PILOT PROJECT ON A LAND USE/LAND COVER DATABASE IN THE AREZZO PROVINCE (TUSCANY, ITALY)

Pierpaolo NAPOLITANO[*], Giancarlo CARBONETTI[*], Enrico ZINI[**],
Michele BOCCI[**], Marcello MIOZZO[***], Paolo PETRI[****]
[*]ISTAT (National Institute of Statistics), Italy
napolita@istat.it
[**]RSDE (Remote Sensing Data Engineering) srl, Italy
ezini@rsde.com
[***]DREAM, Italy
miozzo@dream-italia.it
[****]GEOS srl, Italy
geos@dada.it

KEY WORDS: Land use/Land cover, Quality control, Classification, Spatial data.

ABSTRACT

ISTAT, the National Institute of Statistics in Italy, promoted in 1998 a pilot project for the realisation of a land use / land cover database at 1:25,000 scale over a test area in the Province of Arezzo, 2000 km^2 wide. The aim of the project was to assess the feasibility of a methodology for the development of a land use/land cover database which could become a component of the ISTAT geographical information system. The legend, derived from Corine Land Cover, was composed by 5 levels. The methodology of interpretation is based on the integrated use of multitemporal TM imagery, ortho-photos and auxiliary data. The sample design of the Forest Area Frame is based on a stratification derived from the 3rd level of the Corine legend and Digital Terrain Model. Ground truth from 140 sites, positioned by GPS, was collected. The project was carried out by RSDE, DREAM and GEOS.

1 INTRODUCTION

A considerable lack of cartographic and statistical data on land use and land cover phenomena is observed in Italy with negative effects on environmental monitoring activities and, more generally, on land management. The principal aims of the ISTAT information system on land use and land cover should be: 1) to overcome the lack of knowledge on land status and dynamics, with particular reference to the fields of urban settlements, industrial activities and services, agricultural and forest areas; 2) to realise an information layer to be integrated with the statistical data acquired by ISTAT in order to further improve the spatial analyses; 3) to achieve the co-ordination with other National Institutions with territorial jurisdiction (in particular, the Ministry of the Environment and the Ministry of Agriculture) and with the Regions - institutions directly engaged in the land planning activities.

For these reasons, in March 1998, ISTAT decided to summon a private auction for the realisation of an experimental database on land use and land cover at a 1:25,000 scale and a sample area-survey on woodlands and forest surfaces. The competition was awarded to a group of private firms composed by RSDE, DREAM and GEOS. The pilot project was completed in June 1999.

2 METHODOLOGY

2.1 Specifications

The test has been carried out on a 2000 km^2 wide area, approximately, that is nearly two thirds of the Arezzo Province including the Casentino area, the high valley of the Arno river and part of the Val di Chiana. The technical specifications have been defined by an ISTAT Commission (ISTAT, 1998), which has taken the geometrical quality of the cartography at 1:25,000 scale as reference data. However, considering the difficulties arisen in defining the limits of the above-mentioned categories – particularly in relation to natural and semi-natural environments – the Commission has set out a 25-meter tolerance limit to define the geometrical accuracy of the linear elements to be acquired. The minimum mapping unit has been set to 1 hectare for urban areas, and to 1.56 hectares for all other surfaces. The ED50 system (International Ellipsoid 1909) has been selected as datum together with the UTM projection, which are in line with ISTAT territorial databases of the CENSUS project (Napolitano, 1997).
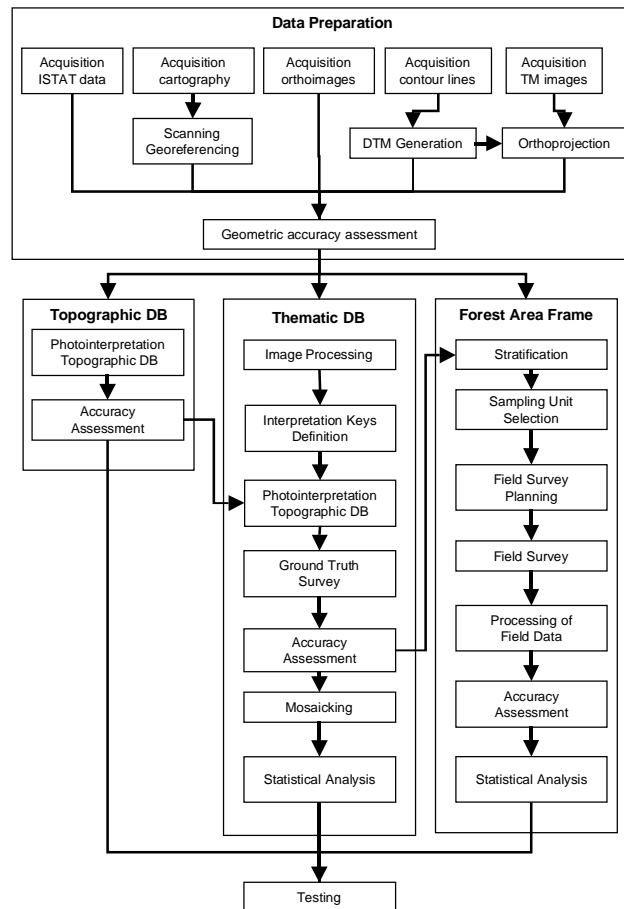
Figure 1. – Process flow

The database is composed by two parts: 1) A topographic component, essentially composed by linear elements; 2) A thematic layer on land use and land cover utilising, where possible, the above linear topographic elements as limits of the areas under classification.

The topographic component includes linear elements relating to highways, national roads, railways and the hydrographic network present on the JOG maps at a 1:250,000 scale, acquired anyhow at the 1:25,000 reference scale form digital orto-photos. Some 360 toponyms – set out on JOG maps - have been added to complete the topographic element. As far as concerns the land use / land cover database, the classification system refers to the Corine Land Cover (European Commission, 1993) project and further integration at the 4th and 5th levels; to the modifications introduced by the ISTAT Commission to match the thematic information and the geometric accuracy of the 1:25,000 scale and to consider the arising phenomena. It is important to remind that the Corine Land Cover legend is divided into three hierarchical levels with 44 classes at the third level, where the classes belonging to the first level are the following: artificial surfaces, agricultural surfaces, forest areas and semi-natural environments, wet areas, water bodies. Any change and integration described below has been added to the procedures carried out in the execution of this project.

A new class aimed at defining the residential widely-spaced structure – with a degree of artificial cover between 10% and 50% - has been added to the artificial surfaces. A distinction between areas with average building height higher and lower than 10 meters, has been made in relation to the continuous and discontinuous residential areas, which are already part of the European project and which are characterized by a >80% degree of cover – a range between 80% and 50%, respectively. As far as concerns the artificial areas for non-residential use the aim was to make a distinction between industrial, commercial and service areas. A transport interchange area (interport) class has been successively introduced. The following two categories have been defined under execution:

1) Dumps, incinerators and waste water plants, including features to be arranged on the basis of available ancillary data;
2) Non-utilised or abandoned areas - inserted in the urban tissue – or areas falling into a transitory stage. This second class has been introduced for social-environmental reasons and for its relevance in specific suburban areas.

As far as concerns agricultural areas, the possibility to discriminate autumn-winter and spring-summer crop areas, respectively, has been tested. The category including nurseries, greenhouses and gardens could be easily distinguished on the ortho-photos and has been consequently introduced in the work. As far as concerns forest areas, the definition of woodland set out by FRA 2000 – based on a 10% cover degree – has been adopted. However, the distinction between different woodlands set out according to the following cover degrees – 10%-20%, 20%-50% and >50% - related to other wood definitions, has also been introduced. A fourth level related to pure woodlands has been introduced to verify the possibility of distinguishing different forest typologies.

2.2   Data used in the activity

At the basis of the photo-interpretation activity several ortho-photos and the relating stereo couples (pairs) were acquired from the national public agency AIMA in 1996 in the frame of an agricultural monitoring activity (AIMA, 1997). Digital ortho-photos referring to the Roma 1940 datum have been referred to the ED50 datum by simply translating their co-ordinates, the parameters of which are supplied by IGM (Istituto Geografico Militare). The geometric quality of each ortho-photo has been checked as well as the continuity of the linear features along adjacent ortho-photos. A Digital Terrain Model (DTM) with pixel 10m has been computed from the contour lines of the new regional technical map of the Arezzo Province supplied by the Tuscany region at a 1:10,000 scale. The DTM accuracy has been checked against the IGM trigonometric points available for that area and with a number of points randomly selected on the IGM maps at 1:25,000 scale. Spectral information consists in three Landsat TM images dated: May 6th 1998, October 26th 1997 and May 19th 1997. 12 ground control points evenly distributed on the reference area – the co-ordinates of which have been drawn from the ortho-photos – have been used to orthoproject the satellite images. An unsupervised classification of the NDVI has been carried out from the TM images taken on May 19th 1997 and May 6th 1998 to analyse the agricultural production areas and urban open spaces. Data from the Forest Inventory of the Tuscany Region and from the forest management plans for regional properties have been used to classify the forest zones. The collected data have provided some information which have proved to be absolutely essential to distinguish the species prevailing in the woodlands. ISTAT territorial bases – related to commune-boundaries, inhabited zones and census sections – as well as associated statistical data have been used as ancillary data. The urban plan has been exclusively used for the Municipality of Arezzo, with particular reference to the map of the planned land use.

2.3   Photointerpretation Activity – Main Criteria and Modalities

The video photo-interpretation/digitalisation process has been directly carried out at a scale included between 1:5,000 and 1:10,000. The ortho-photos have been extremely useful tools for the polygon determination in all the envisaged categories. Their adoption has resulted particularly effective for the identification of both artificial and agricultural areas and to mark the limits between woodlands and non woodlands. In order to determine the classes belonging to urban areas according to the average height of the buildings, it has been necessary to use the stereoscopic couples. The information contained in the map of the planned land use has been considered as the most suitable tool to make a distinction between industrial, commercial and service areas. As far as concerns the evaluation of the agricultural areas, the following three main typologies have been identified on the basis of the NDVI classification calculated on May 19th 1997 and May 6th 1998:

1) highly plant-covered areas on both dates – considered as the areas characterised by repeated spring-summer crop cultivation;

2) low plant-covered areas on both dates – hypothetically considered as the areas characterised by repeated autumn-winter crop cultivation;

3) areas with different values – areas in which a variation in the crop typology might have occurred. The latter presents a percentage weight equal to 40%, approximately, showing a significant variation among the farming practices.

As far as concerns natural environments and the internal classification of wood-covered areas, the Forest Inventory of the Tuscany Region has been constantly used as reference data. The satellite images and the stereoscopic couples are essential instruments capable of identifying those areas where oaks, chestnuts, beeches and silver firs are mainly planted or to make a simple distinction between conifers and broad-leaved forests. However, several problems have arisen in relation to the identification of similar species such as the hop-hornbeam, the hydric plants like poplars and willows and all the conifers except for the silver fir.

3 RESULTS

3.1 Topographic Component – Quality Control

A quality control on each of the database components based on a sample repetition, assisted by evaluators, of the photo-interpretation procedures was set out in the technical specifications. The topographic component has been subjected to several controls in relation to geometry, to the coding of linear structures, to toponymy and to the completeness of the acquisition. A statistic test on the geometrical and thematic accuracy of each element has been carried out, to achieve a quality evaluation, by using the number of incorrect units found in the selected sample. The sample size of the elements to be controlled was equal to 120 and the acceptance threshold was equal to 10. A statistic test on the completeness of the acquisition for each type of element has also been carried out. The reference area has been subdivided into 1 km-wide square cells. A calculation of the number of arcs, nodes and the test application points, has been carried out for each of the square cells. A statistical test has been therefore carried out on a cell sample to compare the real number of elements to be acquired. The completeness evaluation was based on the ratio between the elements that have been acquired and those that were to be acquired, and which we define as completeness ratio. The sample size of the elements to be controlled was equal to 78.

3.2 Thematic Component – Quality Control

Two different procedures have been adopted to monitor the thematic component: the first aimed at directly controlling the polygon photo-interpretation, the second at analysing the overall thematic quality of database by reinterpreting a sample of points. The polygon control has been carried out separately on three different parts of the area, which actually correspond to the processing lots. The control procedure essentially aimed at verifying the interpretation process of the firms. Also in this case, a statistical test of accuracy has been applied. It was based on the same parameters of the homologous test for the topographic component: sample size of 120 polygons, refusal threshold equal to 10. The sample polygons have been re-interpreted with the aim of verifying the limits envisaged by the firms. An accordance of values between the class codes and the inclusion of the polygon referred to in the test in a buffer of 25 m around the re-interpreted polygon has been recorded. The points of a 100 m wide square regular grid have been initially stratified according to the third level of the legend. Then, about 1800 points in total have been randomly selected from these strata, in proportion to strata size. The quality control gave positive results if, for the point under control, a polygon with size at least equal to the mapping unit was capable of matching the code of the class of the database polygon to which belonged the point. The work acceptance was based on the same criteria mentioned above, that is, in a class of 120 analysed elements no more than 10 incorrect elements were allowed.

3.3 Testing of a Sample Area Survey for Woodland Surfaces

Another activity carried out in the frame of the pilot project was related to the testing of a sample survey of spatial type aimed at evaluating several parameters of forest nature on an area covering three communes in the Arezzo Province: Bibbiena, Talla and Castel Focognano – a surface equal to some 200 square chilometers in total. The sample design was based on the following stages:

1) stratification in four strata of the area referred to in the sample survey, in order to classify the territory according to third-level land use and cover classes and altimetry;
2) creation of 1000m-wide square segments, according to a regular grid;
3) overlapping of the layer map and the segment map;
4) individuation of wood-covered segments;
5) determination of segment strata, according to class prevalence criterion on the segment;
6) the number of selected segments were proportional for each strata to strata extension;
7) on-site individuation of the sample areas defined on the basis of a satellite system (GPS) and land surveys. The individuated sample areas were four circular zones, with 8m radius and distributed in each of the sample segments according to a systematic pattern;
8) synthesis and analysis of the results.

The main parameters under evaluation were referred to surface variables and to the number of plants. In particular, have been carried out estimates concerning: a) the surface of land use and land cover classes; b) the incidence surface of the leaves and the branches and the number of plants related to the main arboreal species of the area; c) number of plants according to phyto-sanitary conditions.

1 – Artificial Surfaces		%	3 – Forests and Semi-natural Environments		%
1111	Continuous, high-density residential fabric	0,16	3111	Woods mainly planted with exotic broad-leaved forests	0,31
1112	Continuous, average-density residential fabric	0,37	3112	Woods mainly planted with holm oaks and/or coark-oaks	0,67
1121	Discontinuous residential fabric	1,02	3113	Woods mainly planted with Turkey oak e/o pubescent oak and/or English oaks	32,23
1122	Low-density urban residential fabric	0,68	3114	Woods mainly planted with durmast, English oak and/or hornbeam	0,03
1211	Industrial areas	0,70	3115	Woods mainly planted with hop-hornbeam and manna ash	0,34
1212	Commercial areas	0,04	3116	Woods mainly planted with chestnuts	10,07
1213	Public and private service areas	0,08	3117	Woods mainly planted with beeches	4,71
1214	Waste disposal plants, incinerators and water depurators	0,00	3118	Woods mainly planted with willows and/or poplars and/or alders	0,78
1221	Road and railways network, technical infrastructures	0,20	3121	Woods mainly planted with conifers	0,70
1222	Interports	0,00	3122	Woods mainly planted with Mediterranean pine-trees and/or cypresses	0,60
123.	Ports	-	3123	Woods mainly planted with pine-trees in the sub-alpine and alpine areas	2,36
124.	Airports	0,02	3124	Woods mainly planted with silver firs	0,83
131.	Extractive areas	0,58	3125	Woods mainly planted with red firs	0,03
132.	Waste-disposal areas	0,04	3126	Woods mainly planted with larches and/or Swiss pine	0,00
133.	Building sites	0,12	3131	Mixed woods planted with conifers and broad-leaved forests; 10-20% cover	0,02
134.	Abandoned lands	0,05	3132	Mixed woods planted with conifers and broad-leaved forests; 21-50% cover	0,09
141.	Urban open spaces	0,08	3133	Mixed woods planted with conifers and broad-leaved forests; >50% cover	2,36
142.	Recreational and sports areas	0,12	3211	Grazing lands and grasslands	0,16
		4,26	3212	Grazing lands and shrub grasslands	0.68
			3213	Grazing lands and scrub grasslands	0,51
2 –Agricultural Surfaces		%	3221	Moors and bushes mainly planted with prunes and/or nuts	3,76
2111	Non-irrigated autumn-winter crops	11,35	3222	Moors and bushes mainly planted with alders and/or willows	0,12
2112	Non-irrigated spring-summer crops	6,71	3223	Moors and bushes mainly planted with junipers	0,04
2113	Nurseries, greenhouses and gardens	0,15	3224	Moors and bushes mainly planted with mugho pines	-
2121	Irrigated autumn-winter crops	-	323.	Sclerophyllous areas	0,03
2122	Irrigated spring-summer crops	0,66	324.	Developing wood and bush covered areas	1,35
221.	Vineyards	1,92	331.	Beaches, dunes e sands	0,01
2221	Orchards and minor fruits, Citrus plantations excluded	0,29	332.	Rocks, cliffs, crags and outcrops	0,04
2222	Citrus plantations	-	333.	Low-density vegetation areas	0,03
223.	Olive-grove	4,67	3341	Areas covered with woodlands subjected to fires	-
224.	Other permanent crops	0,20	3342	Land, shrub and/or herbaceous areas subjected to fires	-
231.	Stable grasslands (permanent forage plants)	1,09	335.	Glaciers and perpetual snows	-
241.	Temporary crops associated to permanent crops	0,30			**62.86**
242.	Cultivation systems and by-plot complexes	3,47			
243.	Areas mainly covered with agricultural crops and natural areas	1,70	**5 – Water Bodies**		%
244.	Agricultural & forest areas	0.10	511.	Waterways and channels	0,12
		32.61	5121	Extensively exploited basins or non-extensively exploited basis for aquaculture productions	0,15
			5122	Intensively exploited basins for aquaculture production	0,00
			521.	Lagoons	-
			522.	Estuaries	-
			523.	Sea and oceans	-
					0.27

Table 1. – Land use and land cover up to the 3rd - 4th levels for the Pilot Project area (percentage data)

Notes: "-" indicates "non-present phenomena"; the Wet Zones (4th class) are not referred.

For each of the variables, the parameters referred to in the survey were represented by the total values – according to which the estimates of the segment averages have been calculated – the variance and standard error of the total values. The results of the sample survey have been compared to the data resulting from the land use and land cover database. A cost analysis has also been provided by the firms with regards to the activities carried out on the field, which will prove to be very useful in the feasibility studies aimed at enlarging the survey to wider areas.

Surface estimation by the Area Frame Survey		Hectares	Percentage error (*)
31	Forest areas	11,390.66	5%
3113	Woods mainly planted with Turkey oak e/o pubescent oak and/or English oaks	3,627.80	12%
3115	Woods mainly planted with hop-hornbeam and/or manna ash	405.88	37%
3116	Woods mainly planted with chestnuts	3,237.65	16%
3117	Woods mainly planted with beeches	2,272.31	25%
3121	Woods mainly planted with conifers	211.81	56%
3123	Woods mainly planted with pine-trees in the sub-alpine and alpine areas	50.19	39%
3124	Woods mainly planted with silver firs	32.12	55%
3133	Mixed woods planted with conifers and broad-leaved forests; >50% cover	1,552.91	18%

Table 2. – Surface Statistics from Area Frame Survey

(*) It is equal to the ratio between the estimate of the surface standard error and the estimate of the surface times 100 (Cochran, 1977).

4 CONCLUSIONS

The pilot project has provided significant data, which will need to be evaluated in detail in order to define the characteristics of the land use and land cover database at national level. According to the results currently achieved, several evaluation elements can already be supplied. If we decided to consider as priority aspect an objective thematic content and high geometrical accuracy, we could assume a minimum map unit, equal to at least 1 hectare, that is common to all classes.

The central element of the database should be, in any cases, a shareable legend provided with logical coherence and simplicity, to be considered as a lasting instrument. The classification principles should be clearly defined, in particular by distinguishing the elements that are specifically associated to land use and cover.

According to the knowledge acquired during the orto-photo interpretation, it is convenient to consider the possibility of inserting other information of high environmental and landscape value, such as rows, natural corridors etc. where no satisfactory geo referenced data are provided and which, according to a preliminary exam, seems to be easily and objectively detectable on the orto-photos.

Finally, the database should be conceived as a full geographic information system with efficient managing tools, supplemented by two fundamental activities:

1) product statistic validation, seen as a separate stage from the database monitoring and testing activities, to be possibly achieved by means of on-site comparisons and to be possibly integrated by other independent surveys, aimed at setting out evaluations of quantitative nature;

2) monitoring strategy on land use and cover to be fixed together whit the planning of the database and the legend.

REFERENCES

AIMA, 1997. L'Italia vista dall'AIMA - Compagnia Generale Riprese aeree , Parma.

Cochran W., G. 1977 Sampling Techniques. Third Edition J. Wiley, New York.

European Commission, 1993. Corine Land Cover-Technical guide, Luxembourg.

Gallego F.J, 1995. Sampling Frames of square segments - JRC, EC, Luxembourg.

Healey M.J., 1991. Economic activity and land use, Longman, Avon.

ISTAT, 1998. Capitolato tecnico per il progetto pilota del database sull'uso e la copertura del suolo in scala 1:25.000 su un'area test. A cura della Commissione ISTAT, Roma.

Napolitano, P. 1997. The territorial database of ISTAT – its application to spatial analysis and possible developments. Proceedings of the seminar on regional statistics. Baden 1996 volume 2. Luxembourg.