



e-science and technology infrastructure for biodiversity data and observatories

---

## **Deliverable 5.1.1**

### **Data & Modelling Tool Structures**

#### **- LifeWatch Showcases -**

Cardiff University

Work Package 5a Construction plan strategy

Life Watch  
e-Science and Technology infrastructures  
for biodiversity data and observatories

EU Seventh Framework Programme (FP7) Infrastructures (INFRA-2007-2.2-01)

Contract Number 211372



<b>Security (distribution level)</b>	<b>Public</b>
Actual date of delivery of v0.1	2008-01-08
Deliverable number	5.1.1
Deliverable name	Data & Modelling Tool Structures – Showcases
Type	Report
Status & version	Working Document v0.2
Number of pages	24
WP contributing to the deliverable	5a
Task responsible	Cardiff University
Other contributors	Fraunhofer IAIS
Author(s)	Jon Giddy  Alex Hardisty

## Revision History

<b>Version</b>	<b>Description</b>	<b>Changed</b>	<b>Date</b>	<b>Done by</b>
V0.1	1 <sup>st</sup> Draft	Based on Appendix of the LifeWatch Reference document	2009-12-13	A. Hardisty J. Giddy
V0.2	2 <sup>nd</sup> Draft	Modified & extended	2009-12-18	J. Giddy
V0.3	3 <sup>rd</sup> Draft	Modified	2010-01-06	J. Giddy
V0.4	4 <sup>th</sup> Draft	Modified to match the standards for LifeWatchh deliverables	2010-01-07	A. Poigné
V0.5	5 <sup>th</sup> Draft	Modified	2010-01-08	J. Giddy

## Table of Contents

<b>1</b>	<b>REQUIREMENTS GATHERING</b> .....	<b>1</b>
<b>2</b>	<b>ANALYSIS</b> .....	<b>3</b>
<b>3</b>	<b>SHOWCASES</b> .....	<b>6</b>
3.1	BIODIVERSITY VALUATION MAPS .....	6
	<i>Show Case 1: Biodiversity Richness Analysis and Conservation Evaluation</i> .....	6
	<i>Show Case 2: Biological Valuation Map</i> .....	6
	<i>Show Case 3: Automated Retrieval</i> .....	7
3.2	ECOLOGICAL NICHE MODELLING .....	7
	<i>Show Case 4: Past Behaviour and Future Scenarios</i> .....	7
	<i>Show Case 5: Bioclimatic Modelling and Global Climate Change</i> .....	9
	<i>Show Case 6: Phylogenetic Analysis and Biogeography</i> .....	10
	<i>Show Case 7: Ecological Niche Modelling</i> .....	11
	<i>Show Case 8: Urban Development and Biodiversity Loss</i> .....	14
	<i>Show Case 9: Renewable Energy Planning</i> .....	16
	<i>Show Case 10: Hierarchical scaling of biodiversity in lagoon ecosystems along environmental gradients</i> .....	16
3.3	REAL-TIME MONITORING .....	17
	<i>Show Case 11: Bird Strike Monitoring</i> .....	17
	<i>Show Case 12: Earth Observation</i> .....	18

## List of Figures

Figure 1: Workflow for automated retrieval ..... **Fehler! Textmarke nicht definiert.**

# 1 Requirements gathering

LifeWatch provides infrastructure biodiversity research in Europe. Biodiversity research themes to be covered include for instance

- Discovery of biodiversity
- Biodiversity patterns - mapping hot spots
- Biodiversity processes - monitoring changes
- Systems biology
- Nature conservation and management

*Showcases* are examples of the kinds of scientific studies that biodiversity researchers would like to be able to undertake within the context of the above themes. They are less categorical than IT use cases and better to communicate to the heterogeneous user group of LifeWatch. Ideally, for every biodiversity research theme a list of showcases should be supplied that describe how typical research questions are tackled using which kind of data and software resources.

In order to gather the functional requirements for the LifeWatch infrastructure, for each research theme, we collected several examples of existing work. We refer to examples as Showcases, as there was some confusion over the ideas of a use case and of a scenario. Hence, these Showcases provide a little of both: an explanation of actual work that has been performed which would hopefully have been easier with the LifeWatch infrastructure in place, and example scenarios for how things might work better with appropriate infrastructure. The example Showcases were primarily gathered through contact with researchers involved with the LifeWatch project, through site visits to major institutions, or through the published work of biodiversity researchers.

Table 1 demonstrates the coverage of the research themes by the Showcases.

No	Showcase	Discovery of biodiversity	Biodiversity patterns	Biodiversity processes	Systems biology	Nature conservation
1	Biodiversity Richness Analysis And Conservation Evaluation	*				*
2	Biological Valuation Map		*			*
3	Automated Retrieval	*	*			
4	Past behaviour and Future Scenarios			*	*	
5	Bioclimatic Modelling and Global Climate Change	*		*		
6	Phylogenetic Analysis and Biogeography	*		*	*	
7	Ecological Niche Modelling			*	*	
8	Urban Development and Biodiversity Loss			*		*
9	Renewable Energy Planning			*		*
10	Hierarchical Scaling of Biodiversity in Lagoon Ecosystems				*	*
11	Bird Strike Monitoring		*			
12	Earth Observation	*	*			*

The intention is not that these Showcases are implemented. Instead we plan to use the Showcases to derive an understanding of the generic capabilities that users require in order to perform typical biodiversity science.

From the Showcases, we extract the common functionality to provide a list of capabilities. These capabilities could be combined to implement the documented Showcases, but are generic enough to implement other biodiversity science. In particular, we hope that these generic capabilities are useful for scientists to implement novel science of which we are not yet aware. These capabilities implicitly state requirements on the underlying ICT capabilities that make them possible and will be used as a guideline for construction of the ICT capabilities provided by LifeWatch.



## 2 Analysis

Following collection of the Showcases, we categorised them into several major groups, in order to reduce the analysis required. We then analysed the documented procedures for services that would be required in order for a user to perform the described processes.

Primarily, the Showcases divide into three major areas:

1. Biodiversity Valuation Maps - Analysis and synthesis of historic data to calculate current biodiversity distributions and functional value.
2. Ecological Niche Modelling - Extrapolation of known species occurrences to determine environmental envelopes and predict future distributions due to species interactions and effects of potential changes (e.g. introduced pests, climate change, urban development).
3. Real-time Monitoring - Fast publishing and analysis of biotic and abiotic observations for relatively short-term prediction and policy enactment (e.g. bird strike, snow cover).

Figure 1 shows a workflow taken from documentation for one of the Showcases: the Automated Retrieval Showcase. While this workflow summarises one particular Showcase, it provides a good idea of the generic workflow evident in most of the Showcases:

- Discover, select and retrieve the required data, and transform to common-scale layers.
- Perform some processing on the data, ranging from the simple combination of the layers to generate a distribution map, to computationally intensive predictive models for generating what-if scenarios.
- Export the data for further processing, storage, or visualisation.
- Analyse and interpret the results to inform further actions, such as additional data gathering or policy development.

The three classes of Showcase demonstrate this workflow with particular attributes:

1. Biodiversity Valuation Maps: Step 1 of the workflow requires complex analysis in order to validate data and synthesise missing data to provide a complete picture of the existing situation.
2. Ecological Niche Mapping: Step 2 of the workflow requires computationally-intensive modelling in order to predict future interactions of multiple parameters
3. Real-time monitoring – the entire workflow is performed repeatedly with a fast turnaround time, implying minimal human interaction and good automated recognition capabilities.

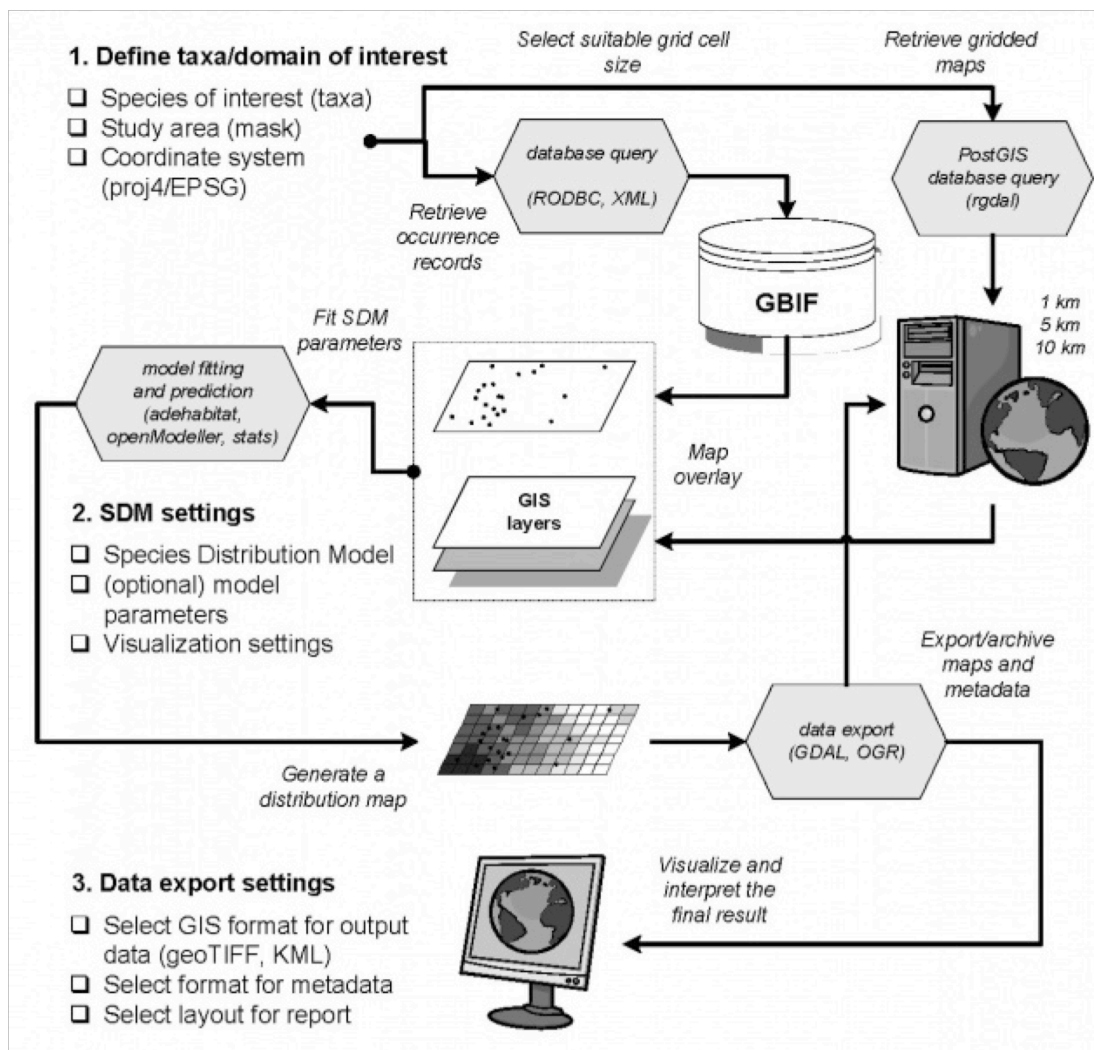


Figure 1: Workflow for automated retrieval

To identify capabilities required for implementation of a Showcase, we obtain a step-by-step procedure for how the work is currently being performed without the existence of a LifeWatch infrastructure. We analyse which generic capabilities each step requires in order to perform the described task. Since the steps in the described procedures are essentially user-driven, not just specified for this exercise, it is quite common for a single step in the procedure to require multiple capabilities.

Capabilities identified for the implementation of these Showcases include:

1. Discovery – the ability to obtain references to a subset of data that matches specified criteria.
2. Metadata – the ability to obtain additional information about the use of data (e.g. licensing restrictions)
3. Retrieval – the ability to retrieve data using a reference
4. Schema mapping – the ability to translate data queries or responses to alternative protocols
5. Format conversion – the ability to translate data to alternative formats
6. Thesaurus – the ability to map queries and results from different vocabularies (languages) without unnecessary loss of information.
7. Gazetteer – the ability to map descriptive terms for geospatial regions to a coordinate reference system.
8. Coordinate Transformation – the ability to map between different Coordinate Reference Systems.
9. Gap analysis – the ability to analyse data to determine where further observations need to be made in order to produce a better understanding of the state of a region

10. Data synthesis – the ability to derive non-observed data points from related existing data (e.g. to determine an estimated value for a missing date based on observed values for similar dates, or for a geographic region based on observed values for sub- or super-regions)
11. Processing – the ability to apply computational processing steps to data in order to transform it into useful analysis.
12. Workflow repository – the ability to store and retrieve transformation
13. Data provenance – the ability to link source and derived data, and the transformational workflows performed
14. Data quality – the ability to evaluate the quality of derived data by reference to the source data and transformations.
15. Citation – the ability to account for contributions to generated results through tracking of the contributions of source data and workflows.
16. Usage tracking – the ability for a data or workflow provider to analyse the usage of their data or tools, for example, to justify further provision or to enforce licensing restrictions.
17. Portrayal – the ability to display data in a manner that assists humans in the understanding of the data (e.g. mapping and visualisation)
18. Annotation – the ability for users to provide information about data without requiring the ability to edit the data at source.
19. Communication – the ability of users to interact with each other
20. Collaboration – the ability of users to work together on a common goal

This provides a basic list of the functional requirements for LifeWatch. There are additional non-functional capabilities that are not identified through the Showcases, such as user authentication and virtual organisations/collaborative networks. These requirements are documented in additional documents including the LifeWatch Reference Model and the Collaborative Networks document.

## 3 Showcases

### 3.1 Biodiversity Valuation Maps

#### Show Case 1: Biodiversity Richness Analysis and Conservation Evaluation

##### User

Biodiversity Scientist

##### Procedure

Step	Capabilities
User supplies taxon name	User Interaction
System retrieves authoritative list of names and synonyms from Species 2000 database	Thesaurus
System composes a distribution data set of specimens or observations from a variety of sources	Retrieval
System uses distribution data set to compute a wide range of diversity measures and displays on a species richness map	Portrayal
User uses species richness map to identify areas of high species richness	User interaction

##### Source

J. S. Pahwa, R. J. White, A. C. Jones, M. Burgess, W. A. Gray, N. J. Fiddian, T. Sutton, P. Brewer, C. Yesson, N. Caithness, A. Culham, F. A. Bisby, M. Scoble, P. Williams and S. Bhagwat (2006). Accessing Biodiversity Resources in Computational Environments from Workflow Applications.

#### Show Case 2: Biological Valuation Map

Creating a Biological Valuation Map requires consensus among scientists on how to interpolate and synthesise the multiple data sources, and how to weight the components in order to create a value index for locations within the region of interest.

##### Procedure

Step	Capabilities
Obtain data on all ecosystem components for which detailed spatial distribution data are available	Discovery
Create map layers based on existing species surveys	Retrieval, Coordinate Transformation, Format Conversion
For point data, interpolate and extrapolate to cover the entire region using GIS, identify areas of poor coverage and conduct further surveys	Data Synthesis, Gap Analysis

For data with incomplete coverage, develop predictive models and use physical habitat data to predict coverage, and conduct targeted surveys to validate model	Processing, Validation
Integrate data with full coverage of region to create a baseline map	Portrayal
Collaborate with all stakeholders to develop a consensus on valuations to be applied to ecosystem components	Communication, Collaboration, Annotation
Evaluate map layers according to set valuing criteria	
Integrate data with incomplete coverage as an "upgrade" for specific sub-regions	Data Synthesis

### Source

<http://www.vliz.be/projects/bwzee/background.php>

## Show Case 3: Automated Retrieval

### Procedure

Step	Capabilities
Obtain list of species names from EBCC Atlas website	Retrieval
Map species name to GBIF ID	Thesaurus, Discovery
Retrieve species counts mapped to one-arcdegree cells in KML from GBIF	Retrieval, Coordinate Transformation
Convert KML format to ArcInfo ASCII grid format, converting points to grid squares	Schema Mapping, Format Conversion
Perform Principal Component Analysis to determine correlations between distributions of different species.	Processing
Produce an X-Y plot of correlations between species	Portrayal
Derive the Biodiversity Index to determine species richness and evenness	Processing
Produce a map showing the Biodiversity Index for species.	Portrayal

### Source

T. Hengl and H Saarenmaa. Automated retrieval and analysis of GBIF records in R: some examples with European breeding birds. Global Ecology and Biogeography, submitted.

## 3.2 Ecological Niche Modelling

### Show Case 4: Past Behaviour and Future Scenarios

The generic question is: "how has {specified biodiversity or ecosystem attribute} been affected by {specified pressure or driver} over a specified period {past or present} and what might happen in the future {e.g. forecasts for different scenarios}?"

Capabilities required include:

- search for data on specified attribute (time, climatic zone, protected area status, soil type, geology, socio-economic parameters etc....)
- search for derived products and knowledge that may be relevant to this enquiry

- search for synonyms of specified attributes (biodiversity or ecosystem data or drivers/pressures)
- specify spatial or temporal constraints on search
- map sites and locations with potentially available data meeting search criteria
- evaluate “fitness for purpose” and data quality constraints of data sources by reporting quality descriptors and methods
- provide other techniques for data visualisation to assist process of data discovery
- provide simple tools for “gap analysis” to aid process of data discovery and assessments of fitness for purpose
- access to log of lessons learned and examples of best practice related to this enquiry
- show constraints on data use and usage (eg licensing)
- simplify and Facilitate data licensing and access arrangements
- show or enable specification of formats of data to be supplied
- send data or enable access to data in specified formats
- provide simple tools to explore and map relationships between attributes and pressure/drivers
- simple on-line capability to model and map relationships between biodiversity attributes and pressures;
- search database of scenarios related to pressures/drivers of interest;
- simple on-line capability to model and map relationships between biodiversity attributes and pressures in relation to forecasts;
- search catalogue of products derived from previous uses of LifeWatch Infrastructure
- show exemplar products
- provide capability to store successful products for use by other users
- provide list of acknowledgements for all data used so that they can be properly acknowledged and cited
- user feedback mechanisms to provide assessments of quality of data and services

## Procedure

Step	Capabilities
search for data on specified attribute (time, climatic zone, protected area status, soil type, geology, socio-economic parameters etc....)	Discovery
search for derived products and knowledge that may be relevant to this enquiry	Provenance
search for synonyms of specified attributes (biodiversity or ecosystem data or drivers/pressures)	Thesaurus
specify spatial or temporal constraints on search	Discovery, Geolinking
map sites and locations with potentially available data meeting search criteria	Portrayal
evaluate “fitness for purpose” and data quality constraints of data sources by reporting quality descriptors and methods	Data quality
provide other techniques for data visualisation to assist process of data discovery	Portrayal
provide simple tools for “gap analysis” to aid process of data discovery and assessments of fitness for purpose	Gap Analysis
access to log of lessons learned and examples of best practice related to this enquiry	Provenance, Workflow Repository
show constraints on data use and usage (eg licensing)	Annotation
simplify and Facilitate data licensing and access arrangements	Usage tracking
show or enable specification of formats of data to be supplied	Schema Mapping
send data or enable access to data in specified formats	Scheme Mapping, Format Conversion

provide simple tools to explore and map relationships between attributes and pressure/drivers	Processing
simple on-line capability to model and map relationships between biodiversity attributes and pressures;	Processing
search database of scenarios related to pressures/drivers of interest;	Discovery
simple on-line capability to model and map relationships between biodiversity attributes and pressures in relation to forecasts;	Processing
search catalogue of products derived from previous uses of LifeWatch Infrastructure	Citation, Workflow repository
show exemplar products	Workflow Repository
provide capability to store successful products for use by other users	Workflow Repository
provide list of acknowledgements for all data used so that they can be properly acknowledged and cited	Citation
user feedback mechanisms to provide assessments of quality of data and services	Annotation, Data Quality

### Source

Provided by Terry Parr, WP7

## Show Case 5: Bioclimatic Modelling and Global Climate Change

Using models of expected changes in climate, the effect on species in the future can be predicted. For example, new conditions may lead to populations becoming more threatened, or to particular variations dominating.

### User

Biodiversity Scientist, Ecologist

### Procedure

Step	Capabilities
User supplies taxon name	
System retrieves authoritative list of names and synonyms from Species 2000 database	Discovery, Thesaurus
System composes a distribution data set of specimens or observations from a variety of sources	Format conversion
System uses distribution data set to compute a wide range of diversity measures and displays on a species richness map	Portrayal
User uses species richness map to identify areas of high species richness	Processing
Evaluate the impact of past climate change on the evolution of	

species	
Model the envelope of climatic and ecological conditions under which a single species lives, deducing the known features of places where it is recorded.	Processing
Calculate a wider set of areas where the species might occur, or predict its future distribution under changing climatic conditions and project these onto a map of the world.	Discovery
Predict the responses of species which may become endangered, or a pest presenting a new or increased threat	Processing

### Source

J. S. Pahwa, R. J. White, A. C. Jones, M. Burgess, W. A. Gray, N. J. Fiddian, T. Sutton, P. Brewer, C. Yesson, N. Caithness, A. Culham, F. A. Bisby, M. Scoble, P. Williams and S. Bhagwat (2006). Accessing Biodiversity Resources in Computational Environments from Workflow Applications.

### Show Case 6: Phylogenetic Analysis and Biogeography

Using phylogenetic analysis, historic distributions of species lineages can be determined and integrated with historic climate data to understand the impact of climate change on distributions of related species in the past.

### User

Paleoecologist

### Procedure

Step	Capabilities
User supplies taxon name	
System retrieves authoritative list of names and synonyms from Species 2000 database	Thesaurus
System composes a distribution data set of specimens or observations from a variety of sources	Retrieval, Format Conversion
System uses distribution data set to compute a wide range of diversity measures and displays on a species richness map	Processing
User uses species richness map to identify areas of high species richness	
Search the EBML DNA sequence database for sequences	Discovery
Use the DNA sequences to produce a phylogeny using parsimony or maximum likelihood techniques	Processing
Estimate the age of species and lineages using temporal calibration	Processing
Develop ancestral bioclimatic models for a particular species lineage	Processing
Combine with (pre-)historical climate models	Schema mapping, Coordinate Transformation
Evaluate the impact of past climate change on the evo-	



Evolution of species	
----------------------	--

## Source

J. S. Pahwa, R. J. White, A. C. Jones, M. Burgess, W. A. Gray, N. J. Fiddian, T. Sutton, P. Brewer, C. Yesson, N. Caithness, A. Culham, F. A. Bisby, M. Scoble, P. Williams and S. Bhagwat (2006). Accessing Biodiversity Resources in Computational Environments from Workflow Applications.

## Show Case 7: Ecological Niche Modelling

This Showcase consists of the exact steps quoted from the source paper. Note that Steps 1 and 9 are normally purely manual steps not done within the system under consideration.

## Procedure

### *Step 1 – problem definition*

Problem definition consists of defining the modelling experiment, which includes the specification of a set of questions that should be answered by the model, considering the selected species, the area to be studied and the geographic resolution of the environmental layers. The ecological niche model should help to understand the environmental situation, indicating possible answers or ways to consider the problem.

The researcher must position the study area, considering the availability and quality of information about the species to be studied and depending on the modelling purpose. This region can be imagined as a two-dimensional matrix of cells with specific size. Thus, the definition of the size of each cell, known as resolution, is mandatory. Decisions about the resolution to use in different situations are directly connected to the quality of the species point data available (e.g. georeference quality of the species data: georeferenced at point level, or at area, locality or county level), size of the selected area, resolution of the environmental data available, time that a model takes to be generated and questions that should be answered by the model. Lower resolution data will furnish models that are useful for big areas and for more general questions about species distribution. Higher resolution data will provide models that are suitable for answering questions in local (small) areas. Since algorithms usually verify each cell of the region, higher resolution implies reduced model throughput by each period of time. However, if the resolution is not high enough (the cell size is not small enough), the result may lack accuracy in the generated model. In the following steps (2 and 3), data acquisition and preparation are discussed. If the researcher is not sure about what to consider as input data, these steps may be performed prior to step 1.

### *Step 2 – species points treatment*

Species points' treatment consists of preparing the species points for the modelling process. Some common problems usually emerge at this step, e.g. digitization errors, points outside the limits of the region and points with absurd values. So, georeferencing tools (such as gazetteers) may be applied to position the occurrence points in the geographical space. Both species presence and absence points must be in the form of georeferenced coordinates. The quality, number and distribution of these points are directly related to the accuracy of the model.

If the researcher does not have the biotic data, an alternative is to access web sites where biological collection data are available, such as GBIF [<http://www.gbif.org/>] and SpeciesLink [<http://splink.cria.org.br/>]. Once the data is found, it must be downloaded and treated, using specific software packages (e.g. Spreadsheets, GISs). Some possible inaccuracies in the species points can be minimized by using the tools available at the SpeciesLink web site, which offers: Data Cleaning (to facili-

tate the detection of possible errors and to help in the process of standardizing data in collections); Geoloc (to georeference data based on locality descriptions); Info X Y (to get locality information using geographic coordinates); Spoutlier (to detect outliers in latitude and longitude); Converter (to convert different types of representations of geographic coordinates and datum).

openModeller desktop (or later versions) offers the “search for locality data tool” option, so as to look for biotic data by means of GBIF or SpeciesLink plug-ins. Species data (species name, longitude and latitude) can be furnished to openModeller in a plain text format, separated by the <tab> character.

### *Step 3 – environmental data treatment*

Environmental data treatment consists of identification, acquisition and conversion (if necessary) of the environmental data required to generate a model, e. g. one or a set of layers for climate, topography, and vegetation index, usually formatted as environmental raster layers.

Preparation of environmental layers is one of the most time-consuming and computer intensive areas of modelling (Chapman et al., 2005). In order to use environmental data that are usually available in different data formats, georeference coordinate systems, projections and resolutions, a previous work is required to make them compatible. The aim is to put them all in the same format and metadata, so as to be compatible with other software packages, such as Desktop Garp [<http://nhm.ku.edu/desktopgarp>] and MaxEnt [<http://www.cs.princeton.edu/~schapire/maxent>]. These requirements can demand mechanical efforts, involving different software packages, such as GIS packages, for data conversion.

Many problems related to format and resolution are already solved by openModeller, which has integrated the GDAL library [<http://www.gdal.org/>] to its source code. So, openModeller can automatically perform the same data conversions furnished by the library without extra effort from the user.

### *Step 4 – data viability analysis*

Data viability analysis refers to the analysis of the data obtained in the previous steps, so as to evaluate the conditions to proceed with the modelling process. The aim is to decide if they suffice to answer the scientific questions or if more data must be included.

Decisions must be taken that will affect the posterior steps, such as:

- “Is the resolution of the available (biotic and abiotic) data compatible with the question to be answered by the modelling process?” “Is it possible to adopt the resolution required for a more precise model?” These questions must consider the resolution of the data available, the characteristics of the species and the question to be answered by the ecological niche model.

- “Which layers should be actually considered?” Some layers may not be useful, considering the complete modelling scenario and the environmental and ecological requirements of the species’. The complexity involved in such decision must be deeply understood and refinements represent one of the most time-consuming steps in the process.

### *Step 5 – algorithms choice*

Algorithms choice consists of selecting algorithms to generate ecological niche models, considering available biotic and abiotic data. Different algorithms have different parameters and requirements, and the data available may be more adequate to the generation of a model using some specific algorithm. Also, different applications present different challenges for modelling algorithms (Peterson et al., 2007b).

Basically, one algorithm represents an implementation of a mathematical method to analyze data in order to obtain an ecological niche model. In essence, algorithms extrapolate from associations between point occurrences and environmental data sets to identify areas of predicted presence on the map, producing the so-called “correlative approach” to ecological niche modelling (Soberon & Peterson, 2005).

The following algorithms are available at openModeller: Distance to Average (Munõz 2004a); Minimum Distance (Munõz, 2004b); Environmental Distance [<http://openmodeller.sourceforge.net>]; Bioclim (Nix, 1986); Climate Space Model (Sutton and Giovanni, 2004); GARP– Genetic Algorithm for Rule-set Production ([Stockwell and Peters, 1999] and [Stockwell and Noble, 1992]); GARP BestSubsets – Genetic Algorithm for Rule-set Production with Best Subsets Procedure (Anderson et al., 2003 R.P. Anderson, D. Lew and A.T. Peterson, Evaluating predictive models of species' distributions: criteria for selecting optimal models, *Ecological Modelling* **162** (2003), pp. 211–232) desktop and openModeller implementation, and Support Vector Machine SVM (Giovanni and Lorena, 2007).

#### *Step 6 – parameter definitions*

Once an algorithm is chosen, it is necessary to define the required parameters to its execution. Usually, parameters are numeric values passed to algorithms as inputs, so as to control some aspects of their behaviour. Each algorithm has its own parameters, and it is not possible to generalize their application. However, a wrong choice may result in a model completely different from what it should be; therefore, it is strongly recommended to understand the meaning of each parameter.

openModeller suggests default values for each algorithm, based on the literature and on the modelling expertise of its users. For a beginner, using such defaults could represent a starting point to generate a model but probably performing further refinements will be necessary, changing the values of the parameters and repeating the experiment.

#### *Step 7 – model(s) generation*

In the model(s) generation step, the software package runs the experiment using the input data furnished to the algorithm and then generates a model.

While many packages offer a single algorithm choice only, openModeller is able to generate several models using different algorithms in the same experiment. openModeller also offers the possibility of generating more than one model with the same algorithm but different parameters in a single experiment.

#### *Step 8 – automatic post analysis*

Automatic post analysis consists of the evaluation of each generated model, in order to assess whether it is adequate for the species. Statistical measures may be calculated in order to indicate the performance (accuracy) of the model. For instance, presence and absence points not applied for the model generation or obtained in field may be used to validate the model, integrated to the use of certain statistical strategies.

openModeller furnishes a ROC analysis. A *Receiver Operating Characteristic (ROC)*, or simply *ROC curve*, is a graphical plot of the sensitivity vs. the specificity for a binary classifier system according to its discrimination threshold variation. ROC can also be represented equivalently by plotting the fraction of true positives (true positive rate) vs. the fraction of false positives (false positive rate) (Fielding and Bell, 1997). It provides tools to select possibly optimal models and to discard suboptimal ones independently of the cost context or the class distribution. This analysis is related in a direct and natural way to a cost/benefit analysis of diagnostic decision-making. Widely used in medicine, radiology, psychology and other areas for many decades, it has been introduced relatively recently in other areas such as machine learning, data mining and ecological modelling. The AUC (Area Under Curve) in ROC analyses is also calculated, representing a value which is used to validate the resulting model and to compare the performance among different algorithms (Phillips et al., 2006).

#### *Step 9 – researcher validation*

In this last step the researcher must decide if the model is acceptable. The knowledge of the researcher about the study species distribution is the main criterion, combined with the statistical results, for validating the model.

If the model is satisfactory, the process is finished. Otherwise, it may be necessary to return and restart the process, partially or fully, until a good sample is produced. openModeller does not present restrictions about a place to restart and the information applied in the generation of a model is stored, so the researcher can go back to any previous step. Among the main reasons for doing that are lack of data, and algorithm or parameter choices. Modelling algorithms are usually based on statistical methods, probability or heuristics (e.g. GARP), so the same input can produce different results in different model generations. Despite the sequence suggested by the presented reference business process, a different order can be adopted although some obvious logical restrictions apply.

## Source

F.S. Santana, M.F. de Siqueira, A.M. Saraiva and P.L.P. Correa. A reference business process for ecological niche modelling. *Ecological Informatics* 3(1):75-86, 1 January 2008.  
doi:10.1016/j.ecoinf.2007.12.003

## Show Case 8: Urban Development and Biodiversity Loss

### Policy Issue – User Need

The relationships between urbanisation and biodiversity impacts as a result of the peripheral growth of the city on the surrounding hinterland is critical to the conservation of biodiversity in Europe today. The EEA report on urban sprawl in Europe demonstrated for the first time, based on new pan-European intelligence on these issues, the major impacts that unparalleled growth and sprawl of the city is having on biodiversity loss. Natura 2000 sites in Benelux are suffering from the encroachment of urban development spilling out from the densely populated regions of northern Europe. At the same time similar impacts are evident around the costas of the Iberian peninsular, and the special protection areas of the new member states.

In addition to these biodiversity impacts the sprawling development of the city is fuelling primarily car based transport growth which forms a major contribution to the growth of greenhouse gas emissions driving climate change. New evidence demonstrates clearly that, car based transport growth associated with the sprawling cities of Europe is a main driver of the growth of transport related greenhouse gas emissions. In 2005 transport emissions accounted for almost 20% of greenhouse gas emissions in EU25. Road transport is responsible for 93% of all transport emissions and urban transport accounts for half of these emissions. The correlation between the growth of urban land and the growth of transport related greenhouse gas emissions is very strong, much stronger than for other factors including population growth and GDP.

The above indicates that appropriate policies to address the growth of transport related emissions in the context of the land-use - transport - environment nexus can secure a big positive impact on climate change and biodiversity priorities without limiting the economic or population growth of the city. Effective management of the city with the control of urban sprawl can create the “win win” opportunity to both address both the climate change impacts of urban development, and at the same time secure a major opportunity to conserve biodiversity.

### Common Problems and Common Solutions

The fundamental drivers of urban change and development are common throughout Europe. Indeed many of these drivers are global in scope. As a result cities throughout Europe are witnessing a range of development issues that have economic, social and environmental character that have common origins and common impacts. Typically the drivers of change create multiple impacts as for example the economic development of the city creates sectoral impacts in relation to the social and environmental characteristics of the city which are evident in differential spatial impacts.

The commonalities of these development issues invites a common approach to their management and resolution. However the complexity of the interactions and interconnections in both sectoral and spatial dimensions demands both deep seated analysis and evaluation, as well as coordinated action by a variety of agencies to secure appropriate and effective responses. Above all the requirement is for integrated approaches to the management of the complexity of the problems facing the city today.

### **Integrated Intelligence and Integrated Management**

Policy integration is universally viewed as a prime requisite for successfully addressing the complex interrelated problems at the urban level, and the interlinked concerns for the development of the appropriate urban model, and its economic and social, and environmental specification. The GMES and INSPIRE programmes aim to secure greater horizontal integration across the sectoral boundaries of the substantive urban policy themes, allied with vertical integration between the levels of governance.

The central proposition is that integrated intelligence, based on policy defined information need provides a key to unlock the prime concerns of the policy end user for a more integrated approach to policy formulation and decision making. The challenges faced by cities in the pursuit of these objectives include not only the definition of the technical specifications for the data models, but also the need to respond to the structural, procedural, professional and political implications of this process.

Many initiatives have been made to address these problems of urban development and management. integrated management systems have been developed in the context of the implementation of the Urban Thematic strategy (6EAP). On this basis and the principles of integrated urban management have been developed and many of the key elements of the data model specified in terms of datasets, indicators and models necessary to secure the policy related information to support effective decision-making have been specified.

### **Opportunity for Lifewatch**

The opportunity for Lifewatch in this field of urban management is substantial. Critical relationships have been identified between urban land use planning, transport, climate change and biodiversity loss that affect all cities of Europe today to a greater or lesser extent.

The problem for the urban manager today is that while many of the principles of integrated urban management have been developed, the necessary integration of data sets, indicators and other information that links land use change to transport development and biodiversity impacts is not available.

This fragmentation of the information and intelligence essential to support integrated policy definitions and policy implementation seriously impedes the effective response of the cities and regions of Europe to these challenges. Improved integrated intelligence offers a major opportunity to address and overcome these deficiencies in policy responses necessary to secure the outcomes that combines the delivery of sustainable urban development, climate change amelioration, and biodiversity conservation. The major opportunity for Lifewatch is to fill this gap and to create synergies with associated initiatives in this field to secure more effective and sustainable urban management.

### **Source**

David Ludlow, University of Western England

## Show Case 9: Renewable Energy Planning

Wind towers need to be located in areas with a consistent reasonably strong air flow, but also need to be placed for economic reasons near transmission facilities and access roads, and for conservation reasons may need to be placed away from particular areas. Finding suitable locations requires GIS mapping of climate, geographical, regional planning and ecological data.

This Showcase provides a simple example of the combination of data from multiple sources. An additional requirement is to conduct what-if explorations that are aware of the potential interactions and conflicts between geospatial objects from different databases.

### Source

Based on a use case suggested to WP7 by Oliviero Spinelli 17/04/09

## Show Case 10: Hierarchical scaling of biodiversity in lagoon ecosystems along environmental gradients

Transitional waters are selective ecosystems filtering species trait and determining species distributions at different level of scale. Landuse changes, habitat loss, chemical anthropogenic pollution as well as global climate changes are likely to affect the transitional water ecosystems potentially threatening their functions, as well as biodiversity conservation. To what extent species ranges and biodiversity are affected by the most selective environmental filters and, which scale(s) is(are) more relevant to biodiversity conservation in European transitional waters are still open questions, mainly due to a lack of organized data at the EU scale, preventing effective analyses and tests of the underlying hypotheses. Here, sharing the facilities of the LifeWatch infrastructure and the existing data platform realized by TWReferenceNET, the 4 research networks on transitional waters that have already expressed their interest to contribute to LifeWatch are working as data providers and users to address these questions. The Showcase is focusing on benthic macroinvertebrates, since most data are available, but can be extended to phytoplankton and phytobentos.

### Data Providers

- TWReferenceNET (Benthic invertebrates, phytoplankton, environmental data)
- Balloon (Benthic invertebrates, phytoplankton, environmental data, landuse data)
- Elnet (Benthic invertebrates, phytoplankton, environmental data, landuse/pressure data)
- LaguNet (Benthic invertebrates, phytoplankton, environmental data)

### Added LW Values

1. Supplying the infrastructure to merge the Network data
2. Providing support to taxonomic standardisation and QA of data
3. Providing support to spatial mapping of data
4. Making available different sources of data on the marine environment affecting colonization of lagoons by benthic macroinvertebrates

### Emerging results from LW

1. to address issues on biodiversity organization and conservation of macro-invertebrate and phytoplankton guild in lagoon ecosystems at large scale and sharing facilities that are rarely

- available even at the scale of large integrated project;
2. to create supporting evidences to refine national criteria of classification of transitional water within the WFD scheme
  3. Procedure
  4. inform the Networks and ask for members involvement as data providers and members of the Study Case working group
  5. data may be either on phytoplankton and/or macrobenthos, as species lists, but they have to be collected at the habitat type level;
  6. data should be described at the metadatabase level
  7. collect participation to the Study Case on the base of the requirement of point (1)
  8. collect and compile the Study Case data-base and allocate tasks (as lines of activities) to partner Networks
  9. revise taxonomic lists and map spatially explicit information
  10. develop data analysis and exchange preliminary results
  11. organize a side evening event inside the Montpellier to discuss the Study Case results and finalise the work
  12. finalise the Study Case as final presentation and paper submission

### Source

Provided to WP5 by LifeWatch Lagoon Networks TWReferenceNET, BALLOON, Elnet, LaguNet

## 3.3 Real-time Monitoring

### Show Case 11: Bird Strike Monitoring

#### Users

Airports/airlines

#### Procedure

- Data transfer from sensors (3 MPRs, WX Radars, SOI radar, Individual birds, ECMWF)
- Verify data files
- Process data
- Store in spatial database
- Integration of data sources
- Extract data for:
- Visualization
- Analysis & modelling
- Generate predictions (store in DB)
- Archive data files
- Provide book keeping services

### Source

FlySafe Bird Strike presentation at Institute for Biodiversity and Ecosystem Dynamics, Universiteit van Amsterdam, 2008-09-04

## Show Case 12: Earth Observation

Earth Observation data can provide:

- Near real time services
  - Snow monitoring, snow on ice (snow melt determines planting seasons)
  - Water quality and temperature
  - Algae blooms
  - Seasonal vegetation monitoring
  - Oil spill detection
- Long term services
  - Land cover, land cover change

### Procedure

- Raw EO data is retrieved
- Data calibration is performed: atmospheric correction, geospatial rectification, cloud masking (all automated, except cloud masking requires visual check)
- Cleaned data is provided within 3 hours of satellite overpass
- Data is combined and provided to users through standard interface

### Source

SYKE, Helsinki

<end of document>