

***DATA MINING & DEVELOPMENT POLICY:  
WHICH HELP ON BUILDING TERRITORIAL INDICATORS ?***

---

**Yann BERTACCHINI**

Maître de Conférences S.I.C

Expert près l'U.E

Laboratoire LePont, Université de Toulon & du Var

Dépt "services & Réseaux de Communication"

200, Avenue Victor Sergent, 83.700 Saint-Raphaël, France

<mailto:bertacchini@univ-tln.fr>

**Sumário** : Agora em, o ambiente de atores territoriais está composto de local, redes internacionais nacionais. Além do mais, por causa do novo sistema de telecomunicações e multimedia, a tecla a novas estacas tornou-se uma prioridade. Agradece a redes e para suas aplicações, uma nova inteligência coletiva será feita possível para resolver problemas provados mais e mais complexo devido a informação massiva. Em hoje mundo de s, quantia de informação sabido é essencial, e dá esses, territórios incluído, quem tem uma abundância de informação, uma borda acima outros. Todos atores locais, organizações públicas privadas, existem numa idade onde todas espécies diferentes de conhecimento são essenciais e importantes sobreviver e prosperar. Sem conhecimento, nós nao saberíamos que o que acontecia ao redor do próximo canto, nem ria't entende a jamais-mudança, mundo jamais crescendo. Um território local que competitivamente pode colecionar dados com êxito pode permanecer em topo de o que's ir em ao redor deles, e alcança níveis altos de entendimento que leva a êxito. É aceitado em parte toda, isso se um sabe, que um receberá, e isto segue tudo e todo o mundo. Se obter a informação muito necessária era tão fácil quanto pode soar, ter havido muito mais êxito, regular mais cedo que falamos, por sobrepujar qualquer barreiras que podem vir no meio. Os territórios em hoje mundo de s também deve lidar com níveis altos de competição, e seu êxito anda de alto nos ombros dos atores locais eles servem. Por cavar fundo em seus fregueses' meio de vida, meio de pensar, territórios podem são capazes de usar esta informação como uma ferramenta positiva ajudar prediz e "lê" seu freguês' mentes de s, baseado em experiências de passado. O pensado que é proposto neste artigo aponta num processo de mediation, como nós também chamamos Sistema Territorial de Inteligência de Informação. Para encorajar o synergy entre atores locais. De fato, novos câmbios -parcialmente virtual- vai ir cohabit com outros lados de relações rompidas. Existir e renovar, esses espaça devem apresentar redes e meio de transferência de conhecimento por disposalsType educacional, econômico democrático ou texto de pasta aqui. Embarcamos para explicar como complexo o local e seu desenvolvimento é e, por quê dá o meio à definição e para o escrito de um procedimento de informational baseado em mineração de Dados para fingir a territoriality como para um território. Também explicamos por quê é a construção territorial uma questão para uma aproximação de systemic cujo propósito central, em nosso interesse, sobra a capacidade do território organizar o relational rede local de vistoria pericial e fazer seu meio a um projeto territorial de inteligência.

**Summary** : From now on, the environment of territorial actors is composed of local, national and international networks. Moreover, because of the new system of telecommunications and multimedia, the key to new stakes has become a priority. Thanks to networks and their applications, a new collective intelligence will be made possible in order to solve problems proved more and more complex due to massive information. In today's world, amount of information known is essential, and gives those, territories included, who have an abundance of information, an edge above others. All local actors, private and public organisations, exist in an age where all different kinds of knowledge are essential and important to survive and thrive. Without knowledge, we would not know what was happening around the next corner, or wouldn't understand the ever-changing, ever growing world. A local territory which can competitively collect data can successfully stay on top of what's going on around them, and achieve high levels of understanding which leads to success. It is accepted everywhere, that if one knows, than one will receive, and this follows everything and everyone. If obtaining the much needed information was as easy as it may sound, there would have been much more success, even earlier than we speak, by overcoming any barriers that may come into the way. Territories in today's world must also deal with high levels of competition, and their success rides high on the shoulders of the local actors they serve. By digging deep into their *customers'* way of life, way of thinking, territories might be able to use this information as a positive tool to help predict and "read" their *customer's* minds, based on past experiences. The thought which is proposed in this article aims at a process of mediation, as we also called Territorial Information Intelligence System. In order to encourage the synergy between local actors. Indeed, new exchanges -partly virtual- will cohabit with other sides of disrupted relations. To exist and renew, those spaces should present networks and means of transfer of knowledge through democratic, economic and educational disposals. We set out to explain how complex the local and its development is and, why it gives the way to the definition and the writing of an informational procedure based on Data mining in order to pretend to territoriality as a territory. Also we explain why is the territorial construction a matter for a systemic approach whose central purpose, in our concern, remains the ability of the territory to organize the relational local expertise network and to make its way to a territorial intelligence project.

**Palavras-chave:** Os dados; Informação; Conhecimento, Local; Apólice; Processo.

**Keywords** : Data; Information; Knowledge, Local; Policy; Process.

# **Data mining & Development policy:**

## **WHICH HELP ON BUILDING TERRITORIAL INDICATORS ?**

### **INTRODUCTION**

Data mining methods

### **ASSOCIATION**

Investigates growth differences  
Private School Location  
The evaluation of socio-economic and cultural activities of cities possessing a university.

### **SEQUENCE-BASED ANALYSIS**

Shifts in Regional Employment  
Territorial Development and Networking

### **CLUSTERING**

Knowledge, innovation and collective learning  
Classification  
Innovation and New Technologies

### **FUZZY LOGIC**

The creation and development of incubators for business ideas in the ICT sector  
Estimation  
Measuring the level of quality of life of local governments

### **GENETIC ALGORITHM**

Spatial Planning, Structural Change and Regional Development policies

### **NEURAL NETWORKS**

Spatial Interaction Modeling of Interregional Commodity Flows  
Data Mining of Advanced Database

### **INTELLIGENT AGENTS**

The regional system of innovation and regional development  
Multidimensional analysis tools  
Immaterial resources and regional development : a territorial approach

### **CONCLUSION**

### **BIBLIOGRAPHIC REFERENCES**

### **INTRODUCTION**

What is data mining? The process of automatically extracting valid, OR "You press the data until they confess".

In today's world, amount of information known is essential, and gives those who have an abundance of information, an edge above others. We all exist in an age where all different kinds of knowledge are essential and important to survive and thrive. Without knowledge, we would not know what was happening around the next corner, or wouldn't understand the ever-changing, ever growing world and the fast pace around us which increases everyday with new coming events and new ways of solving problems. Territories level as organisation need to see the need to acquire knowledge to keep up with the glocal (global-local) competition. A territorial level which can competitively collect data can successfully stay on top of what's going on around them, and achieve high levels of understanding which leads to attractivity.

Data mining is compatible with territorial competitiveness. We will associate them in our communication. The territory is carrying direction and establishing an activity of territorial watch before which would use amongst, other things average, data mining, can reinforce its attractivity.

### **Data mining methods**

By digging deep into their inhabitants' way of life, way of thinking, their habits to do something, territories might be able to use this information as a positive tool to help predict and "read" their inhabitants' minds, based on past experiences. Much of the knowledge that can help out and assure success is hidden, not readily available to the right person at the right time. With the past limiting capabilities that have existed, emphasis has been placed on finding and sorting out this information to benefit the organization at hand<sup>1</sup>. The main problem here that needs to be overcome is the difficulty of extracting this hidden information about any particular system that has been studied from collected data. But the main and whole idea and act of discovering relationships and connecting

variables is a database sum of the main idea and point of Data Mining.

There are various capabilities that are provided by data mining techniques. There are many techniques that Data Mining uses to do exactly that, and these techniques used depend on what type of information one is trying to develop and extract. In the present context of the emergence of the knowledge economy, the role of immaterial resources in regional or local development, in particular as a factor of attractivity and competitiveness, is widely acknowledged. However anyone who tries to consider both resource management and regional development gets confronted to a variety of content, to elements of different nature and origin, moreover to circular definitions<sup>2</sup>. The objective of this paper is to enter the black box of regional resources, more specifically of immaterial resources, and understand its territorial dimension in association with Data Mining methods. It proposes a conceptual framework allowing us to distinguish the various aspects of this immaterial 'magma' and to seize its dynamic through time and space.

## ASSOCIATION

This technique approaches to address a problem that concerns a space-based analysis, treating the variables which are investigating on first step growth differences in the European urban system. The main goal is to locate trends and similar behaviors across a large number of indicators to be used later to understand. This information may be very useful to organisations to adjust aspects of their activities.

### Investigates growth differences

There is evidence strongly suggesting that equilibrating flows between cities are highly constrained in the EU. Models in which growth of real GDP p.c. are the dependent variable perform well and make it possible to test significant hypotheses. Evidence is found which is supportive of a spatial adaptation of the endogenous growth model with the relative size of the university sector having a highly significant role in explaining growth differences<sup>3</sup>.

### Private School Location

A simple logic specification produces evidence of the importance of substitutes to private schools in deterring entry to a community. These are represented by public school spending per pupil and the extent of local library lending.

### The evaluation of socio-economic and cultural activities of cities possessing a university.

The focus is to test socio-economic and cultural capacities of cities in which are located university

to measure whether the university administration necessity of receives or not. On two successive steps is shown student attractive strength of this cities.

## SEQUENCE-BASED ANALYSIS

This is another *space*-basket analysis dealing with the total factors in one professional *transaction*. This type of an analysis can identify a set of variables that could forecast the next anticipated behavior in labour market.

### Shifts in Regional Employment

The U.E has undergone deep structural changes in the 1980s, 1990s with strong implications for the labor markets. The general move with overall decrease (80s) then overall increase (90s) in employment levels in U.E is evident from official statistics.

However, the sectoral and regional mapping of this shift has not been properly studied yet. Which appropriate model is able to measure the impact of output change, productivity gain and other non-labor factors on employment change in the period ?

### Territorial Development and Networking

The purpose is to explore the main features and the significance of the SME sector development for approaching and to identify the signs and the perspectives of SME territorial networking phenomenon<sup>4</sup>. It also seeks to identify phenomena characteristic to territorial networking in which SMEs can bring a significant contribution.

## CLUSTERING

This Data Mining method addresses problems with segmentation and grouping. Data with a big group of attributes are automatically assigned by clustering algorithms identifying unique characteristics into smaller sets of groups. This is often one of the first steps in a data mining analysis, identifying related record groups. These groups may then be used for further exploration and analysis of relationships. When combined with other data mining techniques, other types of characteristics may be determined.

### Knowledge, innovation and collective learning

Innovative capacity of firms has traditionally been explained through intra-firm characteristics, being firms size the most important. A wave of empirical studies identifies small firms as the engines of technological change and innovative activity, at least in certain industries. In the recent literature much emphasis has been put to determinate which are external to the firm; these external factors are called knowledge spillovers, and refer to the positive

externalities that firms receive in terms of knowledge from the environment in which it operates. Both industrial and regional economists underline the importance of knowledge spillovers. The concept of relational capital is fundamental in this respect. Relation capital is in fact defined as the set of all relationships - market relationships, power relationships, co-operation - established between firms, institutions and people, which stem from a strong sense of belonging and a highly developed capacity of cooperation typical of culturally similar people and institutions. It is, indeed, reasonable to expect that relational capital will play a different role in different regional, sectoral and firm's contexts.

### **Classification**

Classification is the most commonly applied data mining method, and is used to develop models that can classify large amounts of records. Pre-classified examples are used to determine a set of parameters needed to properly separate other data.

### **Innovation and New Technologies**

Many European Regions with a high autonomous government degree have done an effort to promote research and technological development (R&D) to achieve welfare and economic growth. Several Regions have organized their efforts in science and technology<sup>5</sup>, but, is really this budget effort been transferred to new technologies adopted by firms? Are science-technology flows increasing? Is there a regional gap between science and technology? Which are the factors determining these differences? What is the role of scientific knowledge to promote technological development in several European Regions.

How is the basic science used by industrial firms? Which are the most dynamic sectors using scientific knowledge? Which are the most demanded scientific fields by private firms? What type of institutions do use the scientific knowledge? Which is the lag time to translate science to technology? The used methodology to explain the link between science and technology is based on scientific inputs in patents, or non-patent citations (NPC).

### **FUZZY LOGIC**

Fuzzy logic is a superset of conventional (Boolean) logic that can handle a concept of partial truth whether values are "completely true" or "completely false", or anything in-between.

### **The creation and development of incubators for business ideas in the ICT sector**

How to increase the number of business ideas inside an University, then to select the best ones and then to fill the incubators of companies ?

### **Estimation**

Estimation deals with the classification technique changed around to involve assigned "scores" based on a pre-scored training set. This replaces assigning a binary classifier as used in classification.

### **Measuring the level of quality of life of local governments**

How to construct the set of indicators to measure the level of quality of life<sup>6</sup> of local governments with emphasis on sustainability and to apply it to local governments. In addition, it attempts to identify whether it can be applied to other foreign countries or not. In measuring the level of quality of life, many statistical methods including factor analysis and AHP technique are employed.

### **GENETIC ALGORITHM**

Basically this technique is an optimisation method using processes such as genetic combination, mutation and natural selection, based on the design of the concepts of natural evolution.

### **Spatial Planning, Structural Change and Regional Development policies**

As numerous European regions, they have been shaped through old industrial manner. These areas were ordinary built from steel and coal. These elements and the related industrial activities shaped the region<sup>7</sup>, brought its growth and also, after almost a century its decline between the 1950s and the 1970s. From that moment on, it took a few years to the regional authorities to get a full perception of the scale of the structural crisis, as it can be seen with the first reports which have made the diagnosis of the economical evolution of the region, published at the beginning and middle of the 1960s.

### **NEURAL NETWORKS**

Stated earlier were a collection of the most commonly used techniques and methods in Data Mining. One widely used technique, Neural Networks, is described in detail below. Neural Networking provides a new and different approach to using computers in the workplace today. By comparing it to the structure of neurons in a human brain, one can visualize the computer's architectural pattern, a huge web of electronic neurons, which send signals to each other through hundreds and thousands of connections. Neural Networks are used to determine patterns and relationships between various parameters and characteristics in information that is presented. This data, collected from various places such as market research, business decisions, is used to

apply a neural network, stepping away from the traditional approaches to information processing.

### **Spatial Interaction Modeling of Interregional Commodity Flows**

Understanding the determinants of interregional commodity flows is critical for both transportation infrastructure planning (highways, railroad tracks, river/port facilities) and regional development policies (location of activities, reducing regional disparities)<sup>8</sup>. Unfortunately, limited data availability has, in the past, hindered empirical research in this area. Based on input-output considerations and in order to differentiate intermediate from final commodity demands, the variables include more detailed descriptions of the economies of the origin and destination states, such as employment and value added for the commodity sector at the origin state, wholesale employment at both ends, manufacturing employment at the destination state, and population and per-capita income at both ends. In addition, the average establishment size for the commodity at the origin is intended to measure scale or diversification effects. The competitive or agglomerative effects of the economic spatial structure are captured with competing destination and intervening opportunities variables.

### **Data Mining of Advanced Database**

Due to the amount of information companies store on their databases, the databases has gone through an advancement in technology that effect the size of the database, the speed need to operate the database, and a way to organize the data. Advancement on the database correlated with the advancement required on the data mining tools. Such tools are intelligent agents and multidimensional analysis tools. Because of the huge amount of data, the data has to be organized and any errors should be eliminated to increase process speed.

### **INTELLIGENT AGENTS**

A huge database needs a tool that would be able to prioritize and/or filter the data that's being updated frequently. Intelligent agents do such as that and act as a secretary in filing and fetching data. There are several categories of intelligent agents available. Some agents are designed to launch automatically to perform specific queries or to search for patterns in data. Others automatically predefined intervals, performing a task or monitoring a condition in the background and returning an alert as required. IBM has proprietary, patent-pending techniques to analyse gigantic data sets for cross marketing or affinity marketing and look for patterns. These patterns may not just be the obvious, but rather ones that an analyst might

miss. Other intelligent agents use fuzzy logic and neural network algorithms to help analyze and control real time technical processes. Yet another agent is able to look for correlation by forming, testing, and modifies its own hypotheses.

### **The regional system of innovation and regional development**

Innovations and the capacity to innovate are crucial factors in the development of a firm and its ability to adapt to changes in the external environment. Growing attention has been paid to the mechanism facilitating innovation in firms, both in large, small and medium-sized enterprises. As a consequence, increasing attention has been on the role of innovation policy in regional development<sup>8</sup>.

The purpose of this project is to analyse the linkages between the business advisory systems efforts to promote innovation and the innovative firm with special attention to small and medium sized enterprises. The analysis of this paper deals with entities and relations of the innovative environment. The focal point is the interaction between the analysed business entity and the external environment as a part of a broader network of innovative relations covering intra-firm as well as extra-firm relations and processes<sup>9</sup>.

### **Multidimensional analysis tools**

There are two types of multidimensional analysis tools and they are multidimensional analysis (MDA) and On-Line Analytical Processing (OLAP)<sup>10</sup>. The idea is to load a multidimensional server with data that is likely to be combined. MDA represents data as n-dimensional matrices called hypercubes. OLAP let the user calculate metrics over a subset of available data, by exploring combinations of one or more dimensions of data. Delta Airlines, for example, has use OLAP to gain insights into its frequent-flyer program. It is able to consolidate data from a 100-GB Teradata into six far more accessible multidimensional databases totalising 6 GB. OLAP servers are great for time-series analysis, recursive calculations, and data with up to about 15 dimensions.

### **Immaterial resources and regional development : a territorial approach**

The analytical framework<sup>11</sup> is based on the idea that resources do not constitute a stock but rather a process. A process through which the objects (immaterial and material) of our surrounding environment are at a certain moment and in certain locus identified as being potentially useful for the production system<sup>12,13</sup>. The main questions here are the following: what are immaterial resources<sup>14</sup>? What are their characteristics? How do they evolve/move through time and space? How are they created? Destroyed? How are they identified? Managed? By whom? Where? How is learning (in it's different aspects) organised in time and space?

## CONCLUSION

Yet one of the common problems is that databases storing company's information are infested by errors, duplicate data, and unnecessary information pertaining to the data-mining applications. For data mining to produce valid results, data has to be cleaned and structured for consistency<sup>15</sup>. Such consistent format could be as simple as assigning 0's and 1's, instead of 'female' and 'male'. The process of cleaning up may be a slow one, but this would pay off in the end. The diversity and integration requirements of data are still a problem and solutions are being developed. Many organizations are still grappling with how to make the whole process work and how to fit it in with other initiatives such as data warehousing and other decision support areas. We have in this paper tried to show the footbridge which existed between dated mining and local development through studies specific. Other applications are under studies.

## BIBLIOGRAPHIC REFERENCES

- <sup>1</sup>CHOO, C W & BERGERON, (2001), P  
"Issue editorial." [Special issue on  
Environmental scanning and  
competitive intelligence.]  
*Information Research*, 7(1).
- <sup>2</sup>BERTACCHINI, Y.,(2002), Territoire &  
Territorialités; Vers l'intelligence  
territoriale; *volet 1*, 190 pages,  
Collection *Les E.T.I.C.*, Saint-  
Raphaël (Var-France).
- <sup>3</sup>BERGERON, P., HILLER, C.,(2001),  
«Competitive intelligence in  
Cronin, B., » (ed). *Annual Review  
for Information Science and  
Technology* (ARIST). Vol 36.
- <sup>4</sup>BERTACCHINI, Y.,(2001), DOU, H., «The  
Territorial competitive  
intelligence: a network concept»,  
Colloque Proceedings VSST  
'2001 p 101, Barcelone 15-19.
- <sup>5</sup>VON KROGH, G., ICHIJO, K., &  
NONAKA,(2000), I. Enabling  
knowledge creation. How to  
unlock the mystery of tacit  
knowledge and release the power  
of innovation. New York: *Oxford  
University Press*.
- <sup>6</sup>BERTACCHINI, Y., LEBRETON, M.,  
(2002),«Acteurs locaux, Réseaux  
et Territoire: représentation du  
potentiel d'action local», 70<sup>e</sup>  
Congrès de l'Acfas, Montréal,  
Canada.
- <sup>7</sup>BARQUERO, A. (eds.),(1998), Organisation  
of production and territory: local  
models of development, Gianni  
Iuculano Editore, Pavia.
- <sup>8</sup>LUNDVALL, B. A.,(1998), National systems  
of innovation, Pinter Publisher,  
London.
- <sup>9</sup>NEMETI, F. AND PFISTER, M.,(1998),  
Aspects de la compétitivité de  
l'industrie microtechnique suisse,  
EDES, IRER, University of  
Neuchâtel, Neuchâtel.
- <sup>10</sup>MORIN & AL., Base de données et  
statistiques,(2002), *Collection  
Sciences Sup*, 384 p, Dunod,  
Paris.
- <sup>11</sup>BENOIT, J.,(1999), «Une communauté de  
pratique en réseau : Le forum de  
discussion et la base de  
connaissances des inspecteurs de  
la Commission de la santé et de la  
sécurité du travail» (CSST).  
(Collection infomètre). Québec :  
CEFRIO.
- <sup>12</sup>BROWN, J. S. & DUGUID, P.,(2000), «The  
social life of information. »,  
Boston, MASS : *Harvard  
Business School Press*.
- <sup>13</sup>HILDRETH, P., KIMBLE, C. & WRIGHT,  
P.,(2000), Communities of  
practice in the distributed  
international environment.  
*Journal of Knowledge  
Management*. 4 (1), 27-38.
- <sup>14</sup>BUMBO, N., COLEMAN, D.,(1998), Case  
study : Building a knowledge  
community at Viant : the case for  
using a holistic approach.
- <sup>15</sup>HAN, J., KAMBER, M.,(2000), Data  
Mining: Concepts and

Techniques, édition *Morgan*  
*Kaufmann Publishers.*