

A framework for data quality for Mediterranean sustainable ecosystem management

Rui Pedro RIBEIRO*, José Guilherme BORGES, Vanda OLIVEIRA

Departamento de Engenharia Florestal, Instituto Superior de Agronomia, Universidade Técnica de Lisboa,
Tapada da Ajuda, 1349-017 Lisboa, Portugal

(Received 16 September 2003; accepted 3 March 2004)

Abstract – Information requirements for Mediterranean ecosystem management have vastly increased and yet a framework for data quality is not available. This article focuses on data quality issues to enhance Mediterranean natural resources management planning. A framework to achieve high data quality standards is presented. Components of a natural resources data quality framework are identified and characterized. Emphasis is on the design of a data lifecycle within an information system such that data quality categories are accounted for along the stages of data acquisition, storage, organization and interpretation. Mediterranean forest ecosystems extending over circa 40 800 ha in Southern Portugal were used as test cases. Results from the application of the proposed framework to data acquired in 397 forest inventory plots, in 141 wildlife transects, in 157 bird observation points and in 138 questionnaires to landowners are reported. Results suggest that the framework does effectively address current Mediterranean ecosystem management data quality needs.

data quality / information system / Mediterranean ecosystem management

Résumé – Un cadre pour des données de qualité pour la gestion durable de l'écosystème méditerranéen. Bien que les besoins d'information nécessaires à la gestion de l'écosystème méditerranéen se soient accrus, il n'existe pas de cadre qui permette de garantir la qualité des données nécessaires à une gestion durable. Cet article se focalise sur la qualité des données permettant une amélioration de la planification de la gestion des ressources naturelles méditerranéennes. Un cadre permettant d'obtenir des données de qualité normées est présenté. Les composantes de ce cadre sont identifiées et caractérisées. L'accent est mis sur la conception d'un cycle de vie des données dans un système d'information, de telle façon que les catégories de qualité des données soient expliquées dans le cadre des étapes d'acquisition, de stockage, d'organisation et d'interprétation. Un écosystème forestier méditerranéen qui occupe approximativement 40 800 ha dans le Sud du Portugal a été utilisé comme test. Les résultats de l'application du cadre proposé et concernant les données recueillies dans 397 placettes forestières, dans 141 transects de faune et flore, dans 157 points de surveillance des oiseaux et 138 questionnaires de propriétaires forestiers sont rapportés. Les résultats rapportés suggèrent que le cadre proposé répond aux nécessités de qualité des données nécessaires pour la gestion de l'écosystème méditerranéen.

données de qualité / système d'information / gestion de l'écosystème méditerranéen

1. INTRODUCTION

Mediterranean forest ecosystems are generally characterized by abundant biological diversity and by a fragility that derives from a harsh climate, difficult socio-economic conditions and a history of natural resources over-exploitation [9, 25, 26, 38]. The complexity of the Mediterranean ecosystem management problem and the variety of Mediterranean natural resources data call for the evolution of a specific approach for data integration and for the development of data standards. Data is an extremely valuable resource as it may translate into information needed to support decision-levels ranging from technical aspects to policy instruments for sustainable practices.

Current technology provides powerful tools for information generation and transfer and Mediterranean natural resources researchers and practitioners have to design their own and unique data-to-information strategies. Organizations bedeviled by redundant, fragmented and inconsistent databases can hardly have an accurate understanding of the Mediterranean ecosystem management problem. Poor quality data thus undermines the effectiveness of strategic, tactical and operational decision-making [33]. Lack of information in a usable form and inefficient and/or untimely transfer of data to appropriate users have been barriers to utilizing the best available knowledge in Mediterranean ecosystem management and to identifying priorities for information generation through research. A framework

* Corresponding author: ruipeiro@isa.utl.pt

for data quality may be a valuable tool for designing adequate data-to-information strategies thus contributing to overcome current problems faced by users of Mediterranean natural resources data and information systems.

There is no rigorously defined set of data quality categories and dimensions [43]. Nevertheless, it is consensual to consider fitness for use as the best criteria to assess data quality [17, 40, 41, 43, 45]. Therefore, the traditional natural resources inventory concern of achieving data accuracy is not enough. Data quality should not be treated just as an intrinsic category independent of the context in which data is used [40]. A framework for data quality must consider accessibility, contextual and representational data categories as well. Information systems can be viewed as data manufacturing systems [44]. Thus a framework for data quality must further consider the data lifecycle within an information system such that all data quality categories are adequately accounted for along the stages of data acquisition, storage, organization and interpretation.

Research focusing on the design and implementation of natural resources information systems has expanded substantially [5, 10, 14, 18, 19, 22, 23, 31, 34, 37, 47]. Several studies stressing Mediterranean ecosystem management [1, 3, 7, 11, 15, 16, 20, 21, 24, 28–30, 35, 36, 42] have either focused on the data acquisition stage or on the generation of knowledge about specific issues through data interpretation. Yet no study has developed a data quality framework for Mediterranean ecosystem management.

In this article, a framework to achieve high natural resources data quality standards is presented. Components of a natural resources data quality framework are identified and characterized. Emphasis is on the design of a data lifecycle within an information system such that data quality is accounted for along the stages of data acquisition, storage, organization and interpretation. A local development organization and a forest landowners association defined as a priority natural resources inventory and assessment of both Serra de Grândola and Serra de Portel. These Mediterranean forest ecosystems extend over circa 40 800 ha in Southern Portugal and were thus used as test cases. The inventory should provide accurate estimates of current cork production, current game potential, current biodiversity levels and of current landowners management objectives. Biometrical, wildlife, bird (as a surrogate for biodiversity) and landowners data was deemed crucial to provide those estimates and for providing technical and management assistance to landowners and information to develop policy instruments for sustainable practices to central and local government agencies. Results from the application of the proposed framework to data acquired in 397 forest inventory plots, in 141 wildlife walked transects, in 157 bird observation points and in 138 questionnaires to landowners are thus reported.

2. MATERIALS AND METHODS

2.1. Test cases

Two representative Mediterranean ecosystems were selected to develop and test a data quality framework. The first, Serra de Grândola, is located in western Alentejo and it extends over approximately 23 000 ha. Its ecological importance is highlighted by its classification

as a CORINE Biotope (C-108) and its integration in the set of sites proposed to be part of the EU network Natura 2000. The climate is characterized by pronounced water deficits after April. Serra de Grândola rises up to 326 m and slopes extend from 9 to over 25%. The main covertypes are dominated by the forest species cork oak (*Quercus suber*) and *Pinus pinea*. These species may occur in pure or mixed stands, on even-aged or uneven-aged stands. Spacings also vary. Higher densities are found at higher altitudes. In the past, land use has led to erosion and soils are generally thin. Medium-sized (5–10 ha) private property prevails. Agroforestry activities, namely range management, are conducted in most stands. Cork is the most important market product.

The second case study, Serra de Portel, is located in eastern Alentejo and it extends over approximately 17 800 ha. Serra de Portel is characterized by a climate harsher than Serra de Grândola's, and it is classified as an area under severe drought and erosion risks. The mean maximum temperature in the summer increases to over 32 °C. Very warm and dry summers contrast with cold winters. Serra de Portel rises up to 421 m with very steep slopes on its eastern side, where the Degebe and Gadiana rivers flow in entrenched valleys. Covertypes are dominated by the forest species holm oak (*Quercus rotundifolia*) and to a lesser extent by *Quercus suber*. Like in Serra de Grândola, these species may occur in pure or mixed stands, on even-aged or uneven-aged stands. Spacings also vary. Serra de Portel has been classified as a site where traditional management systems have high ecological value thus contributing to the conservation of rare and endangered species. Past policies aiming at the expansion of cereal culture had a severe impact on soils that are very thin and stony. This area has undergone a considerable depopulation process since the 60's, which is reflected in a low population density and a high percentage of elderly population. Large-sized (> 100 ha) private property prevails. Agroforestry activities, namely wildlife and range management, are conducted in most stands.

2.2. Data

The data from the study areas was acquired in 397 forest inventory plots, in 141 wildlife walked transects, in 157 bird observation points and in 138 questionnaires to landowners. Plot and transect location was based on a stratified random sampling approach. The definition of land-use strata was based on standards defined by the Portuguese Forest Service [13]. Aerial photographs and a geographical information system were used to map the strata and to locate both plots and transects. Biometrical data was measured in the forest inventory plots according to procedures defined by Barreira et al. [4]. For example, up to 19 variables were measured in each of the 7 999 cork oak trees in the inventory plots (Tab. I). Linear walked transects with up to 250 m were used to acquire data about big game (*Sus scrofa*). The field work on the bird observation points concentrated in spring on a 4-h period after sunrise and on a 3-h period just before sunset as these are the periods when the focal species considered in this study to assess biodiversity develop most of its activity. The number of observations totalled 5 092 individuals from 77 species. The questionnaires were designed and conducted by a landowners association to characterize the land management systems (e.g. cork oak, holm oak, wildlife, range and agriculture management practices).

2.3. A framework for Mediterranean natural resources data quality

Natural resources spatial and aspatial data is used to develop, evaluate, apply and monitor Mediterranean ecosystem management plans and policies. Data quality may thus be assessed by how well it serves these purposes. Yet this assessment is by no means a trivial exercise. Mediterranean ecosystem sustainability depends on complex spatial

Table I. Example of variables measures in cork oak trees.

Variable	No. of measures	Variable	No. of measures
dbh	6024	Number of branches	5638
Stem diameter	232	Debarking code	6087
Crown base height	6063	Cork thickness	6036
Crown diameter	6054	Debarking year	2885
Stem height	5677	Defoliation severity	6114

Table II. Processes and correspondent activities of the information production life cycle.

Processes	Activities
Data acquisition	Development data acquisition protocols. Data acquisition
Information system analysis	System analysis and requirement analysis. Development of information system logical and conceptual model
Information system development	Development and implementation of the information system physical model. Development of interfaces for data insertion
Data insertion	Data insertion in the information system
Data transformation	Development of data transformation algorithms
Information presentation and distribution	Development of interfaces for data querying, exporting and information presentation

and temporal interactions of management decisions. The human dimension ever-present in a constructed landscape further complicates Mediterranean management planning and policy making. For example, sustainability paradigms may change within Mediterranean ecosystem management typically extended time horizons [6]. Data deemed essential twenty years ago may be irrelevant today. Adaptive ecosystem management to address a changing environment and our incomplete knowledge thus requires the assessment of the impact of time on both data and metadata quality [27]. A systematic approach to Mediterranean natural resources data quality must therefore be broadly based to focus on increasing the usefulness of data for better decisions.

2.3.1. Framework elements

Ecosystem management planning is supported by data within information systems. Thus a data quality framework must consider firstly both the processes for natural resources information systems design and implementation and the stages of the data lifecycle within these systems (Tab. II). Data quality is highly dependent on the unfolding and sequencing of these processes and stages. Secondly, the development of a data quality framework encompasses the identification of criteria that may be used to assess the usefulness of data. Jarke et al. [17] underlined the research effort needed to identify and structure such criteria into views, categories and dimensions. No standard set of criteria is available yet. We propose a hierarchical structure (Tab. III) where a distinction is made between external and internal data views as defined by Wand and Wang [43]. The former is directly concerned with the usefulness of data. The latter is use-independent and it is concerned with the design of the data lifecycle and its impact on functionalities required by the external view.

The internal view includes an intrinsic data quality category (Tab. III). It is a data-related category as it focuses on the impact of data acquisition and storage design on data errors. Accuracy is a traditional quality dimension addressed in natural resources inventory. The believability dimension is used to assess data quality when, for example, no documentation is available about inventory sampling strategies. The reputation dimension may be used to classify data sources that have systematically proved to be inaccurate. Strong et al. [40] add objectivity

Table III. Data quality views, categories and dimensions.

Views	Categories	Dimensions
Internal	Intrinsic (data related)	Accuracy
		Believability
		Reputation
		Objectivity
		Availability
	Accessibility (system related)	Accessibility
		Understandability
		Traceability
		Timeliness
		Minimality
External	Contextual (data related)	Security
		Availability
		Completeness
		Relevancy
		Timeliness
	Contextual (system related)	Timeliness
		Interpretability
		Conciseness
		Consistency
		Consistency
Representational (system related)	Interpretability	
	Conciseness	
	Consistency	

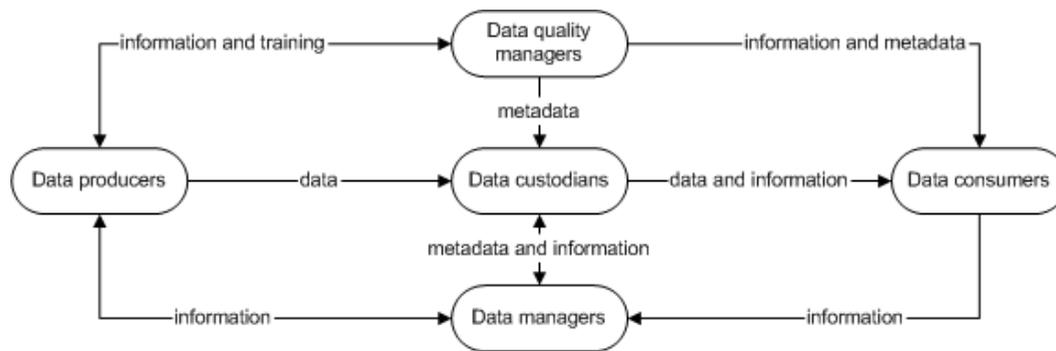


Figure 1. Data user roles and the data/information fluxes between them.

as a fourth intrinsic dimension to assess potential errors associated with derived data (e.g. basal area per hectare as opposed to basal area per tree in a plot). The internal view also includes an accessibility data quality category. It is a system-related category as it focuses on the impact of available computing resources on the potential for data querying and on data security.

The external view includes a contextual data quality category (Tab. III). Some of its dimensions are data-related as they focus on the impact of data lifecycle stages on decision-making. For example, the completeness dimension checks if all data needed to address current management objectives was acquired. Relevancy further builds on the former dimension to assess how well the data stored may address new management requirements. The amount of data stored may complicate the sorting out of relevant data. Thus, data acquisition design should consider both this dimension and the value-added by each data item to avoid costly and ineffective natural resources inventories. Timeliness is a contextual system-related dimension as it focuses on the computing resources needed to make data promptly available when needed by researchers and decision-makers. The external view further includes a representational data quality category. It is a data-related category as it focuses on the impact of data lifecycle stages on decision-making. Interpretability and ease of understanding assess how well data interpretation provides an accurate view of reality, i.e., of the ecosystem management problem. Conciseness and consistency dimensions may be used to assess the usefulness of the combined interpretation of similar data from different sources (e.g. wildlife inventory and socio-economic questionnaires). They may further be used to assess the usefulness of data time-series.

An information system to support ecosystem management planning encompasses data, technology and people [12]. The third element needed to develop a data quality framework thus consists of the assignment of data user roles. These roles are defined such that responsibilities for checking whether data quality criteria are being met are adequately distributed. Jarke et al. [18] assigns responsibilities for checking specific quality dimensions to the information system administrator, the programmer and the decision maker. Ecosystem management planning activities rather suggest that five distinct user roles are considered (Fig. 1). Strong et al. [40] identified three roles: the data producer that acquires or supplies data, the data custodian that stores and processes data within the information system and the data consumer that uses the data. Wang [45] identified a fourth role: the data manager that assists all stages of the data lifecycle. We propose a fifth role: the data quality manager that would be responsible for defining and implementing a data quality policy, for establishing data quality controls and for managing the data quality system in order to achieve the high quality standards defined by Wang et al. [44] based on the ISO 9000 standards.

The distribution of responsibilities for checking whether data quality criteria are being met may be summarized as follows. The data pro-

ducer main focus is on the dimensions of the data quality intrinsic category. The data custodian and the data manager focus on the accessibility and the contextual categories, namely the on the latter's timeliness dimension. The data consumer focuses on dimensions of the contextual and representational categories. The data quality manager is concerned with meeting all criteria and with the distribution of responsibilities among all users. The proposed set of user roles may be seen as knowledge-based human network involving informational interactions to better support decision-making.

2.3.2. Proposed framework

Achieving and implementing high quality standards for natural resources data calls for an effective integration of the data quality framework elements (Fig. 2). The specification of the Mediterranean ecosystem management planning objectives sets the stage for all activities leading to that integration. Firstly, data quality requirements are derived from the management objectives. Secondly, a policy to fulfil them is translated into a quality system by the data quality manager.

Data acquisition protocols are defined according to those requirements. Thus they provide the initial basis upon which the data quality manager may evolve and monitor the data quality policy (Fig. 2). They may further suggest research and development activities as defined by Wang et al. [44] to identify technical specifications for the data lifecycle within the information system (e.g. unambiguous criteria and tests to verify, accept or reject data in each lifecycle stage). The data quality manager is responsible for these activities and for the communication of its results to the other data users.

Protocols interpretation by the data producer has a great impact on data quality. Misinterpretation will lead to data acquisition errors that may impact intrinsic, contextual and representational data quality categories. It will thus constrain the overall quality system. Spurr [39] summarized issues to be addressed at this stage of the data lifecycle. The peculiarities of the objects to be measured may lead to data acquisition errors (e.g. tree height from a tree with irregular bole or stem forms). The characteristics of the measurement devices, the measurement context (e.g. measuring a tree height in a windy day), and human parallax and data insertion errors must also be taken into account. The specificity of Mediterranean natural resources data further complicates data acquisition thus requiring its supervision by the data quality manager. The data quality manager should train, qualify and motivate the data producer so that the measuring equipment is adequately used and data acquisition meets intrinsic data quality dimensions (Fig. 2).

Protocols adequate interpretation by the data custodian is also crucial as it provides the basis both for data modelling and for information system design. The development of data transformation algorithms, data validation routines and user interfaces to insert, validate and

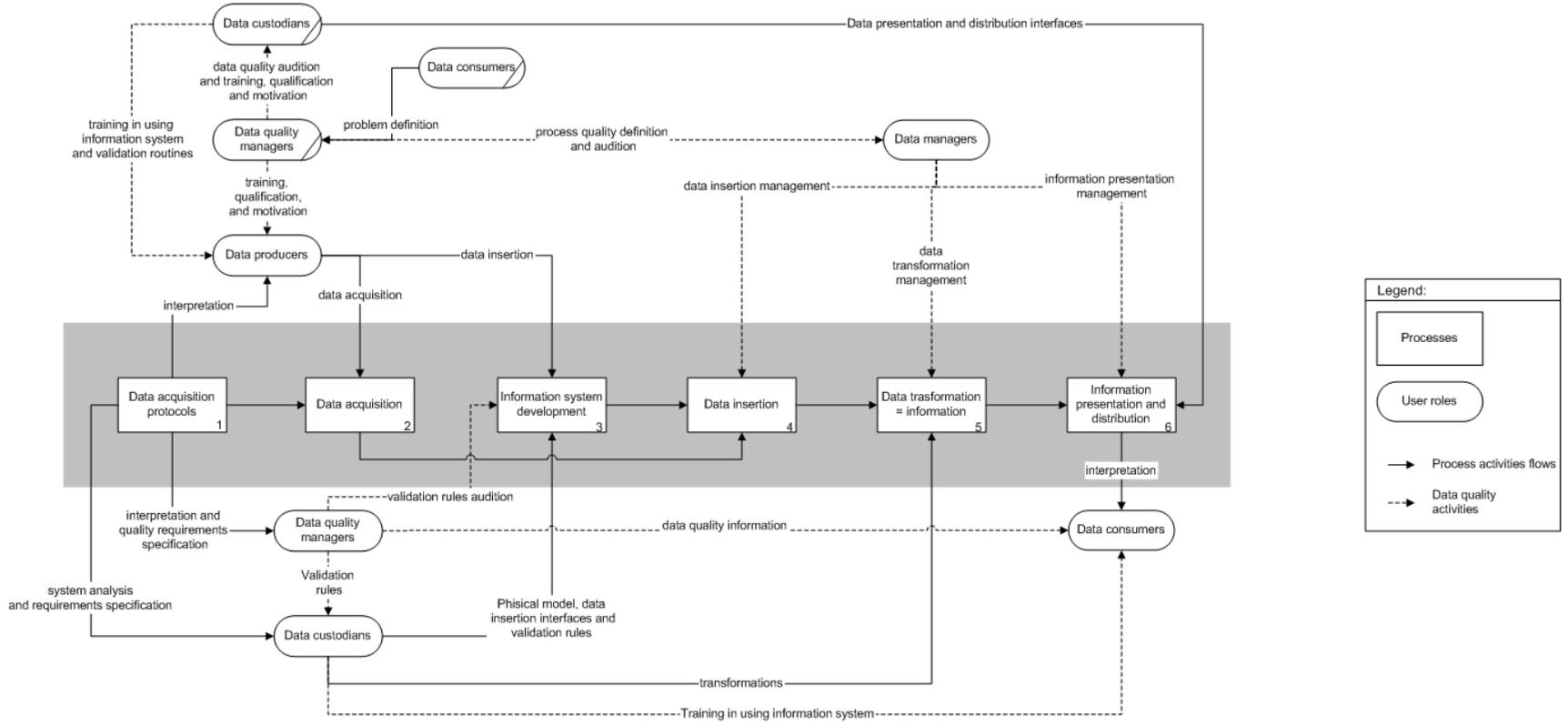


Figure 2. Graphical representation of the relationship between elements of the natural resources data quality framework proposed.

manipulate data is of great importance for data quality. Interfaces should be user friendly, easy and intuitive to use. The impact of poor intrinsic data quality dimensions resulting from failures at the data acquisition stage may be amplified by the lack of adequate validation rules and by a poor system design. The data custodian is responsible for training both the data producer and the data consumer in using validation routines and in using data interpretation interfaces, respectively (Fig. 2). It should also be responsible for production and distribution activities as defined by Wang et al. [44]. The former activities focus on data quality verification and registration along all stages of the data lifecycle. It further set rules to deal with defective data items. The data custodian executes production activities in cooperation with the data manager. Distribution activities focus on the maintenance of the quality system. The data custodian is thus responsible for creating data dictionaries, database documentation and a user-friendly help system. The data quality manager should assist (e.g. definition of validation rules) and audit (check whether rules are being implemented) production and distribution activities by the data custodian (Fig. 2).

The data manager assists all stages of the data lifecycle in order to implement the policy defined by the data quality manager. Specifically, he is responsible for the management and auditing of the information system (e.g., security, recovery and backup policies). In particular, he is responsible for managing data insertion, interpretation and communication procedures (Fig. 2). He also is responsible for selecting technological options that may best fit the data quality policy (e.g. modularity, easiness to use, rapid development). He cooperates with the data custodian to carry out the production activities. He further manages all the metadata information and the documentation produced by the data custodian. The data lifecycle within the information system should be described to help the data consumer become aware of data quality issues and assess the usefulness of interpreted data. The data quality manager should assist and audit the technological and production options by the data manager (Fig. 2).

The data consumer uses data to help support ecosystem management decisions. It thus ties together all stages of the data lifecycle. The data acquisition protocols and the data model are meant to provide a good representation of the data consumer management problem. The interaction between the data quality manager and the data consumer is therefore crucial to define an effective data quality policy (Fig. 2). Mediterranean ecosystem management needs further highlight the importance of this interaction. Management problems frequently encompass ecological, economic and social dimensions and the data consumer is likely to be plural. Typically, there will be several stakeholders and experts from a variety of disciplines. All should interact with the data quality manager. Data storage, organization and interpretation within the information system are designed to help find good solutions to the data consumer problem. The data custodian should interact with the data consumer to develop good data interpretation interfaces. The data consumer should also be aware of the quality policy. In fact his role in defining this policy should not be restricted to the data acquisition and modelling stages. It may help define validation rules and interpretation processes by the data quality manager. Finally, it is the data consumer that may validate the quality policy by checking whether contextual and representational data quality dimensions are met. The usefulness of interpreted data to help support ecosystem management decisions is the ultimate quality check.

3. RESULTS

A local development organization and a forest landowners association set up the Mediterranean ecosystem management problem for decision-making at both Serra de Grândola and Serra de Portel. These non-governmental organizations (NGO) provide both technical and management assistance to landowners

and information to develop policy instruments for sustainable practices to central and local government agencies. They consulted several private and public institutions with a variety of expertise in Mediterranean ecosystem management and research to support the problem identification. The intelligence phase of decision analysis concluded that natural resources inventory and assessment in both areas was a priority. It further highlighted concerns with data quality. These recommendations prompted the research of a framework for data quality to confront this ecosystem management problem. Research went further to provide a general framework that might be adapted to specific Mediterranean ecosystem management problems.

The application of the proposed framework involved the assignment of data user roles. The NGO that set up the process were the primary data consumers. Yet the other institutions involved in the problem identification were also assigned this data user role. The research institution responsible for the framework development provided the data quality manager who worked out data quality requirements and policy to address the ecosystem problem in cooperation with data consumers. It also provided information systems experts and technicians to fulfil the data custodian and the data manager roles. The landowners association and firms specializing in natural resources inventory were assigned data producer roles.

The informational interactions within this knowledge-based network supported the development of both data acquisition protocols and the corresponding information system. Protocol interpretation by the data custodian with the assistance of the data quality manager provided a data model to represent the Mediterranean ecosystem management problem (Fig. 3). Emphasis was on addressing data-related contextual data quality dimensions. This model included a set of entities to represent real-world objects (e.g. Plot, Cork Oak Tree) that are characterized by a set of attributes (e.g. dbh) and that may be associated with each other (e.g. an instance of Cork Oak Tree may be located in an instance of Plot). This relational data model integrated effectively the data acquired in 397 forest inventory plots, in 141 wildlife walked transects, in 157 bird observation points and in 138 questionnaires to landowners (Fig. 3). The data custodian described all entities, attributes and associations in a data dictionary. It further implemented the model within an information system with four modules that might effectively support the data lifecycle according to the data quality manager. The data manager interacted with both the data custodian and the data quality manager to select currently available development technology (e.g. Ms. Access and ESRI MapObject) that might address modularity and easiness of use concerns.

The first module encompassed data insertion and update functionalities. The data custodian designed user-friendly interfaces to sustain intrinsic data quality dimensions (Fig. 4). Geographical information systems (GIS) based interfaces also helped locate plots and transects. The data custodian further programmed routines to address concerns with the accessibility data quality category and the timeliness dimension. The second module included data validation functionalities to further address intrinsic data quality dimensions. Technical specifications by the data custodian and the data quality manager were translated into subsets of general and species-specific routines to verify intrinsic data quality dimensions. Most routines are self-explanatory. Yet data producers received training to better use

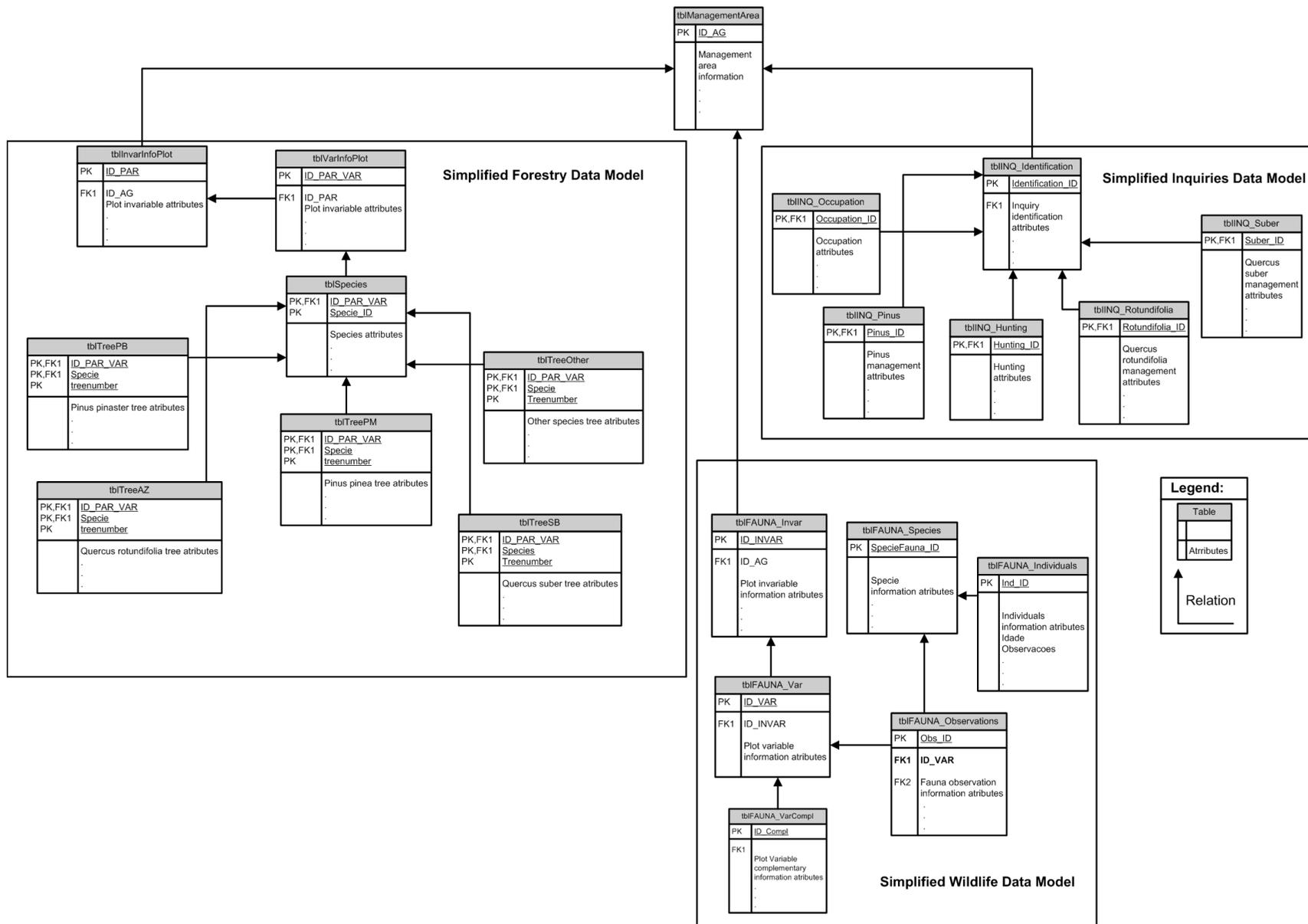


Figure 3. Data model integrating forest, wildlife and socio-economic information.

Figure 4. Example of a data insertion and update interfaces.

them. In some cases, routines provide visual interfaces for better acknowledgment of potential errors (Fig. 5). Data producers contributed to fine tuning both the first and the second module. They extensively used them during the data acquisition and data storage lifecycle stages. The data manager kept a log of this use. Over 41 validation routines were programmed. Over 326 data acquisition and insertion errors were detected and subsequently corrected thus demonstrating the effectiveness of this framework for addressing concerns with intrinsic data quality categories.

The third module encompassed accessibility functionalities. Data consumers other than the NGO required access to spatial and aspatial non-interpreted data to develop research or other activities central to the Mediterranean ecosystem management problem. The data custodian designed this module such that data consumers may use graphical interfaces or process complex structured query language (SQL) procedures for ad-hoc access to all data in the system (Fig. 6). Data consumers may further export accessed data to formats fit for software designed to address specific research goals (e.g. growth and yield modelling, wildlife modelling). This design assisted by the data quality manager addressed concerns with the accessibility data quality category and the timeliness dimension.

The fourth module encompassed data transformation and interpretation functionalities. The data custodian programmed routines to derive data based on available biometrical, wildlife and biodiversity equations and on results from statistical anal-

ysis of responses to landowners' questionnaires. The data export facility in the third module was instrumental to data consumers both to the development of some equations that were programmed within the fourth module (e.g. biodiversity equations) and to statistical analysis of current management practices by landowners. The data manager audited the integration of the routines within the information system to preclude its dependency on the transformation procedures and thus preserve the system modularity and adaptability. It further focused on data transformation timeliness issues. The data quality manager used the export facility to validate results produced by the routines. The data custodian further developed user-friendly interfaces so that data consumers, namely the NGO, might have ready access to useful information about the test areas. These interfaces included visual aids and maps to facilitate data interpretation (Fig. 7). Emphasis was on addressing the representational data quality category. For that purpose, interpretation interfaces were designed to reflect real-time changes on data and on validation and transformation routines.

The system further included documentation and on-line help facilities as guiding tools to data consumers. They were meant to contribute to data consumers' awareness of the data quality policy and to help the NGO assess the usefulness of interpreted data. The system resulting from the proposed data quality framework addressed successfully the NGO requirements. Namely, it provided a consistent approach to integrate a huge amount of data from a variety of sources and disciplines and to bring together

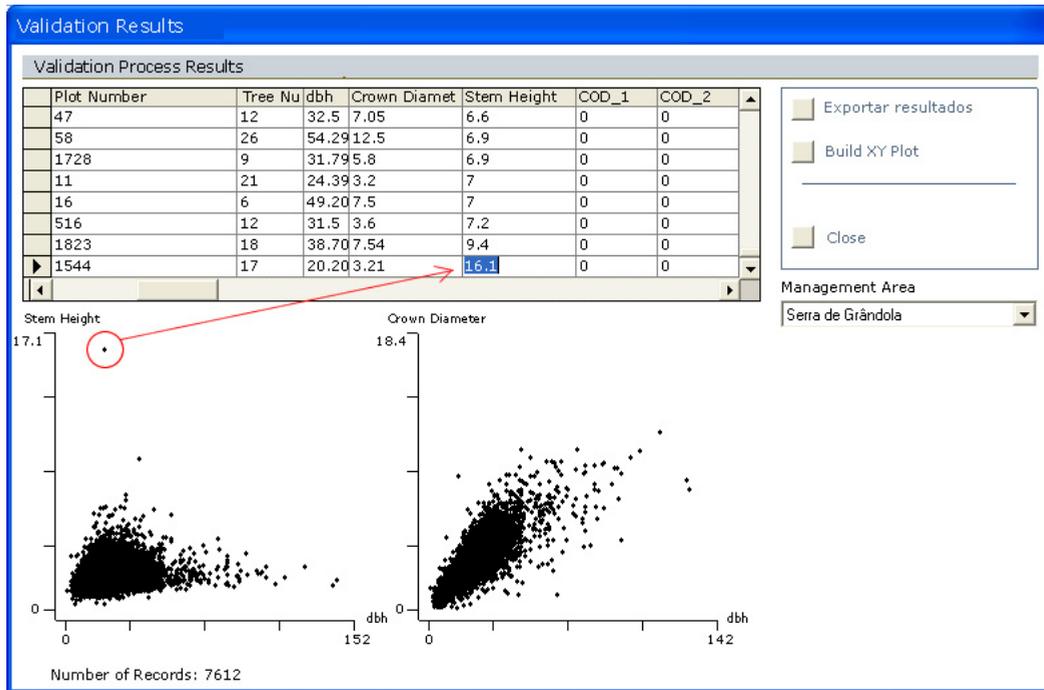


Figure 5. Example of a graphical validation routine.

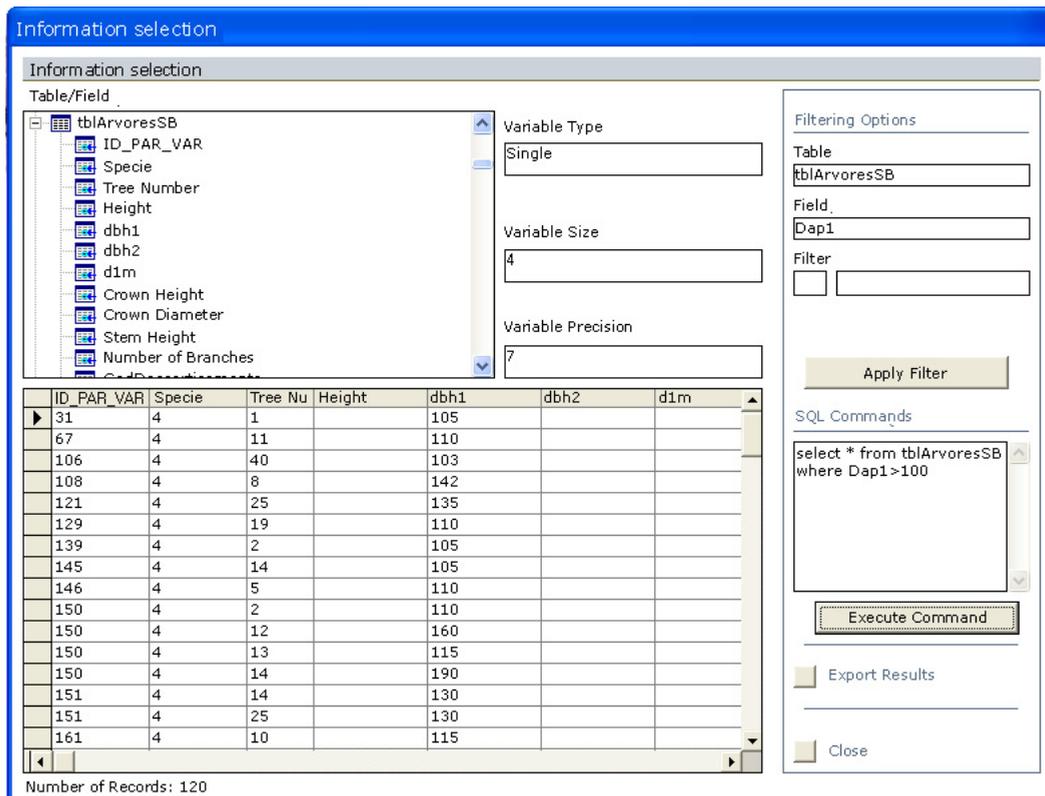


Figure 6. Example of an ad-hoc data access interface.

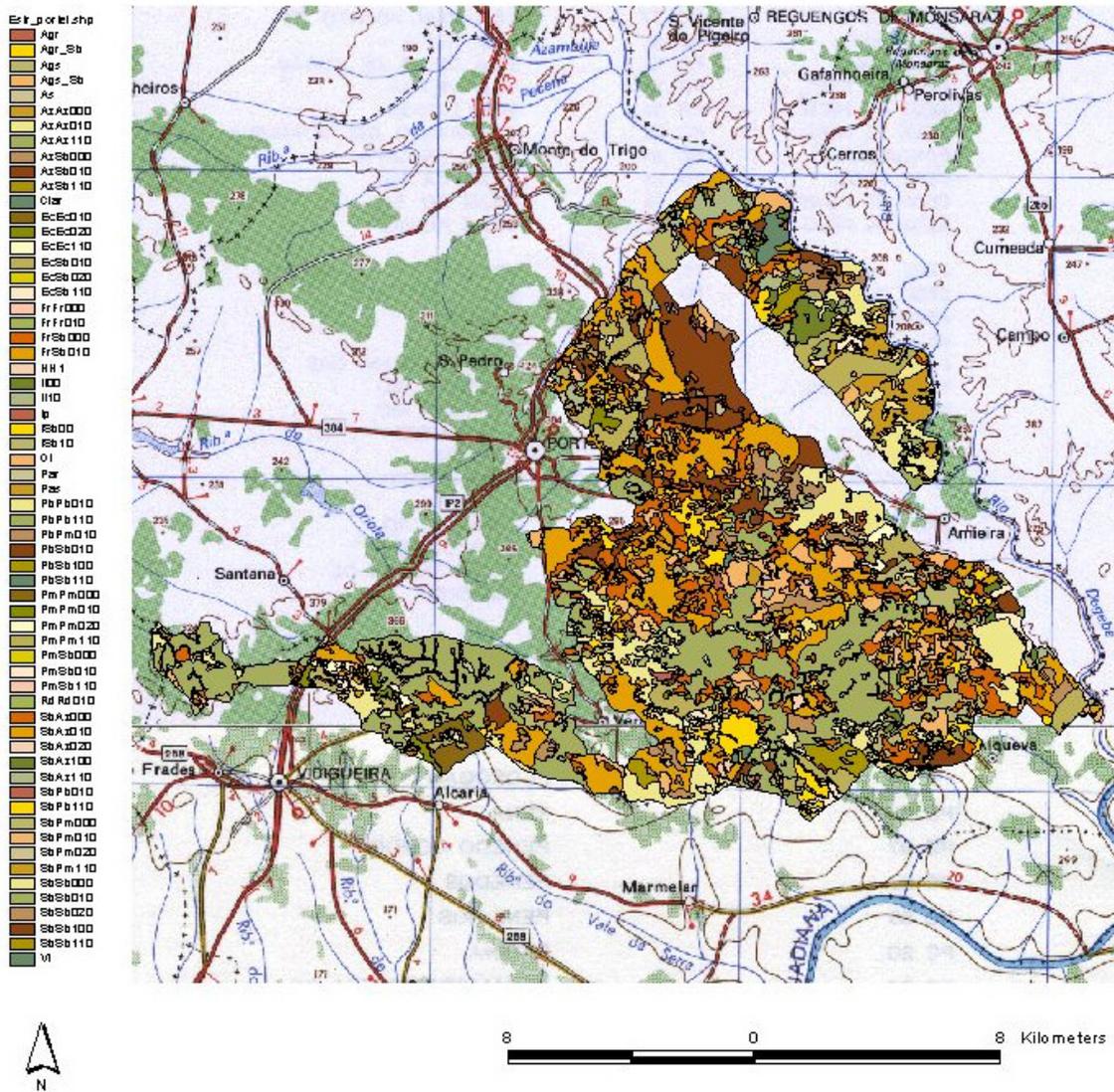


Table 7. Visual aid interfaces to information representation.

data, technology and people for a better assessment of Mediterranean natural resources in the test areas. It was thus used to characterize the test areas by both NGO.

The development of this information system accounted for about 24 percent of the project costs. Data acquisition was the most expensive cost item as it accounted for about 58% of total project costs. The remaining budget was used to fund outreach and demonstration sessions. The system is currently installed in a microcomputer at the headquarters of each NGO where this computing capacity was already available. The documentation and the on-line help facilities enable its operation by current local staff. There is no need for computer science expertise to operate the system. Characterization of the test areas by the NGOs was successful and could not have been accomplished with classical non-automated approaches within an acceptable time frame. The system is also being used to provide data and

information about Serra de Grândola and Serra de Portel to land-owners, researchers and policy-makers. Finally, the system is currently being used also as a tool for demonstrating the framework potential to effectively address current Mediterranean ecosystem management data quality needs.

4. DISCUSSION

The main focus of this article was on developing a data quality framework to address the specificity of Mediterranean ecosystem management. The effectiveness of an information system may be assessed by the usefulness of data it processes. Thus the proposed framework combined three key elements of an information system – data, technology and people – to enhance data quality and fitness for use. This approach was successfully

tested with forestry, wildlife and socio-economic data from two Mediterranean sensitive areas encompassing circa 40 800 ha in southern Portugal.

Results showed that the proposed framework did address concerns with intrinsic, accessibility, contextual and representational data quality categories. Data producers and data consumers understood the importance of technical specifications in each information system module to improve data quality. Results further demonstrated the crucial importance of defining data user roles within such a knowledge-based human network. These results confirmed the assertion that the involvement of data consumers in information systems development is an important factor for its acceptance and subsequent use. It further helped address problems resulting from changes in data producers' teams.

All Mediterranean ecosystem management problems have unique features (e.g. different spatial and temporal scales, different ownerships). In some cases, focusing on specific data quality dimensions may be more important than focusing on others. In other cases, a number of user roles may coalesce (e.g. the data custodian and the data manager) or else we may have several people with the same role. Nevertheless, the proposed framework is general enough to accommodate this variability and to adapt to better enhance data quality. Furthermore, this approach is not tied to a specific technological option. All software that meets the data quality requirements may be a choice for the data manager. Interoperability is just a criterion for selecting technological options.

Research on methods for optimizing the allocation of organizational resources for data quality enhancement may contribute to fine tune and adapt the proposed framework to address specific problems. Ballou and Tayi [2] focused on data integrity maintenance and presented an integer programming formulation for such an allocation problem. These methods might provide a theoretical basis to determine the type and number of validation routines to be considered by the data quality manager and the data custodian. Yet they involve the quantification of costs associated with defective data items and this is not trivial in most Mediterranean ecosystem management problems. Research on more metrics to better assess and quantify Mediterranean natural resources data quality dimensions may also enhance the proposed framework. These additional metrics would help define and further assess the data quality policy. Moreover, research on the impact of providing information about data quality [8] may also help better define interactions between user roles in the proposed framework.

The development of an automated information system as opposed to traditional manual systems is deemed necessary both for efficiency and for effectiveness of data processing, information production and knowledge generation. The diversity of Mediterranean natural resources and the amount of data needed for its assessment can hardly be handled with a manual system. Research on benefit/cost analysis of investment on information systems [46] may help provide a more accurate assessment of thus investment returns.

The complexity of Mediterranean ecosystem management presents a challenge to experts on a variety of disciplines ranging from natural resources inventory to management planning and policy making. The proposed framework may be used to

help confront this challenge. It may be used to address new data needs and the integration of other decision support modules within the information system. To the authors' knowledge, this is the first data quality framework proposed for Mediterranean ecosystem management planning. Hopefully it will provide a basis for further studies on Mediterranean data quality and it will contribute to increase fitness of data for better decision-making.

Acknowledgments: This work was supported by the following projects: (1) Project Sapiens 36332/AGR/2000, with the title "Forest ecosystem management: an integrated stand-to-landscape approach to biodiversity and to ecological economic and social sustainability", funded by FCT, POCTI, and FEDER; (2) Project with the title "Prospective studies of the productive potential of cork oak stands in Serra de Grândola and of *Pinus pinea* stands in Vale do Sado", funded by PAMAF, INIA; (3) Project with the title "Prospective studies of the productive potential of cork and holm oak stands in Serra de Portel", funded by PEDIZA; (4) Project Life with the title "MONTADO – Conservation and Valuation of Montado Forestry Systems for Fighting Desertification" and (5) Project with the title "Definition of a wildlife management information system – inFauna" L-0120, Iniciativa Comunitária PME, funded by the AI.

REFERENCES

- [1] Aussenac G., Ecology and ecophysiology of circum-Mediterranean firs in the context of climate change, *Ann. For. Sci.* 59 (2002) 823–832.
- [2] Ballou D.P., Tayi G.K., Methodology for allocating resources for data quality enhancement, *Commun. ACM* 32 (1989) 320–329.
- [3] Barbati A., Corona P., De Natale F., Tosi V., Marchetti M., Forest remote sensing in Italy in the framework of FRA2000, in: Zawila-Niedzwiecki T., Brach M. (Eds.), *Remote sensing and forest monitoring*, EUR 19530, Office for Official Publications of the European Communities, Luxembourg, 2000, pp. 284–299.
- [4] Barreira S., Soares P., Tomé M., Protocolo de medição de parcelas de inventário da região de intervenção da Direcção de Serviços de Florestas da DRAALG, Relatório Técnico-Científico do GIMREF nº 5/2002, Def, ISA, UTL, Lisboa, 2002, 26 p.
- [5] Borges J.G., Sistemas de apoio à decisão para o planeamento em recursos naturais e ambiente. Aplicações florestais, *Revista Florestal* 9 (1996) 37–44.
- [6] Borges J.G., Paradigmas, tecnologias e equívocos em gestão dos recursos florestais, *Revista Florestal* 12 (1999) 27–34.
- [7] Borges J.G., Oliveira A.C., Costa M.A., A quantitative approach to cork oak forest management, *For. Ecol. Manage.* 97 (1997) 223–229.
- [8] Chengalur-Smith I.N., Ballou D.P., Pazer H.L., The impact of data quality information in decision making: An exploratory analysis, *IEEE Trans. Know. Data Eng.* 11 (1999) 853–864.
- [9] Ciancio O. (Ed.), *The forest and man*, Accademia Italiana di Scienze Forestali, Florence, 1997, 331 p.
- [10] Corona P., Marchetti M., Forest information framework and requirements for designated conservation areas in Italy, in: Morandini R., Merlo M., Päivinen R. (Eds.), *Forest management in designated conservation/recreation areas*, EFI-AISF Conference Proceedings, Firenze, 1998, pp. 255–267.
- [11] Corona P., Marchetti M., Towards an effective integration of forest inventories and natural resources surveys: the Italian perspective, in: Hansen M., Burk T. (Eds.), *Integrated Tools For Natural Resources Inventories In The 21st Century*, USDA, Forest Service, North Central Research Station GTR, NC-212, 2000, pp. 28–34.
- [12] Davenport, Save IT's soul – human centered information management, *Harvard Bus. Rev.* 72 (1994) 119–131.
- [13] DGF, Inventário florestal nacional, 3ª Revisão, 1995–1998. DGF, Lisboa, 2002, 234 p.

- [14] Dykstra D.P., Information systems in forestry, *Unasylva* 48 (1997) 10–15.
- [15] Groupement d'Intérêt Scientifique "Incendies de forêts" État des connaissances sur l'impact des incendies de forêt: mise en place de protocoles expérimentaux pour le suivi des incendies et de la reconstitution des écosystèmes forestiers, *For. Méditerran.* 21 (2000) 342–352.
- [16] Hoff C., Rambal S., An examination of the interaction between climate, soil and leaf area index in a *Quercus ilex* ecosystem, *Ann. For. Sci.* 60 (2003) 153–161.
- [17] Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P., *Fundamentals of datawarehouses*, Springer-Verlag, 2000, 195 p.
- [18] Kangas J., Store R., Leskinen P., Mehtälö L., Improving the quality of landscape ecological forest planning by utilising advanced decision-support tools, *For. Ecol. Manage.* 132 (2000) 157–171.
- [19] Lund G.H., Thomas E., A primer on evaluation and use of natural resources information for corporate databases, USDA Forest Service, Washington Office, General Technical Report WO-62, Washington DC, 1995, 168 p.
- [20] Mendes A, Forest owners' collective action against the risk of forest fire: a game theoretical approach, in: *The Economics of Natural Hazards in Forestry*, 7–10, June, 2001, CTEFC-EFI, Solsona, Catalonia, Spain, IUFRO Division 4. 4.04.02 – Managerial Economics in Forestry, 4.13.00 – Managerial, Social and Environmental Accounting, Padua, Padua University Press, 2001, pp. 137–141.
- [21] Merlo M., Rojas Briales E., Public goods and externalities linked to Mediterranean forests: Economic nature and policy, *Land Use Policy* 17 (2000) 197–208.
- [22] Miragaia C., Borges J., Falcão A., Tomé M., inFlor – a management information system of forest resources, in: Pukkala T., Eerikainen K. (Eds.), *Modeling the Growth of Tree Plantations and Agroforestry Systems in South and East Africa*, University of Joensuu, Faculty of Forestry Research Notes 80, 1998, pp. 131–142.
- [23] Miragaia C., Borges J.G., Tomé M., inFlor, um sistema de informação em recursos florestais. Aplicação em gestão na Mata Nacional de Leiria, *Revista Florestal* 12 (1/2) (1999) 51–56.
- [24] Montero G., Cañellas I., Gestión sostenible del monte Mediterráneo en España, *Unasylva* (número especial sobre bosques Mediterráneos) 50 (1999) 29–34.
- [25] Montgolfier J., Les forêts méditerranéennes et leur aménagement. 1^{re} partie, *For. Méditerran.* 7 (1985) 161–168.
- [26] Montgolfier J., Les forêts méditerranéennes et leur aménagement. 2^e partie, *For. Méditerran.* 8 (1986) 49–60.
- [27] Orr K., Data quality and systems theory, *Commun. ACM* 41 (1998) 66–71.
- [28] Palahi M., Pukkala T., Miina J., Montero G., Individual tree growth and mortality models for Scots pine (*Pinus sylvestris* L.) in the north-east Spain, *Ann. For. Sci.* 60 (2003) 1–10.
- [29] Palahi M., Pukkala T., Optimizing the management of Scots pine (*Pinus sylvestris* L.) stands in Spain based on individual-tree models, *Ann. For. Sci.* 60 (2003) 105–114.
- [30] Papanastasis V.P., Kazaklis A., Land use changes and conflicts in the Mediterranean – Type Ecosystems of Western Crete, in: Rundel et al. (Eds.), *Landscape Degradation and Biodiversity in Mediterranean – Type Ecosystems*, Ecological Studies, 136, Springer-Verlag, Berlin-Heidelberg, 1998, pp. 141–154.
- [31] Pelkki M.H., Rose D.W., Everest G.C., The development of a global data model for natural resources, *The Compiler* 12 (1994) 32–42.
- [32] Pettenella D., Forest resources environmental accounting: two Italian case-studies, in: Buttoud G., Jöbstl H., Merlo M. (Eds.), *Accounting and managerial economics for an environmentally-friendly forestry*, INRA, Economie et Sociologie Rurale, Nancy, Actes et Communications, 15, 1998.
- [33] Redman C.R., The impact of poor data quality on the typical enterprise, *Commun. ACM* 41 (1998) 79–82.
- [34] Ribeiro R., Miragaia C., Borges J., A prototype management information system for plantation forests in eastern and southern Africa, in: Pukkala T., Eerikainen K. (Eds.), *Establishment and management of tree plantations in South and East Africa*, University of Joensuu, Faculty of Forestry Research Notes 120, 2000, pp. 121–131.
- [35] Riera P., Mogas J., Evaluation of a risk reduction in forest fires in a Mediterranean region, *For. Policy Economics* (in press).
- [36] Rondeux J., Forest inventories and biodiversity, *Unasylva* 50 (1999) 35–41.
- [37] Rose D., Information management for improved forestry decisions, in *Proceedings of the 1985 SAF Symposium on Systems Analysis in Forest Resources*, University of Georgia, Athens, 1987, pp. 520–531.
- [38] Scarascia-Mugnozza G., Oswald H., Piussi P., Radoglou K., Forests of Mediterranean region: gaps in knowledge and research needs, *For. Ecol. Manage.* 132 (2000) 97–109.
- [39] Spurr S.H., *Forest Inventory*, Ronald Press, New York, 1952, 488 p.
- [40] Strong D.M., Lee Y.W., Wang R.Y., Data Quality in Context, *Commun. ACM* 40 (1997) 103–110.
- [41] Tayi G.K., Ballou D.P., Examining data quality, *Commun. ACM* 41 (1998) 54–57.
- [42] Tomé M., Coelho M.B., Pereira H., Lopes F., A management oriented growth and yield model for cork oak stands in Portugal, in: Amaro A., Tomé M. (Eds.), *Empirical and Process-Based Models for Forest Tree and Stand Growth Simulation*, Edições Salamandra, Novas Tecnologias, Lisboa, Portugal, 1999, pp. 271–289.
- [43] Wand Y., Wang R.Y., Anchoring data quality dimensions in ontological foundations, *Commun. ACM* 39 (1996) 86–95.
- [44] Wang R.Y., Storey V.C., Firth C.P., A framework for analysis of data quality research, *IEEE trans. know. data eng.* 7 (1995) 623–640.
- [45] Wang R.Y., Total Data Quality Management, *Commun. ACM* 41 (1998) 58–65.
- [46] Ward J., Taylor P., Bond P., Evaluation and realisation of IS/IT benefits: an empirical study of current practice, *Eur. J. Inf. Syst.* 4 (1996) 214–225.
- [47] Varma V.K., Ferguson I., Wild I., Decision support system for the sustainable forest management, *For. Ecol. Manage.* 128 (2000) 49–55.