NATURE-GIS: DEVELOPING A PAN-EUROPEAN GEOSPATIAL DATA MODEL FOR PROTECTED AREAS

David R. Green¹, Stephen D. King¹, Allessandro Annoni², Serge Margouliès³, and Jean-Phillipe Humblet³

¹Centre for Marine and Coastal Zone Management, Department of Geography and Environment, School of Resources, Environment and Society, Elphinstone Road, University of Aberdeen, Aberdeen, AB24 3UF, Scotland. Tel. +44(01224)272324; Fax. +44(01224)272331; Internet: http://www.abdn.ac.uk/cmczm

²JRC, Via E.Fermi 1, I-2102 Ispra (VA), Italy. Tel. +39(0332)78 9393; Fax. +39(0332)78 9222; Internet: http://www.jrc.it

³IONIC Software, 18, Rue de Wallonie, 4460 Grâce - Hollogne, Belgium. Tel. +32 4 364 0 364; Fax. +32 42 53 47 37; Internet: http://www.ionicsoft.com

∗corresponding author

1. INTRODUCTION

Nature-GIS is a European funded project of three years duration to develop a pan-European network for geospatial data for protected areas. It comprises ten workpackages to examine current and future stakeholder requirements for geospatial data, the development of a geospatial data model for protected areas, a demonstrator web service, and finally guidelines and a pan-European network. The project began in April 2002. This paper describes the initial research work and preliminary results of three of the workpackages, namely workpackages 2, 3, and 4 that are responsible for ascertaining the current and future requirements for geospatial data and information and the development of the data model. In conclusion, the paper briefly examines the throughput of these results into workpackage five that will use the data model as the basis for developing a Pan European Protected Areas Network.

2. CONTEXT

Today many organisations and individuals across Europe are responsible for the monitoring, management, and mapping of protected areas for biodiversity conservation, preservation of our natural and cultural heritage, and the development of sustainable recreation. In the context of the EU and its future evolution this is a complex problem. The rapid development of the geospatial technologies and Internet communications and networking technology offer a range of potential solutions to this problem. In practice a solution will involve technologists and end-users from many disparate organisations, in different countries, from different historical, language and cultural backgrounds.

However, whilst all stakeholders have a common goal, there are still many practical problems currently faced in achieving a common interface to GI at the pan-European level. Firstly, the management of protected areas requires access to geospatial data and
information. Realistically such data and information must be widely available to and usable by all stakeholders. Furthermore, a mechanism for sharing information is vital for such information to become a part of everyday activities [1]. This requires that geospatial data is collected to certain standards, is fully documented (metadata), and can be made widely available to all those who can potentially use and would benefit from such widespread access. Furthermore, there is a need for geographical data and information at a wide variety of different scales to provide a basis upon which to support the local requirements for implementing European policies and to inform decision-makers. At present, there is currently considerable difficulty in linking the local operator to EU policies. Additionally, for example, the Convention of Biological Diversity (http://www.biodiv.org/doc/publications/guide.asp) and the European Community Biodiversity Strategy (http://www.esaiweb.org/workshops/MartinNovella.ppt) require research, identification and exchange of information to ease and promote the conservation of biodiversity through the development of implementation plans, and tools and other management strategies. Furthermore, the nature protection issue is of key importance in, for example, Central and Eastern Europe where much of the countryside remains unspoilt with natural habitats and landscapes of considerable ecological value. The growing awareness of environmental issues and global change management has also emphasized the need for rich, timely and flexible information systems capable of supporting decision-making based on the analysis of spatio-temporal data. Another reason is that at the same time GIS tools are rapidly changing, downsizing from large legacy systems to desktop GIS promoting the idea of stronger user involvement in interactions with GIS. As a result, data modeling, data languages and user interfaces emerge as fundamental issues, where significant advances are now needed to meet the requirements of users whose profile and need have drastically changed (http://lbwww.epfl.ch/col/conferences/DEXA/cfp.html).

3. STATEMENT OF THE PROBLEM

At present, the widespread availability of geospatial data and information for protected areas in Europe is still relatively limited. In fact much of the existing geospatial data and information is poorly catalogued and documented and is generally not widely available to those who could benefit most. Furthermore, with the expansion of Europe there is now a need to provide a firm basis upon which future geospatial data and information resources for protected areas can be more widely used within Europe, and also integrated at a wide variety of different scales in the context of planning and decision-making. The rapid evolution of the Internet and communications and networking technology has opened up many avenues to facilitate the wider availability and use of data and information for protected areas. However, currently realising the vast potential of such technology in practice is often restricted by the huge gulf that exists between technical experts, the technology, and end-user communities. Whilst it is widely recognised by many organisations across Europe that geospatial data and Geographical Information System (GIS) technology is vital as the basis for developing future environmental management strategies for protected areas, it is clear that there is still a lack of awareness, knowledge and understanding about the availability of, requirements for and uses of the geospatial technologies, data and information for environmental monitoring and management.

4. THE NATURE-GIS PROJECT

Nature-GIS is an EU funded project that began in April 2002 extending for a period of three years. The project, comprising a total of ten workpackages, involves nineteen partners from across Europe and is co-ordinated by the Geographical Information Systems International Group (GISIG - http://www.gisig.it). The primary goal of the project is to improve the data and information flow to the European policy maker. This will be achieved through improving reporting, raising awareness, and supporting public access to geospatial data and
information about protected areas. It will also contribute toward the development and broadening of dialogue amongst all levels of protected areas responsibility at the European community, national, regional, and local level. In the long term Nature-GIS will seek to support public access to data and information, including non-governmental organisations and furthermore the stakeholders of the new European accession countries.

To achieve these goals, guidelines will be developed for the geospatial data infrastructures for protected areas. These guidelines will help to implement GIS in protected areas management, will demonstrate how Internet web access to information is applicable, raise European awareness for a supra-national approach in GI management, and highlight the requirements for data and information integration. The result will be the development of a pan-European Nature-GIS group. Development of Thematic Networks is considered to provide an effective way to answer many of the questions on how to address and solve problems involving a wide community of actors or stakeholders, and to provide a common framework.

In a wider context, the development of the Nature-GIS network will pay particular attention to the EU Natura 2000 network (http://www.natura2000benefits.org/uk/euro.htm) and to its interaction with other existing territorial entities for nature protection as well as with local actors (administrators, technical operators, scientists, non-governmental and citizens associations). The value of the Nature-GIS network lies with information management issues, and in particular the management of geographic information, that are not specifically addressed within the Natura 2000 framework. Currently there is a lack of consistent information at the local level and the Nature-GIS project will help to make this information available for the Natura 2000 network. Furthermore Nature-GIS has considerable relevance to a number of other European initiatives, and has, for example, recently been selected as one of the demonstrator projects for INSPIRE (http://inspire.jrc.it).

Fig. 1 shows the Nature-GIS framework. For further details of the Nature-GIS project readers are directed to Roccatagliata et al. [2] and the Nature-GIS website (http://www.gisig.it/Nature-GIS).

5. NATURE-GIS: STAGE 1 – WORKPACKAGES TWO, THREE, AND FOUR

This particular paper focuses on the initial stages of the Nature-GIS project, covering the preliminary research work undertaken to date by workpackages two, three, and four. These comprise the following:

- WP2: USER NEEDS ASSESSMENT
- WP3: DATE REQUIREMENTS AND DATA POLICY
- WP4: FUNCTIONAL REQUIREMENTS

The work to date comprises the design, testing and delivery of a survey questionnaire as the basis for the development of a data model for protected areas, that will ultimately be implemented as a pan-European web service, the Pan European Protected Areas Network.
6. QUESTIONNAIRE

A major component used as the basis for defining a data model for protected areas in the NATURE-GIS project entailed designing a survey questionnaire. The questionnaire comprised three components, each the responsibility of the three workpackages:
Mini-session 2.2 Data Policy

- SECTION 1 - WP2: ABOUT YOU AND YOUR ORGANISATION
- SECTION 2 - WP3: DATA REQUIREMENTS
- SECTION 3 - WP4: FUNCTIONAL REQUIREMENTS

To aid in the implementation and successful completion of the questionnaire, a
typology of stakeholders was devised to identify appropriate recipients of the questionnaire.
The typology is shown in Fig. 2.
The stages in the development of the Nature-GIS survey questionnaire included a pilot
survey, paper hardcopy delivered by mail, an electronic web-based form, and face-to-face
interviews usually as a follow-up to a completed paper questionnaire.
The questionnaire was independently developed in sections by each workpackage,
and integrated and revised prior to the pilot that was circulated to Nature-GIS partners.
Subsequently, trial questionnaires were used to refine and then produce a final version. The
project is now awaiting the receipt of the completed questionnaire survey that has been
circulated to a list of stakeholders for each country e.g. French, Italian, Portuguese, and
Estonian. To aid in their successful completion, the questionnaires have been translated into
the language of the country. Representative partners of Nature-GIS in each country are
responsible for questionnaire translation, delivery and receipt. The results from the
questionnaires will be input to an ORACLE database for analysis by the respective
workpackages. The online questionnaire is to be found at http://krown.uoeora.pt/~palma/Inquirer/
The overall purpose of this questionnaire has been to develop a clear understanding of
the current and future requirements for geospatial data by all pan-European stakeholders.
The questionnaire has been designed to provide insight into the current and future needs of
stakeholders upon which it will be possible to provide a comparative analysis of the
potentially different views of data by the technologists and the end-user communities as the
basis for developers of the data model. The completed survey questionnaire will also help to
provide information about what data types are missing, and should be included, the
relationships that should be included, what should be optional versus required, and will help
to clarify what Halpin and Treml [3] define as the need to consider the development of a
generic data model versus theme specific models. For example, in the final questionnaire a
list of data relevant to Protected Areas was included under the following themes or
headings: Land use, Ecosystems, Habitat/Species, Social/Economic/Political, Physical
Features, Built and Cultural Heritage. The questionnaire also asked whether people
Manage, Use or Need these types of data. These themes will likely form the main categories
of the data model, namely the database tables. Furthermore, the questionnaire asked
directly how people describe or model their data now i.e. the relationships between features,
and if they already use a reusable model/schema.

7. DATA MODELS
A data model can be defined as an abstract concept or a set of constructs for
describing and representing parts of the real world (essential features of a class of data) in a
computer system [4]. It provides technologists and end-users with a basis to develop a
common understanding and also as a reference point. Data models or schemas can also be
used to give semantics (meaning) to digital objects. For technologists a data model is the
basis for translating an application into a design and for the implementation of the system.
For end-users, it provides a description of the structure of the system, independent of
specific items of data or details of the particular application. Maguire and Grise [4] consider
data models to be vital for GIS because they define the way that data are stored and
determine the analytical operations that can be performed (see Fig. 3)
Traditionally a data model is described by identifying certain data entities, their
distinguishing attributes, and their relationships to one another
(http://www.cni.reston.va.us/home/cstr/arch/data-model.html). As many data models are now object-
oriented, besides attributes, specification must also include the essential operations on
these objects as well as defining the key attributes that are necessary to describe and interpret the data. These key attributes are effectively "data about data" or metadata.

Nature-GIS: Classification of Stakeholders

Fig. 2 Typology of Stakeholders (University of Aberdeen)
Reference to Halpin and Treml [3] indicates that the definition of a GIS data model is a basic template for implementing any GIS projects whether the activity relates to data input, format, geoprocessing, sharing, maps, analysis and so on. As such for GIS developers and suppliers a data model provides a basic framework for writing program code and maintaining applications [3]. As an example, in the context of ESRI data models (and in this case a coastal and marine data model) this involves defining the Primary ESRI Feature Classes (Points, Lines, Areas, Rasters) and Feature-Class relationships [3].

As noted by Maguire and Grise [4] data modelling is not easy as many people from different backgrounds use geographic data for different purposes, and there is no single type of data model that is best for all circumstances.

Usually in the data modeling process, end-users and technologists (so-called system developers) need to jointly participate in the process. A series of steps or stages for creating a data model are shown below [5]:

- SCIENTIFIC DATA MODEL AND USERS VIEW OF DATA
- DEFINE OBJECTS AND RELATIONSHIPS
- MATCH TO GEODATABASE ELEMENTS (SPECIFY RELATIONSHIPS, BEHAVIOURS)
- ORGANIZE GEODATABASE STRUCTURE

With reference to Fig. 4 definition of the main types of objects (entities) to be represented in a GIS and a conceptual description of the main types of objects and relationships between them is the first step [4] Subsequently this leads to the creation of
diagrams and lists that describe the names of the objects defined, their behaviour, and the type of interaction between objects. Known as a logical data model this provides a basis for defining what the GIS will do and the type of domain over which it will extend [4]. Finally a model showing how the objects under study can be digitally implemented in a GIS is created. Physical models describe the exact files or database tables used to store data, the relationships between object types and the precise operations that can be performed.

In summary, this entails modeling the user's view of the data, and describing the basic features required to solve the problem. MSIGM [5] consider this to involve the selection of the geographical representation (points, lines, areas, rasters, TINS). The data model is usually presented as a 'graphic model' using a graphical modeling language e.g. Microsoft Visio that provides a diagram of the data structure as well as UML (Universal Modeling Language) code, the latter being 'ported' to the GIS to create an 'empty' database structure defining all the necessary relationships and dependencies [3].

8. NATURE-GIS AND DATA MODELS

The development of a data model or structure to be used as the basis for characterising geospatial data for protected areas data is the main goal of the NATURE-GIS Workpackage Three. The data model developed will form the basis of a demonstration of the potential of a pan-European web service in Workpackage Five.

Given the requirement for data from multiple sources effective use and sharing will need to follow a set of common content standards or common data model designs. The first task is to determine the set of GIS themes to address the particular application and information requirements and to provide a definition of each thematic layer in greater detail. Bibr and Kukn [1] suggest that consideration must be given to an ontological foundation to help minimise the problem of semantic heterogeneity.

![Fig. 4 Levels of abstraction relevant to GIS data models [4]](image-url)
As with other applications e.g. coastal and marine data e.g. (see [3]; and also ESRI's homepage (http://www.esri.com/software/arcgisdatamodels) describing different data models) there is a very wide range of data collected e.g. Nature Conservation data for managing 'protected areas'. These are collected using a wide range of methods, instruments, time scales, periodicity, precision and accuracy [3]. Additional complexity results due to the sampling media, data types, and data dependencies involved. ESRI has been actively involved in the development of common content standards (i.e. common data model designs) and provides a framework for key layers of GIS information [6]. The OpenGIS Consortium (OGC) Simple Features Specification [7] considers spatial data management in a database management system (DBMS) tables.

9. BEYOND THE DATA MODEL

Handling large volumes of data and information requires quick access, service delivery, and integration [8]. This can be achieved via the development of interoperable web services, where a web service is defined as a 'software application defined by a URL whose interfaces and binding are capable of being defined, described, and discovered by XML artefacts and supports direct interactions with other software applications using XML based messages via internet-based protocols' [9]. Within this Service, Metadata describe both the data and the services. Furthermore a Service Registry is required to register both data and services metadata.

An interoperability framework provides a coherent vision of interoperable¹ services with a common pattern, to deliver a consistent framework of geospatial web services. It allows description, discovery, negotiation, chaining and use for Web Access for user communities and the development of new services. It must be extensible to provide advanced services for specialised thematic domains, and for communities to define types and schemas to allow semantic interoperability across nations and domains.

10. PROTECTED AREAS NETWORK SPECIFICATION

For the Nature-GIS project the proposed specifications are as follows:

- WMS : FEDERATES THE MAP SERVERS AROUND CAPABILITIES, METADATA, GETMAP AND GETLEGEND (+SLD)
- WFS : SERVER OF FEATURES: CAPABILITIES, METADATA, FEATURESHEMA (MODEL, STRUCTURE, SCHEMA), GETFEATURE AND GML
- WFST : TRANSACTIONAL WFS (CREATE, UPDATE, ANNOTATE, ...)
- ISO 19115 : METADATA
- ISO 19119 : SERVICE METADATA
- WRS : CATALOGS AND WEB REGISTRY SERVICES
- SLD : STYLE DESCRIPTION FOR PORTRAYAL OF FEATURES
- CONTEXTS : SAVE AND SHARE USER'S CURRENT CONTEXT
- POSSIBLY OTHERS : WCS, WTS, ...

The end result will be the provision of a federation of interoperable services for maps (raster), a federation of WFS services for features and for search and uses of European Protected Areas Features, Metadata description of available/usable services and datasets, cataloguing/register of compliant online services, a web client with dynamic search and

¹ Interoperability is defined by Brandmeyer and Karimi ([10], p. 480) who quote (Howie et al. [11]) as being the capability with which two or more programs can share and process information irrespective of their implementation language and platform.
binding of services to match user requirements and to provide ad hoc services (reporting, analysis etc.), enable transactional behaviour if needed (update or annotation). Each stakeholder will register their services to the registry.

The challenges facing the development of this service, which are the responsibility of Workpackage 5 are to (a) define Pan European Semantic Types for protected areas, (b) to establish the availability of metadata, (c) the multi-cultural dimension, (d) security and access, (e) dissemination, (f) training and service, and (g) to accommodate the likely shift in the technology and IT paradigm. Such a service may publish its capabilities with: service metadata, service bindings, and data metadata; a set of services available through a common interface; a registry to discover them and to get information on how to use them; what kind of data they may provide and how to interpret them; a network-wide common set of objects with their properties; define the network i.e. what’s inside, how to access it, and where to start searching for data. Ultimately, together with the registry/metadata mechanism, the servers are web services that can be published by stakeholders, found by users, bound together to build a typical application (Fig. 5) [12].

11. SUMMARY AND CONCLUSIONS

This paper has briefly outlined the overall scope of the EU funded Nature-GIS project. As part of the research work involved in achieving the goals of this research project, a major component of the work to date will involve the development of a geospatial data model for protected areas. This paper will outline the preparatory work involved to date towards the development of the data model, set within the context of a technology-stakeholders questionnaire, and the future goal of implementing an operational demonstration of the data model in the form of a web service. Following completion of the survey questionnaire in January-March 2003, and analysis of the results, the ‘picture’ will be used as the basis to confirm the requirements for refining the development of the protected areas data model, which will subsequently be implemented.

![Diagram](image)

Fig. 5 Specification for NATURE-GIS
12. REFERENCES


