Improving the Usability of the Land Cover and Use Information System of Spain (SIOSE): A Proposal to Distribute New Thematic Layers and Predefined Reclassifications

Benito Zaragozí[1] [a], José Tomàs Navarro-Carrión[2] [b], Jesús Javier Rodríguez-Sala[3] [c], Sergio Trilles[4] [d] and Alfredo Ramón-Morte[2] [e]

[1] *Departament de Geografia, Universitat Rovira i Virgili, C/Joanot Martorell, Vilaseca, Spain*

[2] *Instituto Interuniversitario de Geografía, Universidad de Alicante, C/ San Vicente s/n, Alicante, Spain*

[3] *Centro de Investigación Operativa, Universidad Miguel Hernandez de Elche, Av. de la Universidad, Elche, Spain*

[4] *Institute of New Imaging Technologies, Universitat Jaume I, Av. Vicente Sos Baynat s/n, Castellón de la Plana, Spain*

Keywords: Land Use, Land Cover, Geodatabase, Object-oriented, Siose.

Abstract: Information on land use and land cover (LULC) is fundamental in the study and planning of human activities. In recent years, accessibility to quality geographical information has significantly increased, and this is also true for the case of LULC datasets. In Spain, the Land Cover and Use Information System of Spain (SIOSE) is concerned with harmonising access to this type of information through an object-oriented model and a series of technical specifications that regional administrations must follow. However, the information from SIOSE is so rich and complex that there is a usability gap that makes this data not exploited to its full potential in some contexts. In this communication, we analyse the context in which this *usability gap* occurs, its causes and consequences. Among other possible improvements, we suggest that enriching the SIOSE database with new thematic information would make its use more attractive and reduce the *usability gap* for less expert users. We propose an extension to the SIOSE object-oriented data model that will make it possible to enrich the LULC data with new data that are useful for various types of studies.

1 INTRODUCTION

The information on land use and land cover (LULC) is essential for the planning of human activities. This type of information has the virtue of agglutinating biophysical information and socio-economic uses of a territory. The design of LULC data warehouses is more often related to the concept that geographers have of the landscape configuration as an object of study (Antrop, 2006), which makes them suitable for many landscape science studies. This approach has been applied in numerous studies that show the close relationships between LULC and essential ecosystem services (Foster et al., 2003; Polasky et al., 2011), which casts LULC data as a strategic source of information for natural resource management and land management (Valcarcel and Castaño Fernández, 2013).

1.1 LULC Data Production in the EU

Land use information has been systematically collected in Spain since the creation of the National Geographic Institute (IGN) in 1870, although initially only as a fundamental source for the preparation of the National Topographic Map. The development of Geographic Information Technologies, in the second part of the last century, promoted the creation of large official repositories of digital geographic information compiled through different means, such as: (1) the digitisation of official cartographic information, (2) the incorporation of GPS field data, (3) the interpretation of aerial photography and (4) the development of Remote Sensing. In a broader context, the Coordination of Information on the Environment programme (CORINE) was initiated in 1985 by the Eu-

[a] https://orcid.org/0000-0003-2501-484X
[b] https://orcid.org/0000-0002-8718-218X
[c] https://orcid.org/0000-0002-3796-0692
[d] https://orcid.org/0000-0003-0557-6319
[e] https://orcid.org/0000-0002-9304-0719

Zaragozí, B., Navarro-Carrión, J., Rodríguez-Sala, J., Trilles, S. and Ramón-Morte, A.
Improving the Usability of the Land Cover and Use Information System of Spain (SIOSE): A Proposal to Distribute New Thematic Layers and Predefined Reclassifications.
DOI: 10.5220/0009579502940301
In *Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2020)*, pages 294-301
ISBN: 978-989-758-425-1

ropean Commission in order to dynamically collect, coordinate and structure the inventory of environmental information of the European Union. The Corine databases and several of its programmes were taken over by the European Environment Agency (EEA). Among these, there is the inventory of land cover at a 1:100,000 scale for the whole European territory. This "Land-Cover" project (CORINE Land-Cover; CLC) has set a global precedent (Büttner et al., 2002).

In a global scope, LULC information is used for strategic issues within the United Nations Framework Convention on Climate Change (1992) and the revision of the Kyoto Protocol (1998). In this context, many users are using these data for different interests, which has led to the development of a specific service in Europe to meet this heterogeneous demand, the "Copernicus Land Monitoring Service" (CLMS), within the Global Monitoring for Environment and Security (GMES) initiative of the European Environment Agency (EEA), which in turn depends on the Earth Observation Program "Copernicus" of the European Space Agency (ESA).

In Spain, the Ministry of Development and Ecological Transition addresses the general need for LULC data, under the legal framework of the European Directive INSPIRE (2007/2/CE). The National Geographic Institute (IGN) takes the role of coordinator in the management and creation of data on land occupation according to the National Territory Occupation Plan (PNOT), yielding millions of downloads on the open data platforms of this body and of the regional spatial data infrastructures. The main users of these data are the General State Administration itself and the regional administrations, almost always for issues regarding agrarian policy, environmental management, urban development or cartographic elaboration. Applications to projects in University research laboratories, research officer centres and public and private companies are also remarkable.

1.2 The Appearance of Object-oriented LULC Classifications

Throughout the development of these programs to produce quality LULC data, many technical and technological changes took place. In this sense, the paradigm of hierarchical classification of LULC used at the end of the 20th century by the CORINE Program – and by others in the same period, such as the "land use and land cover classification system" of the United States Geological Survey (Anderson et al., 1976) – showed inadequacies at the beginning of the current century. The baseline of said limitations is that mutually exclusive classes were often more oriented to the realisation of maps than to the analysis and diagnosis of reality. The need for managing more complex and bulky data sets motivated the emergence of trends aimed at applying an object-oriented paradigm (Villa et al., 2008).

Following these trends, the EAGLE group within the European Network of Information and Observation on the Environment (EIONET) was created to define techniques that allow optimising the integration and harmonisation of LULC data from the official repositories of each country at pan-European scale (Arnold et al., 2013). EAGLE proposes an object-oriented data model (OODM) which takes into account standards or reference code lists, such as Corine LC and technical specifications driven by INSPIRE (2007/2/CE) and ISO standard 19144-2 (LCML-Land Cover Meta Language). In 2005, the object-oriented database of the Information System on land occupation in Spain (SIOSE), emerged as an integrated initiative in the EIONET Network (Del Bosque González et al., 2005). SIOSE is developed in Spain through the PNOT, under the coordination of the IGN and the National Geographic Information Center (CNIG). SIOSE is backed by a data model that conforms to INSPIRE technical specifications, and its design follows the indications of the EAGLE group, ensuring compatibility and comparability with pre-existing databases such as Corine CLC90, CLC00, Murbandy/Moland or LCCS of the United Nations FAO.

Since there are no mutually exclusive classes, such an object-oriented model will not incur information losses during the labelling process, thus making feasible to store LULC statistical observations at levels of detail that hierarchical classification models are not capable of due to its dichotomic nature (Omrani et al., 2015). It has an impact on the economic savings in the production of data sets. The elements or variables of the landscape are unique in their definition, which allows the possibility of obtaining custom thematic outputs, according to user needs. As a consequence, the object-oriented approach allows generating dynamic and extensible classifications to respond to future needs. In addition, new types of parameters can be included in different versions of the database without conflict with the previous data (Valcarcel et al., 2008).

2 USABILITY IN THE SIOSE DATABASE

Regardless of the significant advantages of applying the object-oriented paradigm to SIOSE, the SIOSE

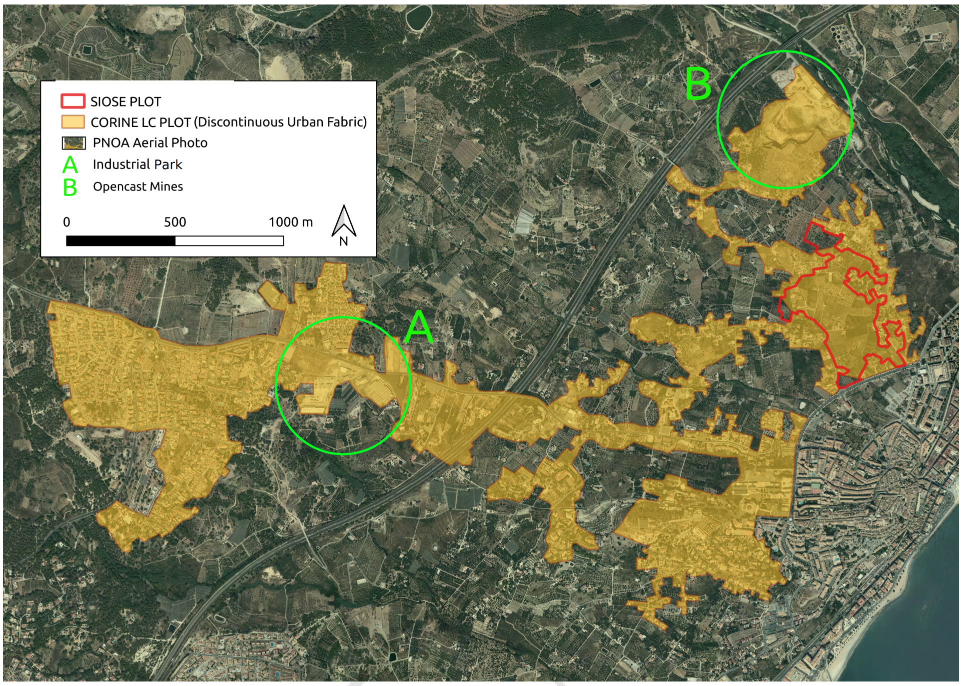

Figure 1: Geometric comparison between a CORINE LC plot (orange limit) and a SIOSE plot (red limit) shows a clear difference in size and scale (1:100.000 and 1:25.000). Furthermore, the classification of the CORINE plot (Discontinuous Urban Fabric) hides other realities such as those found in the SIOSE plot (different types of agriculture) or the cases indicated by green circles: A (Industrial Park) or B (Opencast Mines).

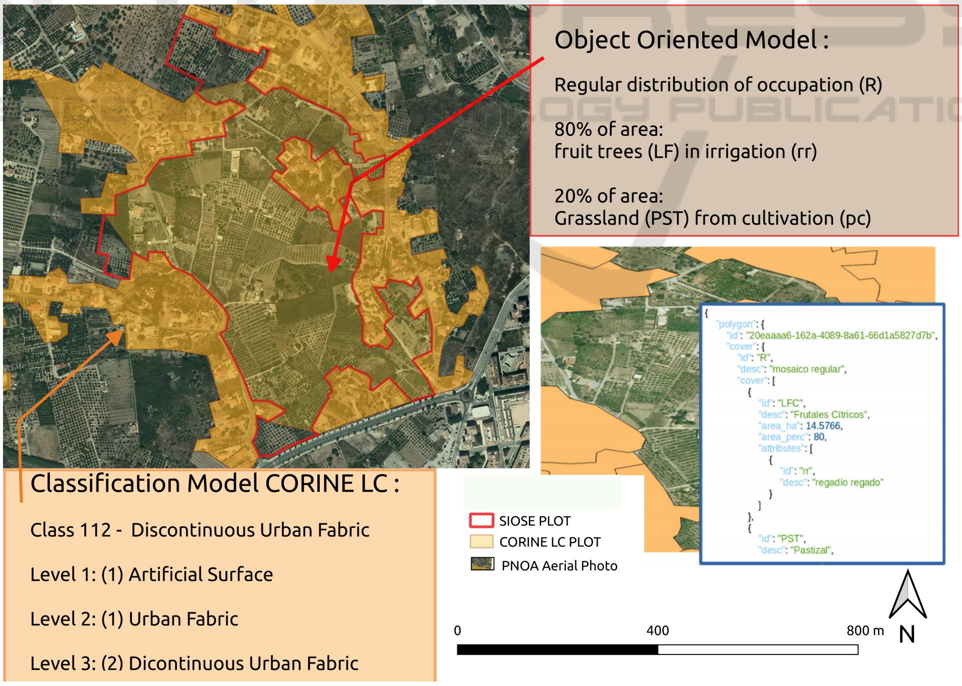

Figure 2: The SIOSE Object-Oriented Model offers more wealth of information than the CORINE LC Classification Model. SIOSE provides a higher level of detail as to the types of crops and the area quantifiers. In this example, the plot is composed of 2 covers, 80% of irrigated fruit trees and 20% of cultivated pastures, both regularly distributed).

data model has also certain drawbacks, and it is often difficult to understand for non-expert users due to differences in operation concerning traditional hierarchical classifications which may arguably be regarded as more straightforward. Hierarchical data models, such as CORINE LC, facilitate the interpretation of information by reducing thematic resolution so that directly observable data mostly fit the model requirements. On the other hand, the object-oriented model must be adapted to relational database management systems with spatial capabilities, and database administrators have to deal with this incompatibility at the conceptual level. SIOSE is a case of the object-relational impedance mismatch that has been clearly identified in the literature as a problem of data structure due to paradigm differences (Ireland et al., 2009). Another drawback derived from the complexity of the system is the analysis of the evolution of Land Occupancy and the detection of changes, which must be carried out not only in geometry but also in semantics (Valcarcel and Castaño Fernández, 2013). These types of problems have already been faced in different studies, where the authors point to the usefulness of the data collected by SIOSE, but also to the difficulty in handling those data. As for example, some effort was necessary to use SIOSE LULC data in studies related to areas such as climate change (Ropero et al., 2019; Olaya-Abril et al., 2017), flood risk and mapping of flood areas (Morte et al., 2019), farmland abandonment (Peña-Angulo et al., 2019), Wildland–Urban Interface (Badia et al., 2014) or purely cartographic studies (García-Álvarez, 2018).

Nevertheless, one of the most outstanding aspects of SIOSE goes beyond the object-oriented data model and instead resides in the fact that it has led to the development of a coordinated and participatory production system that integrates data from all interested public administrations. This results in an economy of effort and an increase in the quality of the outcomes which has earned it the 2013 UN Public Service Award.

In this general context, the SIOSE-INNOVA project (http://siose-innova.es/) aims to address both issues, the technical problems related to the SIOSE object-oriented data model and the usability drawbacks derived from it. The main objectives of this work are as follows:

1. To define the SIOSE usability gap by analysing the ecosystem where these data are used (actors, contexts and use cases).
2. To evaluate different solutions to overcome the usability gap.
3. To propose a “soft” solution to increase the attractiveness of the SIOSE database, which complements other measures based on alternative technologies or mediating platforms.

3 THE SIOSE USABILITY GAP

In this work, *usability* is considered as a measure of quality that evaluates how easily the SIOSE database is used. The word *usability* also refers to methods to improve usability during the design process. Among the factors that determine the usability, we can mention accessibility, readability, navigability, ease of learning, speed of use, user efficiency and error rates (Ben Ramadan et al., 2017). According to this definition, there may be different actors and use cases in which the SIOSE database could be more challenging to manage, query or update, so finding a general solution for these many scenarios would require a broader perspective from the design phase. Thankfully, the SIOSE object-oriented model favours extensibility (Valcarcel et al., 2008).

Usability is also a characteristic of a system that is intended to be used (1) by a specific type of user, (2) to complete the task for which the system has been designed, and (3) in the context in which the interaction occurs (Ben Ramadan et al., 2017). In this sense, it is essential to perform a detailed description of these three components of the SIOSE *environment* of actors, applications and contexts. In this section, we have attempted to do so.

In figure 3, we show a use case diagram in which we perform an analysis to identify, clarify, and organise system elements. Of course, this is only a model that may not be applicable in some cases but, to the best of our knowledge, this model represents most of the uses of SIOSE present in the scientific literature and the uses that we know from our own experience. The SIOSE project must take into account three different phases that are needed for leveraging the use of the LULC data. These phases are (1) the production phase, (2) the data integration phase and (3) the usage phase. In each of these phases, we can see different actors – individuals involved with some parts of the system according to their roles – that accomplish one or more tasks for the SIOSE environment. On the bottom left-hand side of the diagram, there are the main producers of SIOSE, who carry out their tasks from an official request (see section 1). On the bottom right-hand side, we have placed sources of Volunteer Geographic Information, who are encouraged to contribute *in situ* data to increase the SIOSE data quality. In the usage phase, we can distinguish those users with little or no knowledge about spatial databases — but who use data derived from SIOSE for their spe-

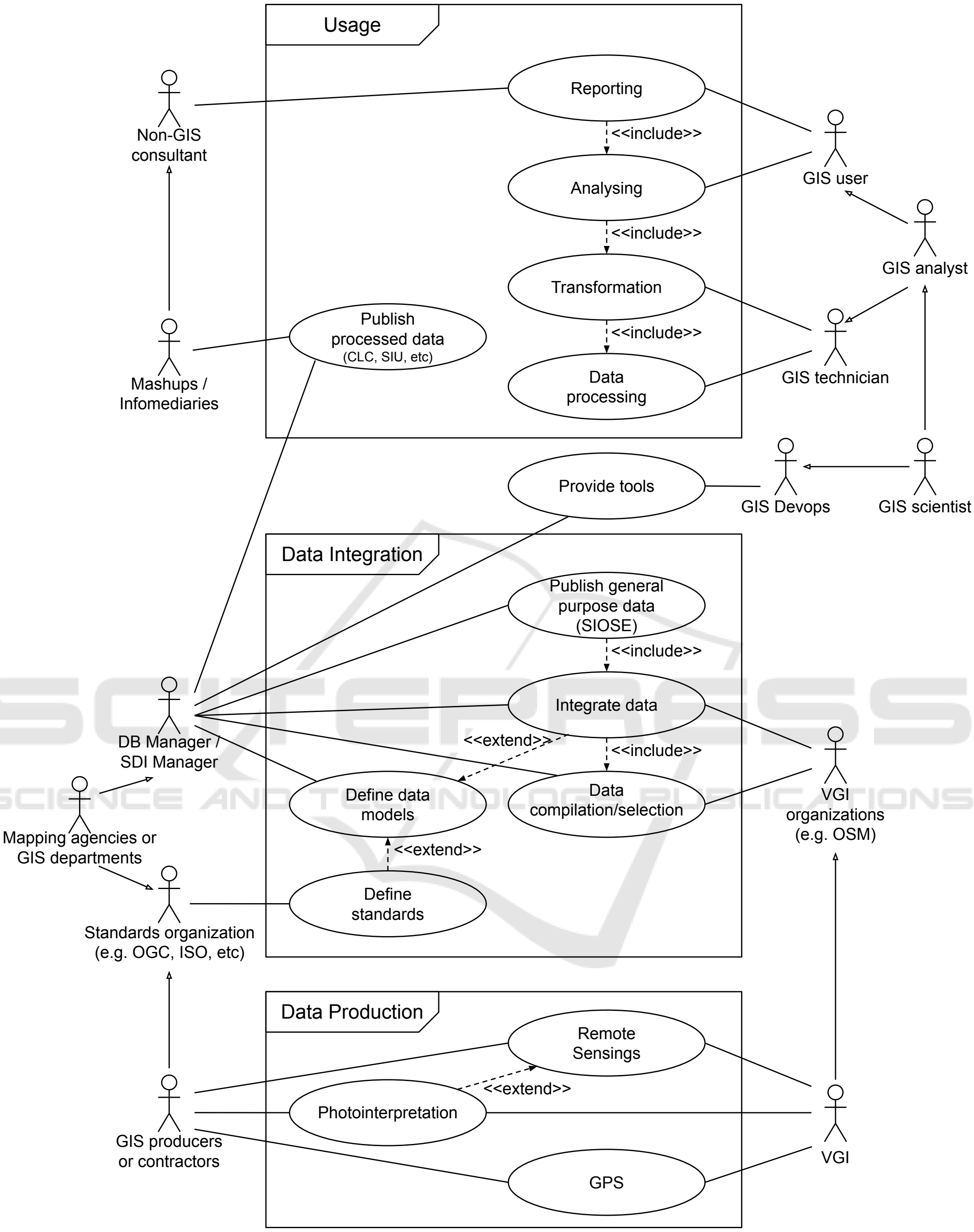

Figure 3: Use case diagram showing the actors, applications and contexts where the SIOSE is developed.

cific objectives — from other users with varying experience in exploding the SIOSE database.

The diagram displays actors who can perform different tasks related to their context, so they are more or less attached to a particular phase. For example, a GIS technician may be interested in processing or transforming some SIOSE data — for mapping or analytical purposes — but they usually will not have any interest or resources for developing new tools. Another example would be that of the Non-GIS consultant, who may be interested in reporting some information. Still, they lack experience in working with geospatial data — so they would need the collaboration of a GIS analyst. However, once the database manager publishes the general-purpose SIOSE data sets or the required specific processed data, there are no specific actors taking care of providing new tools or facilitating the access to this information (e.g. in the form of operating environments with custom reclassifications and pre-configured equivalence maps). This is represented as the *Provide tools* task being out of the main production contexts.

The usability problems of SIOSE may require the development of new tools that help less expert users to convert LULC information to custom classifications that are more useful in their fields of study. Another possibility would be to strengthen those aspects of training that the different actors require in each case. However, this would affect a significant number of users and does not seem necessary in a context where most official GIS data is distributed in a way that is most usable for users. To date, few studies have attempted to solve this problem by using different approaches including the development of new interoperable web services (Fernández Villarino et al., 2012), the use of alternative database technologies (Navarro-Carrión et al., 2016) or the design of an ArcGIS extension for performing the reclassification of hierarchical LULC information (Fernández Noguerol, 2017). Apart from these technical solutions, other possibilities may involve a reinterpretation of the SIOSE data model to allow more specific information to be attached to the database. From this perspective, it would be necessary to (1) identify the main applications that users make of the SIOSE data, and (2) propose at least one mechanism for allowing any user to add new thematic information – or useful reclassifications – that can be reused by the community. In this paper, we address the second question (an extensibility mechanism) in a generic way, considering that users may have very different needs.

4 AN EXTENSION TO THE SIOSE DATABASE MODEL

Given its object-oriented design, the SIOSE data model and other similar models mentioned in the introduction are well prepared to add new elements to a LULC description without the previous information being affected.

The SIOSE data model uses the *composite pattern*, which is a partitioning design pattern which treats a group of objects in the same way as a single instance of the same type of object (Gamma et al., 1994). In figure 4 this is shown in the three central classes (*LandCoverComponent*,*LandcoverComposite* and *LandCoverLeaf*) and their relationships. This model can be easily extended through the *Attribute* interface, which represents the land-use part of each land cover component or unit.

We show our proposal in figure 4 and we can use the SIOSE polygon described in figure 2 to set an example on how to read the class diagram. That figure is showing LULC information from the SIOSE-2011 database, which is the *LandCoverDataSet* in this example. The red polygon can be seen as an instance of the *LandCoverUnit* class, which inherits from a geometry object. A composite land cover composes this polygon with an 80% of the surface covered by fruit trees and a 20% of grassland. The composite pattern would allow any of these land covers to be further decomposed (e.g. 90% percent are old orange trees, but a 10% is planted with young avocados, which are more productive). As seen in figure 2, in the current SIOSE data model, land use information (*Attributes*) can only describe a full land cover (e.g. irrigation for fruit trees) but, in our proposal a more precise description could be made, for example, to explain if different irrigation systems coexist (*Presence*), which is the predominant irrigation system (*Percentage*) or store the exact length of the irrigation network (*Countable*). In this model, it is not necessary that the geometry be a polygon, we could also use other types of geometries (for example, points to represent trees, linear chains for power lines, among other possibilities).

The main difference from the current data model is the greater semantic richness of the attributes, which have been so far less important according to the production scale. This extension or similar modification is necessary to import or to combine new data from other sources. For example, it would be possible to build some thematic information from OpenStreetMap or save a custom reclassification and distribute it without losing the link to SIOSE.

5 CONCLUSIONS AND FUTURE APPLICATIONS

The SIOSE database is an important resource for performing many different geographical studies. The database is modelled using an object-oriented approach, which has many technical advantages but also a few usability problems. This work has shown that, despite the aforementioned usability problems of

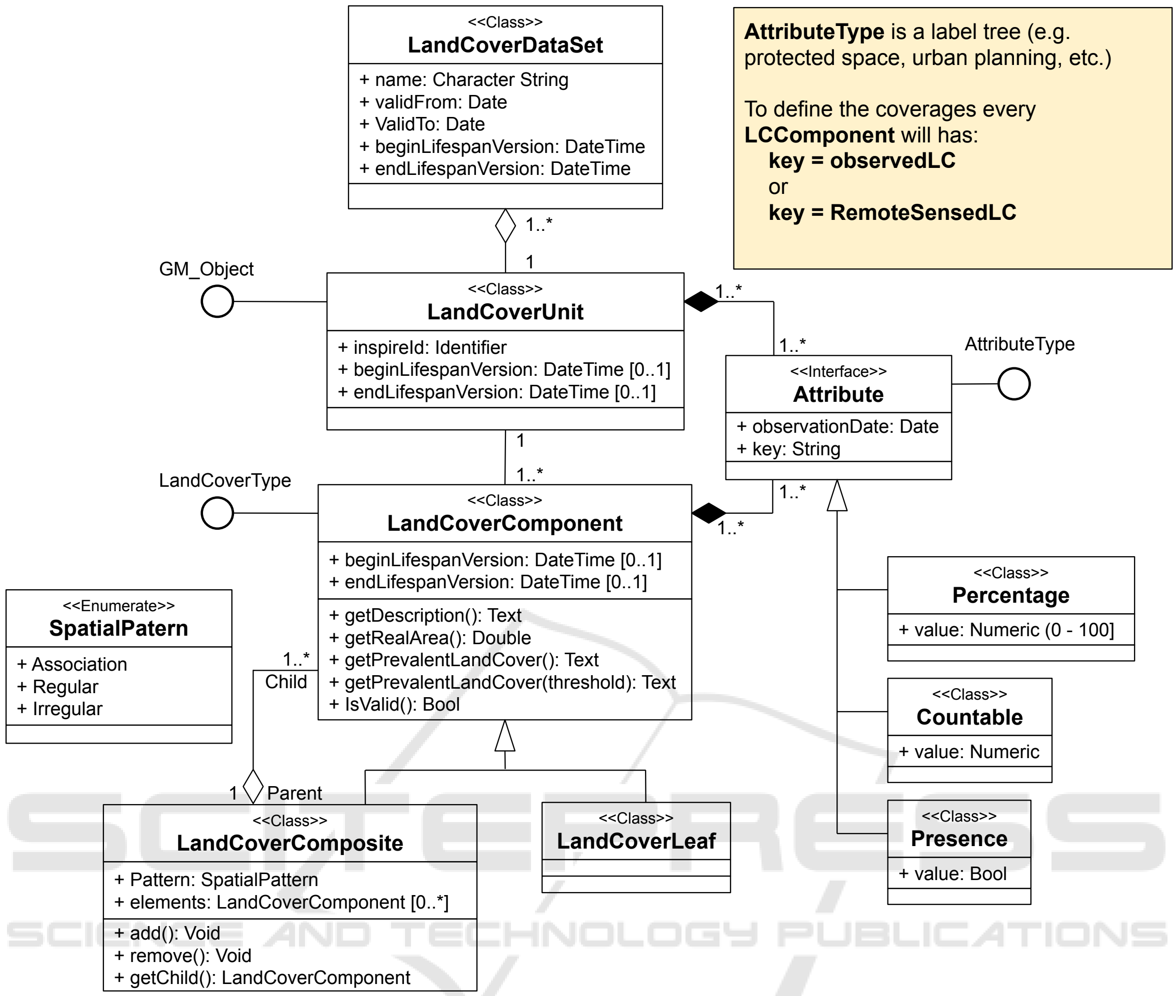

Figure 4: Class diagram for extending the SIOSE data model.

SIOSE, the object-oriented model (a *composite pattern*), can be easily extended to enrich the LULC database further.

We have proposed a simple solution for extending the model, but other possibilities could be evaluated. We consider it attractive not to lose sight of this possibility since the fact of enriching the database would generate certain positive effects: (1) it would compensate even more the effort to work with SIOSE, and (2) it would open new possibilities in the distribution of geographic information (e.g. distribute thematic reclassifications of interest in some fields). In other words, it would be possible to produce and distribute specific data sets for different applications (e.g. fire risk, planning, tourism, etc.).

ACKNOWLEDGEMENTS

This work was supported by grants from the Spanish Ministry of Economy and Competitiveness, project SIOSE-INNOVA (CSO2016-79420-R AEI/FEDER UE). Sergio Trilles has been funded by the postdoctoral programme PINV2018-Universitat Jaume I (POSDOC-B/2018/12) and research stays programme PINV2019-Universitat Jaume I (E-2019-31)

REFERENCES

Anderson, B. J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E. (1976). A Land Use and Land Cover Classification System for Use with Remote Sensor Data. Technical report, U.S. Geological Survey, Washington.

Antrop, M. (2006). Sustainable landscapes: contradiction, fiction or utopia? *Landscape and Urban Planning*, 75(June 2005):187–197.

Arnold, S., Kosztra, B., Banko, G., Smith, G., Hazeu, G., and Bock, M. (2013). The EAGLE concept – A vision of a future European Land Monitoring Framework. In *EARSeL Symposium proceedings 2013, Towards Horizon 2020*, pages 551–568.

Badia, A., Pèlachs, A., Vera, A., Tulla, A. F., and Soriano, J. M. (2014). Land use and land cover change and the effects on vulnerability to forest fire of counties in the mountains of Catalonia : from managing the land to managing a threat. *Pirineos. Revista de Ecología de Montaña*, 169:1–13.

Ben Ramadan, A. A., Jackson-Thompson, J., and Boren, S. A. (2017). Geographic Information Systems: Usability, Perception, and Preferences of Public Health Professionals. *Online Journal of Public Health Informatics*, 9(2).

Büttner, G., Feranec, J., and Jaffrain, G. (2002). Corine land cover update 2000. Technical report, European Environment Agency, Copenhagen.

Del Bosque González, I., Arozarena Villar, A., Villa Alcázar, G., and Valcárcel Sanz, N. (2005). Creación de un Sistema de Información Geográfico de Ocupación del Suelo en España. PROYECTO SIOSE. In *XI Congreso Nacional de Teledetección*, pages 255–262, Puerto de la Cruz, Tenerife. Universidad de La Laguna.

Fernández Noguerol, S. (2017). Desarrollo de herramientas para el tratamiento de la información y el análisis con SIG de los usos del suelo utilizando el SIOSE. Una aproximación al caso de Asturias. *GeoFocus Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, 20(20):233–251.

Fernández Villarino, X., Delgado Hernández, J., Valcárcel Sanz, N., Caballero, M. E., Benito Saz, M. A., and Porcuna Fernández-Monasterio, A. (2012). Geoservicios web SIOSE: un sistema de información como servicio público. In *VI Jornadas de SIG Libre*, page 8, Girona. Servei de Sistemes d'Informació Geográfica i Teledetecció (SIGTE). Universitat de Girona.

Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D., and Knapp, A. (2003). The Importance of Land-Use Legacies to Ecology and Conservation. *BioScience*, 53(1).

Gamma, E., Helm, R., Johnson, R., and Vlissides, J. M. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional, Boston, 1 edition.

García-Álvarez, D. (2018). The Influence of Scale in LULC Modeling. A Comparison Between Two Different LULC Maps (SIOSE and CORINE). In Camacho Olmedo, M., Paegelow, M., Mas, J., and Escobar, F., editors, *Geomatic Approaches for Modeling Land Change Scenarios. Lecture Notes in Geoinformation and Cartography*, chapter 10, pages 187–213. Springer, Cham.

Ireland, C., Bowers, D., Newton, M., and Waugh, K. (2009). A classification of object-relational impedance mismatch. *Proceedings - 2009 1st International Conference on Advances in Databases, Knowledge, and Data Applications, DBKDA 2009*, pages 36–43.

Morte, A. R., Carrión, J. T. N., and Botella, E. G. (2019). Objective assessment of land use in hydrographical studies. *WIT Transactions on Ecology and the Environment*, 234(2019):41–51.

Navarro-Carrión, J. T., Zaragozí, B., Ramón-Morte, A., and Valcárcel-Sanz, N. (2016). Should EU land use and land cover data be managed with a NOSQL document store? *International Journal of Design & Nature and Ecodynamics*, 11(3):438–446.

Olaya-Abril, A., Parras-Alcántara, L., Lozano-García, B., and Obregón-Romero, R. (2017). Soil organic carbon distribution in Mediterranean areas under a climate change scenario via multiple linear regression analysis. *Science of the Total Environment*, 592:134–143.

Omrani, H., Abdallah, F., Charif, O., and Longford, N. T. (2015). Multi-label class assignment in land-use modelling. *International Journal of Geographical Information Science*, 29(April):1023–1041.

Peña-Angulo, D., Khorchani, M., Errea, P., Lasanta, T., Martínez-Arnáiz, M., and Nadal-Romero, E. (2019). Factors explaining the diversity of land cover in abandoned fields in a Mediterranean mountain area. *Catena*, 181(May):104064.

Polasky, S., Nelson, E., Pennington, D., and Johnson, K. A. (2011). The Impact of Land-Use Change on Ecosystem Services , Biodiversity and Returns to Landowners : A Case Study in the State of Minnesota. *Environmental and Resource Economics*, 48:219–242.

Ropero, R. F., Rumí, R., and Aguilera, P. A. (2019). Bayesian networks for evaluating climate change influence in olive crops in Andalusia, Spain. *Natural Resource Modeling*, 32(1):1–18.

Valcarcel, N. and Castaño Fernández, S., editors (2013). *Cartografía de Ocupación del Suelo en España. Proyecto SIOSE*. Number May 2014. Centro Nacional de Información Geográfica (CNIG), Madrid.

Valcarcel, N., Villa, G., Arozarena, A., Garcia-Asensio, L., Caballlero, M., Porcuna, A., Domenech, E., and Peces, J. (2008). SIOSE , a successful test bench towards harmonization and integration of land cover/use information as environmental reference data. In *Remote Sensing and Spatial Information Sciences*, volume XXXVII, pages 1159–1164, Beijing.

Villa, G., Valcarcel, N., Caballero, M. E., Porcuna, A., Domenech, E., and Peces, J. J. (2008). Land cover classifications: an obsolete paradigm. In *ISPRS Archives – Volume XXXVII Part B4*, pages 619–614, Beijing.