Information Management

Building the European Alien Species Information Network (EASIN): a novel approach for the exploration of distributed alien species data

Stelios Katsanevakis*, Konstantins Bogucarskis, Francesca Gatto, Jochen Vandekerkhove, Ivan Deriu and Ana Cristina Cardoso

European Commission, Joint Research Centre, Institute for Environment and Sustainability, Ispra, Italy

E-mail: stelios@katsanevakis.com (SK), konstantins.bogucarskis@jrc.ec.europa.eu (KB), francesca.gatto@jrc.ec.europa.eu (FG), ivan.deriu@ext.jrc.ec.europa.eu (ID), jochenvdk@hotmail.com (JV), ana-cristina.cardoso@jrc.ec.europa.eu (ACC)

*Corresponding author

Abstract

The European Alien Species Information Network (EASIN; http://easin.jrc.ec.europa.eu) aims to facilitate the exploration of existing alien species information from distributed sources through a network of interoperable web services, and to assist the implementation of European policies on biological invasions. The network allows extraction of alien species information from online information systems for all species included in the EASIN catalogue. This catalogue was based on an inventory of reported alien species in Europe that was produced by reviewing and standardizing information from 43 online databases. It includes information on taxonomy, synonyms, common names, pathways of introduction, native range in Europe, and impact. EASIN catalogue entails the basic information needed to efficiently link to existing online databases and retrieve spatial information for alien species distribution in Europe. Using search functionality powered by a widget framework, it is possible to make a tailored selection of a subgroup of species based on various criteria (e.g., environment, taxonomy, pathways). Distribution maps of the selected species can be produced dynamically and downloaded by the user. The EASIN web tools and services follow internationally recognized standards and protocols, and can be utilized freely and independently by any website, while ownership of the data remains with its source, which is properly cited and linked.

Key words: biological invasions; online databases; interoperable web services; widgets; standardization

Background information

Europe is severely affected by biological invasions, which are considered as one of the most important direct drivers of biodiversity loss, and a major pressure to several types of ecosystems, with both ecological and economic impacts (MEA 2005). A conservative estimate of the annual damage caused in the EU by alien species is € 12 billion (EC 2008a). Recognizing the need for robust action to control biological invasions and thus mitigate their impacts on biodiversity, ecosystem services and human activities, the European Commission has recently adopted a Communication presenting policy options for an EU Strategy on Invasive Species (EC 2008a). Currently a dedicated legislative instrument is being developed by the Commission (to be proposed in 2012) as dictated by Action 16 of the Biodiversity Strategy (EC 2011). The latter explicitly requests that "by 2020, Invasive Alien Species and their pathways are identified and prioritized, priority species are controlled or eradicated, and pathways are managed to prevent the introduction and establishment of new IAS".

There is a need for accurate, detailed, and timely information on alien species occurrence, distribution and impacts to implement the European policies for the efficient prevention, early detection, rapid response, and management of biological invasions and also to evaluate management measures (Lee et al. 2008; Simpson et al. 2009; Hulme and Weser 2011). In recognition of this need, a large number of human networks and online databases have been created to provide information on biological invasions on a national, supranational, or global scale. These information systems provided the basic knowledge for various landmark assessments of alien invasions in Europe or in its regions

235

Figure 1. Schematic illustration of the relationships among EASIN (European Alien Species Information Network) and data providers. EASIN aggregates data from all linked data providers and offers tools and web services to the providers and other interested hosts. All information provided by EASIN's services are linked to the source data, where the user should seek more detailed and disaggregated information.
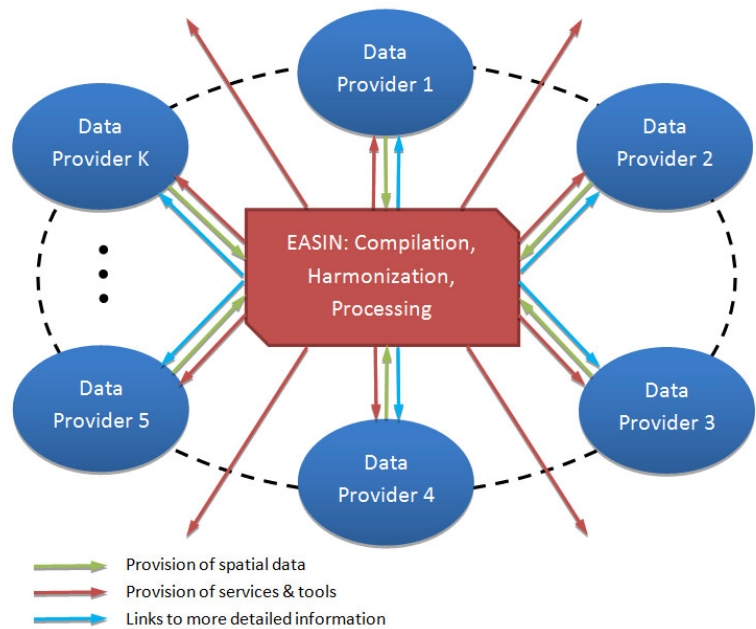
(e.g. Chiron et al. 2009; Vilà et al. 2010; Pyšek et al. 2010) and have been effective at raising awareness and improving surveillance of biological invasions.

However, various issues such as variation in data quality, inadequate updating of information, lack of common and agreed definitions of 'alien' and 'invasive species', inconsistencies among databases, lack of geo-referenced species occurrences, and gaps in the spatial and taxonomic coverage (Vandekerkhove and Cardoso 2011; Hulme and Weser 2011; Gatto et al. 2012), in many cases are bottlenecks for the effective assessment and management of biological invasions. In a comparative assessment of the existing online information systems on alien species (Gatto et al. 2012) large differences among databases were found that were only partially explained by variations in their taxonomical, environmental, and geographical scopes. Even DAISIE (Delivering Alien Invasive Species Inventories for Europe; http://www.europe-aliens.org), the most comprehensive online resource of country-level alien species occurrences in Europe, did not include >30% of species reported in online databases (Gatto et al. 2012). Hence, there is a strong need for

integration and harmonization of existing distributed information on alien species.

One approach to improve the quality of information on alien species, increase its availability and accessibility, and ultimately support a cost-efficient invasive alien species policy is to create a network of online interoperable web services through which information in distributed resources can be accessed (Vandekerkhove and Cardoso 2011; Gatto et al. 2012). The successful implementation of such a concept relies on the continued engagement at national and regional scale to collect and provide data; the willingness of database managers to harmonize their information; the development of a set of interoperable web services through which the information can be explored; and appropriate and sustainable funding. The European Commission's Joint Research Centre (JRC) has put efforts towards building such a network of online interoperable web services by developing its central platform (Figure 1). This paper describes the first stage of development of the European Alien Species Information Network (EASIN). The EASIN website (http://easin.jrc.ec.europa.eu) was made available to the public on May 15th 2012.

Figure 2. Schematic illustration of the information contained in the EASIN catalogue and how species-specific information is retrieved from data providers. Each valid name of a species is linked to 'original names' (valid or invalid names that appeared in one or more online databases) and 'synonyms'. When spatial data for a specific species are sought, a search is made to data providers using not only the valid name but also all 'original names' and synonyms.
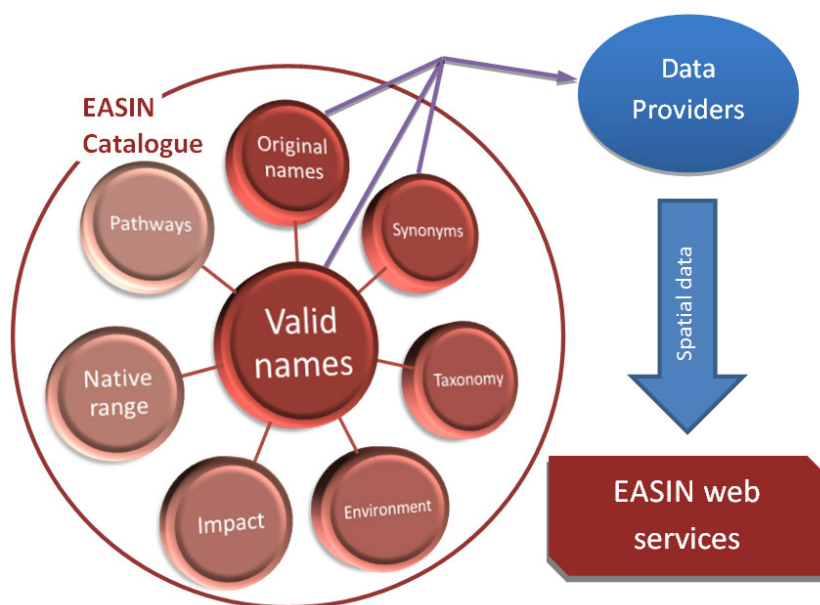
Scope

EASIN aims to enable easy access to data and information on alien species in Europe from existing on-line databases (Figure 1) to assist policy makers and scientists in their efforts to prevent and control alien invasions. It facilitates the exploration of existing alien species information from a variety of distributed information sources by developing and making freely available tools and interoperable web services compliant with internationally recognized standards. EASIN seeks to provide an integrated view of biological invasions on a European scale, utilizing regional and national information that is collected and assessed in global, European, supranational and national databases, thereby building on a large network of experts and citizens.

EASIN aims also to provide web tools and services that can be utilized freely and independently by any host. Data retrieved by EASIN can then be explored in various ways. Basic functionalities can be accessed using web widgets (a web widget is a stand-alone application that can be embedded into third party sites by any user on a page where they have rights of authorship) having interactive alien species data querying, GIS-based mapping and reporting interfaces. These widgets can be installed into third party web-sites and in most cases spare users from seeking alien species information among multiple web interfaces. EASIN offers a rich web services API (Application Programming Interface) to provide sufficient flexibility to achieve special or specific tasks, allowing advanced users to build custom applications tailored to particular needs. In every case, ownership of the data remains with its source, which is properly cited and linked. The current list of cooperating providers of geographical data is given in EASIN's website (http://easin.jrc.ec.europa.eu/Partners).

EASIN catalogue of alien species

The EASIN catalogue is in the core of the system. It entails the basic information needed to efficiently link to existing online databases and retrieve spatial information for alien species distribution in Europe (Figure 2). It is based on an inventory of all species and subspecies reported to be alien in (part of) Europe by one or more of 43 online information systems: 7 with global coverage, 2 with European coverage, 5 with supranational coverage, 26 with national coverage, and 3 with sub-national coverage (Table 1). Some of these databases are not specifically targeting alien species but serve a more general purpose of biodiversity monitoring; among these only those in which it was possible

Table 1. The 43 online species databases used to create the EASIN catalogue. Numbers of alien species may slightly deviate from what was reported in the source databases (retrieved in August 2011), as some species were excluded based on the methodology described herein.

Name (and country)		N. of alien species	Website	Notes
Global databases				
Bird Checklists of the World (AviBase)		62	http://avibase.bsc-eoc.org	Birds only
A Global Information System on Fishes (FishBase)		195	http://www.fishbase.org	Fishes only
Global Invasive Species Database (GISD)		317	http://www.issg.org/database/welcome/	
European and Mediterranean Plant Protection Organization - Plant Quarantine Data Retrieval System (EPPO-PQR)		725	http://www.eppo.org/DATABASES/pqr/pqr.htm	
FAO Database on Introductions of Aquatic Species (FAO-DIAS)		476	http://www.fao.org/fishery/introsp/search/en	Fisheries related
Global Invasive Species Threat Assessment		205	http://conserveonline.org/workspaces/global.invasive.assessment	Marine species only
Invasive Species Compendium (CABI-ISC)		664	http://www.cabi.org/isc/	
European databases				
Delivering Alien Invasive Species Inventories for Europe (DAISIE)		10526	http://www.europe-aliens.org	
European list of Alien Birds (by SEO/Bird life)		597	http://www.seo.org	Birds only
Supranational databases				
Atlas of Exotic Species in the Mediterranean (CIESM)		326	http://www.ciesm.org	
European Network on Invasive Alien Species (NOBANIS)		7383	http://www.nobanis.org	
Baltic Sea Alien Species Database		113	http://www.corpi.ku.lt	
Regional Euro-Asian Biological Invasions Centre (REABIC)		12	http://www.reabic.net	
Caspian Biodiversity Sea Database		36	http://www.caspianenvironment.org/biodb/	
National databases				
Austrian Biodiversity Information System (OASIS 2.0)	Austria	1132	http://www.umweltbundesamt.at	
Aquatic Invaders of Belarus	Belarus	59	http://aliensinbelarus.com	Aquatic species only
Harmonia	Belgium	55	http://ias.biodiversity.be	
Checklist for aquatic alien species in the Belgian part of the North Sea and adjacent estuaries	Belgium	67	http://www.vliz.be	Aquatic species only
Danish Database of Introduced Species	Denmark	108	http://www.naturstyrelsen.dk	
Estonian Alien Species Database	Estonia	912	http://eelis.ic.envir.ee	
Marine Alien Species of Estonia	Estonia	12	http://www.sea.ee	Marine species only
Aquatic alien species in German Inland and Coastal Waters	Germany	139	http://www.aquatic-aliens.de	Aquatic species only
Ellenic Network on Aquatic Invasive Species database (ELNAIS)	Greece	317	https://services.ath.hcmr.gr	Aquatic species only
Flora of Iceland	Iceland	301	http://www.floraislands.is	Plants only
National Invasive Species Database	Ireland	83	http://invasives.biodiversityireland.ie	
Database of Alien Plants in Ireland	Ireland	405	http://www.biochange.ie	Plants only
Fish of Latvia	Latvia	21	http://latvijas.daba.lv	Fishes only
Lithuanian Invasive Species Database	Lithuania	661	http://www.ku.lt	
Neophyten in Luxemburg	Luxembourg	22	http://www.mnhn.lu	Plants only
Overview of Dutch biodiversity	Netherlands	1752	http://www.nederlandsesoorten.nl	
Norwegian Alien Species Database	Norway	215	http://www.artsdatabanken.no	
Alien Species in Poland	Poland	1225	http://www.iop.krakow.pl/ias/	
Invasive Plants Species in Portugal	Portugal	32	http://www1.ci.uc.pt/invasoras/	Plants only
BioDat Database Introduced Animal Species	Russia	574	http://www.biodat.ru	Animals only
BioDat Database Introduced Plant Species	Russia	404	http://www.biodat.ru	Plants only
Invasive Exotic Species of the Iberian Peninsula (InvasIber)	Spain	56	http://invasiber.org	
Alien species in Swedish seas and coastal areas	Sweden	77	http://www.frammandearter.se	Marine species only
Invasive Alien Plants	Switzerland	44	http://www.cps-skew.ch	Plants only
Non-Native Species Information Portal (NNSIP)	United Kingdom	285	https://secure.fera.defra.gov.uk/nonnativespecies	
Non-native marine species in British waters: a review and directory	United Kingdom	69	http://www.jncc.gov.uk/page-2597	Marine species only
Sub-national databases				
Alien Species in the Canary Islands	Canary Islands	1810	http://www.interreg-bionatura.com	
Audit of Non-Native Species in England	England	2688	http://www.brc.ac.uk/resources.htm	
Invasive Alien Species in Northern Ireland	Northern Ireland	49	http://www.habitas.org.uk	

Table 2. The main steps of the procedure to compile the EASIN catalogue. The progress achieved so far for each of the three environments (terrestrial, freshwater, and marine) is indicated, referring to version 1.9 of the EASIN catalogue that was uploaded on May 15th 2012.

Steps for the compilation of the EASIN catalogue	terrestrial	freshwater	marine
1. Collection of original names of 'alien species' from online databases			
2. Exclusion of 'extinct', 'potential aliens', 'reintroduced', 'excluded', etc.			
3. Modification of original names to conform with international standards			
4. Spelling errors identified semi-automatically			
5. Synonyms checked in WoRMS, ITIS, CoL and replaced by valid names			
6. Removal of duplicates			
7. Steps 1-6 repeated with offline resources (databases, literature, etc)			
8. Validation of names by experts			
9. Retrieval of information on pathways (expert contribution)			
10. Retrieval of information on native range in Europe (expert contribution)			

to retrieve specific information about the alien/native status of each species by country were considered. Alien species names were extracted for all European countries, considering Europe in a broad sense (as mentioned in http://europa.eu/about-eu/countries/), i.e. including the 27 EU Member States, the 5 Candidate countries and 17 other European countries. To have full coverage of the four seas surrounding Europe, alien marine species reported from the entire Mediterranean Sea were included, i.e. also from North African and Near East Mediterranean countries.

From each of the 43 online information systems, all names of species listed as 'alien', 'cryptogenic', 'introduced', 'casual alien' or 'invasive' were extracted. Species listed as 'potential aliens' (watch lists), 'reintroduced', 'excluded' or 'extinct' were excluded. Marine species with type locality within the same regional sea (e.g. NE Atlantic species reported as aliens in the North Sea) were excluded. Vagrant marine species that have entered the Mediterranean via Gibraltar (mostly tropical Atlantic fish and decapods) or Mediterranean planktonic species that entered the Black Sea via the Dardanelles Straits were also excluded. For alien pests from the EPPO (European and Mediterranean Plant Protection Organization) database we included only those with at least some reported records (X1-X3 pests) and excluded pests whose occurrence in a given country was unclear (X0 pests). From the Audit of Non-Native Species in England, 'aliens' categorized as 'formerly native' and 'native with large addition from domestic or non-native stock' were also excluded. The inventory of all extracted species names was further processed in a multi-step procedure to harmonize notations across databases and identify duplicate taxon entries (see Appendix I for a detailed description).

The EASIN catalogue contains additional information for each species included (Figure 2). A taxonomic tree (kingdom, phylum, class, order, and family) is associated with every species in the catalogue. Information on pathways of introduction is being included by largely following the framework proposed by Hulme et al. (2008) (see Appendix I for details). Species that are recognized to have a high impact (i.e. present in the 'high-impact' or 'worst invasive' species lists of DAISIE, GISD and SEBI-2010) are also highlighted. For those species that are alien in some regions of Europe and native in others (e.g. Ponto-Caspian species invading the Baltic and the North Sea through inland canals), the native range has been defined (on a country level for terrestrial and freshwater species and on a marine basin level for marine species). Some of this information (i.e. pathways, native range) is currently available only for marine species.

Compilation of the EASIN catalogue is an ongoing process and includes several steps to achieve high quality standards. For the three final steps (see Table 2) contribution from external contracted experts is being sought (this has been currently done only for marine species). Table 2 shows the progress accomplished so far (by May 15th 2012; version 1.9 of the EASIN

Figure 3. Example of the output of the simple search widget after searching for 'Siganus'. The control button for *Siganus luridus* has been pressed and a drop-down window with some basic information for the species has been opened.

catalogue) for each of the three environments (terrestrial, freshwater, and marine).

In the EASIN catalogue a unique identification code (R_ID) is assigned to each valid species and subspecies name and is used to connect it with related features (e.g. environment, synonyms, common names, taxonomy, pathways, impact), which were allocated in different tables. A formal procedure is followed for updating the catalogue so that all modifications and updates are traceable. A versioning mechanism allows tracking any change in the catalogue. This mechanism is based on the LSID (Life Science Identifiers) approach (TDWG 2011). The LSID concept introduces a straightforward approach to naming and identifying data resources stored in multiple, distributed data sources in a manner that overcomes the limitations of naming schemes in use today. A unique LSID is generated by the system for each collection of species selected by the user. This LSID will allow the user to reach the latest revision of the EASIN catalogue data, when accessing the EASIN system again. If updates took place in the meantime, the system will provide notifications for these species that were updated.

EASIN Main Products

Widgets Framework

An important feature of EASIN is the capability to make a tailored selection of a subgroup of alien species from the Catalogue and request relevant output. For this purpose the Widget framework has been developed, providing users with relevant stand-alone applications and services. It is a fast and flexible way to export pre-configured EASIN functionality on other websites focusing on biodiversity issues. The framework has two main types of web widgets, search widgets and widgets to display maps and download queried data.

In the simple search widget (Figure 3), one or more species may be selected by name (full or part of the names, scientific or common names). Suggestions are automatically shown to assist the
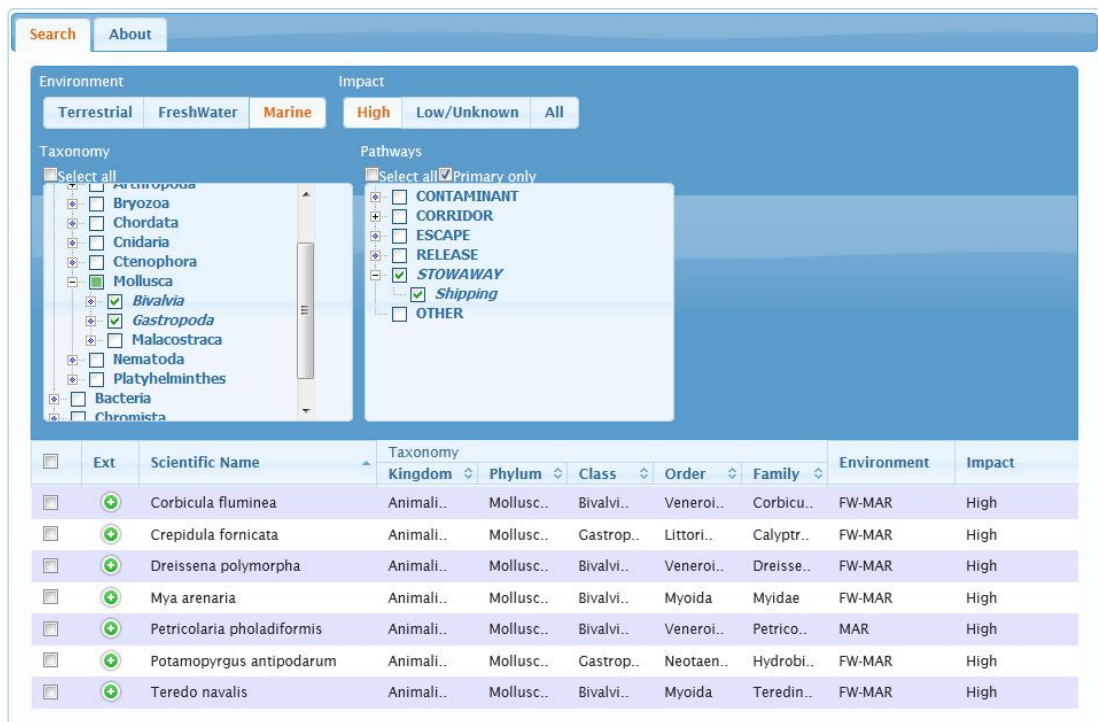
240

Figure 4. Example of the combined search widget: high-impact marine bivalves and gastropods introduced in Europe by shipping.

user to select the desired species based on scientific or common names. Two filters, one on environment (terrestrial, marine, freshwater) and another on impact (high, low, all) can be also jointly used to further specify species selection.

The combined search widget (Figure 4), contains two additional filters: taxonomy (based on an hierarchical selection on the taxonomic tree, from genus to phylum) and pathways (selecting one or more of the 21 pathway/sub-pathway categories). After filtering the catalogue, the user may refine the selection by selecting/deselecting species.

The results of the search widgets may be saved, and then retrieved using data display widgets in tabular form or used for the production of distributional maps based on a choice of two possible grids: 10×10 km or 10×10 minutes of a degree. When a single species is selected, the cells of the map grid will be coloured based on a three-colour pattern corresponding to four possible states: (0) absence or no information (no colour); (1) presence as alien; (2) previously present but now absent (e.g. due to eradication); (3) presence as native. When many species are selected, the

number of present species in each cell will be depicted by a colour gradient. For each cell of the produced maps and for each species reported as present, links to the original sources of information will be available.

Widgets can be integrated into any webpage, have no link to a particular technology, support various layouts, and require zero programming skills from webmasters to be installed. Instead of writing code, web enablers can configure and place bits of functionality in a predefined grid. The reliance on ready-made features and functionality greatly reduces a project development time allowing performers to address scientific issues better. The Widget Framework embodies best practice methods and may offer valuable support to the developers of similar projects.

EASIN Web Services

All the data indexed by EASIN from distributed data sources e.g. used by the widget framework are made available via the interoperable web-services. The services were implemented respecting various open standards, i.e. freely

Table 3: Timetable of EASIN development.

	2012				2013				2014			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
EASIN catalogue												
Initial compilation of the alien species inventory (steps 1-6; Table 2)	▓											
Improve data quality (steps 7-10; Table 2) for marine species		▓										
Improve data quality (steps 7-10; Table 2) for freshwater species			▓	▓	▓							
Improve data quality (steps 7-10; Table 2) for terrestrial species					▓	▓	▓	▓				
Updating EASIN catalogue (continuous process)	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓
Web widgets												
Simple search widget	▓	▓										
Combined search widget												
Mapping occurrence of selected species		▓	▓	▓								
Partnerships with online data providers												
Link to GBIF, GISIN, REABIC	▓	▓	▓									
Link to CIESM, ELNAIS			▓	▓								
Workshop with all major database administrators												
Link to most major databases					▓	▓	▓	▓				
Modeling spatial distribution												
Predictive distribution maps									▓	▓	▓	

available specifications and rules, such as OpenSearch (DeWitt 2011), OGC (2011), and TDWG (2011). Using open standards ensures interoperability among different software products made by different manufacturers. These principles mean that users can implement custom solutions without having any knowledge of the service providers. This facilitates just-in-time integration and allows users to implement new ideas easily and dynamically.

Most of the EASIN web services use the RESTful approach (Fielding 2000) and return requested data as XML. The mapping service is built upon OGC WMS 1.3.0 standard (OGC 2011), which allows easy integration into users' mapping applications. In addition, each map layer is accompanied with INSPIRE compliant metadata (EC 2008b).

The web-services display the data retrieved from original data providers in an aggregated way. EASIN uses a well-known brokering approach similar to what is used by EuroGEOSS (2011) and also investigated in the context of GISIN (Graham et al. 2011) to index the data from the data providers. It includes three main steps: analysing the data source, capturing changes, and transforming the source data model to the EASIN data model. The latter preserves important elements from all data sources/providers e.g. geo-referenced elements and links to the original data collections. The model also employs the mechanism for data harmonization among different sources. It allows the user to search multiple databases at once in real time, and arrange the results into a useful form for further reporting operations, such as mapping.

Current Status - Future Plans

EASIN is now in the initial phase of development and not all features have been made available to the public. The current status and the timeline for future development is depicted in Table 3. The marine part of the catalogue is at a more improved stage of development in relation

to freshwater and terrestrial environments, as all ten steps (Table 2) for its compilation have been completed. Steps 7–10 (as in Table 2) have been scheduled to be completed by the first quarter of 2013 for freshwater species and by the end of 2013 for terrestrial species. The search widgets are functional and open to the public, while the mapping widgets will be available to the public by the end of 2012. By that time, spatial data would be retrieved from multiple sources (Table 3). This initial set of source databases was purposely selected to have variable attributes in terms of spatial coverage, interoperability, and format of stored spatial information, and will thus serve as a pilot project to validate and improve the performance of EASIN. Based on the experience gained with this initial set of online databases, a targeted workshop will be organized for other potential data providers by the end of 2012, aiming to expand the EASIN network to all major related information systems by the end of 2013. A further step will be the development of a modelling platform, allowing users to produce predictive distribution maps based on historical and contemporary distributions of alien species and anticipated changes in the environment (e.g. climate change, land use).

Acknowledgements

References

Cavalier-Smith T (2004) Only six kingdoms of life. *Proceedings of the Royal Society of London-B* 271: 1251–1262, http://dx.doi.org/10.1098/rspb.2004.2705

Chiron F, Shirley S and Kark S (2009) Human-related processes drive the richness of exotic birds in Europe. *Proceedings of the Royal Society of London-B* 276: 47–53, http://dx.doi.org/10.1098/rspb.2008.0994

DeWitt C (2011) OpenSearch description document – Open Search. http://www.opensearch.org (Accessed September 2011)

EC (2008a) Towards an EU strategy on invasive species. COM/2008/789, European Commission, Brussels, 10 pp

EC (2008b) Commission Regulation No 1205/2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata. European Commission, Brussels, 19 pp

EC (2011) Our life insurance, our natural capital: an EU biodiversity strategy to 2020. COM/2011/244, European Commission, Brussels, 16 pp

EuroGEOSS (2011) EuroGEOSS Broker - EuroGEOSS. http://www.eurogeoss-broker.eu (Accessed September 2011)

Fielding TR (2000) Architectural styles and the design of network-based software architectures. PhD Thesis, University of California, Irvine, 90 pp

Gatto F, Katsanevakis S, Bogucarskis K, Vandekerhove J, Cardoso AC (2012) European Alien Species Information Network (EASIN): The need of an integrated European information system on alien species. Proceedings of the ECSA 50 Conference "Today's Science for tomorrow's management", June 3-7, 2012, Venice, Italy

Graham J, Jarnevich CS, Simpson A, Newman GJ, Stohlgren TJ (2011) Federated or cached searches: Providing expected performance from multiple invasive species databases. *Frontiers of Earth Science* 5: 111–119, http://dx.doi.org/10.1007/s11707-011-0152-7

Hulme PE, Weser C (2011) Mixed messages from multiple information sources on invasive species: a case of too much of a good thing? *Diversity and Distributions* 17: 1152–1160, http://dx.doi.org/10.1111/j.1472-4642.2011.00800.x

Hulme PE, 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: 403–414, http://dx.doi.org/10.1111/j.1365-2664.2007.01442.x

Lee II H., Reusser DA, Olden JD, Smith SS, Graham J, Burkett V, Dukes JS, Piorkowski RJ, McPhedran J (2008) Integrated monitoring and information systems for managing aquatic invasive species in a changing climate. *Conservation Biology* 22: 575–584, http://dx.doi.org/10.1111/j.1523-1739.2008.00955.x

MEA (Millennium Ecosystem Assessment) (2005) Ecosystems and Human Well-being: Biodiversity Synthesis. World Resources Institute, Washington, DC, 86 pp

OGC Open Geospatial Consortium (2011) OGC® Standards and Supporting Documents - OGC. http://www.opengeospatial.org/standards/ (Accessed September 2011)

Pyšek P, Jarošík V, Hulme PE, Kühnd I, Wilda J, Arianoutsou M, Bacher S, Chiron F, Didžiulis V, Essl F, Genovesi P, Gherardi F, Hejda M, Kark S, Lambdon PW, Desprez-Loustau ML, Nentwig W, Pergl J, Poboljšaj K, Rabitsch W, Roques A, Roy DB, Shirley S, Solarz W, Vilà M, Winter M (2010) Disentangling the role of environmental and human pressures on biological invasions. *Proceedings of the National Academy of Sciences of the United States of America* 107: 12157–12162, http://dx.doi.org/10.1073/pnas.1002314107

Simpson A, Fournier C, Sellers E, Browne M, Jarnevich C, Graham J, Mehrhoff L, Madsen J, Westbrooks R (2009) Invasive species information networks: collaboration at multiple scales for prevention, early detection, and rapid response to invasive species. *Biodiversity* 10: 5–13, http://dx.doi.org/10.1080/14888386.2009.9712839

TDWG Biodiversity information standards (2011) GUID and Life Sciences Identifiers Applicability Statements - TDWG. http://www.tdwg.org/standards/150/ (Accessed September 2011)

Vandekerkhove J, Cardoso AC (2011) Online information systems with alien species occurrence records in Europe: coverage, complementarity and compatibility. JRC Scientific and Technical Reports, Publications Office of the European Union, EUR 24752 EN, ISBN 978-92-79-19665-2, 63 pp, http://dx.doi.org/10.2788/65974

Vilà M, Basnou C, Pyšek P, Josefsson M, Genovesi P, Gollasch S, Nentwig W, Olenin S, Roques A, Roy D, Hulme PE (2010) 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): 135–144, http://dx.doi.org/10.1890/080083

Appendix 1. Protocol for the initial compilation of the EASIN catalogue of alien species in Europe.

Compilation of species names – correction of spelling errors - standardization

This protocol describes the different steps taken to harmonize the taxon name notations, identify duplicate taxon entries in the alien species reference list, and retrieve all relevant information (taxonomy, pathways, impact etc) for all alien species.

Duplicate names in the initial list of alien species (produced by merging inventories from all online sources) were removed using Excel's *'Remove Duplicate'* function. Because of variations across information systems in the notation of the taxon names, many taxa still had multiple entries in the table.

To facilitate the process of identification of additional duplicates and also to standardize species notation according to accepted nomenclature, the following steps were followed:

- Species names were split over multiple columns using the Excel "text to column" function.
- Names higher than species rank were deleted.
- Conformation with international nomenclature (International Code of Botanical Nomenclature, the International Code of Zoological Nomenclature and the International Code of Nomenclature of Bacteria): The following changes were made: *ssp.*, *sub.* and *subspecies* were replaced by *subsp.*; *variant* was replaced by *var.*; *forma* was replaced by *f.*; *sensu lato* was replaced by *s.l.*; *cfr.* was replaced by *cf.*; × (used for hybrids and nothotaxa) was replaced by *x* . Spaces between x and epithets for nothotaxa were removed.
- Taxon rank specifications (*agg.*, *cf.*, *s.l.*, *s.str.* and *hyb.*) were removed from the species name and added to an extra column ('Taxon rank').
- Author's names were removed from the species name and added to an extra column ('Author').
- Special characters were substituted, according to international nomenclature codes.
- Supplementary acronyms, numbers, and text were deleted.
- Dots were added or deleted where necessary.

For taxon names with multiple taxon ranks (e.g. *subsp.* and *var.*), the correct taxon rank was identified through cross checking with the following international systems in a hierarchical way: first in World Register of Marine Species (WoRMS) and in the Integrated Taxonomic Information System (ITIS), then in the Catalogue of Life (CoL) and in the Global Biodiversity Information Facility (GBIF).

After the standardization of the species name notation, spelling errors (max. 3 characters) were identified and corrected with the Excel add-in *'Fuzzy duplicate finder'* (AbleBits, Homel, Belarus; http://www.ablebits.com). Pairs of names obtained this way (potential duplicates) were cross-checked in a hierarchical way with WoRMS, ITIS, CoL, Encyclopedia of Life (EoL), and GBIF, to assess which entry was valid. If the questioned names were both valid, they were both kept as different taxa, otherwise the invalid name was replaced by the valid one.

Checking synonyms

Synonyms present in the list were identified through a cross-check with ITIS, WoRMS, and CoL, and a further correction of invalid names and removal of duplicate records was done. Synonyms were checked in WoRMS using the "match taxa" tool provided in the WoRMS website, in ITIS by extracting them from the downloadable version of the databases, and manually in CoL.

Compilation of the taxonomic tree

The taxonomy (Kingdom, Phylum, Class, Order and Family) of each alien taxon was retrieved in a hierarchical way from WoRMS, ITIS, CoL, EoL and GBIF. Note that not all species names could be matched with WoRMS and ITIS, but often the taxonomy could be resolved through these sources by using the genus name instead. The taxonomy of the few alien taxa that remained unresolved after cross-checking with WoRMS, ITIS, CoL, EoL and GBIF was completed by following the source database or through literature search. The six-kingdom (Bacteria, Protozoa, Chromista, Plantae, Fungi, Animalia) taxonomic system for cellular organisms proposed by Cavalier-Smith (2004) was followed, with two additional categories, i.e. 'viruses' (non-cellular forms) and 'unresolved' (taxa with yet unresolved taxonomy).

Identification of the environment

For each species, the environment(s) where it can be found were identified following this procedure:

Marine:

- Present in marine databases (World Register of Marine Species: WoRMS; Flanders Marine Institute: VLIZ; Främmande arter i svenska hav, Alien species in Swedish seas; JNCC report on Non-native marine species in British seas; Marine alien species of Estonia database)
- Listed as present in a sea in DAISIE
- Listed as "estuaries and brackish area" or "marine habitat" in NOBANIS
- Listed as marine in Ellenic Network on Aquatic Invasive Species (ELNAIS); Fish of Latvia; Aquatic-aliens.de

Appendix 1 (continued)

Freshwater:

● Present in freshwater databases (Freshwater Animal and Diversity Assessment: FADA; Register of Freshwater Plants by CEH)

● Listed as present in "estuaries and brackish area", "lakes", "watercourses" or "wetlands" in NOBANIS

● Listed as aquatic inland in DAISIE

● Listed as freshwater in Ellenic Network on Aquatic Invasive Species (ELNAIS); Fish of Latvia; Aquatic-aliens.de

Terrestrial:

● Listed as Terrestrial in DAISIE (excl. aquatic inland)

● Listed as present in "agricultural areas", "arctic/alpine habitat", "boreal forest", "coastland", "disturbed areas", "grass and heathlands", "greenhouses or compost heaps", "mixed conifer/broadleaf forest", "riparian zones", "rocks and lavafields", "shrublands", "temperate broadleaf forest", "under human management" or "urban areas" in NOBANIS

For unresolved names, the environment was added based on literature and expert judgement. The 'marine' category includes offshore, coastal and estuarine species; the latter have been included in both the 'marine' and the 'freshwater' environments.

Identification of pathways
Identification of pathways is based on expert contribution. An adaptation of the framework proposed by Hulme et al (2008) has been followed. Five main pathways have been included (each one divided into more specific sub-categories): Release (Biocontrol; Game animals; Landscaping-Erosion control; Pets, Terrarium-Aquarium species; Other); Escape (Cultivation and Livestock; Aquaculture; Ornamental planting; Use of live food-bait; Pets, Terrarium-Aquarium species; Zoos, botanical gardens); Contaminant (Trade of contaminated commodities; Packaging materials; Aquaculture); Stowaway (Shipping; Aviation; Land transport); Corridor (Lessepsian migrants; Inland canals; Railroads and Highways).

Impact classification (high, low/unknown)
The impact of species present in the 'high-impact' or 'worst invasive' species lists of DAISIE, GISD and SEBI-2010 was classified as 'high', while for the rest species as 'low/unknown'.