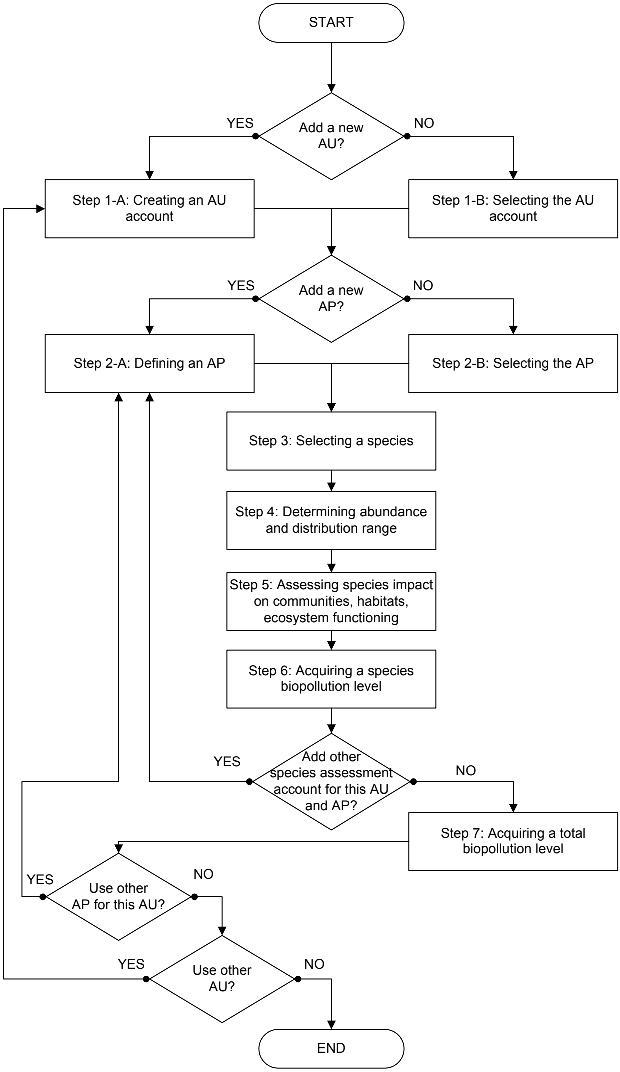


Figure 19. The biopollution level assessment procedure: AU – assessment unit account; AP – assessment period. The sequence of steps may vary according to the purpose of a study.

Step 1: Creating an assessment unit account

The Biopollution level assessment procedure starts from defining an assessment unit account (AU) with clear geographical boundaries (Figure 19, Step 1-A). If another species should be assessed in the same AU, it is selected from the list of assessment units already registered in the system Figure 19, Step 1-B.

AU can cover natural or semi-natural aquatic (marine, inland waters) or terrestrial environments. The size of the AU should relate to a manageable area decided by a contributor based on a purpose of the assessment. For example, in the aquatic environment an AU may be a coastal area, estuary, lake, marina, fish farm, marine protected area, etc. In a terrestrial realm this may be a wooded region, wetland, mountain valley, nature reserve or garden. The prerequisite for the assessment is sufficient data on abundance and distribution of an alien species present in the assessment unit. A basic knowledge of the native biota and the bioinvasion environmental impacts are also required.

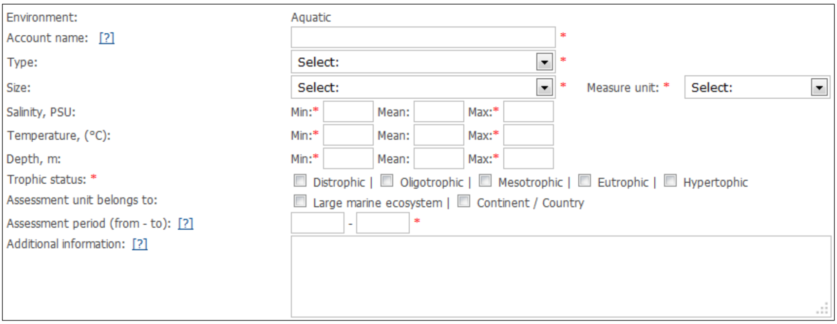


Figure 20. Print screen from BINPAS showing extraction of the Assessment unit account registration form. There data on its type, size, basic environmental information and geographical location are stored.

In BINPAS an assessment unit account contains multiple fields (Figure 20). There AU name is unique name of an area for which the assessment is made (e.g. Saginaw Bay, Lake Plateliai, Venice Lagoon, Danube River from Melk to Vienna). Depending on selected environment (aquatic or terrestrial) its type is considered (Table 8).

Table 8. Environmental types used in BINPAS.

Aquatic environment	Terrestrial environment
Marine region	Artificial habitats
Open sea	Coastal habitats
Coastal waters	Cultivated habitats
Coastal lagoon	Grasslands
Inland region	Heathland / scrub / tundra
Canal	Inland shoreline
Lake	Inland un-vegetated habitats
River	Mires, bogs and fens
Pond	Woodland

AU size is selected from predefined values ((0-1], (1-10], (10-100], (100-1000], etc.), where measure unit is identified separately. For lakes, expansive areas square kilometers and for rivers, coastlines linear kilometers should be used.

For aquatic environment data on salinity (measured in *psu*), temperature (measured in *°C*) and depth (measured in *m*) are collected. Such attributes contain minimum, mean and maximum values of basic environmental data. For further analysis data on AU trophic status are included. There are multiple selections from dystrophic (e.g. high levels of humic matter; brown- or tea-coloured waters), oligotrophic (e.g. low primary productivity; low nutrient content; low algal production; very clear waters; high drinking-water quality), mesotrophic (e.g. intermediate level of productivity, which is greater than oligotrophic ones; commonly clear water with beds of submerged aquatic plants; medium levels of nutrients), eutrophic (e.g. high biological productivity; excessive nutrients; dominating aquatic plants or algae; algae blooms) and hypertrophic (e.g. nutrient-rich; frequent and severe nuisance algal blooms; low transparency; large amounts of plants, fish and other animals) values.

The geographical affiliation of assessment unit is organized by selecting continent and one or more countries. If the assessment unit is within marine waters, one or several LME should be identified. Additionally, more precise location is defined using Google maps (Figure 21), where coordinates (latitude and longitude) are stored. There aquatic and terrestrial environments are identified by different markers – "A" and "T" respectively.

Figure 21. Locations of aquatic assessment units registered in BINPAS.

Step 2: Defining an assessment period

The assessment period (AP) is determined for every assessment performed to store the years over which the study took place as well as additional information on basic aquatic environmental parameters (trophic status, salinity and temperature for aquatic environments), which could change over AP. The earliest AP usually is considered to be the baseline from which further assessments can be compared.

Normally, an AP would range from one year to a decade and may be decided by, for example, an appropriate national or international authority. To simplify data input, initial period is defined during AU registration form (Figure 20). The separate data entrance form is used if the assessment is performed for additional AP (Figure 19, Step 2-A). Step 2-B is undertaken if an additional species is to be assessed in the same AU for the same AP.

Step 3: Selecting a species

A species can be selected from a drop-down menu, which is organized according to its taxonomic position (Figure 19, Step 3). The list of scientific names is prepared according to the selected taxonomic group, but if there is uncertainty about a species taxonomic position, a special tool, which retrieves species list according to a pattern of one or more criteria (species scientific name, genus or taxon position), should be used.

BINPAS has been grouped taxonomic data in the following way: aquatic inland, aquatic marine, terrestrial plants, terrestrial vertebrates, terrestrial invertebrates and fungi. Subdivisions then classified these to an organism group, for example, birds, fish, insects, reptiles or amphibians, plants and fungi. According to the type of assessment unit, corresponding lists are generated.

Step 4: Determining the abundance and distribution range

Each selected species is ranked according to its abundance and distribution range (ADR). There values for abundance and distribution levels are identified separately and the total ADR automatically derived (Figure 22).

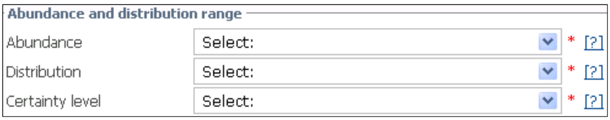


Figure 22. Print screen from BINPAS showing a species ADR determination.

The abundance (i.e. numbers per area unit, biomass or percentage of coverage) of an alien species is ranked in relation to the abundance of the relevant ecological group (i.e. phytoplankton, macroalgae, zoobenthos or fish), to which the alien species belongs to. It is classified as “Low” if the species makes up only a small part of a relevant community (e.g. an introduced fish forms only a small portion of the entire fish community), “Moderate” if the species makes-up less than a half of abundance of a community, and “High” when it exceeds half (i.e. quantitatively dominates in the invaded community). The distribution scale ranges from one locality to all localities.

Each contributor is required to provide a level of certainty for an ADR entry. It is expressed as either “High”, if the species ADR was studied in the entire assessment unit, “Medium”, if the species ADR was studied in a part of the assessment unit and extrapolated to the entire system by expert judgment, and “Low”, if it is known that species is recorded in the in the assessment unit, but the species ADR was not studied. Also underlying information, which includes explanations and relevant references, should be provided. At this step a species assessment account is being created, which stores an assessment for a species in a defined AU for a certain AP.

Step 5: Assessing a species impacts on communities, habitats and ecosystem functioning

At this step the magnitude of the species impacts is to be assessed and information added to a species assessment account. The known impacts are evaluated by selecting one of five classes (none, weak, moderate, strong and massive) for three categories (native species and communities, habitats and ecosystem functioning).

For each impact assessment certainty level is identified. In this case “High” means that impacts were documented by field and/or experimental studies for the given assessment unit; “Medium” – the impacts were documented by field and/or experimental studies for a part of the assessment unit and extrapolated to the entire system by expert judgment; and “Low” – the impacts were not documented neither by field nor by experimental studies, expert knowledge of the species impact based on data from studies made elsewhere was applied. Moreover, free text comments and documented references should be provided for each impact.

Both quantitative and qualitative data may be used for impact assessment. According to BPL algorithm the system will exclude unlikely impact magnitude levels by the obtained ADR class. For example, if an alien species occurs in low numbers in one or several localities (ADR class "A") it is unlikely that it may cause moderate-to-massive impacts on communities, habitats or ecosystem functioning in the entire assessment unit. In opposite, the ADR class "E" (an alien species occurs in high numbers in all localities) implies that at least some changes in the community structure and/or ecosystem performance, hence, situation of "No measurable impact" is very unlikely (Olenin et al., 2007).

Step 6: Acquiring a species biopollution level

After assessments of a species impacts are completed, BINPAS calculates the BPL for a species in a particular AU for the defined AP according to the entries made, thus completing a species assessment account. The calculation for a single species is based on the ADR class and the highest impact level in any one category (C, H or E). This procedure starts if at least one impact on communities, habitats or ecosystems is known.

Step 7: Acquiring the overall biopollution level for an assessment unit

The assessment procedure is repeated for each NIS known in a particular assessment unit, if the overall BPL is to be calculated for an AU for the defined AP. It is determined according to the greatest BPL in at least one species assessment account. For example, if for a 3-year assessment period, BPL was 1 (weak biopollution) for 10 NIS, but at least one species once showed BPL 3 (strong biopollution), then the overall BPL for the assessment unit would be 3.

Data search and presentation

BINPAS provides reports in lists, free text, tables and diagrams. Dynamically generated charts show species number for each ADR, BPL class, and number for impacts on communities, habitats, ecosystem functioning (Figure 23).

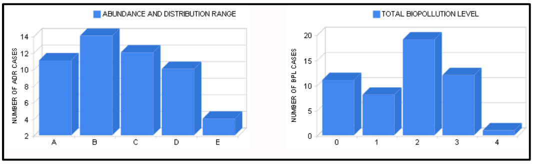


Figure 23. Example of a graphical report on BPL assessment for the dinoflagellate *Prorocentrum minimum* in the Baltic Sea. Account author: A. Zaiko.

Published reports are available to all users of the system. Also they can be exported and downloaded. All underlying data for the above studies may be stored in a form of the biopollution assessment bulletin, an automatically generated PDF file available online (Figure 24). The bulletin displays basic information on the assessment unit accounts, assessment periods and data contributors. If references and comments were provided, they are also shown.

Biopollution Assessment Bulletin

Biopollution assessment bulletin was generated by Biopollution Assessment System web site (http://www.corpi.ku.lt/databases/binpas)

Assessment unit: Lough Derg; Country/State/Territory: Europe, Northern Europe, Ireland, Ireland

Assessment period: 1998-2006

Author: Dan Minchin

Date of submission: 2010-06-26

Species	ADR	Imp.Com.	Imp.Com.Conf.L	Imp.Hab.	Imp.Hab.Conf.L	Imp.Eco.	Imp.Eco.Conf.L	Tot.BPL
Gammarus tigrinus	A	0 (C0)	Low	0 (H0)	Low	0 (E0)	Low	0
Chelicorophium curvispinum	A	0 (C0)	Low	0 (H0)	Low	0 (E0)	Low	0
Elodea canadensis	B	1 (C1)	Low	2 (H2)	Low	1 (E1)	Low	2
Lemna minuta	C	1 (C0)	Low	1 (H0)	Low	1 (E0)	Low	1
Hottonia palustris	C	1 (C1)	Medium	2 (H1)	Medium	N/A (N)	Medium	2
Anguillicola crassus	A	1 (C1)	Medium	0 (H0)	Medium	N/A (N)	Medium	1
Elodea nuttallii	D	2 (C2)	Medium	3 (H3)	Medium	3 (E3)	Medium	3
Dreissena polymorpha	D	2 (C2)	Medium	3 (H2)	Medium	2 (E2)	Medium	3

Figure 24. Example of Biopollution Assessment Bulletin. Basic and underlying information is displayed. Assessment author: D. Minchin.

4.3. Integration of AquaNIS and BINPAS

As it was stated above, an ideal NIS information system should contain data on species taxonomy, their biological traits, introduction events, pathways and vectors of introduction, and should be equipped with a tool to assess the magnitude of bioinvasion impacts (see Section 2.2.3 for details). Both information systems, AquaNIS and BINPAS, have been developed individually for different purposes (see Section 4.1.1 and 4.2.1), however, integration of these two tools was considered essential for enhancement of their functionality and applicability for solving research and managerial questions.

AquaNIS and BINPAS have been integrated using the following data blocks: "Species taxonomy", "Geography" and "Users" (Figure 25).

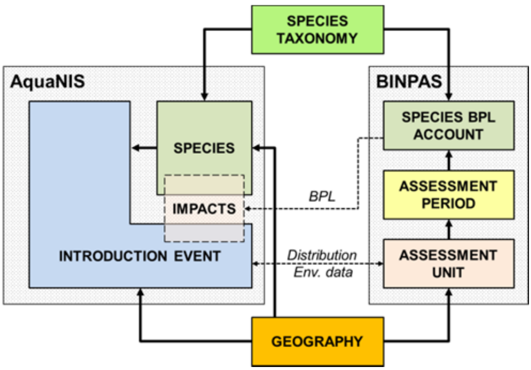


Figure 25. Integration concept of AquaNIS and BINPAS. External data blocks ("Species taxonomy", "Geography") provide validated data on species taxonomy and geography. "Users" data block is not visualized as it is connected to all blocks and contains data on registered users.

In the integrated system the "Species taxonomy" data block stores validated data on species names, full taxonomy, authority, etc. Data in this block is regularly updated using the external taxonomic data providers (currently WoRMS). The "Geography" data block is described in Section 4.1.2.3, additionally including lists of big lakes, large rivers as they are used in BINPAS.

"Users" data block stores information about registered users to identify users and their control levels of access according to whether it is being used by guests, demo-users, contributors, editors or administrators (Table 9). All these roles are created by applying different combination of rights, such as: Add records, Delete own/other record, Edit own/other record, Manage access to own/other records, Manage own/other user accounts, Manage supporting data, Manage user activities, Manage user requests, Confirm records, etc.

Table 9. Matrix of user rights.

	Review	Enter	Edit own	Edit all	Confirm
Administrator	+	+	+	+	+
Editor	+	+	+	+	+
Contributor	+	+	+		
Demo user	+	+	+		
Guest	+				

Guests are only permitted to browse the databases, and those records which are available for public domain. Also there is the demo implementation of BINPAS, where users can try-out its capabilities without any risk of compromising the underlying database. They use a separate "sandbox" application, where data are stored temporarily (http://www.corpi.ku.lt/databases/binpas/demo).

Records can be added and managed by registered users only (administrators, contributors and editors). An administrator has control over the entire system, and is responsible for technical issues, the administration of user rights and for the development of AquaNIS and BINPAS.

Data flow between AquaNIS and BINPAS ensures integrated search (see Section 4.1.3) and continuous data entrance procedure. AquaNIS provides data on NIS introduction events, for which biopollution levels should be assessed. If BPL was assessed for a species within assessment unit, where introduction event is not recorded in AquaNIS, its primary data are automatically filled in AquaNIS. As a result, BPL indexes are transmitted from "Species account" (BINPAS) to "Impacts" data block (AquaNIS). Also environment data (salinity, temperature ranges, etc.) are exchanged between both systems.

Although AquaNIS and BINPAS have been integrated using direct queries, the availability to produce data for other systems (e.g. GBIF network, Lane and Edwards, 2007) is very important for ensuring effective data search and being maximally reusable in a variety of contexts. There the Darwin Core standard, which provides stable semantic definitions, was applied. An archive of several files is arranged in a star-like manner (Figure 26). The structure of stored data is described within the *meta.xml* file. The core file *species.csv* contains list of registered NIS and additional taxonomic information. Other files are extensions, which include data with species unique identifier. These and other Darwin Core archives can be easily prepared according to requests from potential data receivers.

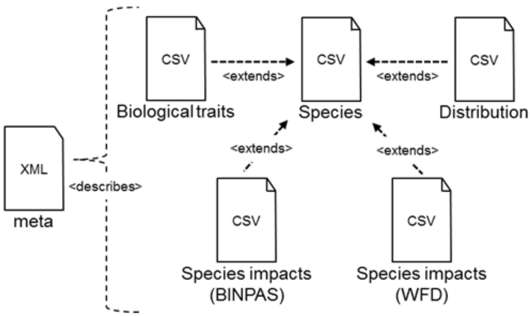


Figure 26. Contents of the archive for data exchange. The file "meta.xml" describes the content and structure of files, while CSV files contain full datasets.

4.4. Application of AquaNIS

4.4.1. Recent data availability check

AquaNIS contains multiple attributes within "Species" and "Introduction event" data blocks (see Section 4.1.2 for details), which are saturated with data at different extent. The system is equipped with data entering function, where almost all attributes can be selected as "Not entered", meaning that its value is not considered at current moment, and as "Unknown", if there are no studies or evidences for the exact data entry. In other cases a value must be selected, checked from listed ones, or entered as free text.

AquaNIS has an opportunity to check data completeness for "Species" (Table 10) and "Introduction event" (Table 11) data blocks identifying records where values of attributes are not entered. Also it is designed to detect other data features, e.g. wide date periods, broad salinity ranges, etc. Results for "Introduction event" data block may be organized according to the responsible contributors. As it may be seen (Table 10, Table 11) the recent data completeness for most attributes doesn't exceed 30%. However, it is important that such data availability check is recommended before making any analysis.

Totally AquaNIS contains data on 1232 aquatic NIS and cryptogenic species introductions into 52 recipient regions in Europe. The registered NIS represent a broad spectrum of free-living and parasitic multicellular and unicellular organisms including 34 phyla, 68 classes, 187 orders, 518 families and 863 genera. These numbers are revised with the inclusion of newly recorded NIS and their spread into new regions, with changes to nomenclature and taxonomy. Because of the dynamic nature of the database, the species numbers, figures and all other calculated outputs are relative, reflecting actual situation of present knowledge.

Table 10. Data coverage of "Species" data block (March, 2013, at total 1232 records).

Attribute	Not entered, %	Unknown, %	Known, %
Association with vessel vectors	83	8	9
Bioaccumulation association	100	0	0
Characteristic feeding method	80	0	20
Developmental trait	85	0	15
Habitat modifying ability potential	92	1	7
Life form	78	0	22
Mobility	79	0	21
Molecular information	84	-	16
Native origin	84	2	14
Reproductive frequency	85	1	14
Reproductive type	82	0	18
Salinity tolerance range	90	-	10
Sociability	83	0	17
Sub-species level	97	1	2
Toxicity	90	2	8

Table 11. Data coverage of "Introduction event" data block (March, 2013, at total 3596 records).

Attribute	Not entered, %	Unknown, %	Known, %
Date of first record	-	15	85
Habitat type	82	0	18
Migration pattern	97	1	2
Pathway / Vector	13	69	18
Population status	69	4	27
Recipient region (LME sub-region)	7	-	93
Reproductive duration	98	1	1
Reproductive seasonality	98	1	1
Salinity range	95	-	5
Source region	70	13	17
Species status	6	-	94
Temperature range	99	-	1
Wave exposure	84	0	16
Zonation	78	-	22

4.4.2. Temporal trends of non-indigenous and cryptogenic species at pan-European and regional scales

Pan-European scale: The CBD indicator "Cumulative number of alien species in Europe since 1900th"

The CBD indicator "Cumulative number of alien species in Europe since 1900th" (see Section 3.6 for details) was calculated for decadal scales. Only metazoan marine NIS in EEA-Europe (EU27 plus Iceland, Norway, Turkey and the former Yugoslavia countries) introduced after 1900 were included (Figure 27). To avoid species duplicate counting only the first occurrences in the entire EEA-Europe region were counted.

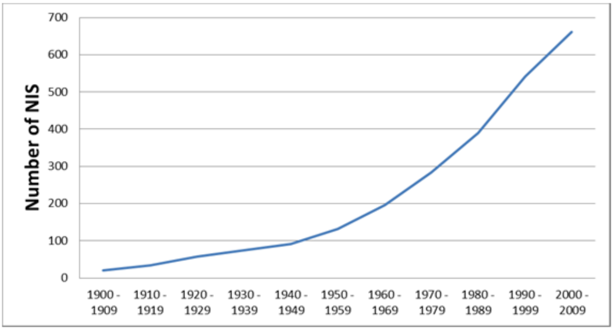


Figure 27. Cumulative number of metazoan marine NIS in EEA-Europe introduced after 1900 based on AquaNIS data.

This pressure indicator showed increasing trend in terms of introduction of new NIS. Also it suggested that efforts to stem biodiversity loss have clearly been inadequate with increasing pressures and it is highly unlikely that the CBD "2010 target" (CBD, 2010) has been met.

Regional scale: Temporal trends of NIS and cryptogenic species in the Baltic and North seas

The CBD indicator was also applied at the regional scale. It was noticed, that there is a difference in the cumulative number of NIS recorded in the two adjacent regional seas of the northern Europe, the North and Baltic seas. The North Sea hosts a higher number of NIS over time since 1900; however the increase pattern is rather similar (Figure 28).

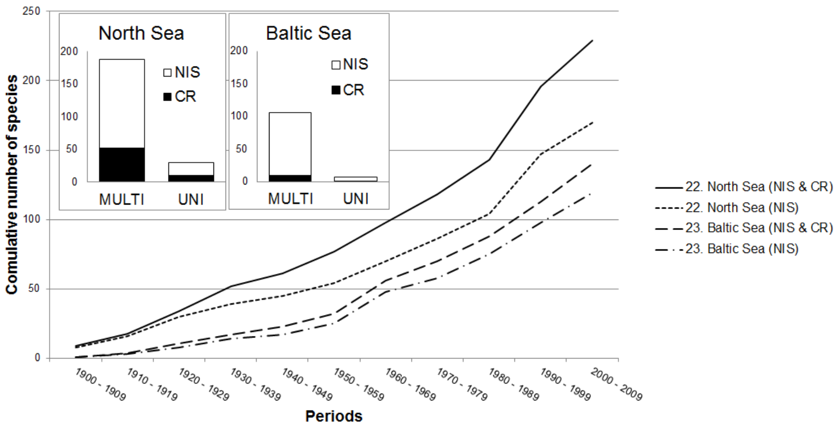


Figure 28. Cumulative numbers of non-indigenous (NIS) and cryptogenic species (CR) occurring in the North and Baltic seas since 1900. Inset: the difference in the numbers of metazoan (MULTI) and unicellular (UNI) NIS and cryptogenic species for both seas.

The numbers of cryptogenic species demonstrated a similar pattern. However, the reported first observations for almost 10% of NIS and cryptogenic species still remain unknown. Using the AquaNIS database it was revealed that the ratio of cryptogenic species to recognised NIS is lower in the Baltic Sea (10% *versus* 38% in the North Sea), probably because it is easier to notice a new species arrival in a naturally species low diversity system like the Baltic Sea. This ratio is likely to decline in the future due to increased attention which is being paid to new NIS detection in recent years. For example, in the Baltic forty-one new NIS arrivals were recorded since 1970, but only one of these, the dinoflagellate *Prorocentrum minimum*, is indicated as a cryptogenic species (Olenina et al., 2010).

The number of unicellular NIS and CR is relatively low comparing with the metazoan numbers for both seas. The former are almost certainly underestimated, due to their small size and difficulties in recognition and identification. For example, in AquaNIS out of 107 species, which were recorded in association with vessel vectors, 100 are metazoans and only 7 are unicellular organisms.

4.4.3. Metazoan NIS richness in different sea regions of Europe by phyla

The total number of multicellular NIS recorded in the Baltic Sea, Celtic Sea - Biscay Shelf, Mediterranean Sea and North Sea is 868. The Mediterranean Sea hosts the highest number of NIS (63%), followed by the Celtic Sea (14%).

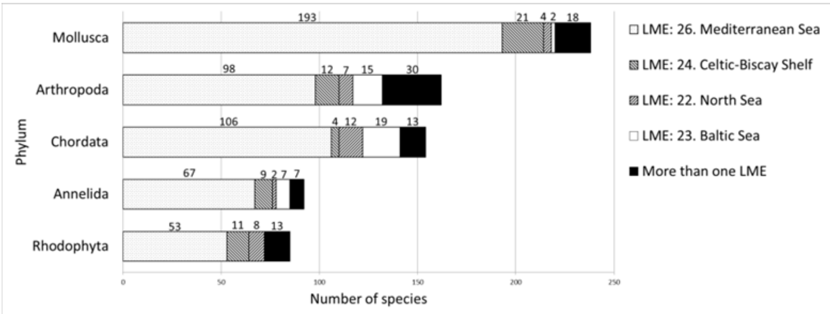


Figure 29. NIS richness of the largest phyla in four LMEs. The last segment on the right-hand side (coloured black) indicates the number of NIS found within more than one of these four LMEs. Only multicellular organisms are included.

Five phyla – Mollusca, Arthropoda, Chordata, Annelida and Rhodophyta – are by far the most species rich taxonomic groups in these LMEs, comprising 28, 19, 18, 11 and 10 % of the total NIS, respectively (Figure 29). The highest number (30) of widespread species, which have spread in more than one LME belongs to arthropods, were mostly represented by Malacostraca (71%).

4.4.4. New NIS arrivals

NIS are recorded continuously with approximately two new records annually within each LME during the past decade (Table 12, Figure 28). The eastern part of the Mediterranean Sea (LME 26E) is an exception with an average of a dozen NIS records annually, mostly due to invasion through the Suez Canal (Galil, 2009).

For the risk assessment of newly arrived NIS, it would be important to know the "unique" NIS, i.e. those NIS only found to occur in a single LME region, as these species may spread further to neighbouring seas. Such list of "unique" NIS for each recipient region may be retrieved using the Comparison of Search Results tool.

The greatest number (80%) of newly recorded NIS since the beginning of this century has been reported from the eastern Mediterranean Sea (Table 12). For other parts of the Mediterranean Sea (Adriatic Sea and the western Mediterranean) the numbers are basically lower.

Table 12. Metazoan NIS recorded in the Large Marine Ecosystems of Europe and neighbouring regions since the XXI century. The Mediterranean Sea sub-regions: western, eastern parts and Adriatic.

LME / LME sub-region	Number of NIS	Number of NIS since 2000	Number of unique NIS since 2000
22. North Sea	143	22	8 (36%)
23. Baltic Sea	97	22	8 (36%)
24. Celtic – Biscay Shelf	145	22	10 (45%)
26. Mediterranean Sea	656	241	219 (91%)
26W. W. Med. Sea	174	32	6 (19%)
26A. Adriatic Sea	107	28	5 (18%)
26E. E. Med. Sea	546	209	168 (80%)

4.4.5. The most invasive species

The most invasive species were identified on the basis of the number of invaded LMEs using the following search criteria: "multicellular" AND "non-indigenous" AND "introduced since 1900" (Figure 30; see Appendix). It is obvious that most NIS occur only in one of the LMEs, while two LMEs are invaded by 56 species and even less species spread into 3 or more LMEs.

The importance of introduction pathways was identified for the most invasive species, which have spread in three or more LMEs, as follows: natural spread from neighbouring countries (141 introduction events), vessels (116) and culture activities (99). It is interesting to note that most of LMEs invaded by the same species are located next door to each other, for example, Celtic-Biscay Shelf and North Sea; Baltic and North Seas; Mediterranean and Black Seas.

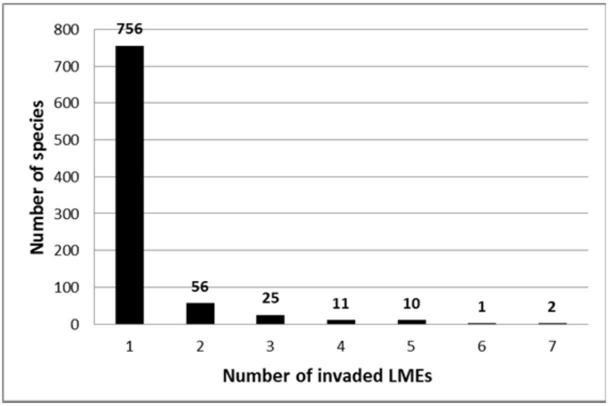


Figure 30. The number of NIS occurring in one or more LMEs.

4.4.6. Bio-geographical regions by NIS compositions

NIS compositions can significantly identify bio-geographical regions according to species presence or absence in each country (Paini et al., 2011).

List of 1003 NIS registered in 40 countries within five LME's (Baltic Sea, Celtic-Biscay Shelf, Iberian Coastal, Mediterranean Sea and North Sea) was used for cluster analysis. The null hypothesis, that there are no differences between clusters at 10, 20, and 30% similarity levels, was tested using ANOSIM procedure. According to generated *p* and *R* values, the null hypothesis was rejected identifying that clusters were significantly separated at all three tested levels.

At the highest hierarchical level, the Large Marine Ecosystems (LMEs) of the Atlantic coast of Europe together with the Baltic Sea are clearly separated from the Mediterranean Sea (Figure 31, Figure 32).

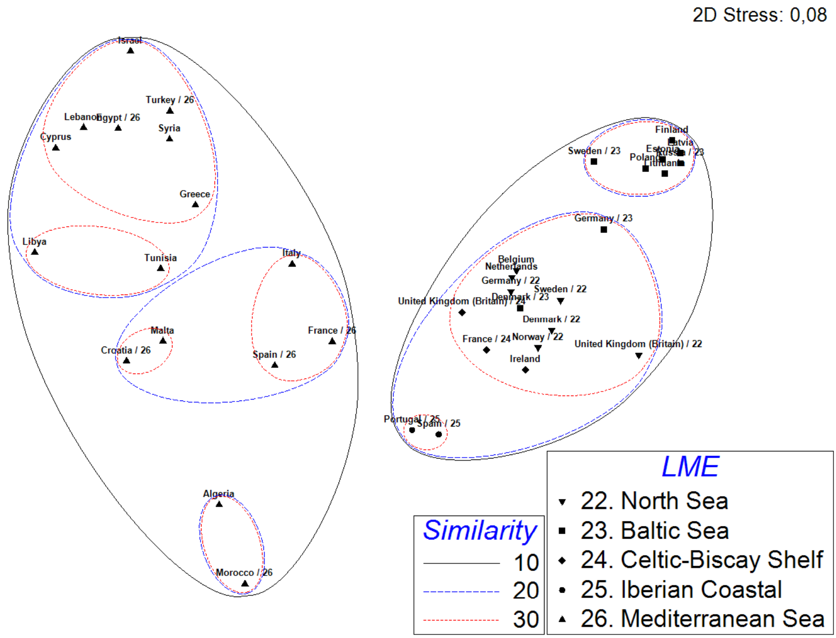


Figure 31. MDS plot of similarities between country regions according to registered NIS.

Figure 32. Clusters' visualisation on the map. B – inner part of the Baltic Sea; C - European counties' costs on the northern and central parts of the Mediterranean Sea; E - the eastern Mediterranean part including European, Asian and African coastal countries; G - western African coast in the vicinity of Gibraltar strait; I - Iberian coast countries; N – Celtic Sea - Biscay Shelf and the North Sea.

However at the lower levels, the Mediterranean Sea region is divided into three areas: 1) the eastern Mediterranean part including European, Asian and African coastal countries; 2) European counties' costs on the northern and central parts of the Sea and 3) western African coast in the vicinity of Gibraltar strait. Also, the second major cluster splits into three separate groups at the lower hierarchical level: 1) Iberian coast countries, 2) Celtic Sea – Biscay Shelf and the North Sea, and 3) inner part of the Baltic Sea. It is important to note that the western part of the Baltic (Denmark/Baltic and Germany/Baltic) belong to the second cluster, while the inner Baltic joins coasts of Estonia, Finland, Latvia, Lithuania, Poland, Russia/Kaliningrad, Russia/Sankt-Petersburg and Sweden.

This analysis showed that regional environmental conditions play an important role in shaping the composition of NIS. In the case of the Baltic Sea, the areas with salinity less than 10 psu (i.e. inner Baltic) form one distinct group, while its western part belongs to the Atlantic coast of Europe (i.e. the marine part). The higher water temperature may be an important factor distinguishing the Iberian coast countries from the rest of Atlantic coast.

Another important reason for different NIS composition may be the specific pathways operating in certain regions. This is especially obvious in the case of Eastern Mediterranean, where the Suez Canal is causing the intensive flow of NIS, so called Eritrean invasion (Galil, 2009).

Interesting aspect, resulting from the analysis, is that the clusters based on presence/absence of NIS do not fully coincide with conventional division into LMEs (Figure 11), for example, the LME22 and LME24 are clustered together including the western part of the Baltic Sea, and the rest of LME23 forms a separate group.

4.4.7. Taxonomic distinctness analysis of the Baltic Sea region

The above cluster analysis roughly indicated the major bio-geographical regions of European and neighbouring seas based on NIS composition. For further more detailed analysis it is feasible to apply the taxonomic distinctness approach based on Linnaean classification. Such approach was applied to a separate group of the LME23 Baltic Sea (inner Baltic countries), revealed in the previous analysis, in order to measure the degree to which NIS composition in each country differs from the regional composition.

According to the taxonomic distinctness analysis the master species composition was retrieved from NIS list recorded in all inner Baltic countries up to 2013. Figure 33 shows that Estonia, Finland, Lithuania, Poland and Russia fall inside 95% funnel. The null hypothesis that there are no differences of NIS registered in Sweden and Latvia comparing with the master NIS composition was rejected with *p* value 1.0 and 0.8%. It means that NIS compositions in these two countries have specific features, which differ them from the regional NIS composition.

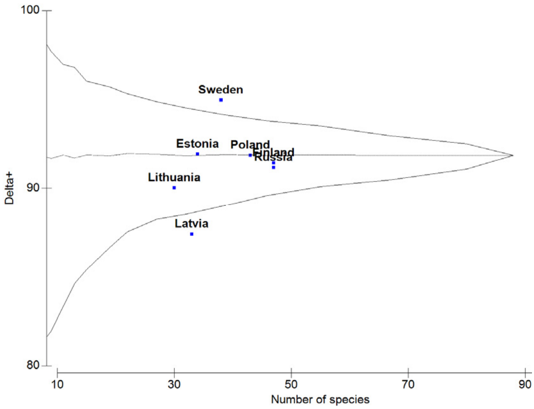


Figure 33. Taxonomic distinctness analysis of the Baltic Sea region up to 2013. Latvia and Sweden fall outside 95% funnel (*p* value 0.8 and 1.0% accordingly).

Taxonomic distinctness analysis also may be applied to examine how differences in NIS composition are changing over time. This implies that the higher are the differences between countries the less is homogeneity of the region in general.

Two periods were selected for the analysis (1950-1979 and 1980-2009), because data for these periods is more reliable than in previous ones. The master NIS composition lists were formed accordingly for each period.

The number of introduced NIS was 38 in the period 1950-1979 and 48 in 1980-2009. The analysis showed that in the first period four countries differ from the master list, i.e. fall outside 95% funnel, while in the second period there are no such countries (Figure 34).

The analysis indicated the increasing homogeneity of the region in terms of NIS compositions in recent decades. Such increase may be caused by intensification of sea trade, movements of good and people within and outside the region. On another hand, this implies the policy message that the management of invasive species should be harmonized on the level of the entire region.

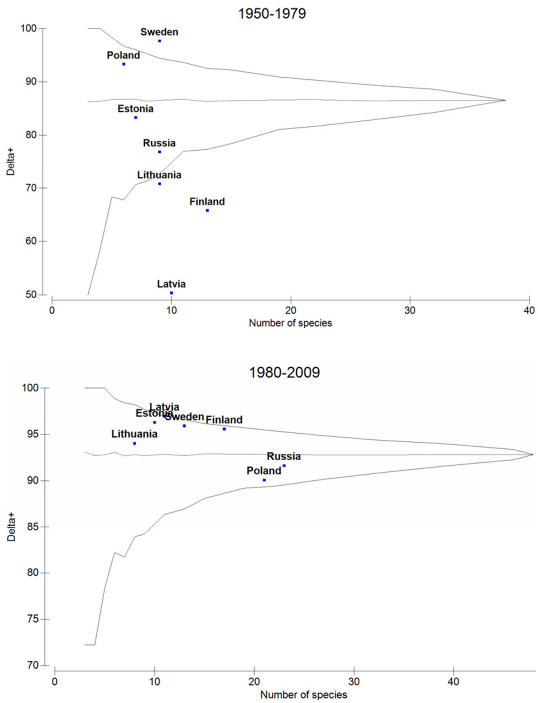


Figure 34. Taxonomic distinctness analysis of the Baltic Sea region for periods 1950-1979 (above) and 1980-2009 (bellow). In period 1950-1979 Finland, Latvia, Sweden and Lithuania falls outside 95% funnel plot (*p* values 0.2, 0.2, 0.2 and 0.24 accordingly).

4.5. Application of BINPAS

4.5.1. Visibility and recent data coverage

Since the open access version of BINPAS was launched in spring 2010, connections from more than 540 different cities have been registered (Figure 35). Although no analysis was performed, it seems that these connections largely correspond to locations of major research centers involved in biological invasion studies. This shows a great interest in the BPL method application. However, it is obvious that not all these connections have resulted in BPL assessments for aquatic or terrestrial environments. The total number of assessments is less than it can be expected from the browsers activity. In principle, this may mean that either the system (BINPAS) is not yet sufficiently user friendly, or the BPL method is too time consuming or problematic (e.g. Wittfoth and Zettler, 2013). Ideas on how to improve both BINPAS and BPL method itself are discussed in Section 4.5.3.

Figure 35. Geographical locations of connections to BINPAS summarized at cities level from 2010-01-01 to 2013-01-01.

At present (March, 2013) BINPAS contains more than 500 species assessments for aquatic and terrestrial environments (Table 13). Unique number of assessed species is around 190 within 230 different localities. For the further analysis only species assessments for aquatic environment will be considered.

Table 13. BINPAS data coverage (March, 2013).

	Aquatic environment	Terrestrial environment	Total
Species	148	38	186
Species assessments	396	116	512
Assessment units	132	97	229
Assessment periods	176	107	283

The data availability analysis showed that ADR assessments have been performed in all cases, as it should be according to the methodology (Olenin et al., 2007). However, assessment of the impacts varies in great extent: around 30% of impacts on habitats and ecosystem functioning are unknown, while impacts on communities for 90% cases are provided.

For the further BINPAS data analysis, 207 instances with known nominal values on ADR, IOC (impacts on communities), IOH (impacts on habitats), IOE (impacts on ecosystem functioning) and BPL for aquatic environment were extracted. Other supporting data (e.g. tropical status, temperature range) for most records were missing and not involved in the analysis.

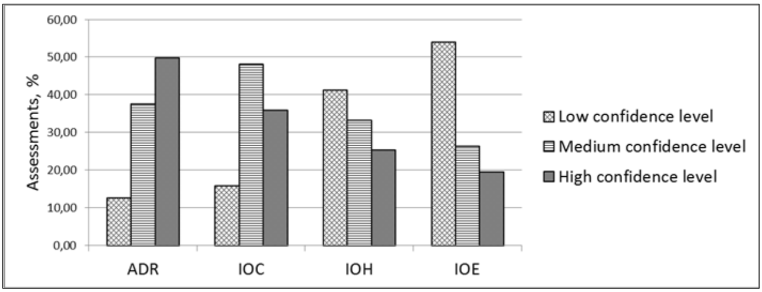


Figure 36. The ratio between impact assessments and confidence levels for ADR, IOC, IOH and IOE.

Figure 36 shows, that confidence levels for ADR assessment mostly are High (50%) and Medium (38%), and for IOC – Medium (48%) and High (36%). However, confidence levels for IOH and IOE are ranged as Low (41% and 54%), Medium (33% and 26%) and High (26% and 20%) respectively. It means that assessing impacts on habitats and ecosystem functioning are problematic tasks, because most confidence levels were identified to be Low.

4.5.2. Biopollution level assessment scenarios

The discussed biopollution level assessment steps (see Section 4.2.3 for details) have been realized in a computerized system so that existing information of an alien species together with known impacts can be converted into uniform biopollution measurement units. BINPAS allows eight possible assessment scenarios, where Scenario I is compulsory, while others may be applied depending on the purpose of a study:

I. One NIS in one AU for one AP.
II. One NIS in one AU for several AP.
III. One NIS in several AU for one AP.
IV. One NIS in several AU for several AP.
V. Several NIS in one AU for one AP.
VI. Several NIS in one AU for several AP.
VII. Several NIS in several AU for one AP.
VIII. Several NIS in several AU for several AP.

According to the needed scenario, the proper sequence of steps (1 – 7) is preceded. For example, **Scenario VI** was used to illustrate the application of the BPL method in the initial publication (Olenin et al., 2007), where all NIS found in the Curonian Lagoon were assessed for two periods.

In the first study performed by **Scenario IV**, the most detailed information was gained for the dinoflagellate *Prorocentrum minimum* (Olenina et al., 2010). The research was based on joint 1980–2008

HELCOM and national phytoplankton monitoring program for eleven sub-regions of the Baltic Sea. Samples were analyzed and species identified by phytoplankton experts, using the mandatory Baltic Sea international monitoring methods. Data on abundance, relative biomass and distribution of species at monitoring stations within a sub-region was used. Different yearly periods were studied in the analysis (1980–1984, 1985–1989, …, 2005–2008) in each sub-region as such periodicity usually is being used to report the environmental status in the HELCOM area. Each assessment period in a particular sub-region was considered as a case study; and in total there were 66 such case studies (6 time periods for 11 assessment units). The only truly invasive species, that has increased in abundance and has spread over large areas during the study period, was the dinoflagellate *Prorocentrum minimum*. The assessment of bioinvasion impact was based on the analysis of this species alone.

A further study was aimed by **Scenario VII** at the overall biological invasion impact assessment of multicellular plant and animal species in the same large marine geographical region – the Baltic Sea, as defined by the Helsinki Commission (Zaiko et al., 2011). The assessment was performed for nine Baltic sub-regions for one assessment period (1990–2010). There only established multicellular aquatic species with documented impacts listed in the Baltic Sea Alien Species Database (http://www.corpi.ku.lt/nemo) were assessed (seaweeds, polychaetes, crustaceans, mollusks, etc.). Among the 119 registered alien species in the Baltic Sea, 79 were defined as established (sustaining self-reproducing populations) and 43 of the last were those having any documented ecological impact. Others were considered as having no or negligible effect and consequently BPL = 0. During this study, it was identified that the highest biopollution (BPL = 3, strong impact) occurs in coastal lagoons, inlets and gulfs, and the moderate biopollution (BPL = 2) in the open sea areas (Figure 37).

Figure 37. Assessment units defined in the Baltic Sea for the biopollution assessment procedure and their biopollution level (lighter regions are those with BPL = 2, darker – with BPL = 3). Numbers in parentheses indicate impacting alien species within an assessment unit (with BPL > 0) (Zaiko et al., 2011).

Another local scale study was performed in terrestrial environment according to **Scenario III** (Butautytė-Skyrienė et al., 2011). BINPAS was applied to assess the impact of the muskrat *Ondatra zibethicus* on native species, communities, habitats, and ecosystems. This species was introduced to Lithuania from Archangelsk in 1954 and from Kazakhstan in 1956, and have spread over almost all rivers of Lithuania. The authors examined abundance and distribution range and impacts of the muskrat *Ondatra zibethicus* in 11 forest enterprises in Lithuania for the assessment period 1986-2011. The abundance of muskrats was assessed by the numbers of individuals, lodges and burrows. It was identified that distribution and abundance during the last 10 years has been highly variable – ADR ranges from class A to E. The impact of muskrats varied between different regions of Lithuania. Generalization of performed assessments of distribution of invasive muskrats and impacts on ecosystems was published in additional paper (Skyrienė and Paulauskas, 2012).

4.5.3. Patterns extraction from accumulated data

Estimation of BPL values from ADR classes

The exploration of biopollution assessment data showed that most common ADR values are C and A (32% and 26% of cases, respectively); and common BPL values are equal to 2 (Moderate impact) and 1 (Weak impact) for 41% and 27% of assessments (Table 14).

Table 14. The frequency of BPL values according to ADR.

ADR	BPL0	BPL1	BPL2	BPL3	BPL4	
A	47%	53%	-	-	-	26%
B	-	59%	41%	-	-	19%
C	-	7%	90%	3%	-	32%
D	-	-	22%	78%	-	17%
E	-	-	-	67%	33%	6%
	12%	27%	41%	18%	2%	

From Table 14 it is seen that BPL values vary between 0 (47%) and 1 (53%) for ADR class A. For ADR class B, BPL may be equal to 1 (59%) or 2 (41 %); and for class C – 1 (7%), 2 (90%) or 3 (3%). If ADR classes are D or E, BPL values are accordingly 2, 3 (22%, 78%) or 3, 4 (67%, 33%).

In general, higher ADR values result in upper BPL assessments. This implies that BPL value may be predicted by ADR class without going into details about impacts on community, habitats and ecosystem functioning. For example, if ADR of a NIS is assessed as class C, there is a 90% probability that BPL value will be 2 (Moderate biopollution).

Simplification of biopollution level assessment

As it was shown above (see Section 4.5.1), a NIS impacts on habitats and ecosystem functioning are often unknown. Moreover, for known ones, low or medium confidence levels usually are indicated.

To simplify the BPL assessment procedure classification methods were applied. Different algorithms were compared by a number of incorrectly classified instances (named as *error*). An algorithm was considered as a suitable, if the error rate was lower than 4%. In all cases cross-validation with 10 folds was used.

Classification results using full dataset (ADR, IOC, IOH and IOE attributes) for BPL values are displayed in Table 15. Although the error rates were low (average 2%), the result of all correctly classified instances was not achieved.

Table 15. Classification results for BPL values.

Classifier name	Error, %
Bayes Network Classifier	0.97
Decorate	2.42
FT	2.42
Logistic Regression	1.93
Multilayer Perceptron	1.45
NNGE	2.42
Radial basis function network	1.45
Random Committee	2.90
Random forest trees	3.38
Rotation Forest	1.45
Average	2.08

For the further analysis, the relationship between the response variable (BPL) and four explanatory variables (ADR, IOC, IOH and IOE) was explored using classification trees. The generated classification rules at the first division split the dataset according to the value of ADR (A, B *vs* C, D, E), and after that IOC is considered (Table 16). Application of these rules on the current dataset resulted in 2% incorrectly classified instances.

Table 16. Classification rules for assessment of a species biopollution level (BPL): ADR – abundance and distribution range; IOC – impact on communities; IOH – impact on habitats; IOE – impact on ecosystem functioning.

<table>
<tr><th>1</th><th>2</th><th>3</th><th>#</th></tr>
<tr><th>ADR</th><th>IOC</th><th>IOE</th><th>BPL</th></tr>
<tr><td rowspan="3">A</td><td rowspan="2">C0</td><td>E0</td><td>0</td></tr>
<tr><td>E1</td><td rowspan="2">1</td></tr>
<tr><td rowspan="2">C1</td><td>E0, E1</td></tr>
<tr><td rowspan="3">B</td><td>E2</td><td>2</td></tr>
<tr><td>C0</td><td></td><td>1</td></tr>
<tr><td>C2</td><td></td><td>2</td></tr>
</table>

<table>
<tr><th>1</th><th>2</th><th>3</th><th>#</th></tr>
<tr><th>ADR</th><th>IOC</th><th>IOH</th><th>BPL</th></tr>
<tr><td rowspan="2">C</td><td rowspan="3">C1, C2</td><td>< H3</td><td>2</td></tr>
<tr><td rowspan="2">H3</td><td rowspan="3">3</td></tr>
<tr><td rowspan="4">D</td></tr>
<tr><td rowspan="2">C1, C2</td><td>H2</td></tr>
<tr><td>H1</td><td>2</td></tr>
<tr><td>C3</td><td></td><td rowspan="2">3</td></tr>
<tr><td>E</td><td>C2, C3</td><td></td></tr>
</table>

According to classification rules, the dataset was divided into two groups: 92 instances with ADR class A or B (further called as ADR_AB), and 115 instances with ADR class C, D or E (further called as ADR_CDE). For these datasets the same classifiers were run, and the average number of incorrectly classified instances was identified to be lower: 1.2 and 1.4% (Table 17). Some algorithms (Bayes Network Classifier, Radial basis function network, and NNGE for ADR_AB dataset; Multilayer Perceptron for ADR_CDE dataset) arranged assignments without errors. In general, Bayes Network Classifier performed as best in all cases (Table 15 and Table 17).

Additionally error rates were tested for such situations, when NIS impacts on habitats or ecosystem functioning are unknown ("Error no IOH" and "Error no IOE" respectively), while data on ADR and IOC are available. For ADR_AB dataset with missing data on IOE the error rate for assessing BPL value was higher comparing when IOH was unknown (6% and 3%). For ADR_CDE dataset, missing data on IOH generated higher probability of error than missing data on IOE (8% and 2%). From Table 15, Table 16 and Table 17 it is seen that for ADR_AB it is important to have data on IOC and IOE, while for

ADR_CDE – IOC and IOH. In other case, accuracy of BPL assignment decreases.

Table 17. Error rates (%) from classification results using ADR_AB and ADR_CDE datasets. There "Error no IOH" identifies error rates for situations with unknown IOH, while data on IOC and IOE are available.

Classifier name	ADR_AB dataset			ADR_CDE dataset		
	Error	**Error no IOH**	**Error no IOE**	**Error**	**Error no IOH**	**Error no IOE**
Bayes Network Classifier	0.00	2.17	4.35	0.87	6.09	0.87
Decorate	1.09	2.17	6.52	2.61	8.70	2.61
FT	2.17	2.17	6.52	0.87	7.83	0.87
Logistic Regression	2.17	2.17	6.52	1.74	7.83	2.61
Multilayer Perceptron	2.17	3.26	6.52	0.00	8.70	0.87
NNGE	0.00	4.35	5.44	1.74	8.70	1.74
Radial basis function network	0.00	2.17	5.44	0.87	8.70	0.87
Random Committee	1.09	2.17	6.52	1.74	9.56	2.61
Random forest trees	2.17	2.17	6.52	1.74	11.30	1.74
Rotation Forest	1.09	3.26	5.44	1.74	6.96	1.74
Average	1.20	2.61	5.98	1.39	8.44	1.65

In general, it was shown that more than 10 classifiers may be used to generate classification rules for determining BPL values with low error rates (2% average error rate). The output of these classifiers is simple IF-THEN statements (e.g. NNGE classifier) or more difficult modules (e.g. Multilayer Perceptron, Bayes Network Classifier). Error rates may be minimized if different algorithms are applied according to ADR values (ADR_AB and ADR_CDE). Also these modules showed that missing data on impacts may involve 2-8% error rate for assigning BPL values.

The possibility to reduce the BPL scale

Cluster analysis was performed to generate clusters on primary dataset (207 instances with known ADR, IOC, IOH and IOE values) to minimize the total number of BPL classes. A number of clusters was selected by cross validation method (CLOPE, Cobweb, DBSCAN and EM algorithms), or was defined to be 4 (Farthest First, Hierarchical Clusterer and kMeans algorithms). CLOPE suggested that the number of clusters should be equal to 40, while Cobweb, and DBSCAN – 92, and 14 respectively. EM and Farthest First generated 4 clusters. Classifiers resulted that the average error rate using clusters from EM was lower than from Farther First (1% and 2%).

Table 18. The output generated by Prism and LMT classifiers. The first is based on logical conditions, while another on weights.

Prism rules, 1.45% errors	**LMT, 1.93% errors**
If ioc = C1 and ioe = E1 then cluster0 If ioc = C1 and adr = A then cluster0 If ioc = C1 and adr = C and ioe = E0 then cluster0 If ioe = E1 and ioc = C0 and adr = C then cluster0	Class 0 : -3.19 + [ioc=C1] * 4.59 + [ioh=H1] * 0.91 + [ioe=E2] * -3.29 + [ioe=E1] * 1.69
If ioe = E3 then cluster1 If ioc = C3 and ioe = E2 then cluster1 If ioh = H3 and adr = D then cluster1 If ioc = C3 and ioh = H2 then cluster1 If adr = D and ioc = C1 and ioh = H2 then cluster1	Class 1 : -0.81 + [adr=C] * -1.29 + [adr=D] * 1.14 + [ioc=C3] * 2.81 + [ioh=H1] * -1.81 + [ioh=H3] * 1.67 + [ioe=E1] * -1.33 + [ioe=E3] * 2.16
If ioc = C2 and adr = C then cluster2 If ioc = C2 and ioh = H1 then cluster2 If ioc = C2 and adr = B then cluster2 If ioc = C2 and adr = E then cluster2 If ioc = C2 and ioe = E1 then cluster2 If ioc = C2 and ioh = H2 and ioe = E2 then cluster2 If ioe = E2 and adr = C then cluster2 If adr = D and ioh = H1 and ioc = C3 then cluster2 If ioe = E2 and adr = B then cluster2	Class 2 : -0.54 + [adr=A] * -1.27 + [ioc=C2] * 3.59 + [ioc=C0] * -1.42 + [ioh=H1] * 1.08 + [ioh=H3] * -1.32 + [ioe=E3] * -1.41
If ioc = C0 and ioh = H0 then cluster3 If ioc = C0 and ioe = E0 then cluster3 If ioe = E0 and ioh = H0 then cluster3 If ioe = E0 and ioh = H2 and ioc = C1 then cluster3	Class 3 : -2.78 + [adr=C] * -1.12 + [ioc=C0] * 1.5 + [ioh=H1] * -0.96 + [ioh=H0] * 2.98 + [ioe=E0] * 4.04

For these clusters Radial basis function network and Decorate classifiers performed without errors, while DTNB and NB Tree resulted with some errors – 0.5%. The available output of Prism and LMT classifiers is presented in Table 18.

Clusterization results showed that biopollution assessment method may be revised and simplified to 4 BPL classes. However, discussions and experiments should be taken by researchers and experts of biological invasion science for further interpretation of generated clusters and theirs applicability in real situations.

5. DISCUSSION

5.1. Added heuristic value of the NIS information systems

There is an increasing demand to develop information systems, storing data on NIS and its introduction events, taking into account the effects of NIS as an inevitable component for impacts on ecosystem goods and services (e.g. Oguz et al., 2008). Moreover, nowadays management of requests from researchers and policy makers is not reasonable without verified, updated datasets, provided by global repositories, which are gathering and managing data from unlimited number of sources, adding additional value to compare and analyze datasets in regional and global scales, and finding various patterns. However, currently only few cases were published, where information on NIS has been involved in ecosystem assessments (e.g. Wallentinus and Nyberg, 2007; Ojaveer and Eero, 2011; Wittfoth and Zettler, 2013).

During the last decade NIS databases are increasingly being used for research. Below there are listed how information derived from these online sources was used in the peer-reviewed literature (Olenin et al., 2013). Additionally, they are stressed by existing opportunities of AquaNIS and BINPAS to cover such needs.

a) To aid the compilation of NIS lists for specific areas (e.g. Gollasch and Nehring, 2006; Zaiko et al., 2007; Westphal et al., 2008; Galil, 2009, 2012; Occhipinti-Ambrogi et al., 2011).
The analysis of data storage in AquaNIS showed that the amount of NIS and cryptogenic species is increasing over time in regional and global context; either there is a substantial difference in the cumulative number of NIS occurring in different LMEs (Figure 27, Figure 28). Also it was noticed, that it is highly unlikely that the CDB “2010 target” (CBD, 2010) has been met. The origins of some NIS and cryptogenic species may be deduced using molecular tools, where AquaNIS allows gathering molecular availability and other relevant data.

b) To prioritize the most invasive NIS (e.g. Cambray, 2003; Olenina et al., 2010; Savini et al., 2010; Zaiko et al., 2011).

In this study the most invasive multicellular NIS were identified on the basis of the number of invaded LMEs. Only 13 such species have spread into 5 or more LMEs (Figure 30). The most common phyla within these species are Mollusca and Arthropoda.

c) To define the major pathways and vectors, responsible for NIS introductions, and assess their risk (e.g. Gollasch, 2006; Hulme et al., 2008; Marchini et al., 2008; Savini et al., 2008; Minchin et al., 2009; Galil, 2012).

In this study the importance of introduction pathways was identified for the most invasive species, which have spread in 3 or more LMEs, as follows: natural spread from neighbouring countries, vessels and culture activities. This information, comprised with the most common vectors, can be easily generated for different regions and used as primary criteria for NIS risk assessment.

d) To identify, quantify and summarize the ecological impacts of specific taxa (e.g. Butchart, 2008; Vilà et al., 2009; Occhipinti-Ambrogi and Galil, 2010).

Inclusion of impacts assessed by AquaNIS or BINPAS allows prioritizing species according to impacts on WDF ecological quality elements, impacts on MSFD qualitative descriptors, species toxicity, or BPL values on communities, habitats and ecosystem functioning.

AquaNIS differs from the existing NIS information sources in its structure, functionality, maintenance principles and output potential for end-users. The system is integrated with BINPAS, which facilitates usage of biopollution assessment method. The developed features and functionality provide an added heuristic value to multiple data, allow covering various requests, comprehensive analyses, such as identifying most invasive species, important introduction pathways, predominant taxonomic groups, analyzing species traits and ecological preferences, assessing the risks posed by alien species on economies and ecosystem functioning, providing recommendations for management measures, etc.

In general, BINPAS is the first information service on NIS, which integrates both data submitted by experts and active rule-sets to produce ecologically meaningful assessment of bioinvasion impacts. The integrated BPL approach enables objective comparison between diverse invaded ecosystems, monitoring of the level of bioinvasion impacts in the same ecosystem over different assessment periods, and evaluation of the same invader impacts in various regions even if a limited amount of information is available (e.g. Olenin et al., 2007; Olenina et al., 2010; Zaiko et al., 2011; Butautytė-Skyrienė et al., 2011). BINPAS evaluates impact assessment for aquatic and terrestrial environments. This is the main reason why AquaNIS and BINPAS are developed as separate systems.

At present moment (March, 2013) the insufficient completeness of the datasets within AquaNIS and BINPAS is limiting research and management activities (see Section 4.4.1). However, even with such level of data completeness, interesting conclusions were derived on bio-geographical implications based on Pan-European analysis of NIS compositions, increasing homogeneity of the Baltic Sea region in terms of NIS introductions, differences in invasion trends in various seas and among NIS and cryptogenic species, etc. Performed studies indicated, that the developed NIS information systems can cover a wide variety of case studies on bioinvasion management, which are currently being taken by exploring several cross referenced NIS databases.

5.2. Data quality control

The lack of reliable information on NIS is a common problem in biological invasion studies (Vilà et al., 2009; Olenin et al., 2010, 2011). To ensure the proper data coverage, AquaNIS and BINPAS foresee the use a variety of data sources, ranging from "grey literature", official environmental reports, and unpublished field survey data to materials published in peer-reviewed journals.

User support is a weak point in many projects (Bach et al., 2011; Narščius et al., 2012). To cover this lack, AquaNIS involved documented frequently asked questions, online help by e-mails, video and voice conferences, or step by step screenshots, prepared video records, feedback, workshops and phone calls during the development. It was noticed that the communication between users and database managers becomes difficult when technical problems are discussed (e.g. Bach et al., 2011). However, most researchers and end-users pay attention not only on the number of existing records, but also on data quality and maintenance.

Data quality is an important issue resulting the acceptance and long-term usage of databases, and it is essential to improve the data quality management as a crucial task (Bach et al., 2011; Hulme and Weser, 2011; Olenin et al., 2013). Below are presented methods, used to ensure data quality of the developed NIS information systems. However, experience showed that quality control is very time consuming, expensive and complex process.

Manual data quality check. The NIS information systems involve the editorial board consisting of the chief editor and managing editors (Figure 38).

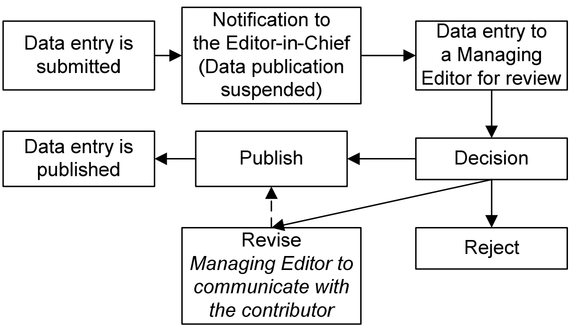


Figure 38. AquaNIS and BINPAS data quality assurance procedure.

Any expert can be granted editor rights under mutual agreement with the chief editor. Editors are responsible for verifying entries according to their taxonomic or geographical region experience (e.g. series of assessment units) and for supervising new data entries and for those that have been already entered.

Application of automatic or semi-automatic routines. AquaNIS and BINPAS automatically check data completeness while a record is being registered. They consider if confidence levels, comments and references are provided, avoid duplicates, initiate to upload supporting files, exclude theoretically not possible choices, forbid not permitted actions, etc. After a new or modified entrance has passed automatic check procedure, it is forwarded to the editorial board.

Relation on a reputation of data providers. This approach was used in special cases, such as, the primary data importation from datasets provided by database developers, or external sources, e.g. WoRMS.

To ensure data quality assurance a quorum of volunteers is needed in order to achieve the long-term maintenance and reliability of the database, because it requires frequent updating and corrections, and thus becomes a living instrument (Bach et al., 2011). One of the options for AquaNIS would be to involve relevant expert groups, such as the Working Group on Introductions and Transfers of Marine Organisms (WGITMO) of the International Council for the Exploration of the Sea (ICES), and the Working Group on Ballast and Other Ship Vectors (WGBOSV), which has ICES, the Intergovernmental Oceanographic Commission (IOC), and the International Maritime Organization (IMO) as parent organizations.

5.3. Long-term maintenance

For most projects funding is a serious and crucial problem (Bastow and Leonelli, 2010; Enke et al., 2012; Olenin et al., 2013). Many NIS databases have been developed within the framework of short term national or international projects, where funding was available only for their duration, while repositories should offer a service on a long-term scale. Emphasis has often been placed upon elaborate design involving sophisticated web-technologies (Olenin et al., 2013).

During development stage greater investments usually are made in implementation of the database structure, than populating it with data and elaborating ecologically meaningful data output functions. As a result, the most productive period normally occurs near the end of a project with several scientific publications and reports for managers.

However, the usefulness of the developed information system depends not only on deliverables obtained by the project, but whether the system is being maintained after the project termination. Unfortunately, it is easier to obtain funding for developing a new database, than for improving existing ones and for database improvement and maintenance (Simpson et al., 2006; Olenin et al., 2013). If continuing financial support is not secured, the database functioning can be maintained only by volunteering, while the usefulness diminishes overtime. Moreover, it may disseminate outdated and misleading information. In ideal situations the funding of a database should be secured at a basic level for technical support and for data management. The benefits of a "living" database grow as it accumulates updated entries by various contributors (Figure 39).

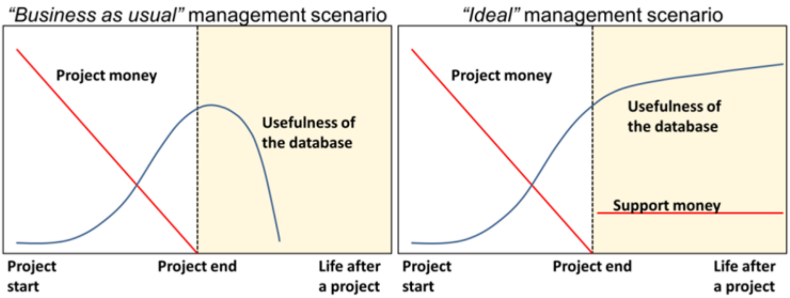


Figure 39. Usefulness of the database under "Business as usual" and "Ideal" scenarios.

In this study it was demonstrated that AquaNIS and BINPAS have to be useful for managers and researchers. It is important that the maintenance funding should be provided by governmental and/or intergovernmental institutions. However, currently it is not clear how support will be ensured after current is terminated, but some new project initiatives are being developed (e.g. Joint European Database on Invasive Alien Species in Baltic, Black and North Sea Regions).

5.4. Future perspectives

In the last decades biodiversity research has grown to include various types of data (McCann, 2000). There is a tendency that only on taxonomy based datasets are incomplete for sufficient analysis, and additional information is needed to be integrated for decision making (e.g. Steele et al., 2007, Moustahfid et al., 2010, Certain et al., 2011). However, traditionally most databases take just one aspect of ecological information without managing the full suite of data (Edwards et al., 2000; Peterson et al., 2010).

It is suggested here, that the developed NIS information systems (AquaNIS and BINPAS) not only allow integrative analyses including data from different fields, but also provide tools, services and opportunities to make analysis at different scales (see Section 4.4 and 4.5 for details). Moreover, AquaNIS is capable to integrate data from different existing blocks to derive information to support management. For example, mapping of the spread of NIS between and within LMEs, identifying principle pathways and vectors within countries and LMEs, defining the most invasive species as well as "next pests" (sensu Hayes et al., 2005) to provide target lists for monitoring. Supplementary information on biological traits, environmental ranges, and habitat preferences of NIS will serve to increase the accuracy of selecting NIS of management consequence. Bioinvasion impact assessment tool provides an opportunity to assess the magnitude of environmental impacts caused by NIS in recipient regions or localities within regions.

Although the NIS information systems fulfils most of the raised requirements (see Section 2.2.3 for details, Table 19) to get results with a minimum of efforts, but end-users have to think about what they are doing to understand what the results mean.

Table 19. Proposed requirements (see Section 2.2.3) and their implementation in AquaNIS.

Requirement	Implementation in the NIS information systems
Clear definitions for important elements of attributes	Definitions of biological traits, pathways, vectors, impacts and other related information on species and introduction events are made available for all attributes and presented in pop-up form.
Interactive tools for data analysis	Search results can be compared using the Search comparison tool and displayed in a form of matrixes with the set of attributes selected by a user.
Information on data origin and version control	References and Comments fields for attributes; names and contact details of contributors and editors; version control for data changes.
Feedback from users	Users have an opportunity to provide their opinion and suggestions on quality of data, functionality of the systems or propose changes to records using build-in interactive tools.
Usage of common standards for data	Data are formulized in order to unify the biological information domain. Species names, geographical data, data exchange are organized according to existing standards.
Easy and user-friendly access to the information, retrieving it by browsing or searching.	Access is based on precise and dependable control, which ensures that users see and manage only the data they are entitled to have access to. Data are presented as lists or retrieved in matrixes, which can be exported into spread sheet files.
Technical support and specialized courses at usage	Technical support is ensured using online chats, video tutorial on how to enter/search/modify data. Specialized courses were arranged via Skype conferences or at meetings to acquaint them with functionality of the system at various stages of its development.

Requirement	Implementation in the NIS information systems
Data quality, sustainability and the protection of intellectual property assurance	The content is reviewed and managed by regional and taxonomical researchers. Data quality is ensured by editorial board, feedback and changes by non-registered users. Data are made publicly accessible after the authors' assent. The copyright notices are available online.
Provide various services and ensure that the functionality of the system is of sufficient value to users	Electronic catalogue; output availability for print media and for the Internet; analysis on the large amount of data; search system; comparison of search results; automatic quality checks; editorial board; chat forums; annotations; Web 2.0 functionalities; information about changes on records.
Link together digital data from disparate sources to answer complex questions	Currently AquaNIS and BINPAS are integrated with WoRMS to reuse information without duplication of efforts. Advanced search is performed so that simple inputs lead to rich outputs of interconnected data.

The plans of future work are further development of the NIS information systems to serve numerous purposes, such as: to help the industry and local authorities in managing the problem of NIS; to prevent the spread of NIS; to provide tools for the risk assessment; to share data more effectively; to be a base for education and public awareness; to gather data using automated web search tools. It is intended that these new changes and intensive communication with researchers will cover other world regions, but will not change the current usage of the system. Also they will enable research of invasiveness determinants (van Kleunen et al., 2010) through traits comparisons between invasive and non-invasive NIS, successful invasive NIS and common/expanding native species, and non-invasive NIS and rare/declining natives. The systems are free for all specialists working in the field of invasion biology, managers dealing with biodiversity loss caused by biological invasions and other professionals who are willing to share their own data on impacts and distributions of alien species.

6. CONCLUSIONS

1) According to data accumulated in the information system, there are 1232 aquatic non-indigenous (NIS) and cryptogenic species known to be introduced into 52 recipient regions in Europe and neighboring areas. The recorded species are represented by a broad spectrum of free-living and parasitic multicellular and unicellular organisms including 34 phyla, 68 classes, 187 orders, 515 families and 851 genera.

2) The CBD (Convention on Biological Diversity) indicator “Cumulative number of alien species in Europe since 1900th” shows the sharp increase in recent decades, meaning the increasing pressure resulting from biological invasions. Inclusion of cryptogenic species in the cumulative number does not change the shape of the curve essentially. On the regional level, the ratio of cryptogenic species to recognised NIS is lower in the Baltic Sea than in the North Sea (10% *versus* 38%, respectively). Most probably this is because it is easier to notice a new species arrival in a naturally species poor system like the Baltic than in the North Sea.

3) At the pan-European scale, specific environmental conditions and pathways are important factors shaping NIS compositions in the regional European seas. At the highest hierarchical level, the Large Marine Ecosystems (LMEs) of the Atlantic coast of Europe together with the Baltic Sea are clearly separated from the Mediterranean Sea.

4) The taxonomic distinctness analysis of the Baltic Sea region at the level of countries’ coasts indicates the increasing homogeneity of the region in terms of NIS compositions in two recent decades. This implies the policy message that the management of invasive species should be harmonized on the level of the entire region.

5) Application of the developed NIS information systems for specific bioinvasion case studies showed, that these systems can be used as multipurpose tools, providing a flexible platform for bioinvasion data storage, extraction and analysis. Machine learning methods can facilitate optimization of the BPL method.

7. REFERENCES

Alimov A.F., Bogutskaya N.G., 2004. Biological Invasions in Aquatic and Terrestrial Ecosystems. KMK Scientific Press, Moscow, 436 pp (in Russian with English summary).

Andow D.A., 2005. Characterising ecological risks of introductions and invasions. Pages 84-103 in Mooney H.A., Mack R.N., McNeely J.A., Neville L.E., Schei P.J., Waage J.K., editors. Invasive alien species: a new synthesis. Island Press, Washington, D.C.

Appeltans W., Bouchet P., Boxshall G.A., Fauchald K., Gordon D.P., Hoeksema B.W., Poore G.C.B., Soest van R.W.M., Stöhr S., Walter T.C., Costello M.J. (Eds.), 2010. World Register of Marine Species. http://www.marinespecies.org (accessed 2013-03-27).

Arbačiauskas K., Semenchenko V., Grabowski M., Leuven R.S.E.W., Paunović M., Son M.O., Csányi B., Gumuliauskaitė S., Konopacka A., Nehring S., van der Velde G., Vezhnovetz V., Panov V.E., 2008. Assessment of biocontamination of benthic macroinvertebrate communities in European inland waterways. Aquatic Invasions, 3: pp. 211-230.

Bach K., Schäfer D., Enke N., Seeger N., Gemeinholzer B., Bendix J., 2011. A comparative evaluation of technical solutions for long-term data repositories in integrative biodiversity research. Ecological Informatics 11, pp. 16-24.

Baker R., Cannon R., Bartlett P., Barker I., 2005. Novel strategies for assessing and managing the risks posed by invasive alien species to global crop production and biodiversity. Annals of Applied Biology 146, pp. 177-191.

Baltic Sea Alien Species Database, 2013. Olenin S., Daunys D., Leppäkoski E., Zaiko A., (editors). http://www.corpi.ku.lt/nemo (accessed 2013-03-27).

Barry C. and Lang M., 2001. A Survey of Multimedia and Web Development Techniques and Methodology Usage, IEEE Multi Media, April-June 2001, Vol. 8 No. 2, pp. 52-61.

Bastow R., Leonelli S., 2010. Sustainable digital infrastructure. EMBO Reports 11, pp. 730-734.

Baziukaite D., Naršcius A., 2010. The modified Renyi-ClipX conceptual clustering algorithm. Applied computer science, Institute for Environment, Engineering, Economics and Applied Mathematics. ISSN: 1792-4863, ISBN: 978-960-474-225-7, pp. 396-400.

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

Berendsohn W.G., Döring M., Gebhardt M., Güntsch A., 2002. BioCase – A biological collection access service for Europe. Alliance News.

Bertzky M., Kleemann S., 2009. Multilevel discrepancies with sharing data on protected areas: what we have and what we need for the global village. Environmental Management, 90, pp. 8-24.

Blum S., Vieglais D., Schwartz P.J., 2001. DiGIR – Distributed Generic Information Retrieval. 2001 TDWG Meeting, Sydney, Australia.

Bowker G.C., 2000. Biodiversity data diversity. Social Studies of Science 30, pp. 643–683.

Branton R., Bajona L., Bond S., Kennedy M., Ricard D., 2007. Methods for Standardizing, Validating and Enriching Taxonomic Metadata. NAFO SCR Doc. 07/08, Serial No. N5349, pp. 8.

Butautytė-Skyrienė G., Paulauskas A., Ulevičius A., 2011. Assessment of invasive muskrat Ondatra zibethicus distribution and impacts on ecosystems. 8th European Vertebrate Pest Management Conference, DOI: 10.5073/jka.2011.432.014.

Butchart S.H.M., 2008. Red List Indices to measure the sustainability of species use and impacts of invasive alien species. Bird Conservation International 18(S1), pp. 245-262.

Cambray J.A., 2003. Impact on indigenous species biodiversity caused by the globalisation of alien recreational freshwater fisheries. Developments in Hydrobiology 171, pp. 217-230.

Campbell M.L., Galil B., Gollasch S., Occhipinti-Ambrogi A., 2008. Guidelines for controlling the vectors of introduction into the Mediterranean of non-indigenous species and invasive marine species. First GloBallast Regional Task Force Meeting Dubrovnik, Croatia, 11-12 September 2008. REMPEC/WG.29/INF.6 UNEP RAC-SPA Tunis, pp. 24.

Cardoso A.C., Free G., 2008. Incorporating invasive alien species into ecological assessment in the context of the Water Framework Directive. Aquatic Invasions 3 (4), pp. 361-366.

Carlton J.T., 1989. Man's role in changing the face of the ocean: biological invasions and implications for conservation of near-shore environments. Conservation Biology 3, pp. 265-273.

Carlton J.T., 2001. Introduced species in U.S. coastal waters: environmental impacts and management priorities. Pew Oceans Commission, Arlington. http://www.pewoceans.org/reports/introduced_species.pdf (accessed 2013-03-27).

Carlton J.T., 2009. Deep invasion ecology and the assembly of communities in historical time. In: Rilov G., Crooks J. (Eds.), Biological Invasions in marine ecosystems: Ecological, management and geographic perspectives. Ecological Studies 204. Springer, Heidelberg, Germany, pp. 16-56.

Carlton J.T., Geller J.B., 1993. Ecological roulette: the global transport of nonindigenous marine organisms. Science 261, pp. 78-82.

Carlton J.T., Ruiz G. M., 2005. Vector science and integrated vector management in bioinvasion ecology: conceptual frameworks. Pages 36-58 in Mooney H.A., Mack R.N., McNeely J.A., Neville L.E., Schei P.J., Waage J.K., editors. Invasive alien species: a new synthesis. Island Press, Washington, D.C.

Carlton, J.T., 2002. Bioinvasion ecology: assessing invasion impact and scale. In: Leppäkoski, E., Gollasch, S., Olenin, S. (Eds.), Invasive Aquatic Species of Europe – Distribution, Impacts and Management. Kluwer Academic Publications, Dordrecht, Boston, London, pp. 7-19.

Casal C.M.V., 2006. Global documentation of fish introductions: the growing crisis and recommendations for action. Biological Invasions 8, pp. 3-11.

CBD, 2002. Guiding principles for the prevention, introduction and mitigation of impacts of alien species that threaten ecosystems, habitats or species; annex to COP 6 decision VI/23 of the Convention on biological diversity. http://www.biodiv.org/decisions/default.asp?lg=0&dec=VI/23 (accessed 2013-03-27)

CBD, 2010. 2010 Biodiversity Indicators Partnership. 2010. Biodiversity indicators and the 2010 Target: Experiences and lessons learnt from the 2010 Biodiversity Indicators Partnership. Secretariat of the Convention on Biological Diversity, Montréal, Canada. Technical Series No. 53, pp. 196.

Certain G., Skarpaas O., Bjerke J.W., Framstad E., Lindholm M., Nilsen J.E., Norderhaug A., Oug E., Pedersen H.C., Schartau A.K., van der Meeren G.I., Aslaksen I., Engen S., Garnasjordet P.A., Kvaloy P., Lillegard M., Yoccoz N.G., Nybo S., 2011. The nature index: a general framework for synthesizing knowledge on the state of biodiversity. PLoS ONE 6 (4), 12.

Chew M.K., Laubichler M.D., 2003. Natural enemies – metaphor or misconception? Science, 301, pp. 52-54.

CIESM, the Mediterranean science commission. Atlas of Exotic Species in the Mediterranean, 2013. http://www.ciesm.org/online/atlas (accessed 2013-03-27).

Clarke K.R., Warwick R.M., 1998. A taxonomic distinctness index and its statistical properties. J Appl Ecol 35, pp. 523-531.

Colautti R. I., MacIsaac H. J., 2004. A neutral terminology to define 'invasive' species. Diversity and Distributions, 10, pp. 135-141.

Constable H., Guralnick R., Wieczorek J., Spencer C., Peterson A.T., 2010. VertNet: a new model for biodiversity data sharing. PLoS Biology 8 (2), e1000309. doi:10.1371/journal.pbio.1000309.

Costello M.J., 2009. Motivating online publication of data. BioScience, 59, pp. 418-427.

Costello M.J., Bouchet P., Boxshall G., Arvantidis C., Appeltans W., 2008. European Register of Marine Species. http://www.marbef.org/data/erms.php (accessed 2013-03-27).

Crooks J.A., 2009. The role of exotic marine ecosystem engineers. In: Rilov G., Crooks J.A. (Eds.), Biological Invasions in Marine Ecosystems: Ecological, Management, and Geographic Perspectives, Ecological Studies, vol. 204 (XXVI). Springer-Verlag, pp. 215-238.

DAISIE European Invasive Alien Species Gateway, 2008. http://www.europe-aliens.org (2008) (accessed 2013-03-27).

Davis M. A., Thompson K., 2000. Eight ways to be a colonizer; two ways to be an invader: A proposed nomenclature scheme for invasion ecology. Bulletin of the Ecological Society of America, 81, pp. 226-230.

Deacon J., 1995. Model-View-Controller (MVC) Architecture. https://techsimplified2.com/Uploads/Agendas/October28,2011.pdf (accessed 2013-03-27).

EASIN, 2013. The European Alien Species Information Network. http://easin.jrc.ec.europa.eu/ (accessed 2013-03-27).

EC, 2008. Commission of the European Communities. Brussels, 3.12.2008 SEC(2008) 2887 Commission staff working document. Annex to the Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions towards an EU strategy on invasive species impact assessment.

ECCHM, 2005. Biodiversity monitoring and indicators. SEBI2010 Expert Group 5 - invasive species. Convention on Biological Diversity, Montreal.

ECPGR, European Cooperative Programme for Plant Genetic Resources, 1980. http://www.ecpgr.cgiar.org (accessed 2013-03-27).

Edwards J.L., Lane M.A., Nielsen E.S., 2000. Interoperability of biodiversity databases: biodiversity information on every desktop. Science 289 (5488), pp. 2312-2314.

EEA, 2003. European Environment Agency. European environmental indicators. EEA, Copenhagen. http://themes.eea.europa.eu/indicators/ (accessed 2013-03-27).

Elliott M., 2003. Biological pollutants and biological pollution - an increasing cause for concern. Marine Pollution Bulletin 46, pp. 275-280.

EllisLab, 2002. CodeIgniter. http://ellislab.com/codeigniter (accessed 2013-03-27).

Elton C.S., 1958. The Ecology of Invasions by Animals and Plants. Methuen, pp.181.

Enke N., Thessen A., Bach K., Bendix J., Seeger B., Gemeinholzer B., 2012. The user's view on biodiversity data sharng – Investigating facts of acceptance and requirements to realize a sustainable use of research data. Ecological Informatics, 11, pp. 25-33.

Euler F., 1999. An objective indicator of functional integrity in avian communities. For Ecol Manage 115, pp. 221-229.

European Commission, 2001.Common implementation strategy for the Water Framework Directive (2000/60/EC). Strategic document as agreed by the Water Directors under Swedish presidency, 2 May 2001. European Commission, pp. 81.

European Union, 2008. Directive 2008/56/EC (Marine Strategy Framework Directive). Off J Eur. Union. L164: pp. 19-40.

Executive Order, 1999. Invasive Species. The White House, Washington, D.C.

Fegraus E.H., Andelman S., Jones M.B., Schildhauer M.P., 2005. Maximizing the value of ecological data with structured metadata: an introduction to Ecological Metadata Language (EML) and principles for metadata creation. Bulletin of the Ecological Society of America 86, pp. 158-168.

Frame M. T., Cotter G., Zolly L., Little J., 2002. Information Science and Technology Developments Within the National Biological Information Infrastructure. Science & Technology Libraries 23(4), pp. 59.

Frank E., Hall M., Trigg L., Holmes G., Witten I.H., 2004. Data mining in bioinformatics using Weka. Bioinformatics 20(15).

Galil B.S., 2009. Taking stock: inventory of alien species in the Mediterranean Sea. Biological Invasions 11(2), pp. 359-372.

Galil B.S., 2012. Truth and consequences: the bioinvasion of the Mediterranean Sea. Integrative Zoology, 7, pp. 301-313.

Galil B.S., Gollasch S., Minchin D., Olenin S., 2009. Alien marine biota of Europe. In: DAISIE. Handbook of Alien Species in Europe. Invading Nature – Springer Series in Invasion Ecology. Springer, Dordrecht, vol. 3, pp. 93-104.

GenBank, 2009. www.ncbi.nlm.nih.gov/genbank (accessed 2013-03-27).

Genovesi P., 2001. Guidelines for eradication of terrestrial vertebrates: a European contribution to the invasive alien species issue. Other Publications in Wildlife Management 24, pp. 28.

GISD, Global Invasive Species Directory, 2012. http://www.issg.org/database (accessed 2013-03-27).

GISIN (Global Invasive Species Information Network). List of Invasive Alien Species (IAS) Online Information Systems, 2008. http://www.gisin.org/WebContent/WS/GISIN/Documents/draftiasdbs.htm (accessed 2013-03-27).

Gollasch S., 2006. Overview on introduced aquatic species in European navigational and adjacent waters. Helgoland Marine Research 60(2), pp. 84-89.

Gollasch S., Macdonald E., Belson S., Botnen H., Christensen J.T., Hamer J.P., Houvenaghel G., Jelmert A., Lucas I., Masson D., McCollin T., Olenin S., Persson A., Wallentinus I., Wetsteyn L.P.M.J., Wittling T., 2002. Life in ballast tanks. In: Leppäkoski, E., Gollasch, S., Olenin, S. (Eds.), Invasive Aquatic Species of Europe – Distribution, Impact, and Management. Kluwer, Dordrecht, pp. 217-231.

Gollasch S., Nehring S., 2006. National checklist for aquatic alien species in Germany. Aquatic Invasions 1(4), pp. 245-269.

Grosholz E., 2002. Ecological and evolutionary consequences of coastal invasions. Trends in Ecology and Evolution, 17 (1), pp. 22-27.

Halevy A., 2005. Why your data won't mix. ACM Queue 3, pp. 50-58.

Heger T., Saul W.C., Trepl L., 2013. What biological invasions ‘are’ is a matter of perspective. Journal for Nature Conservation 21, pp. 93-96.

Hulme P.E., Bacher S., Kenis M., Klotz S., Kühn I., Minchin D., Nentwig W., Olenin S., Panov V., Pergl J., Pyšek P., Roques A., Sol D., Solarz W., Vilà M., 2008. Grasping at the routes of biological invasions: a framework for integrating pathways into policy. Journal of Applied Ecology 45(2), pp. 403-414.

Hulme P.E., Pyšek P., Nentwig W., Vila M., 2009a. Will threat of biological invasions unite the European Union? Science, 324, pp. 40-41.

Hulme P.E., Roy D.B., Cunha T., Larsson T.B., 2009b. A pan-European inventory of alien species: rationale, implementation and implications for managing biological invasions. Handbook of alien Species in Europe (ed. by DAISIE), pp. 1-14, Springer, Dordrecht.

Hulme P.E., Weser C., 2011. Mixed messages from multiple information sources on invasive species: a case of too much a good thing? Diversity and Distributions 17(6), pp. 1152-1160.

IABIN, Inter-American Biodiversity Information Network, 2002. http://iabin.databasin.org/ (accessed 2013-03-27).

IMO, 2004. International Convention for the Control and Management of Ships' Ballast Water and Sediments. International Maritime Organization. http://www.imo.org (accessed 2013-03-27).

IUCN, 2000. IUCN Guidelines for the Prevention of Biodiversity Loss Caused by Alien Invasive Species. Species Survival Commission (SSC). Invasive Species Specialist Group (ISSG), Auckland, New Zealand, pp.14 http://www.issg.org (accessed 2013-03-27).

IUCN, 2011. ww.iucn.org/about/union/secretariat/offices/iucnmed/ iucn_med_programme/species/invasive_species (accessed 2013-03-27).

Johnston E.L., Piola R.F., Clark G.F., 2009. The role of propagule pressure in invasion success. In: Rilov G., Crooks J.A. (Eds.), Biological Invasions in Marine Ecosystems: Ecological, Management, and Geographic Perspectives, Ecological Studies, vol. 204 (XXVI). Springer-Verlag, pp. 134-151.

Jones C.G., Lawton J.H, Shachak M., 1994. Organisms as ecosystem engineers. Oikos 69(3), pp. 373-386.

Karatayev A.Y., Burlakova L.E., Padilla D.K., Mastitsky S.E., Olenin S., 2009. Invaders are not a random selection of species. Biological Invasions 11 (9), pp. 2009–2019.

Kenis M., Bacher S., Baker R.H.A., Branquart E. et al., 2012. New protocols to assess the environmental impact of pests in the EPPO decision-support scheme for pest risk analysis. Bulletin OEPP/EPPO Bulletin, 42 (1), pp. 21.

Krutchen P., 2003. The Rational Unified Process: An Introduction. Addison-Wesley, Boston, MA.

Kuipers T., Hoeven J., 2009. Insight into digital preservation of research output in Europe. Survey Report FP7-2007-223758 PARSE.Insight. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf (accessed 2013-03-27).

Lane M.A., Edwards J.L., 2007. The Global Biodiversity Information Facility (GBIF). Biodiversity Databases: Techniques, Politics, and Applications (Systematics Association Special Volume). Gordon Curry (Editor), Chris Humphries (Editor). ISBN-10: 0415332907.

Leonard D.R.P., Clarke K.R., Somerfield P.J., Warwick R.M., 2005. The application of an indicator based on taxonomic distinctness for UK marine biodiversity assessments. Journal of Environmental Management 78, pp. 52-62.

Leppäkoski E., Gollasch S., Gruszka P., Ojaveer H., Olenin S., Panov V., 2002. The Baltic – a sea of invaders. Canadian Journal of Fisheries and Aquatic Sciences 59, pp. 1175-1188.

Leung K.M.Y., Dudgeon D., 2008. Ecological risk assessment and management of exotic organisms associated with aquaculture activities. In Bondad-Reantaso M.G., Arthur J.R., Subasinghe R.P. (eds). Understanding and applying risk analysis in aquaculture. FAO Fisheries and Aquaculture Technical Paper. No. 519. Rome, FAO. pp. 67-100.

Lynch C.A., 1991. The Z39.50 information retrieval protocol: an overview and status report. ACM SIGCOMM Computer Communication Review Homepage archive, Volume 21 Issue 1, Jan. 1991, pp. 58-70.

Maindonald J., Braun W.J., 2007. Data Analysis and Graphics Using R: An Example-Based Approach (Cambridge Series in Statistical and Probabilistic Mathematics). ISBN-10: 0521762936.

Malcolm J.S.,Walter G.B., 2007. Networking Biological Collections Databases: Building a European Infrastructure. Biodiversity Databases: Techniques, Politics, and Applications (Systematics Association Special Volume). Gordon Curry (Editor), Chris Humphries (Editor). ISBN-10: 0415332907.

Marchini A., Savini D., Occhipinti-Ambrogi A., 2008. Aquaculture and alien species in the EU-Med region. Part II: dispersal risk into the wild. Biologia Marina Mediterranea 15 (1), pp. 234-235.

Mastitsky S.E., Karatayev A.Y., Burlakova L.E., 2013. Aquatic invaders of Belarus: Alien species database. http:// www.aliensinbelarus.com (accessed 20.09.12).

McCann K.S., 2000. The diversity-stability debate. Nature 405 (6783), pp. 228-233.

McDonald A. and Welland R., 2001. Agile Web Engineering (AWE) Process, Department of Computing Science Technical Report R-2001-98, University of Glasgow, Scotland.

McDonald A. and Welland R., 2005. Agile web engineering (AWE) process: perceptions within a fortune 500 financial services company. Web Engineering, Vol. 4, No.4 pp. 283-312.

McGeoch M.A., Chown S.L., Kalwij J.M., 2006. A globalindicator for biological invasion. Conservation Biology, 20, pp. 1635-1646.

Meng C., 2005. Biological information standards. Bulletin of the American Society for Information Science and Technology 30, pp. 15-17.

Michener W.K., Allard S., Budden A., Cook R.B., Douglass K., Frame M., Kelling S., Koskela R., Tenopir C., Vieglais D.A., 2012. Participatory design of DataONE - enabling Cyberinfrastructure for the Biological and Environmental Sciences 11, pp. 5-15.

Minchin D., 2001. Introduction of exotic species. In: Steele J., Thorpe S., Turekian K. (Eds.), Encyclopedia of Ocean Sciences. Academic Press, London, pp. 877-889.

Minchin D., 2007. Rapid coastal survey for targeted alien species associated with floating pontoons in Ireland. Aquatic Invasions 2 (1), pp. 63-70.

Minchin D., Galil B.S., David M., Gollasch S., Olenin S., 2006. Overall introduction. In: Golash, S., Galil, B.S., Cohen, A.N. (Eds.), Bridging Divides. Maritime Canals as Invasion Corridors. Monographiae Biologicae 83. Springer, Dordrecht, pp. 1-4.

Minchin D., Gollasch S., Cohen A.N., Hewitt C.L., Olenin S., 2009.Characterizing vectors of marine invasions. In: Biological Invasions in Marine Ecosystems. Rilov G. and Crooks J.A. (eds) Springer, Ecological Studies 204, pp. 109-116.

Moustahfid H., Tyrrell M., Link J., Nye J., Smith B., Gamble, R., 2010. Functional feeding responses of piscivorous fishes from the northeast US continental shelf. Oecologia 163 (4), pp. 1059-1067.

MySQL A.B., 1997. MySQL Developer Zone. http://www.mysql.org/ (accessed 2013-03-27).

Narščius A., Olenin S., Zaiko A., Minchin D., 2012. Biological invasion impact assessment system: from idea to implementation. Ecological Informatics 7(1), pp. 46-51.

Nentwig W., Kühnel E., Bacher S., 2009. A Generic Impact-Scoring System Applied to 515 Alien Mammals in Europe. Conservation Biology 24(1), pp. 302-311.

NEOBIOTA, 2012. "Halting Biological Invasions in Europe: from Data to Decisions". 7th European Conference on Biological Invasions. Pontevedra, Spain, September 2012. http://neobiota2012.blogspot.com/

NISBASE, an International Nonindigenous Species Database Network, 2003. http://www.nisbase.org (accessed 2013-03-27).

NOAA, 2012. Large Marine Ecosystems of the World (online). http://www.lme.noaa.gov (accessed 2013-03-27).

NOBANIS, The European Network on Invasive Alien Species, 2000. http://www.nobanis.org (accessed 2013-03-27).

Occhipinti-Ambrogi A., Galil B., 2004. A uniform terminology on bioinvasions: a chimera or an operative tool? Marine Pollution Bulletin 49, pp. 688-694.

Occhipinti-Ambrogi A., Galil B., 2010. Marine alien species as an aspect of global change. Advances in Oceanography and Limnology, 1 (1), pp. 199-218.

Occhipinti-Ambrogi A., Marchini A., Cantone G., Castelli A., Chimenz C., Cormaci M., Froglia C., Furnari G., Gambi M.C., Giaccone G., Giangrande

A., Gravili C., Mastrototaro F., Mazziotti C., Orsi-Relini L., Piraino S., 2011. Alien species along the Italian coasts: an overview. Biological Invasions 13, pp. 215-237.

Office of Technology Assessment, 1993. Harmful non-indigenous species in the United States. OTA-F-565. U.S. Congress, Office of Technology Assessment, U.S. Government Printing Office, Washington, DC, pp. 397.

Oguz T., Fach B., Salihoglu B., 2008. Invasion dynamics of the alien ctenophore Mnemiopsis leidyi and its impact on anchovy collapse in the Black Sea. Journal of Plankton Research 30, pp. 1385-1397.

Ojaveer H., Eero M., 2011. Methodological Challenges in Assessing the Environmental Status of a Marine Ecosystem: Case Study of the Baltic Sea. PLoS ONE 6(4): e19231. doi:10.1371/journal.pone.0019231.

Olenin S., 2002. Invasive aquatic species of Europe - distribution, impacts, and management (in book). Leppäkoski E., Gollasch S., Olenin S. (ed.): Kluwer Academic Publishers, pp. 525-528.

Olenin S., Alemany F., Cardoso A.C., Gollasch S., Goulletquer P., Lehtiniemi M., McCollin T., Minchin D., Miossec L., Occhipinti Ambrogi A., Ojaveer H., Jensen K.R., Stankiewicz M., Wallentinus I., Aleksandrov B., 2010. Marine Strategy Framework Directive - Task Group 2 Report. Non-indigenous species. EUR 24342 EN. ISBN 978-92-79-15655-7. ISSN 1018-5593. DOI 10.2788/87092. Luxembourg: Office for Official Publications of the European Communities. pp. 44.

Olenin S., Elliott M., Bysveen I., Culverhouse P., Daunys D., Dubelaar G.B.J., Gollasch S., Goulletquer P., Jelmert A., Kantor Y., Mézeth K.B., Minchin D., Occhipinti-Ambrogi A., Olenina I., Vandekerkhove J., 2011. Recommendations on methods for the detection and control of biological pollution in marine coastal waters. Marine Pollution Bulletin 62(12), pp. 2598-2604.

Olenin S., Minchin D., 2011. Biological Introductions to the Systems: Macroorganisms. In: Wolanski E and McLusky DS (eds.) Treatise on Estuarine and Coastal Science, Vol 8, pp. 149-183.

Olenin S., Minchin D., Daunys D., 2007. Assessment of biopollution in aquatic ecosystems. Marine Pollution Bulletin, 55(7-9), pp. 379-394.

Olenin S., Narščius A., Minchin D., David M., Galil B., Gollasch S., Marchini A., Occhipinti-Ambrogi A., Ojaveer H., Zaiko A., 2013. Making non-indigenous species information systems practical for management and useful for research: an aquatic perspective. *Submitted*.

Olenina I., Wasmund N., Hajdu S., Jurgensone I., Gromisz S., Kownacka J., Toming K., Vaiciute D., Olenin S., 2010. Assessing impacts of invasive phytoplankton: The Baltic Sea case. Marine Pollution Bulletin, 60: pp. 1691-1700.

Olsen L., Major G., Shein K., Scialdone J., Vogel R., Leicester S., Weir H., Ritz S., Stevens T., Meaux M., Solomon C., Bilodeau R., Holland M., Northcutt T., Restrepo R.A., 2007. NASA/Global Change Master Directory (GCMD) Earth Science Keywords.

Olyarnik S.V., Bracken M.E.S., Byrnes J.E., Hughes R., Hultgren K.M., Stachowicz J.J., 2009. Ecological factors affecting community invisibility. In: Rilov G., Crooks J.A. (Eds.), Biological Invasions in Marine Ecosystems: Ecological, Management, and Geographic Perspectives, Ecological Studies, vol. 204 (XXVI). Springer-Verlag, pp. 215-238.

Orendt C., Schmitt C., van Liefferinge C., Wolfram G., de Deckere E., 2009. Include or exclude? A review on the role and suitability of aquatic invertebrate neozoa as indicators in biological assessment with special respect to fresh and brackish European waters. Biological Invasions. doi:10.1007/s10530-009-9448-x.

Paavola M., Olenin S., Leppäkoski E., 2005. Are invasive species most successful in habitats of low native species richness across European brackish water seas? Estuarine, Coastal and Shelf Science 64, pp. 738-750.

Paini D.R., Bianchi F.J.J.A., Northfield T.D., De Barro P.J., 2011. Predicting Invasive Fungal Pathogens Using Invasive Pest Assemblages: Testing Model Predictions in a Virtual World. PLoS ONE, Edited by Chaoyang Xue, vol. 6, issue 10.

Payton I.J., Fennel M., Lee W.G., 2002. Keystone species: the concept and its relevance for conservation management in New Zealand. Science for Conservation, 203. Department of Conservation, Wellington, New Zealand, pp. 29.

Pearson T.H., 2001. Functional group ecology in soft-sediment marine benthos: the role of bioturbation. Oceanography and Marine Biology. An Annual Review, 39, pp. 233-267.

Peters R.H., 1991. A critique for ecology. Cambridge University Press, Cambridge, UK

Peterson A.T., Knapp S., Guralnick R., Soberon J., Holder M.T., 2010. The big questions for biodiversity informatics. Systematics and Biodiversity 8 (2), pp. 159-168.

Pullin A.S., Salafsky N., 2010. Save the whales? Save the rainforest? Save the data! Conservation Biology 24, pp. 915-917.

Pyšek P., Hulme P.E., Nentwig W., 2009. Glossary of the main technical terms used in the handbook. In: DAISIE. The Handbook of Alien Species in Europe. Springer, Dordrecht, pp. 375-381.

Pyšek P., Richardson D.M., 2010. Invasive species, environmental change and management, and health. Ann. Rev. Environ. Res. 35. doi: http://dx.doi.org/10.1146/annurev-environ-033009-095548

Refsgaard J.C., Jørgensen L. F., Højberg A. L., 2007. Data availability and accessibility. State of the art on existing data required for modeling for research purposes and for the implementation of the Water Framework Directive. Harmoni-CA, http://www.wise-itcl.info (accessed 2013-03-27).

Reise K., Olenin S., Thieltges D.W., 2006. Are aliens threatening European aquatic coastal ecosystems? Helgoland Marine Research, 60 (2), pp. 106-112.

Rilov G., Crooks J., 2009. Biological Invasions in Marine Ecosystems: Ecological, Management, and Geographic Perspectives. Ecological Studies 204. Springer, Heidelberg, pp. 641.

Rosenthal H., Gollasch S., Voigt M., 2000. Final Report of the European Union Concerted Action "Testing Monitoring Systems for Risk Assessment of Harmful Introductions by Ships to European Waters". Contract No. MAS3-CT97-0111. pp. 72.

Ruiz G.M., Rawlings T.K., Dobbs F.C., Drake L.A., Mullady T., Huq A., Colwell R.R., 2000. Global spread of microorganisms by ships. Nature 408: pp. 49-50.

Ruusalepp R., 2008. Infrastructure planning and data curation. A comparative study of international approaches to enabling the sharing of research data. http://www.dcc.ac.uk/docs/publications/reports/Data_Sharing_Report.pdf (accessed 2013-03-27).

Savini D., Marchini A., Occhipinti-Ambrogi A., 2008. Aquaculture and alien species in the EU-Med region. Part I: status and trends. Biologia Marina Mediterranea 15 (1), pp. 240-241.

Savini D., Occhipinti A.A., Marchini A., Tricarico E., Gherardi F., Olenin S., Gollasch S., 2010. The top 27 alien animal species intentionally introduced by European aquaculture and related activities: stocking, sport fishery and ornamental purposes. Journal of Applied Ichthyology 26 (2), pp. 1-7.

SEBI 2010 project (Streamlining European 2010 Biodiversity Indicators, http://biodiversity-chm.eea.europa.eu/information/indicator/F1090245995).

Settele J., Penev L., Georgiev T., Grabaum R., Grobelnik V., Hammen V., Klotz S., Kotarac M., Kuhn I. (Eds), 2010. Atlas of Biodiversity Risk. Pensoft. Sofia. pp. 264.

Sherman K., Duda A.M., 1999. An ecosystem approach to global assessment and management of coastal waters. Mar. Ecol. Prog. Ser. 190, pp. 271-287.

Shine C., Genovesi P., Gollasch S., Kettunen M., Pagad S., Starfinger U., 2008. Technical support to EU strategy on invasive species (IAS). Policy options to control the negative impacts of IAS on biodiversity in Europe and the EU (Final module report for the European Commission). Institute for European Environmental Policy (IEEP), Brussels, Belgium (available at http://ec.europa.eu/environment/nature/invasivealien/docs/Shine2009_IAS_Final%20report.pdf)

Simberloff D., 2003. Confronting introduced species: a form of xenophobia? Biological Invasions, 5, pp. 179-192.

Simon K.S., Townsend C.R., 2003. Impacts of freshwater invaders at different levels of ecological organisation, with emphasis on salmonids and ecosystem consequences. Freshwater Biology, 48 (6), pp. 982-995.

Simpson A., Sellers E., Grosse A., Xie Y., 2006. Essential elements of online information networks on invasive alien species. Biological Invasions 8, pp. 1579-1587.

Skyrienė G., Paulauskas A., 2012. Distribution of invasive muskrats (*Ondatra zibethicus*) and impact on ecosystem. EKOLOGIJA, Vol. 58, No. 3, pp. 357-367.

Soberón J., Peterson A.T., 2009. Monitoring biodiversity loss with primary speciesoccurrence data: toward national-level indicators for the 2010 target of the convention on biological diversity. AMBIO: A Journal of the Human Environment 38, pp. 29-34.

Steele J.H., Collie J.S., Bisagni J.J., Gifford D.J., Fogarty M.J., Link J.S., Sullivan B.K., Sieracki M.E., Beet A.R., Mountain D.G., Durbin E.G., Palka D., Stockhausen W.T., 2007. Balancing end-to-end budgets of the Georges Bank ecosystem. Progress in Oceanography 74 (4), pp. 423-448.

Strayer D.L., 2010. Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. Freshwater Biology 55(s1), pp. 152-174.

TDWG, 2010. Historical Darwin Core. http://wiki.tdwg.org/twiki/bin/view/Darwin-Core/DarwinCoreVersions (accessed 2013-03-27).

Tenopir C., Allard S., Douglass K., Aydinoglu A.U., Wu L., Read E., Manoff M., Frame M., 2011. Data sharing by scientists: practices and perceptions. PloS One 6, e21101.

United Nations, 2012. UN Population Division's quinquennial estimates and projections, Department of Economic and Social Affairs (online). http://www.un.org/esa/population/ (accessed 20.09.12).

Valéry L., Hervé F., Levfeuvre J.C., Simberloff D., 2008. In search of a real definition of the biological invasion phenomenon itself. Biological Invasions, 10, pp. 1345-1351.

Van House A.N., 2003. Biodiversity Data Perversity: understanding barriers to sharing data across epistemic communities. Presentation at the Annual Meeting of the Society for Social Studies of Science, Atlanta, GA, Oct. 16, 2003.

van Kleunen M., Dawson W., Schlaepfer D., Jeschke J.M., Fischer M., 2010. Are invaders different? A conceptual framework of comparative approaches for assessing determinants of invasiveness. Ecol. Lett. 13, pp. 947-958.

Vandekerkhove J., Cardoso A.C., 2011. Online information systems with alien species occurrence records in Europe: Coverage, complementarity and compatibility. Publications Office of the European Union, ISBN: 978-92-79-19665-2.

Verling E., Ruiz G.M., Smith L.D., Galil B., Miller A.H., Murphy K.R., 2005. Supplyside invasion ecology: characterizing propagule pressure in coastal ecosystems. Proceedings of the Royal Society of London 272, pp. 1249-1257.

Vilà M., Basnou C., Pyšek P., Josefsson M., Genovesi P., Gollasch S., Nentwig W., 2009. How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment Frontiers in Ecology and the Environment 8 (3), pp. 135-144.

Vilà M., Basnou C., Pyšek P., Olenin S., 2009. How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. Frontiers in Ecology and the Environment 8, pp. 135-144.

Wallentinus I., Nyberg C. D, 2007. Introduced marine organisms as habitat modifiers. Marine Pollution Bulletin 55, pp. 323-332.

Walther G.R., Roques A., Hulme P.E., Sykes M.T., et al., 2009. Alien species in a warmer world: risks and opportunities. Trends in Ecology and Evolution Vol.24 No.12.

Weibel S., 2005. The Dublin Core: A Simple Content Description Model for Electronic Resources. Bulletin of the American Society for Information Science and Technology 01/2005; 24(1), pp. 9-11.

Westphal M.I., Browne M., MacKinnon K., Noble I., 2008. The link between international trade and the global distribution of invasive alien species. Biological Invasions 10(4), pp. 391-398.

White R.J., 2007. Linking biodiversity databases: Preparing species diversity information sources by assembling, merging and linking databases. Biodiversity Databases: Techniques, Politics, and Applications (Systematics Association Special Volume). Gordon Curry (Editor), Chris Humphries (Editor). ISBN-10: 0415332907.

Wieczorek J., Bloom D., Guralnick R., Blum S., Doring M., Giovanni R., Robertson T., Vieglais D., 2012. Darwin Core: An Evolving Community-

Developed Biodiversity Data Standard. PLoS ONE;Jan2012, Vol. 7 Issue 1, pp. 1.

Williamson, M., 1996. Biological Invasions. Chapman and Hall, London, pp. 244.

Wilson J. R. U., Dormontt E. E., Prentis P. J., Lowe A. J., Richardson D. M., 2009. Something in the way you move: Dispersal pathways affect invasion success. Trends in Ecology and Evolution, 24, pp. 136-144.

Witten I.H., Frank E., Hall M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems). ISBN-10: 0123748569.

Wittenberg R., Cock M.J.W., 2001. Invasive alien species: A toolkit of best prevention and management practices. http://www.gisp.org/publications/toolkit/Toolkiteng.pdf (accessed 2013-03-27).

Wittfoth A. K. J., Zettler M. L., 2013. The application of a Biopollution Index in German Baltic estuarine and lagoon waters. Management of Biological Invasions. Volume 4, Issue 1, pp. 43-50.

WSSD, 2002. United Nations Report on the World Summit on Sustainable Development. Johannesburg, South Africa, 26 August-4 September 2002, 173 p.

Yilmaz P., Kottmann R., Field D., Knight R., et al., 2011. “Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications”. Nature Biotechnology Volume: 29, pp. 415-420.

Zaiko A., Lehtiniemi M., Narščius A., Olenin S., 2011. Assessment of bioinvasion impacts on a regional scale: a comparative approach. Biological Invasions 13, pp. 1739-1765.

Zaiko A., Olenin S., Daunys D., Nalepa T., 2007. Vulnerability of benthic habitats to the aquatic invasive species. Biological invasions 9, pp. 703-714.

APPENDIX. The list of NIS, which have spread into 3 or more LMEs. Number of invaded LMEs by a species is indicated in the first column.

#	Species	Family	Order	Class	Phylum
7	Codium fragile fragile	Codiaceae	Bryopsidales	Bryopsidophyceae	Chlorophyta
7	Bonnemaisonia hamifera	Bonnemaisoniaceae	Nemaliales	Florideophyceae	Rhodophyta
6	Callinectes sapidus	Portunidae	Decapoda	Malacostraca	Arthropoda
5	Crassostrea gigas	Ostreidae	Ostreoida	Bivalvia	Mollusca
5	Rhithropanopeus harrisii	Panopeidae	Decapoda	Malacostraca	Arthropoda
5	Neosiphonia harveyi	Rhodomeliaceae	Ceramiales	Florideophyceae	Rhodophyta
5	Colpomenia peregrina	Scytosiphonaceae	Scytosiphonales	Phaeophyceae	Ochrophyta
5	Crepidula fornicata	Calyptraeidae	Neotaenioglossa	Gastropoda	Mollusca
5	Rapana venosa	Muricidae	Sorbeoconcha	Gastropoda	Mollusca
5	Eriocheir sinensis	Varunidae	Decapoda	Malacostraca	Arthropoda
5	Sargassum muticum	Sargassaceae	Fucales	Phaeophyceae	Ochrophyta
5	Ruditapes philippinarum	Veneridae	Veneroida	Bivalvia	Mollusca
5	Gonionemus vertens	Olindiasidae	Limnomedusae	Hydrozoa	Cnidaria
4	Ficopomatus enigmaticus	Serpulidae	Canalipalpata	Polychaeta	Annelida
4	Marsupenaeus japonicus	Penaeidae	Decapoda	Malacostraca	Arthropoda
4	Mnemiopsis leidyi	Mnemiidae	Lobata	Tentaculata	Ctenophora
4	Asparagopsis armata	Bonnemaisoniaceae	Nemaliales	Florideophyceae	Rhodophyta
4	Acartia (Acanthacartia) tonsa	Acartiidae	Calanoida	Maxillopoda	Arthropoda
4	Grateloupia turuturu	Dasyaceae	Ceramiales	Florideophyceae	Rhodophyta
4	Styela clava	Styelidae	Pleurogona	Ascidiacea	Chordata
4	Cordylophora	Cordylophorida	Anthoathecata	Hydrozoa	Cnidaria

#	Species	Family	Order	Class	Phylum
	caspia	e			
4	Heterosiphonia japonica	Rhodomelaceae	Ceramiales	Florideophycea e	Rhodophyta
4	Homarus americanus	Nephropidae	Decapoda	Malacostraca	Arthropoda
4	Oncorhynchus gorbuscha	Salmonidae	Salmoniformes	Actinopterygii	Chordata
3	Hydroides elegans	Serpulidae	Canalipalpata	Polychaeta	Annelida
3	Mya arenaria	Myidae	Myoida	Bivalvia	Mollusca
3	Petricolaria pholadiformis	Petricolidae	Veneroida	Bivalvia	Mollusca
3	Antithamnionell a spirographidis	Ceramiaceae	Ceramiales	Florideophycea e	Rhodophyta
3	Antithamnionell a ternifolia	Ceramiaceae	Ceramiales	Florideophycea e	Rhodophyta
3	Austrominius modestus	Austrobalanidae	Sessilia	Maxillopoda	Arthropoda
3	Lomentaria hakodatensis	Lomentariaceae	Rhodymeniales	Florideophycea e	Rhodophyta
3	Mercenaria mercenaria	Veneridae	Veneroida	Bivalvia	Mollusca
3	Undaria pinnatifida	Alariaceae	Laminariales	Phaeophyceae	Ochrophyta
3	Balanus trigonus	Balanidae	Sessilia	Maxillopoda	Arthropoda
3	Acipenser baeri	Acipenseridae	Acipenseriform es	Actinopterygii	Chordata
3	Anguillicoloides crassus	Anguillicolidae	Spirurida	Secernentea	Nemata
3	Balanus amphitrite	Balanidae	Sessilia	Maxillopoda	Arthropoda
3	Ensis directus	Pharidae	Veneroida	Bivalvia	Mollusca
3	Gammarus tigrinus	Gammaridae	Amphipoda	Malacostraca	Arthropoda
3	Gracilaria vermiculophylla	Gracilariaceae	Gracilariales	Florideophycea e	Rhodophyta
3	Hemigrapsus penicillatus	Grapsidae	Decapoda	Malacostraca	Arthropoda
3	Hemimysis anomala	Mysidae	Mysida	Malacostraca	Arthropoda
3	Potamopyrgus antipodarum	Hydrobiidae	Neotaenioglossa a	Gastropoda	Mollusca

#	Species	Family	Order	Class	Phylum
3	Pseudodactylogyrus anguillae	Dactylogyridae	Monopisthocotylea	Trematoda	Platyhelminthes
3	Spartina x townsendii	Poaceae	Cyperales	Liliopsida	Magnoliophyta
3	Tricellaria inopinata	Candidae	Cheilostomata	Gymnolaemata	Bryozoa
3	Oncorhynchus keta	Salmonidae	Salmoniformes	Actinopterygii	Chordata
3	Diadumene lineata	Diadumenidae	Actiniaria	Anthozoa	Cnidaria
3	Diadumene cincta	Diadumenidae	Actiniaria	Anthozoa	Cnidaria