EnviroInfo 2006 (Graz)
Managing Environmental Knowledge

Contribution of a Spatial Multidimensional Portal for Natural Hazards Data Exploration

Julien Iris[1], Aldo Napoli[1], Franck Guarnieri[1]

Abstract

Management of natural hazards necessitates for the public and the private sector to be aware of the status of all the information at different levels and to build pertinent indicators for the decision making process. The indicators can be built by combining different types of data: descriptive data, numerical data and geographical data. In order to have an overview of the system the exploration of the indicators through various levels would be interesting:
visualization through geographical levels; national, provincial, departmental, communal and cadastral; visualization through temporal levels; analyzed periods; visualization through thematic levels; prevention themes, hazard themes, exposure themes.

In order to navigate through all this information, a flexible structure must be implemented to store all the pre-calculated indicators and to make it easily accessible for reporting (graphics, charts, maps, tables).
Dimensional modeling can help in the decision making process by exploiting urban data and risk data.
This paper gives an overview of the evaluation of the pertinence of dimensional modeling to support decision makers to get better insight on the data relative to natural hazards in France.

1. Introduction

The use of information technology in natural hazards is a key point for every organization that needs to analyze and to explore the whole of available data and knowledge on the subject. In 2003, (*Guarnieri and al 2003*) published a book which gives an overview of experiences and operational tools contributing to a better understanding of natural hazards. It also demonstrates that it contributes to improve prevention process. Three main axes are relevant: formalization of data and knowledge in order to normalize the dialogue between the actors of the system; use of information technology dedicated to earth observation by integrating environmental data, simulation process and spatial representations; access and sharing of spatial information according to the needs of each actor. In 2005 the French association for prevention of natural catastrophes published a white book (*AFPCN 2005*) which makes propositions concerning the use of geographical information to reduce natural risks in France and in Europe. The main suggestions are transparency, accessibility, confidentiality and subsidiarity.

The present paper follows the previous threads of research in that field by presenting the formalization and the conception of a prototype in the context of natural hazards management in France. The first chapter makes a brief overview on the needs of natural hazards management and details the available geographical data. The second chapter describes the concept of Spatial OLAP which corresponds to the multidimensional approach developed by the Center for Research in Geomatics under the direction of Professor Yvan BEDARD *(Bédard 1997*) and presents a case study with the conception of a prototype for one department in south of France. Finally the conclusion will expose the limits and the potentials of this technological approach.

---

[1] Ecole des Mines de Paris – Pôle Cindyniques
http://www.cindy.ensmp.fr
julien.iris@ensmp.fr,
aldo.napoli@ensmp.fr,
franck.guarnieri@ensmp.fr

2. Context and available data on Natural Hazards in France

The principle of protection against natural hazards in France has been built upon a public-private sector partnership between the state and the insurance companies. The law defines a "natural hazard" as an event resulting from the abnormal intensity of a natural agent and defines "damage due to natural catastrophes" as the damage resulting from the abnormal intensity of a natural agent when usual measures to prevent such damage have not been applied. Since 1995 a new tool has been created to control the land use in hazardous zones: the Prevention Risk Plan (PPR). The role of this plan is to integrate into the Possession Land Survey (Plan Local d'Urbanisme – PLU) the specific areas where it is forbidden to build and where it is possible to build under certain conditions. It contains the global objective in regard to the risks, the hazard maps, the risk zone map attached to the PLU and the recommendations and obligations for urban development of risk zone. The prevention and the protection of communities to natural hazards is under the responsibility of the public institutions. The Ministry of Environment and Sustainable Development coordinates the advancement of procedures at the national level. The prefect of the department is responsible of the realization of Prevention Risk Plans for the communities exposed to natural hazards in the department. The mayor of the community is responsible for the application of the recommendations and the obligations relative to the development in risk zones. Today almost 5000 PPR are effective in 2005 but about 20 000 communities are exposed to natural hazards: 11,600 communities are susceptible to flood, 5,500 to earthquake, 4,500 to subsidence and 600 to avalanche (*Sanson, 2001*). To summarize, 60% of the French communities are susceptible to a peril due to a natural event. In this context, the classification of communities through pertinent indicators could help to have better insight on the evolution of communities in regard to natural hazards. These indicators must take into account the spatial distribution of the population and of the dwelling places and the spatial distribution of hazards inside the communities.

In order to lower the degree of uncertainty in regard to natural hazards three axes of exploration have been selected: the hazard data which concerns the cause of the disaster; the exposure data that characterize the object exposed to the hazard; the consequence data which describe the damage and losses resulting from the disaster.

Concerning the hazard, some maps are realized locally in each community where PPR have been prescribed and approved. The communities and the prefectures store some studies but there is some difficulties to access and to exploit these maps because the administrative procedures are not clear and the data formats are varying from one place to another. However some public (or semi public) institutions, specialized in natural phenomena, are involved in publishing vector risk maps directly exploitable. DIREN (Regional Directorates for the Environment) are regional public structures specialized in water management and they are publishing the AZI (Atlases of flood zones) which represents the potential flood plains of main basins. AZI presents flood plains for a 100 years return period or the major event observed (PHEC – Highest water level known). BRGM (France's leading public institution involved in geosciences) (*BRGM 2004*) publishes landslides, underground cavities, subsidence areas everywhere in France. All the maps are accessible online from their specific databases. SISFRANCE is a database published by the BRGM, EDF and Ministry of Environment that contains the localizations earthquake epicenters (*BRGM 2004*). Concerning avalanches, RTM (Renovation of Montain Lands) publishes some avalanche maps called CLPA (Local Maps for Avalanches Prevention) (*ONF 2005*). The maps listed in this paragraph are in vectorial format directly exploitable by a GIS (Geographical Information System).

Concerning the exposure, it exists databases containing information relative to the ground occupancy inside the community: type of buildings, type of activities, number of residents or potential occupancy. The realization guide of Technological Prevention Plan (*PPRT 2005*) details the methodology to identify and qualify the exposures at the scale of the community. In order to qualify the existing urbanization, geographical information databases are commercialized by IGN (National Institute of Geography) like BDOrtho that delimits the elements of ground occupancy with uniform sets (size, height; density) or BDTopo that contains additional information like the function of each set (industrial, commercial, agriculture). Some information about the public infrastructures and their capacity are provided by the Departmental Direction of Equipment (*DDE, 2005*). Concerning the population, the main demographic data sources used in France are provided by INSEE (National Institute of Statistics); indicators are available at the scale of technical perimeters (“îlots” or IRIS for example) or at the scale of the community. It is possible to distinguish the density with the individual habitations and the collective habitations using the number of stages in buildings. (*CERTU 2005*) has published a method of estimation of the population using these criteria. Other database could be used in order to evaluate the employments zones like (*Astrée 2005*) with a geolocalization of the employments at the level of the address. BD SIRENE from updated by INSEE contains the companies but the data are not geolocalized.

Concerning the consequences, very few public databases exist. Only GASPAR (*GASPAR, 2005*) database (Assisted Management of Administrative Procedures relative to Natural Hazards) updated and published by the French Ministry of Environment and Sustainable development contains all the hazard decrees classified per community and per hazard from 1982 to now *(Prim 2005)*.

A community is vulnerable if the population and the infrastructures are included in hazard zones and if there is a delay in the process of prevention in term of urbanism control. It gives an idea of the indicators that could help to support decision making process for policy makers at the national level for the Ministry of Environment or for the Ministry of Equipment but also at the regional level for the DIREN and DDE that needs aggregated data visualization in order to follow up the evolution of their communities facing natural hazards.

3. Spatial OLAP for natural hazard data exploration

To understand Spatial OLAP (SOLAP) it is important to review briefly what is OLAP (Online Analytical Processing). OLAP and Spatial OLAP both exploit the potential of multidimensional structures following the rules of Dimensional Modelling. Kimball (*Kimball 2002*) defines dimensional modelling (DM) as a logical design technique that seeks to present the data in a standard, intuitive framework that allows for high-performance access. DM is adequate for decision support applications: necessary for OLAP tools often used to provide access to a data-warehouse. Data-warehouses store large volumes of data for the analysis process and for conducting the organization’s strategic decisions. DM is inherently dimensional, and it adheres to a discipline that uses the relational model with some important restrictions. The concept of DM and OLAP has been formalized by Ralph Kimball. OLAP has first been defined as "...the name given to the dynamic enterprise analysis required to create, manipulate, animate and synthesize information from exegetical, contemplative and formulaic data analysis models. This includes the ability to discern new or unanticipated relationships between variables, the ability to identify the parameters necessary to handle large amounts of data, to create an unlimited number of dimensions, and to specify cross-dimensional conditions and expressions" *(Codd and al. 1993*). The multidimensional approach is based on *dimensions and measures.* Dimensions represent the analysis axes, while measures are the numerical attributes being analyzed against the different dimensions. A measure can be considered as the dependent variable while dimensions are the independent variables. A dimension contains members that are organized hierarchically into levels, each level having a different granularity going from coarse at the most aggregated level to fine at the most detailed level. The members of one level can be aggregated to form the members of the next higher level. A set of measures aggregated according to a set of dimensions forms what is often called a *data cube* or a hypercube. Inside a data cube, possible aggregations of measures on all the possible combinations of dimension members are pre-computed. This greatly increases query performance in comparison to the conventional transaction-oriented data structures found in relational and object-relational database management systems (DBMS). The common *OLAP* architecture can be divided into three parts: the multidimensionally structured database, the OLAP server that manages the database and carries out the different calculations, and finally the OLAP client that accesses the database via the OLAP server. This access allows the end user to explore and analyze the data using different visualization methods and adapted operators such as drill-down (show details), roll-up (show a more global picture, also called drill-up), drill-across (show another theme at the same level of detail) and swap (change a dimension for another one).

Spatial OLAP can be defined as "*a visual platform built especially to support rapid and easy spatiotemporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays" (Bédard 1997*). SOLAP are meant to be client applications sitting on top of a multi-scale spatial data warehouse *(Bédard et al. 2001*). While designing a SOLAP application and architecture, one has to determine the spatial dimensions needed by the user, the desired geometry for the dimension members, the spatial measures to be included, and the spatial operators to be available in order to perform spatial analysis. Three types of spatial dimensions are recognized: the non-geometric spatial dimensions, the geometric spatial dimensions, and the mixed spatial dimensions. In the first type of dimension, the spatial reference uses nominal data only (e.g. place names) as no geometry or cartographic representation is associated with the dimension members. It is the one currently used for conventional OLAP tools. The two other types of spatial dimensions include geometric shapes spatially referenced on a map to allow its dimension members to be visualized and queried graphically. Another type of dimension used in SOLAP is temporal dimension. For this dimension there is a temporal slider in SOLAP interface that allows the users to navigate through temporal members easily. In the spatial multidimensional context, not only the dimensions can possess a geometric shape, but also the measures. Two types of measures then exist, a conventional measure and a spatial measure. A SOLAP system should be able to manipulate both: the numerical measures as used in conventional OLAP systems and the spatial measures (*Bédard et al.* 2001). Actually SOLAP support numeric measures like other OLAP tools. A spatial representation is possible when a dimension member has a corresponding geometric object (vector or raster pixels). CRG is working on the implementation of topological and metric operators and the results of the computations will be stored in cube cells like surface and distance.

In the frame of an evaluation of web mapping technologies, different prototypes of internet portals for the diffusion of risk data have been developed. After an evaluation of the technology MapServer (*Chaze, 2004* and *Guion, 2005*) the evaluation of the potentiality of SOLAP technology has been started. A first prototype of a SOLAP portal has been developed by Chaze and Napoli (*Chaze, 2005*) with the collaboration of the CRG team in Quebec. An improvement of this portal has been developed in order to specialise the given information with more dimensions and heterogeneous data sources. The new prototype exploits GASPAR database and Atlases of Flood Zones (AZI) for analyse on flood hazard in South of France department. To construct a SOLAP architecture the first step consists in building the multidimensional database (MDDB). The MDDB is created in a relational environment as noticed in the previous chapter. Every MDDB, also named data cube, is composed of one table with a multipart key, called the fact table containing the measures (values of indicators), and a set of smaller tables called dimension tables representing the analysis axes. Each dimension table has a single-part primary key that corresponds exactly to one of the components of the multipart key in the fact table. The list of indicators will serve to define the features of the MDDB. Each indicator must be described with precision because it will give information about where to look up in the source databases. Here are some examples of requested indicators: number of buildings, number of residents in activity zones included in flood zones with past hazard decrees, with prevention plans approved (or prescribed) for the community of Antibes (south of France). Each measure value is calculated for a given combination of dimensions and aggregated for each hierarchical level of the selected dimensions. One value constitutes one cell of the data cube: one cell is an intersection of all dimensions. A measure is necessary numeric. From the list of indicators two measures have been identified: the number of buildings and the number of residents. The measures will result from the combinations of geographical and non geographical information. The dimensions are divided in three categories: spatial dimensions, temporal dimensions, nominative dimensions. The nominative dimensions are the dimensions without spatial representation.

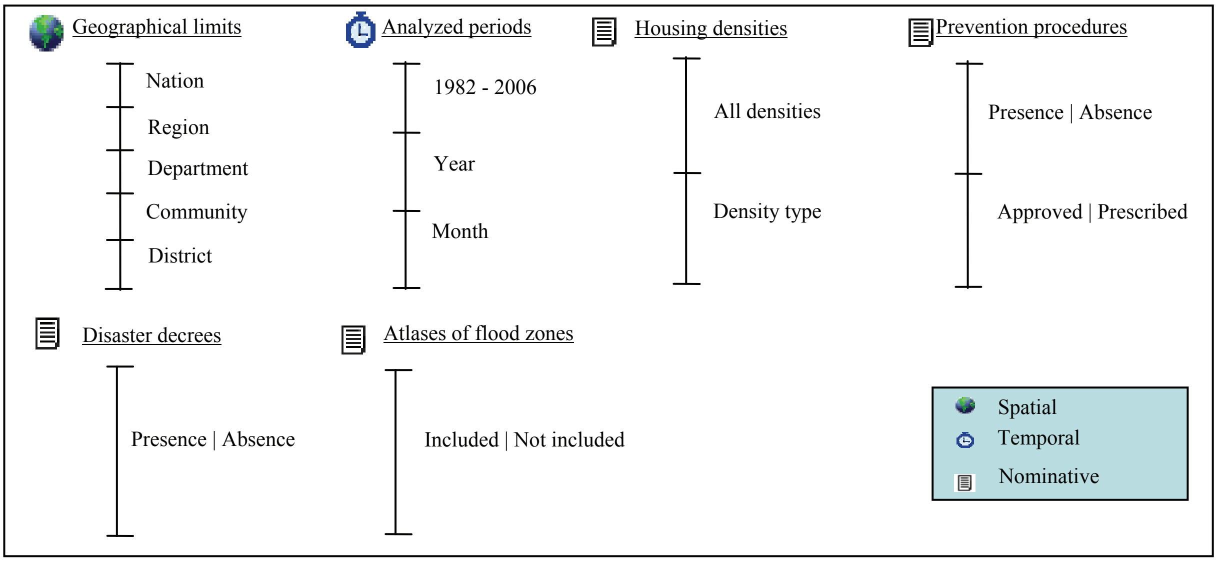

Fig. 1: List of dimensions

Geometrical objects are attached to the spatial dimension. “Geographical limits” uses BD Geofla (IGN) for the polygons relative to national, regional, departmental, communal levels and uses BD Iris (INSEE) for the district level. Both spatial data sources are in vector format. The dimension “Analyzed periods” is the temporal dimension and allows the user to select the periods of analyse from 1982 to now. The dimension “Housing densities” contains the three types of zones in Bd Iris : “Habitat” individual and collective (H), “Activity” industrial and commercial (A), “Diverse” agriculture fields and forest (D). The dimension “Prevention Procedures” concerns the status of the prevention plan at the community level. The dimension “Hazard decrees” represents the presence or absence of decrees in GASPAR per community. The dimension “Atlases of flood zones” concerns the inclusion or not into envelops of flood zones. The multidimensional data model is the scheme of the MDDB. The model has a star architecture with the fact table in the middle and the dimension tables around. As explained previously the fact table contains all foreign keys of the dimension tables and all the measures. The schema is not normalized as becomes obvious, for example, from the dimension table “Analyzed periods”. The attributes *month* and *year* are nested, which implies some redundancies. For small or medium-sized data volumes, such schemas have a sufficient performance because join operations are only necessary between the fact table and the related dimension tables. But for important sized data like “Geographical limits” the snowflake schema is recommended: instead of modelling each dimension by one table, a table is created for each level of a hierarchical attribute. Identifiers play the role either of a primary key in dimension tables or a foreign key in fact table. The “Geographical limits” is decomposed in five tables “Districts”, “Communities”, “Departments”, “Regions” and “Countries”.

4. Discussion and conclusion

The prototype has been developed for one department in south of France: all combinations are calculated from the granular level to the aggregated levels of each dimension and stored in a data cube. The data cube has been created in a relational database environment (Mysql) and a web interface has been developed to calculate the combinations, to load the data into the cube and to maintain the metadata. All geographical computations between Iris zones and Atlases of flood zones have been realized with Arcview (Esri). Once the data loaded into the cube, SOLAP interface gives a lot of flexibility to generate charts, maps and cross tables. The tool is able to display one or more measures simultaneously. The operators drill-up, drill-down, slide and roll-up gives the possibility to set the dimensions with one or more of its members, to consult the aggregations for all types of representations.

Beyond the development of this prototype some other combinations could be very useful to better understand the vulnerability of the territory. For example the computation of BD Astrée with localisations of employments with various hazard maps (flood maps, landslide maps, subsidence maps) and with BD Ortho maps, with more precision on the building function, will give the opportunity to analyse the concentration of activities in risk zones. Also it could be interesting to identify public and private buildings susceptible to welcome people in case of emergency near the dangerous zones (from administrative entities like DDE). All this data could be aggregated and analyzed through various representations in order to detect and prioritize the communities where some measures must be taken: by evaluating the construction rules and of the urbanism rules in the prevention plans, evaluating the alert communication process for these zones and the structural measures at the level of the risk basins.

However Spatial OLAP presents some limitative aspects. The aggregations are not calculated on the fly: the data structure must contain all combinations for all levels of aggregations. It generates some volumetric data problems. The number of dimensions and the number of hierarchical levels per dimension must be restricted in order to limit the amount of data. Otherwise it is consuming a lot of time and a lot of disk space for loading the data in the cube. The Center of Research in Geomatics is actually developing the aggregation on the fly to resolve this problem. It will be operational in 2006. (*Marchand and al. 2001*) have proposed a method that implements a Spatio Temporal Topological Operators Dimension (ST20D). The interest of such operators is to go deeper in geographical analyses by pre-computing or performing on the fly calculations between geometric objects. Nevertheless these functionalities will be useful in the future in order to analyse the distance between a building and the river or to make spatio-temporal relations between geometrical objects. The reader is referred to (*Marchand and al., 2001*) to have more explanations about ST20D. For the moment it is at the state of research and it is not implemented in the commercialized version of SOLAP. According to the results obtained with the prototype Spatial OLAP presents important potentials for the exploration and for the analyses of data relative to natural hazards and other risk management data. Nowadays it is important to make reporting on detailed and aggregated geographical data in order to have better insight on the system and to help the decision process for risk managers by computing hazard, exposure and consequence data. As explained in this paper heterogeneous data sources on hazard, exposure and damage are distributed on the territory. SOLAP is an attractive solution to regroup geographical and not geographical data in a uniform representation (cleaned and aggregated) subject oriented with the pertinent choices of indicators and dimensions. Another advantage is the advanced navigation interface through geographical data; SOLAP is not attached to one specific data format: it can exploit geographical data from a Mapfile, an Arcview file and other format such as raster or satellite photos: for each format, the tool translates it into a web mapping format. This transversal access gives more flexibility in the computation of heterogeneous data sources.

Acknowledgments

Based on a collaboration with the Center of Research in Geomatics (http://www.crg.ulaval.ca) and the company Kheops technology (http://www.kheops-tech.com) in charge of the commercialization of Spatial OLAP.

Bibliography

AFPCN (2005): S'informer pour réduire les risques naturels, Livre blanc, Les publications de l'AFPCN.

Bédard Y. (1997): Spatial OLAP », Vidéoconférence. 2ème Forum annuel sur la R-D, Géomatique VI: Un monde accessible, 13-14 novembre, Montréal.

Bédard Y., Larrivée S., Proulx M-J., Caron P-Y ., Letourneau F. (1997): Geospatial Data Warehousing: Positionnement technologique et stratégique. Rapport pour le Centre de recherche pour la défense de Valcartier (CRDV), 79 p.

Bédard Y., Merrett T. and Han J. (2001): Fundamentals of spatial data warehousing for geographic knowledge discovery. In: Geographic Data Mining and Knowledge Discovery.

BRGM (2004): Rapport d'activité 2004, Presse BRGM.

CERTU (2005) : Méthode d'estimation de la population par la méthode du noyau, Note méthodologique et technique, CERTU et CETE de Rouen, 2005 available on http://www.certu.fr

Chaze X. (2004): Conception et développement d'un portail Internet sur les risques naturels en France à l'aide de la technologie MapServer, 2005, Final internship report for National School of Geographical Science (ENSG) in partnership between Pôle Cindyniques, 80 p.

Chaze X. (2005) : Conception et développement d'un portail Internet sur les risques naturels en France à l'aide de la technologie SOLAP, 2005, Final internship report for National School of Geographical Science (ENSG) in partnership between Pôle Cindyniques and Center of Research in Geomatics.

DDE (2005) : http://www.equipement.gouv.fr/

Guarnieri F., Garbolino E. (2003): Systèmes d'information et risques naturels, Les presses de l'Ecole des Mines de Paris

Guion N. (2005): Conception et développement d'un portail Internet sur les risques naturels en France à l'aide de la technologie CartoWeb 3, Final internship report for National School of Geographical Science (ENSG) in partnership between Pôle Cindyniques, 50 p.

IGN (2005): http://www.ign.fr

Kimball R., Ross M. (2002): The Data Warehouse Toolkit – 2nd edition, Eyrolles editions.

Marchand P., Bédard Y., Edwards G. (2001): A hyper cube-based method for spatio-temporal exploration and analysis. Geolnfomatica, Submitted May 2001.

Prim (2005): http://www.prim.net Prevention and information on major hazards MEDD

Rivest S., Bedard Y., Marchand P. (2001): Toward better support for spatial decision making defining the characteristics of spatial online analytical processing (SOLAP), Published in Geomatica review vol. 55 2001