A Prototype System for monitoring information demand and data availability

G. Borga, S. Castelli, S. Dalla Costa, L. Di Prinzio, S. Picchio, A. Sau

University IUAV of Venice- Faculty and Department of Urban & Regional Planning, Venice, Italy

ABSTRACT: The complexity of the issues connected with the protection and management of the environment involves a need for increasing amounts of knowledge and data in support of planning processes. A research project launched by the University IUAV of Venice, Italy – Department of Planning aims to contribute to a better understanding of the relationship between data supply available from new technology sources and public administration archives and the demand for information from urban policy-makers and managers. The research project also sets out to provide an overview of the various institutions that work in the field of environmental and urban information management with a view to proposing the creation of a new institutional arrangement whereby a National Agency and series of Regional Agencies would be responsible for setting up an integrated system for the collection, processing and dissemination of environmental and territorial information. To this end, a prototype of such a system has been created and tested in an area study.

1 THE DESIGN OF THE RESEARCH PROJECT

The complexity of the issues connected with protection and management of the environment entails a growing demand for knowledge and data concerning changes that have occurred in the past, current states and trends.
New technologies offer extraordinary opportunities for a joint public and private sector effort to improve our understanding of territorial and environmental states and problems.
However, at least as regards Italy, environmental and territorial information systems suffer both from cultural underdevelopment and a lack of public initiative. Political and cultural effort is needed to promote action designed to provide better understanding of the environment, with a view to ensuring its protection and enhancement. A research project launched by the University IUAV of Venice, Italy – Department of Planning has two primary aims:
− to promote reconsideration of the connection between supply and demand with regard to information to support the processes of territorial and environmental management;
− to promote reconsideration of the present organization of the national mapping agencies.
The research project is divided into four sub-projects. The objective of the first is to explore the system of “ongoing” land and environmental management archives (those which are stable and regularly up-dated) commonly kept by public administrations for specific institutional/managerial activities. The second sub-project explores the contribution currently being made by the system of sensors installed on various platforms (satellites, aerial, terrestrial and marine vehicles and networks) and dedicated to monitoring the environment. The aim of this sub-project is therefore to classify all the data collected by these sensors and to make them available for a variety of different users. Together, the first two sub-projects represent the complete structure for the supply of environmental and territorial data. The goal of the third section is to analyse the demand for information on the part of the regulatory-political-institutional system on the one hand and of policy makers and urban managers on the other. This sub-project also sets out to understand the relations between demand segments and supply opportunities. The objective of the fourth sub-project is to investigate and present an overview of the institutions that deal with territorial information (cartographic institutions at state, regional, ministerial levels, etc.) in order to verify the relationship between their respective
institutional duties and the market. The aim of the project as a whole is therefore to undertake a critical description of the state of the art in this sector, compare it with the real situation in other national contexts, and make a constructive contribution to the debate over the reform of the sector. To this end, the work also proposes the establishment of a service to be provided by a national Agency and by a number of regional Agencies, their purpose being to create an integrated system of collection, elaboration and dissemination of environmental and territorial information.

The four research projects effectively perform some of the functions envisaged for the regional Agencies thanks to the development of a prototype system/application that simulates:

− the monitoring, recording and classification of the demand for information expressed by various policy makers and urban managers;
− the monitoring, recording classification of the sources of information constituting the entire range of data supply coming from new technologies and institutional archives;
− a comparison of supply and demand, highlighting the information deficit.

2 “INFORMATION DEPOSITS”: INSTITUTIONAL PRODUCTION OF ENVIRONMENTAL DATA. ENHANCEMENT AND INTEGRATION AS A MEANS TO THE CONSTRUCTION OF KNOWLEDGE FRAMEWORKS

Public administrations produce, collect and elaborate considerable amounts of data, incorporating them into more or less digitally formatted databases/archives for specific institutional/managerial activities. These archives often have territorial and environmental contents but despite their huge potential they remain an unknown and unexploited resource. Many national and international initiatives have been devised for the purpose of adding value to public sector information. These include the European INSPIRE directive designed to lead to the creation of an Infrastructure for Spatial Information in the European Community and numerous directives in the field of e-government and re-use of public documents in the public sector, some of which have been incorporated into Italian regulations. However, in this national and international scenario, the situation as regards PA archives and spatial and other information in Italy remains critical in terms of its availability, quality, organisation, accessibility and sharing. These problems are noticeable both at the level of large-scale institutions such as the National Mapping Agencies and at the local authority level, where access for the general public tends to be difficult.

2.1 Goals and strategies of development

The aim of the research is to classify digital information resources useful for land and environmental management, especially those regularly updated by public authorities, and to contribute to the debate over how to make the best use of them and how best to integrate them with information derived from new technologies. Considering the number of data banks kept by the wide range of local and central public and private organizations working in the environmental sector with various objectives and responsibilities, the task of detailed screening of all information, archives and datasets is an extremely complex operation that requires a considerable effort. A series of strategic decisions have therefore been taken in order to facilitate the pursuit of general aims and to optimize the survey of archives and databases. During the first phase, national and international best practices and codes, webGIS and web-archives were analyzed. An initial screening of national authorities and their databases led to the creation and testing of a specific database (“DBGiacimenti”) for the classification of the contents of public digital archives. In drawing up a list of the features needed to describe heterogeneous databases reference was made to soundly based international experience in the matter and attention was paid to ensuring alignment and integration of this with all the other databases created in the framework of this research project, with a view to constructing an integrated supply-demand information system for online data-sharing. The survey of public archives and databases was carried out both on-line, as regards the great central institutions, but mainly through direct interviews and site visits to some local administrations, chosen in the case-study area within the Veneto Region. This aim of this
strategy is to achieve a detailed characterization of all the existing databases of the various authorities in the same area, and to highlight all criticalities and inconsistencies.

2.1.1 The “DBGiacimenti” database for the classification of public digital archives: main features
Due to content and format differences, the definition of a set of metadata able to characterize the databases was found at an early stage to be a complex activity. We therefore decided to use keywords obtained from the widely used GEMET Thesaurus (General European Multilingual Environmental Thesaurus) to describe information contents and to refer to the metadata guidelines (the ISO Standard 19115:2003) for the design of the databases. The “DBGiacimenti” database also describes all private and public national/local authorities working in the field of environment and territorial management, classified by their juridical status and connected to the produced or distributed information sets.

2.1.2 Survey of public sector digital archives
As mentioned earlier, the survey of the public sector digital archives is being carried out in two phases:
− an on-line survey of national archives (National geographical map Institutes, National Agency For Environmental Protection, National Geological Service, National Marigraphic Service, ...);
− “in the field” investigations of local authorities (municipal and provincial administrations, Cooperatives, Regional Agency for Environment Protection, etc..) and some private organizations with environmental and urban management responsibilities in the case study area within the Veneto Region. Investigations did not include strictly financial and administrative sectors but did include all departments connected with territory/environment, e.g. registry offices, productive activities, pollution, planning, land use, etc.
The survey was sometimes carried out by interviewing the heads of the various administrative departments, and in other cases it was assisted by the availability of archives compulsorily catalogued in compliance with the law on privacy or by centralized information systems.

2.2 Results
Investigations and public archive analyses are still in progress, but the emerging overview already provides stimulating material in many areas:
− it allows a wide range of analysis of different types of data, treatment, accessibility, within different local authorities;
− it highlights relationships, information fluxes and dynamics linking different authorities at different levels and hierarchies working in the same area;
− it provides an effective testbed for the “DBGiacimenti” database in different cases, with different types of data, and its suitability for other research purposes.
Some weak points in the research strategy are also emerging: the classification of practically all the public digital archives with a territorial connection in the same area requires considerable effort but statistical analysis can still not produce a comparison of these data with those of public administrations with similar responsibilities but working in different territorial areas.

3 ACQUISITION, INTEGRATION AND DISTRIBUTION OF DATA FROM AERIAL, SATELLITE AND TERRESTRIAL SENSORS
In recent years there has been a considerable development in new technologies for the management of territorial and environmental information. There are about active 100 sensors lodged on satellites and intended for observation of the Earth and for climatic and meteorological analysis. Aircraft (civil aviation planes, helicopters, ultralight aircraft) have long been used for territorial monitoring and map-making, but, recently, these traditional technologies have been supplemented by other instruments, such as Laser scanners, for the acquisition of DTM/DSM data, or hyper-spectral sensors, which allow very detailed
classification of land use and cover. A series of sensors intended for bathymetric and geomorphic surveys (e.g. the multibeam system) are installed on ships or other craft. Then there are many instruments installed on vehicles and used for assessing and "automatic" survey of roads, signs and signals, territorial elements (integrated gps systems, inertial and video-cameras). Then, finally, there are sensor systems forming networks for monitoring the environment (air and water quality), agriculture and meteorological situations. These enable multi-temporal analysis, as well as forecast models for different kinds of applications.

3.1 Goals and strategies of development

The aim of this survey is to underline the contribution that can be made by technological resources that are currently available on the market and from which data and information on territory and environment can be obtained, comprehensively classified and made available to a variety of different user profiles, according to the different application areas.

The specific objectives of the research are:

− to undertake a comprehensive survey of sensors and their platforms, classifying them in a data base oriented towards a variety of user profiles and fields of application;
− to analyse the possibility of integration between different sensors and by-products, and to evaluate the prospects of improving them with data from the public sector archives, with the aim of building up a single supply system;
− to help define a system of correlating the supply and demand of data intended to support the instruments of territorial and environmental planning.

The survey started from an analysis of the activities and projects of public bodies and research institutes, with a view to obtaining an initial overview of sensors intended for territorial and environmental study and of the means by which data and products are disseminated and distributed. The huge amount of literature available and the contribution Earth observation satellites have been making for many years provided the inevitable starting point of this work. The analysis of data banks and web sites offering access to satellite pictures (e.g. the ITC's database of Satellites and Sensors - International Institute for Geo-Information Science and Earth Observation and the CEOS Data Base - Committee on Earth Observation Satellites), which are still very bound to a scientific and sectorial approach to the subject, made it possible to orientate the subsequent classification of the sensors and the construction of the data bases towards three main objectives:

− to ensure that all public bodies and institutions dealing, either locally or at central level, with territorial planning and management are acquainted with the characteristics and performance specifications of the sensors;
− to orientate classification towards territorial and environmental applications;
− to draw attention to the contribution of sensors installed on other platforms, taking into consideration, however, only those already used systematically and ignoring research projects or prototypes.

3.1.1 The "DB Sensori" database

The analysis phase was preparatory to the planning of the database: it led in fact to a map of the problems associated with classification of the sensors and to definition of the conceptual and logical structure of the "DB Sensori" database. The database, filled with the data and characteristics of the sensors, is the basic instrument, a hinge linking the data "market" and the end users, divided into different areas of application (land risks and defence; planning and land use; mobility; environment). The principal elements of the database are therefore the sensors and their platforms; while secondary elements comprise data, projects and applications. The "DB Sensori" database makes it possible to view and to examine data relative to the instruments (general information, technical details, information about cost per km2 and on data distributors), and to the different platforms on which they are installed (general information, technical details). It also makes it possible to examine the relationship between these data and the information "deposits" for which they may have been the source. A series of queries have therefore been built in in order to allow the user to choose a customised solution: optimum
space resolution, cost, consistency with the application area to be studied, the degree of updating, the local and international distributors and so on.

3.1.2 comparison and relationship with integrated research
Action to deal with the problems concerning the classification of sensors in relation to application areas was taken from the very beginning. Nevertheless, the methodological approach adopted for research on information sources (started in a later phase) inevitably raised questions over classification and the definition of applications. Classification became even more difficult when tackling the problem of how to establish a contact between all the results available in terms of information supply - sensors and information sources - and the system of information demand. Some of the components of the database relating to sensors were therefore redefined in order to make it homogeneous and compatible with the other archives. The same definitions were adopted for information relating to space (scale, resolution) and to time (date of creation, date and frequency of data updating). The system of connecting archives in terms of applications was more difficult to achieve. The description and classification of the applications and contents that can be obtained from sensor products (and also from sources) must comply with specific requirements:
- they must be semantically correct;
- they must not be self-referential;
- they must be complete and comprehensive
- they must above all be in common with the other data banks created in the framework of the research (that is, they must represent the contents of information sources and they must describe the information demand).

The use of a common language for classification of subjects and applications was intended to facilitate exchanges between data banks (via a system of key words). This language was achieved starting from semantically consistent words, already structured into categories, relations and importance hierarchies, derived from a specific and shared dictionary. The GEMET - General Multilingual Environmental Thesaurus - of the European Environment Agency appeared as the most suitable instrument for the fixed objectives. The inevitable modification of the "DB Sensori" database and discussions about the use of GEMET gave rise to an experimental phase designed to put the integration of databases to the test and to calibrate the system of data demand and supply.

3.2 Results
We have completed the database relative to sensors installed on satellites and aircraft; we have also completed a series of alignment operations between databanks comprising the information supply (“DBSensori” and DB Giacimenti).

The next step is an investigation of the instruments installed on vehicles and boats and of the system of stations (monitoring network). It will also be necessary to complete discussion of the best ways and means of integrating data derived from the systems of sensors and sources. After one year of research, it is increasingly evident that the issue of sensors is extremely complex and, above all, is constantly evolving. It is nevertheless a very stimulating subject, especially in relation to the possibilities of integration between different technologies and between data and products derived from technologies with information resources belonging to public bodies and other institutions.

4 THE NEED FOR INFORMATION AS CONTAINED WITHIN NORMS AND AS EXPRESSED BY POLICY-MAKERS/URBAN MANAGERS
The aim of this research is to classify the demand for information connected with territorial management and to relate it with the data supply offered by technological instruments (such as sensors) and by public agencies (information “deposits”). At the same time it is intended to experiment with techniques aiming at integration of sensor-sourced data with information “deposits” so as to obtain an “information surplus” which could be orientated towards specific segments of the demand.
4.1 Goals and development strategies

The first issue to be solved was an analysis of the precise meaning of the expression “information demand”. This was needed both in order to define a way to structure it inside a database, and in order to design the research and interviews necessary to record it. At the same time an “expeditious screening” was carried out in order to select a first data set relating to ongoing governance/management activities, existing prescriptions and norms and the subjects involved. This was indispensable in order to test the functional efficacy of the database. After the initial screening, it became possible to identify the principal elements of the information demand, to classify them, to record them inside the data base and to perform some initial queries by connecting the information supply data and by operating cross-controls. Along with the “refining” and fine-tuning of the interpretative and logical models, the survey of demand will have to be supplemented with more systematic inquiries and with interviews aimed at a significant panel of public and private operators. The analysis will have to concentrate on a geographically limited area so as to cover the largest possible part of application themes. Subsequently, and by exploiting such models, it will be possible to reach a definition of the thematic “packets” which, when the data derived from sensors is integrated with those sourced from “complementary” information bodies, will make it possible to extract new, important information; the packets so defined – or re-defined on the base of other criteria – can be related with the mapping of the information demand to generate a synoptic view which shows the extent to which the demand has been satisfied.

4.2 Results.

4.2.1 The database for classification of the information demand.

The main product of the analysis is a database called “DBDomanda”, the aim of which is to produce a systematic record of the information demand. As is well known, the use of DBMS systems is grounded on the definition of each element included, i.e. on a definition of the features characterizing each of the items which are part of the data base. The features of the information demand are in large part analogous to those of the supply and basically belong to three types:

- semantic features (concerning the contents);
- spatial features (for localization and for degrees of spatial definition);
- time features (for matters concerning the updating of data).

While for space and time features there are established criteria (see ISO – Metadata), for semantic classification it was decided to make use of the GEMET Thesaurus, drawn up by the European Environment Agency (EEA) and described below.

4.2.2 Populating the database

In order to test the efficiency of the classification models chosen, it was decided to create an initial information set using procedures of an expeditious (i.e. non-systematic) nature, by which it was possible to obtain specific information with limited use of resources. This first data set needs later to be integrated with extensive and systematic methods, which would secure a larger coverage for the analysis. For the screening phase, it was decided to opt for a research based on two archives of metadata developed in a web environment, which, as systematic repertories of geographic and non-geographic data banks, have made it possible to extrapolate a first list of plans and tools for territorial management, the legislative frameworks of reference, and the information demand thus generated. It is planned that the picture thus obtained should be supplemented first with a more systematic analysis aimed at completing the regulation framework of reference, and secondly with direct interviews with selected public and private operators, so as to record the demand that may not be considered in the regulations.

4.2.3 Application for consult the database

Finally, an integral part of this research is the design of an information system which can supply a permanently updated picture of the production – diffusion – utilization flows of territorial data. A tool of this kind is based on constant monitoring, which guarantees the feeding of the
integrated database. Furthermore, since it can be interrogated by territorial management operators, it is designed to supply the information necessary for each of them to make the most efficient use of the technological tools available in matters of geographic and environmental information.

5 INTEGRATED DATABASES FOR COLLATION OF INFORMATION SUPPLY AND DEMAND

The permanent monitoring of the activities carried out by operators in charge of territorial administration and protection of the environment, proposed as one of the functions of the Regional Agencies, therefore provides material to three different databases:

- the supply of data coming from technological equipment (“DBSensori”);
- the supply of data supplied by institutional activities (“DBGiacimenti”);
- the demand for territorial-environmental information to carry out the tasks of territorial administration (“DBDomanda”).

A fourth database contains the elements that enable the integration of the first three. Altogether they make up what we have called the “integrated database”, that is the database that constitutes the main ‘data provider’ for the application to consult the databases. Before giving a detailed description of the interface and functionality of the access application, we must expand on the techniques used to integrate the different databases and briefly introduce the use of the General European Multilingual Thesaurus, GEMET, which is the key element of the integration.

The use of the DBMS technology to monitor, record and relate the supply and demand of territorial information presupposes the design of a logical model to identify, among other things, a semantic mediator that will allow demand records to be matched against those of the supply. While it is evident that a normal textual description of information contents does not allow analysis of “matching” between demand and offer, it is not so easy to find another way to describe/present these contents. On reflection, the best solution appeared to be the use of a list of “keywords”, three definitions of which were considered:

- a dynamic list, auto-expanded through free input by the user;
- a static or dynamic list compiled or expanded by an administrator;
- a static list taken from other disciplinary contexts.

At this point, analysis of the GEMET proved interesting, in particular its rich internal system of hierarchies and semantic classifications. This led to the third of the above options being chosen and the GEMET was duly included in the database. GEMET’s main component is the list of descriptors in which elements are explained briefly and translated into various languages. The descriptors are classified by means of two different systems:

- the first type involves “assembling” groups and macro-groups (so the relation is one to many);
- the second type involves direct “assignment” to one or more topics in a list of forty terms (so the relation is many to many).

These two systems “coexist without interfering” with each other and therefore constitute a double system of classification. In addition to being classified, the descriptors are reciprocally related, again with a double system: hierarchical and associative. The difference between the two systems of relations (which in fact are physically similar), is purely informational and consists in the fact that a hierarchical relationship connects a broader-term with a narrower-term, whereas an associative relationship connects two related-terms. At this point, it is clear that the table of the 5208 GEMET descriptors constitutes an excellent list of keywords for the semantic characterization of the information contents also because of the classification and relationship tables that link these descriptors to each other. The supply of data from sensors is recorded in the “DBSensori” database, which is divided into two principal elements: “Sensore” and “Piattaforma”. The “sensor” element provides a material representation of the recorded datum, while “platform” relates to “sensor” and thus represents the plurality of “vehicles” on which sensors can be installed; finally, what can be obtained by the use of these data is inserted into a relationship table that registers the possible applications and the descriptors connected to them. The data supply from public sector archives, on the other hand, is registered in the
“DBGiacimenti” database, which contains the major elements “Dataset” and “Soggetti”: the former is a catalogue of all the information resources registered; the latter is a record of the public/private subjects connected for various reasons to these resources. This database also includes a specific relationship table containing the multiplicity of links between a database and all the descriptors that provide a brief explanation of its information contents. Finally, the ‘DBDomanda’ database is used to register the information demand and contains the principal elements “Attività”, “Provedimento”, “Soggetto”; respectively, these represent:
- the territorial administration actions carried out by subjects;
- the regulatory provisions that form the reference framework;
- the typologies of the subjects involved in the performance of the above “Attività”.

Another element within the “DBDomanda” database – “Domanda” – connects the other three and makes it possible to deal with the demand extrapolated from the regulatory framework of reference and with that expressed directly by public authorities working in the field of territorial governance and management. Consequently the results are associated to one or more descriptors, which characterize their information contents and enable them to be related with the supply. Technically, the relation between demand and supply is established when “joints” linking the GEMET descriptors are inserted into the database and by executing the queries that provide the synoptic framework of the relation between supply and demand. It is interesting to note that the GEMET system of hierarchical relations can be exploited to explore “indirect” as well as “direct” matches. For example, a “direct match” is one that connects a database concerning, say, landslide phenomena with a demand for information about areas that are subject to landslides, using the descriptor “landslide”. By contrast, if we consider a specific sensor used for meteorological investigations and forecasts, associated with the descriptor “meteorological phenomena”, and a demand for information about wind speeds associated with the descriptor “wind”, the two elements can still be linked, in this case through an “indirect match”, in that the descriptors “meteorological phenomena” and “wind” are in turn linked to each other by a hierarchical relationship inside GEMET. If we consider the complexity of the hierarchy of descriptors it is easy to imagine that exploration of the hierarchy will produce many more matches between supply and demand than would be possible through “direct matches”.

It is equally clear that the higher the number of connections matching supply and demand descriptors the less significant the match will be.

The integrated database therefore provides an instrument capable of comparing the information demand with the available data, of relating the monitoring activity and the high or poor quality of the information related with territorial/environmental data. For this reason, inside the database, each element of the demand requires an indicator which is based on the amount of resources connected to the descriptors matched and weighted according to the importance attributable to each individual match.

6 THE APPLICATION TO CONSULT THE INTEGRATED DATABASE

The application providing access to the integrated database has been designed with two necessities in mind:
- to enable the contents to be consulted by a network of operators and stakeholders;
- to supply an innovative technological instrument for the activities of monitoring and support to operators in the field provided by an institutional authority for territorial/environmental information.

To attain these objectives, the web is the natural choice as the platform for primary development of the application and for the conveyance of data. In effect, the first point constitutes the “front-end” of the database, while the second is the “back-end” with which the system administrator can record the permanent monitoring activity. The administrator will also use the “back-end“ functions to develop a parallel activity of devising “packages” that combine data from sensors with information layers to obtain “added information value” and orientates them to specific segments of the demand.

While the way the back-end of the system functions is substantially analogous that of a classic CMS, the public interface features a number of distinctive characteristics.
The navigation system introduces a unique initial access phase that follows two distinct routes: the first invites the user to choose one of the territorial government activities in the database and then gives access to the analytical phase of matching information supply and demand; the second allows the user to explore all data availability, regardless of the use that can be made of it; for both approaches the user may choose a geographical area of reference. After the initial choice of activity and geographical area, the user choosing the analytical demand/supply route enters a page containing an abundance of customizable elements; the upper part of the page shows the list of the demand items connected with the activity chosen, each with a short description, and a score that indicates how much the supply repository offers in response to each single item. The score indicator can be increased to allow the exploration to focus on the detail of a single semantic descriptor, visualizing the related information resources.
The lower part of the page summarizes the supply related to the demand and allows the detailed records of the information resources consulted and visualized in two/three dimensions; in some cases it is possible to download the data directly.

Amongst the more interesting tools assisting the user to explore information supply is a navigable map with selectable content with which the user can visualize one or more supply layers superimposed on a generic two-dimensional cartographic base and the preview of the data on Google Earth through which it is also possible to visualize data in three dimensions. In both the lists of demand and supply it is possible to highlight the matched items, i.e. to indicate which resources satisfy a given demand and which demands can a given resource contribute a response to.

Unlike the first approach, the second leads directly to analysis of the available resources for the chosen territorial area irrespective of the demand in question. The research tools include a first filter through which certain topics can be removed from the complete list inherited from the GEMET dictionary in order to focus the research on those which are of interest. The next page shows the same interface used for the first approach as regards supply; above this section, a series of mutually integrable criteria allows the user to set up personalized filters and thereby reduce a potentially long list of information layers; searches can be conducted on the basis of words contained in titles or descriptions and the database can be filtered by descriptor, minimum and maximum scale, date of updating and type of resource.

7 CONCLUSIONS

The four research projects have so far led to the creation of a first prototype and to its testing inside a study area.

Future targets, in the light of considerations prompted by the test phase, involve passing from the prototype phase of the system to a more operational and consolidated phase.

It is planned in particular:
- to create a simplified interface for access to the information (forms, metadata, information on costs, etc.)
- to focus on automatic methods of updating data, as regards both supply (“DBSensori” and “DBGiacimenti”) and demand (“DBDomanda”).
- to start a debate about the institution of national and regional agencies.

REFERENCES


