Geo-spatial Information Science
Vol. 4, No. 2, p. 40~46 June 2001

# KGI-XGIS SUPPORTING DECISION MAKING WITH KNOWLEDGE-BASED TECHNIQUES

Steffen Vogt

***KEY WORDS*** KGI-XGIS; spatial data; EIA process; expert system

***ABSTRACT*** The concept of the King George Island Expert GIS (KGI-XGIS) is introduced. KGI-XGIS is a combination of a rule-based expert system and a GIS meant to form an intelligent spatial decision support system. The system provides the spatial knowledge necessary for the environmental impact assessment process as dictated by the 'Madrid Protocol' for King George Island (South Shetland Islands, Antarctica). It also serves as information system to the scientific user community. Topographic maps, remote sensing data, thematic maps based on field surveys and other digital data form the input data to the GIS. Given the sparseness of the available data these must be combined and used in the most efficient way. Therefore expert knowledge of different domains will be coded into a rule-based expert system which is coupled to the GIS. To handle the expert knowledge and the spatial data of different types and different quality the KGI-XGIS incorporates knowledge-based techniques and fuzzy reasoning. The selection of a camp site on the ice free area of Fildes Peninsula and data quality management are used as two examples to demonstrate the capabilities of the system.

## 1 Introduction

King George Island (South Shetland Islands, Antarctica) is one of the most densely populated areas in Antarctica. Permanent stations of nine different nations and an airstrip suitable for huge aircrafts are located on its ice free areas. These areas comprise less than 5% of the island but carry most of the fauna and flora. The concentrated human activities result in considerable impacts on the vulnerable ecosystems and to a severe degradation of the environment[1]. Management plans for two major ice free areas have been proposed to ATCM and several protected areas have been established[2]. Given the manifold research activities and the increasing tourism an efficient environmental protection strategy remains an issue of paramount importance to the island.

The Protocol of Environmental Protection to the Antarctic Treaty ("Madrid Protocol") establishes principles of environmental protection in Antarctica. A crucial part of implementing these principles is defined by Article 3, paragraph C. It states the necessity of collecting sufficient information about planned activities "to allow prior assessments of, and informed judgements about, their possible impacts on the Antarctic environment and dependent and associated ecosystems and on the value of Antarctica for the conduct of scientific research".

A vital part of the necessary information is spatial data. As the state-of-the-art tool to handle such data a geographic information system (GIS) should be used to provide the spatial information that relates to the planned activities[3,4]. Although GIS have emerged into a tool that is easy to use through common graphical user interfaces (GUI) there is still a considerable lack of knowledge among potential users about how to use the systems to derive the desired information. Unfortunately, at present

Steffen Vogt, Ph. D candidate, Department of Physical Geography University of Freiburg, Werderring 4, D-79085 Freiburg, Germany
E-mail: steffen.vogt@ipg.uni-freiburg.de

GIS can only be regarded as data processing systems but they largely lack the ability to process knowledge on how to use the data in a meaningful way. The latter step is left to the user but often GIS users are non-experts in handling spatial data.

Common pitfalls when using GIS, not always realised by the users themselves, include unstructured definition of the problem, insufficient or inappropriate data, incomplete information on which data exist, a principal uncertainty about the relevance of data sets to use and last but not least only a basic knowledge about the analysis techniques and their proper use.

To avoid these problems the aim of the KGI-XGIS is to aid non- and semi-specialists in using spatial data sets as decision support in the process of environmental impact assessments (EIA).

The XGIS should therefore:

1) provide spatial data from various sources for environmental assessments,

2) aid in the definition of the specific problem,

3) aid in the selection of appropriate data sets,

4) aid in the selection and use of appropriate analysis techniques,

5) provide reliable and reproducible results,

6) adhere to ISO/TC 211 and the standards defined by SCAR programs,

7) provide a flexible and adaptive architecture for the modification and enhancement of capabilities,

8) have a user-friendly interface for human-machine interaction.

In addition to the use as a decision support system for EIA the KGI-XGIS is also a valuable tool as information system to the scientific user community.

## 2 Expert GIS

To add the capability of processing knowledge about the spatial data an expert system (XS) is wrapped around the GIS like a shell. Fig. 1 shows the model of the XGIS. The XS consists of a knowledge base (KB), an inference machine (IM), a knowledge acquisition module (KAM) and an explanation module (EM). Expert knowledge is fed into the system through the KAM. The user addresses the spatial information stored in the GIS and the data base management system (DBMS) through the XS which supports him by using the expert knowledge of the KB. The EM allows to trace back the system's reasoning. The user is not interacting with the GIS directly but through the expert system which assists him in addressing spatial problems.

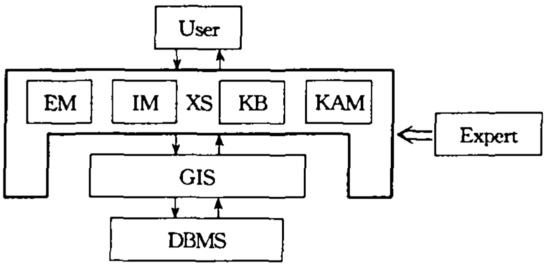



Fig. 1 Model of the XGIS

An expert system can be defined as an intelligent computer system that imitates our reasoning process and uses knowledge to provide human expert-level solutions to complex problems in a specific problem domain. In a rule-based expert system the knowledge is coded into a knowledge base consisting of ordered sets of IF-THEN rules called a production system. The inference machine has the formal capability to use this knowledge in an effective way to derive implicit information from explicit information. Several principles and techniques can be used to handle the rules. As these are described in many standard text books[5,6] on expert systems they will not be discussed here in detail.

The main difference of rule-based expert systems compared to conventional computer programs is that knowledge is stored in a declarative way rather than in a procedural way. In traditional programs problem solving consists of applying a specific algorithm to a precisely defined problem. Only data that give information on an instance of the problem can be uploaded into the program. Any modification of the problem solving requires writing of a completely new algorithm or program. Improvements or changes are then difficult to implement.

Contrary to the concept of a hardcoded algorithm specialised on a neatly defined problem is the concept of declarative knowledge as it is realised in rule-based expert systems. In the latter case the knowledge about the problem domain is stored in the form of the production rules. The abstract problem solving competence remains to the inference machine. New knowledge about the problem domain is added to the system simply by adding new rules to the knowledge base. This makes rule-based systems especially suitable for semi-structured problems and in cases where the knowledge about the problem domain is expanding dynamically. Decision making in environmental impact assessments poses a wide variety of similar but not exactly equal problems. At the same time environmental impacts in Antarctica is a current research topic resulting in a continuously better understanding of many aspects of the implications of human activities on the ecosystems. Thus the application of knowledge-based tools like the KGI-XGIS promises an efficient support in this domain.

Typical properties of spatial data sets in Antarctica are imprecision, uncertainty and vagueness. Data sets are often incomplete, the cartographic reference is not as accurate as desired and the thematic information not always has the desired reliability. To deal with these problems the theory of fuzzy sets and fuzzy logic is incorporated into the XGIS. Fuzzy sets and fuzzy logic can model imprecision and uncertainty in the spatial domain, in the feature domain and in the inference mechanisms including linguistic variables in the formulation of the rules[7~10].

## 3 Model of the KGI-XGIS knowledge base

The conceptual model of the KGI-XGIS modifies the general model of an XGIS to the problem of handling spatial data for King George Island with respect to environmental impact assessments. The knowledge base is divided into modules that represent the different domains of expert knowledge necessary for the EIA process. Each of the modules can be divided into further submodules which represent specific subdomains. This keeps the amount of rules that the system must consider at one time low and ensures efficient knowledge handling. Management and update of the knowledge base is thus kept fairly easy within the modules.

The design of the knowledge base for the KGI-XGIS reflects the different steps of the EIA process as described in COMNAP / ATCM (1999). In the following the terms activity, action and output are used according to the definitions given in this document. The basic steps for the EIA process are to define the planned activity and the area of interest. With this information it should be possible to assess potential impacts.

The KGI-XGIS knowledge base consists of three main modules, the activities and actions (A&A) module, the environments and impacts (E&I) module and the spatial information and spatial analysis (SI&SA) module (Fig. 2), each containing several submodules.

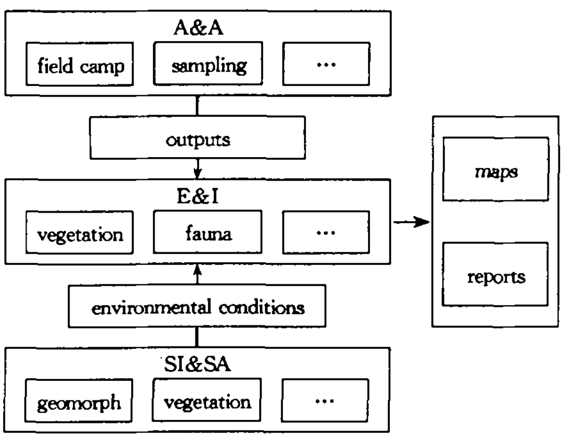



Fig. 2 The knowledge base of the KGI-XGIS

Initial scoping of a project requires to define the planned activity and to identify all related actions. This part of the process is of crucial importance as the actual environmental impact assessment is based on the possible outputs of the detected actions. Outputs are defined as physical changes to the environment. The A&A module contains expert knowledge that helps defining the activity. This knowledge can be derived for example from technical experts in the field of the activities or from logistics personnel. The user may be asked to give details on the principal characteristics of the planned activity, a piece of information which is used by the system to find possible outputs of the activity. The first example described below illustrates the process.

To assess the potential impacts of these outputs it is necessary to know which are the environmental conditions in the area with respect to these outputs. The SI&SA module uses available spatial data sets to produce information on these conditions. This module contains expert knowledge on the relation of spatial properties in the Antarctic environment and on spatial analysis techniques. One submodule of paramount importance is the spatial data quality submodule that traces data quality issues through the process of spatial analysis. This important topic is elaborated further in the second example given below.

## 4 Acquiring the expert knowledge

"Those persons responsible for an Environmental Impact Assessment Process need to ensure that they consult as widely as is reasonably necessary and possible in order that the best available information and professional advice contribute to the outcome. A number of different participants may be involved throughout this process, ranging from those who are involved in the details of nearly all parts of the process (e. g. environmental officer, proponent of the activity) to those who are the technical experts who provide input in particular subjects of the process (e. g. researchers, logistic personnel, others with experience at the location or in a particular activity)." (COMNAP / ATCM, 1999)

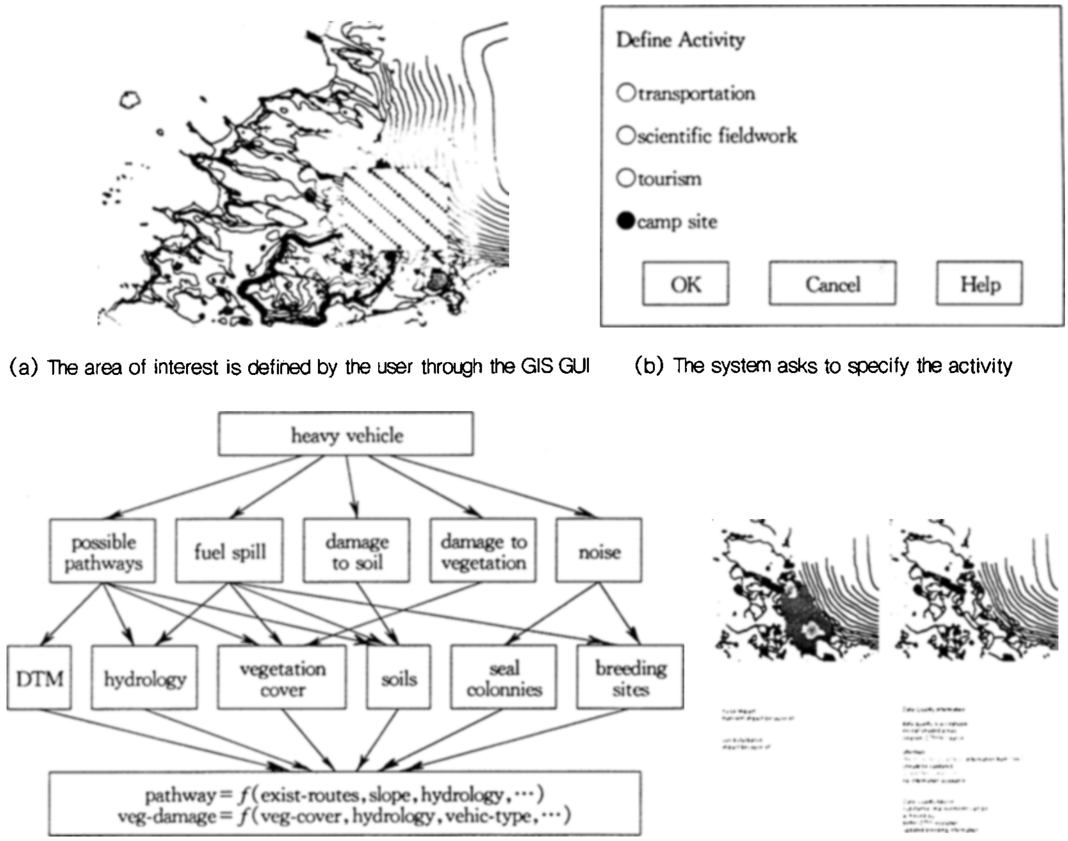
Acquiring expert knowledge even for one domain of interest is often a time consuming and difficult job. Collecting the best available professional advice includes consulting many different experts. An approach to avoid this expense for every new problem again is to collect all the available advice in advance, storing it in some form and then using this data base for the actual cases. The knowledge base of an expert system can serve as such a virtual expert. Obviously the crucial point is to acquire and store the best information possible in this data base. This includes not only finding the best experts in the domain of interest, getting all the relevant information from them and storing it in an efficient way but also updating of the knowledge base if the knowledge expands.

As is elaborated in the quotation above, relevant experts for EIA processes in Antarctica include researchers who can judge the environmental sensitivity and threats, logistics personnel who is able to specify the technical needs and people with field experience in the area of interest. In structured interviews the expert knowledge in the domain of interest will be collected from as many experts as possible. The acquired knowledge then must be formalised and coded into production rules. Acquiring the knowledge is a heuristic process and it will often be necessary to refine the interviews and to ask the experts again to improve the knowledge base. The knowledge acquisition process requires either a knowledge engineer who knows how to handle the knowledge acquisition module of the expert system shell or a module that provides an easy to use and intuitive GUI.

### 4.1 Example 1 Selecting a camp site on Fildes Peninsula

The problem of "Selecting a camp site on Fildes Peninsula" illustrates a simple application of the XGIS when planning a research project. First a preliminary assessment of the possible impacts has to be drawn up. A group of geologists, for instance, want to use the KGI-XGIS to find a suitable camp site in terms of constraints given by logistics and in terms of minimal impact on the environment. They do not have field experience nor a thorough understanding of the ecology of the area. The XGIS supports them in using the appropriate data sets for the spatial analysis that is necessary to support them in finding an appropriate camp site. Figs. 3(a) ~ (d) illustrate how the system leads them through the process.

First the area of interest has to be delimited. This is done by using standard GIS functions through the GIS GUI like zooming in or out and dragging a polygon (Fig. 3(a)). In a second step the planned activity must be defined and specified by selecting buttons of a menu the system presents polygon (Fig. 3 (b)). Once the activity is specified the XGIS checks its knowledge base for possible actions related to the activity. If necessary it may request more details. Now the system automatically identifies all possible outputs of all the actions. This is the result the A&A module presents to the E&I module, which has been shown in Fig. 2.




(a) The area of interest is defined by the user through the GIS GUI

(b) The system asks to specify the activity

(c) The system automatically checks for possible actions related to the activity, identifies necessary spatial data sets and performs the required analysis

(d) Maps and associated reports are presented as output

Fig. 3 An example of the application of the KGI-XGIS system

The E&I module has the capability to identify impacts on the ecosystem by comparing the possible outputs with the actual environmental conditions. The E&I module checks for the necessary data sets it needs to determine the environmental conditions and asks the SI&SA module to provide these data sets. This module now tries to provide the best spatial information available by selecting existing data sets and performing appropriate analysis on them (Fig. 3(c)).

Once the information about the relevant environmental conditions is available to the E&I module it produces the desired output in the form of maps and reports (fig. 3(d)) which then support the decision makers in taking their decision. In this case the maps might show possible pathways to the camp, the area around a breeding site in which noise impacts the animals, freshwater lakes that are sensitive to fuel pollution, etc. The reports give information about which data sets have been used or where the data quality is poor and the results should be treated cautiously. The reports can also indicate non existent data sets that could improve the quality of the results substantially if provided.

### 4.2 Example 2 Guiding through the data quality jungle

Current GIS are used through graphical user interfaces (GUI) which allow easy handling of the systems even to non-specialised users. On one hand this is a desirable effect as everyone can use and produce spatial data sets through these GUIs. On the other hand inexperienced users may produce nonsensical results in applying inappropriate analysis techniques on spatial data sets. This is easily done by using the menu bars and clicking on the buttons of the GIS GUI. Still the system provides excellent looking results like maps or new attributes to spatial entities. The danger is that we trust these meaningless results and base decisions on them.

To avoid this "garbage in - gospel out" effect, data quality information must be included as a vital part of the analysis process and must be clearly indicated on the final outputs of the system. Although data quality information is often complex to understand, let alone to use, the data must comply to quality standards, the user must recognise data quality and the final product should carry data quality attributes. The aim of the data quality submodule, which forms a part of the SI&SA Module, is to use given quality information throughout the analysis processes and to produce a report on data quality that even potential non-experts can understand.

Data quality information will represent an essential part of the data model for the KGIS. The model will comply with the forthcoming ISO TC 211 standard. An overview of this abstract data quality model is given in Fig. 4, in which data quality is described in qualitative form by the data quality overview elements and in quantitative form by data quality (sub-) elements. But reporting the data quality for a data set is not enough. It must be taken into account throughout the analysis process. The XGIS thus combines the information given in the metadata with the information on the intended use to transform the abstract data quality value in an interpretation relevant to the planned use. This will be done by using the expert knowledge in the knowledge base submodule the specified problem pertains to.

Taking a thematic map of vegetation coverage derived from remote sensing data by automatic classification as an example, the reported data quality on thematic accuracy might incorporate percentage values about omission and commission errors in the vegetation classes. Omission errors indicate areas that are excluded from the category to which they belong, and commission errors indicate areas that are incorrectly included into a category. For an intended use of the map as part of assessing the impact of heavy vehicle tracks this information is of different importance than for an intended use of the map to estimate the impact of fuel spill on the vegetation.

Fig. 5 depicts how rules are applied to use the data quality information throughout an analysis process , in which the left branch describes the conven -

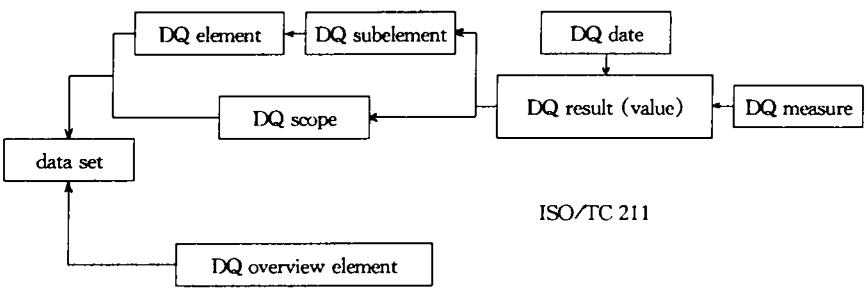



Fig. 4 The abstract data model according to the ISO/TC 211 draft

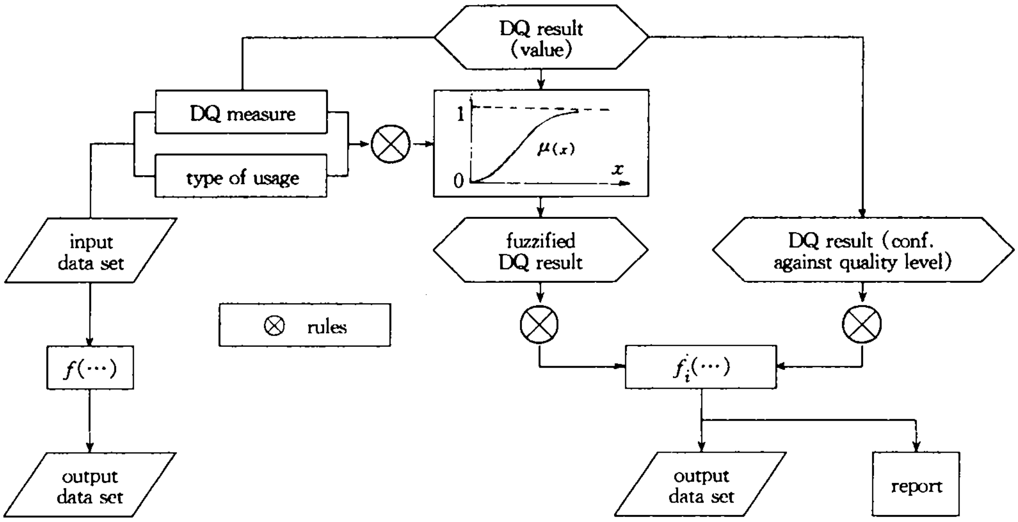



Fig. 5 Using the expert knowledge by applying rules to the data quality information

tional way without using the rules. On the right branch data quality information is taken into consideration by using expert knowledge to select the appropriate analysis technique. First expert knowledge helps in transforming the abstract data quality value by selecting an appropriate fuzzy membership function with respect to the intended analysis. In a next step the now fuzzified data quality information is used for the choice of an analysis technique corresponding to the data quality. If the quality is poor only a simple technique might be employed. If the quality is good enough more sophisticated techniques which might provide better results but need more accurate data selected. Thus the systems avoids producing results that suggest a high reliability and precision where the data quality does not support such assumptions. This prevents decision making based on false premises.

## 5 Implementation

Constructing expert systems from scratch is not necessary. Expert system shells are tools that make things much easier. These shells can be regarded as context independent interpreters for declarative knowledge representation and inference. The implementation will be based on the Java Expert System Shell (JESS) which is very similar to the widely used expert system shell CLIPS. JESS provides all the functions that are necessary for an efficient implementation of the XGIS including fuzzy reasoning by the Fuzzy JESS extension. JESS is written completely in Java which makes JESS platform independent. Java also facilitates programming GUIs and supporting APIs to commercial GIS packages like ESRI's Arc/Info or ArcView products. For non-commercial use JESS is distributed without licence fee and the experience of thousands of CLIPS implementations helps in constructing and in the improvement of the expert system. The XGIS can be constructed by using any GIS software as long as the GIS package provides some means for interapplication control like RPC or APIs. At the moment Arc/Info is used.

## 6 Outlook

To demonstrate the capabilities of the KGI-XGIS a preliminary and rather simple version with expert knowledge for only certain selected domains will be constructed. In the second step expert knowledge for all the submodules will be derived from interviews with experts and coded into the system. After improvement and adjustment of the human-machine interfaces to the needs of potential users the system will support implementing the 'Madrid Protocol'.

### Acknowledgements

*This projected is funded by the Federal State of Baden-Wuerttemberg, Germany through the LGFG program.*

### References

1 Harris C M. Environmental effects of human activities on King George Island, South Shetland Islands, Antarctica. *Polar Record*, 1991, 27: 193~204

2 Foreign & Commonwealth office. List of protected areas in Antarctica, London, 1997. 33

3 Harris C M. Environmental management on King George Island, South Shetland Islands, Antarctica. *Polar Record*, 1991, 27: 313~324

4 Smith R, Lewis I, Walton D H W, *et al*. Developing the Antarctic Protected Area System, 1994. 137

5 Durkin J. Expert Systems. 1995. 800

6 Giarratano J C, Riley G D. Expert systems: principles and programming, 3rd ed. 1998. 597

7 Altman D. Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems*, 1994, 8: 271~289

8 Burrough P A, Frank A U. Geographical objects with indeterminate boundaries, 1996. 345

9 Burrough P A, McDonnell R A. Principles of Geographic Information Systems, 1998. 333

10 Leung Y. Intelligent spatial decision support systems. 1997. 470

11 COMNAP/ATCM (Council of Managers of National Antarctic Programs on behalf of the Antarctic Treaty Consultative Meetings). Guidelines for environmental impact assessment in Antarctica. 1999. 21