How big data might induce learning with interactive visualization tools

*Giuseppe Roccasalva

How big data might induce learning with interactive visualization tools

Key words: **big data, visualization analytic tools, story-telling, we-gov, inductive learning, business analytics.**

Abstract The essay presents some results of an initial collaboration carried out between Polytechnic of Turin and CSI Piemonte (In-house ICT company of Piemonte Region). Different data visualization tools were selected and studied (Gapminder, ManyEyes, Open eXplorer and Fineo) in order to aid inductive learning about BD. The intelligent use of public sector information (PSI) becomes an important player for researches, urban studies and the governance but also for business. The doubling of digital information each two years made Data Visualization discipline crucial for accesses and includes complex info especially for business analysts. "We discover the world through our eyes" wrote Few, one of the most important manager of visual communication. Accordingly, forms of communication such as Graph and tables boosted the visual dimension for creating relation between values, colors, shapes and most of all analysts and customers are exploring new ways of learning ("storytelling" dimension). As planners and citizens, we rely greatly on eyes which are managing most of human sensors (nearly 70%) giving a crucial role to perception and cognitive maps, bias and new thoughts. Among the different tools and approaches, this paper highlights and applies a visualization and analytic technique for a specific case study where it is newly important to try synergies in decision making, planning and inductive learning processes.

INTRODUCTION

Planning, like all other disciplines concerned by geographical aspects, must face the huge information produced by digital communication, open software-sources and social networks. Blog, ranking lists, databases of statistical data, are only some of the available information which consciously or unconsciously reaches us daily. This requires new and simple ways of visualizing and reading digital information. This article starts describing the main features of big data (BD) and sounding opinions on their exploitation through the growing of visual analytic tools. Next it drafts the concept of profitable BD based on customers demand curve in order to choose the feasible visual analytic tool both in the private and public practices. Last, it describes a tool and its interaction role through an experiment carried out in consortium of 6 municipalities in the north east area of Turin.

* Giuseppe Roccasalva, Research Fellow Politecnico di Torino, Facoltà di Architettura (giuseppe.roccasalva@polito.it)

BIG DATA AND THEIR VALUE[1]

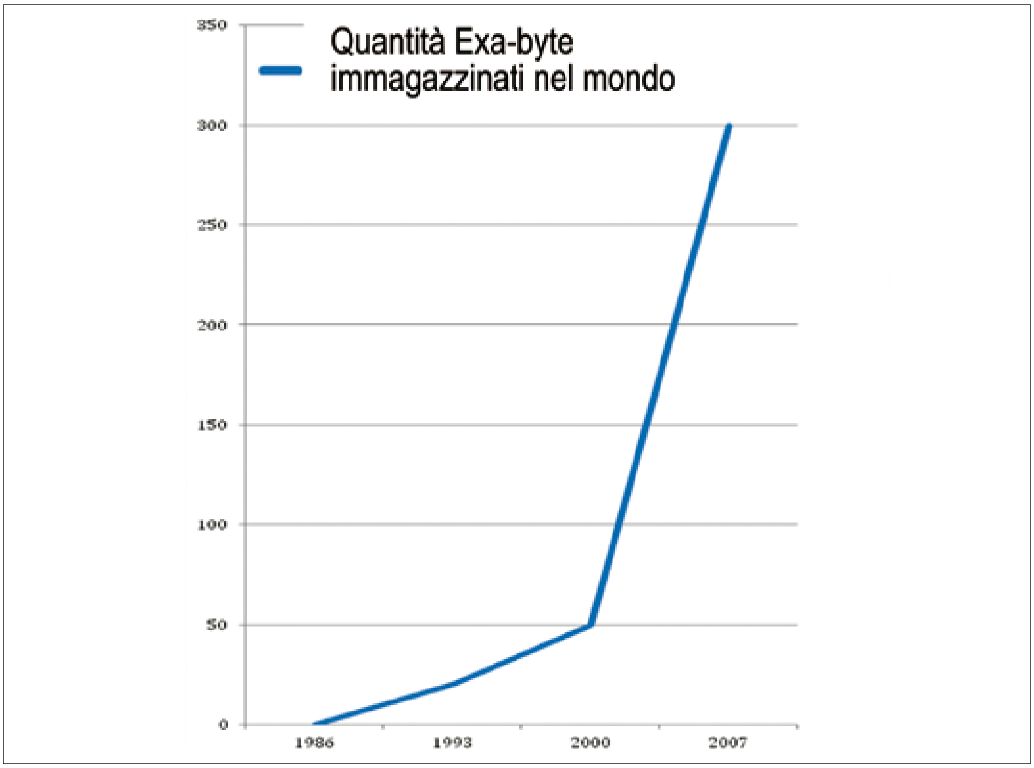
Big Data (BD) is a "buzzword" which stands for the quantity of digital information available on servers, going from some terabytes to dozens of petabytes stored in a dataset. In more explicit terms, eight bit (8B) is the basic dimension of informatics (it is as a letter or a number in a page) and one kilobyte (1KB) is about a text page, one megabyte is about a song while one gigabyte is about a film. So far, these are common and understandable dimensions for all the IT users. When it goes to terabytes or even petabytes these are dimensions which can be respectively comparable to all the books of a main city library or all the mails which are hourly processed by Google provider. The growing of data storage is increasing (100% every two years). McKinsley Global Institute reports that more than 80% of all the American economic sectors consists of companies storing on average more data than the Congress Library. Many of these data gather information about citizens and cities, opening up new opportunities for territorial studies. Batty[2] wrote "Our need to understand how all these dimensions are coalescing, merging, complementing, and substituting for one another has never been more urgent. It constitutes a major challenge for planning and design in the near future". Often, data are collected with temporal information; this has brought urban analysts to focus on tools which represent flows and network of flows. By 2020 thirty-five zettabytes of data will be stored between fixed discs and web-servers. This source can be comparable with the ones we commonly rely on (energy, raw materials, oil and so on…). In the last years, the exploitation of BD has produced application tools that gave benefit to communities, which reduced the functioning costs of local administration and even private practices. BD are mainly seen in this article as information which can benefit many sectors by revealing – with specific analytic tools - from trends to individual habits as urban commuting choices, customers profiles but even urban planning expectancies for dwelling prices or for qualities of neighborhood. The exploitation of digital information can generally improve global economy and public-private productivity but even reward single citizens and customers. According to McKinsley reports, BD can make Ocse Public Authorities save between 150 and 300 billion Euros by managing basic info coming from citizens, arranging automatic answering services, or warning services for fraud and transparency of practices. Public sector productivity can be improved by 20% if we balance both the cost reduction and the augmented service quality. American health care system will save 300 billion dollar a year which counts for 8% of National social assistance, while private companies can increase their revenues till 60%.

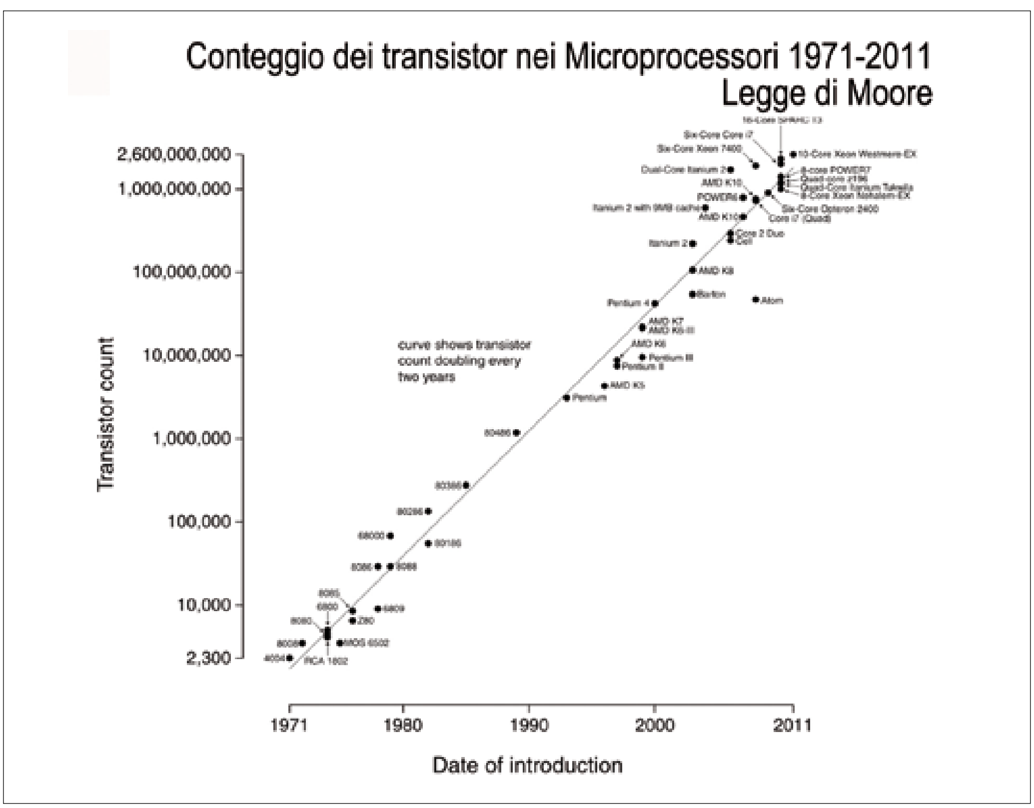
Some threshold dimensions

Within the multidisciplinary studies about BD, it is common opinion that quantity growth is never decreasing. Many researchers have studied the total quantity of generated data, stored data and used data by picturing an exponential trend of growth (see figure 1). In 2010 companies held seven exabytes of data while customers can store only six exabytes on personal computers[3]. This trend is due to the more accessible cost of technology, to the development of digital communication services (blog, wiki, social networks) and to the open and collaborative expansion of free softwares and applications. In 2008 American families were reached by and produced 3.6 zettabytes of information daily (34 gigabytes per person a day); in this ambit, written words account for 0.1% of total[4]. Questioning the limits of BD growth, some studies can help to frame this query. The Moore[5] law shows with a growing curve the increasing number of transistors on each chip produced every year. This increasing is doubling almost every two years. In this regard, it is possible to appreciate a correlation between Moore curve and the growing curve of world stored data (see figure 1).

1 McKinsey Global Institute (MGI) reports
2 Mikael Batty, Environment and Planning, 2012
3 McKinsey Global Institute (MGI) reports
4 University of California, San Diego in *The Economist*
5 Gordon Moore, 1965

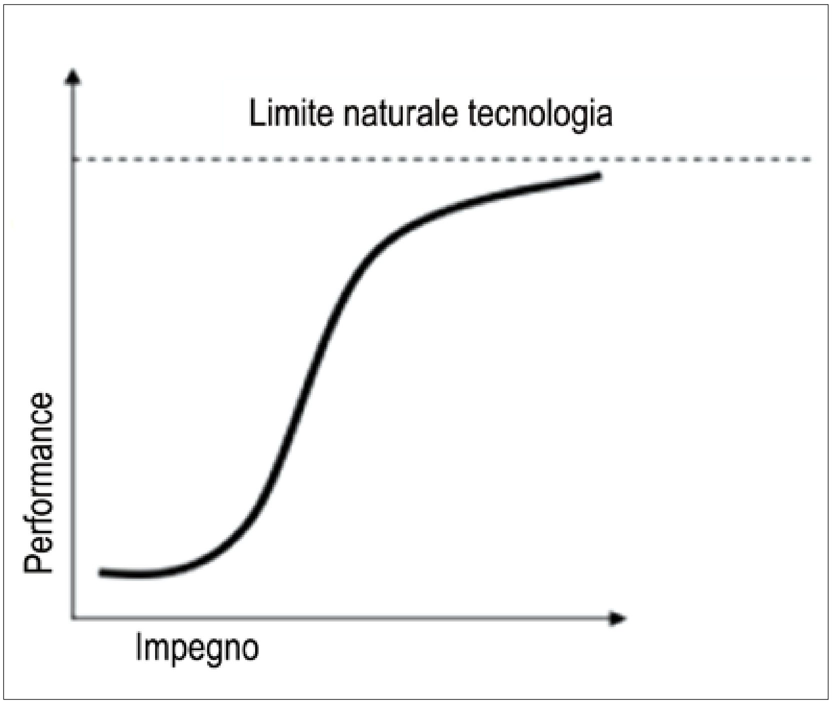
Figure 1 Big Data growing curve (source MGI) and Moore's law





It is a common belief that the future of BD can share some meaning with the S-form technologic curve; the latter describes the threshold of technology with respect of its use (see figure 2). According to Schilling, performance is slowly improving in the initial use because technology is not yet understood. The performance improvement leans up when researchers and organization have made their knowledge of technology more consistent. At a certain point, source and labor committed decrease and technology comes to its natural limit[6].

Figure 2 The S-form technologic curve



A tentative definition of Profitability

Is it possible to consider first the users demand rather than the production of goods and services? According to neoclassical and Keynesian thinking[7], this is possible in an economic system in which interaction of functional operators (family, business, government, etc …) exists.

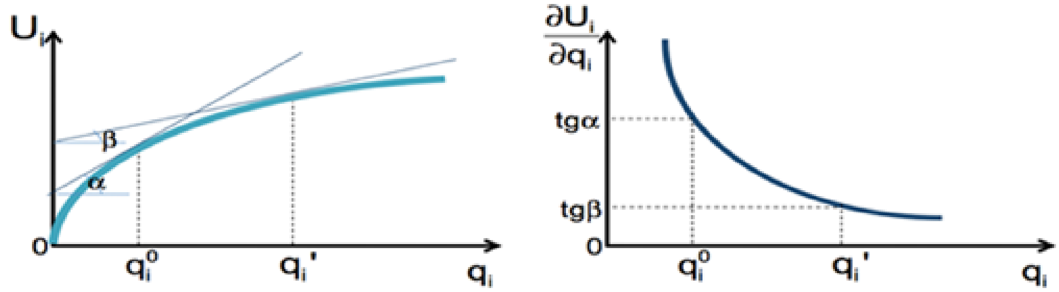
How to use the so-called PSI (Public Sector Information) as an economic source? The concept of big data "profitability" was outlined in order to understand to what extent and in what form public information can create economic values. Referring to the microeconomic theory, the utility of a certain good (in this case digital data) follows a specific function, where each additional good increases the degree of satisfaction (U), but with decreasing values.

6 Schilling M. A. (2009), *Gestione dell'innovazione.*

7 Keynes economic theory, according to Bruno Jossa, have focused more on the consumers rather than the production of products, finding that the critical demand is often not enough and sustainable, this gives chance to the public intervention.

Figure 3 source: liberally adapted from “Urban Economy” (O’Sullivan A.)

$$\frac{\partial U}{\partial q} > 0 \, ; \frac{\partial^2 U}{\partial q^2} < 0$$



Profitability concept was divided in two perspectives (business and socio economic) in order to understand how BD works in economic systems which have different objectives, contexts and stakeholders. It is necessary to profile the users who have the ability to access digital data and who have acceptable levels of digital divide experience. In this regard, some categorizations were studied but their detailed characteristics are omitted for this article.

In the business perspective the chart shows the use of the digital data with respect to their profitability. Three types of users-consumers are described (user A, B and C). They interact differently with digital data and, specifically:

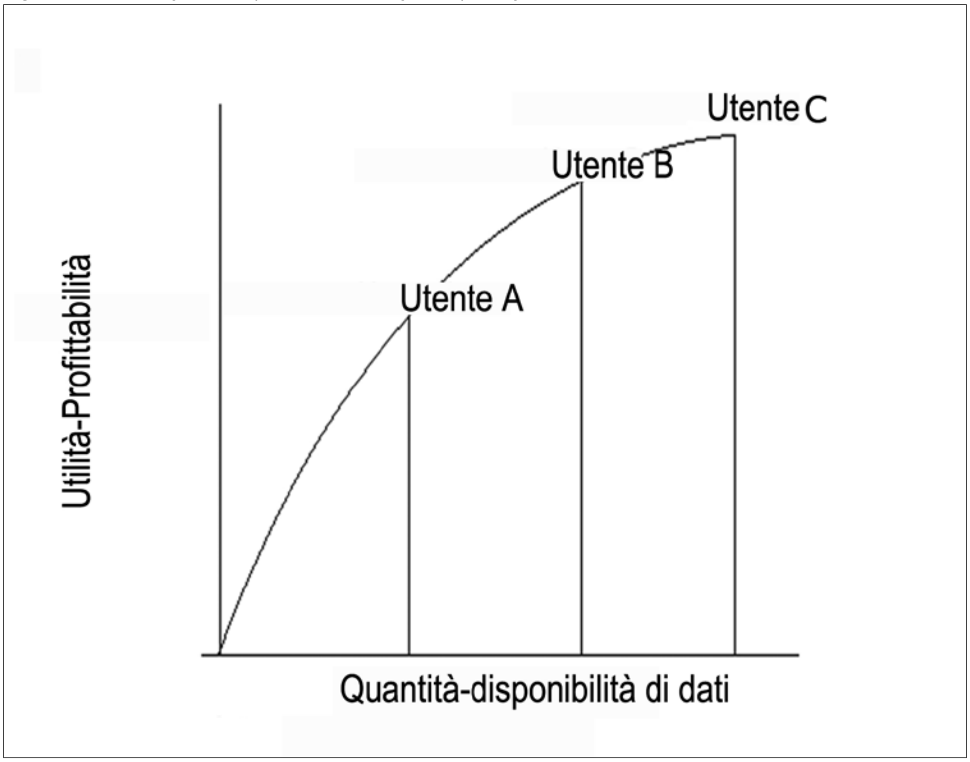
- (0;A) à users-consumers using data for communication, e.g. for chatting, twitting; these persons know different tools for exchanging data and digital information

(low profitability);

- (0;B) à users-consumers using data for getting informed, e.g. for web-searching; these persons manage data and succeed in learning or even make simple analysis (medium profitability);
- (B;C) à users-consumers using data for building some business, e.g.App designer, data scientist, entrepreneur or expert of the public sector; these persons are able to deeply understand digital data by even exploiting their values in many ways

(high profitability).

Figure 4 Profitability with respect to availability and quantity of BD



In the socio-economic perspective, the profitability is represented by a decreasing curve. The function changes with the increasing of BD and it can be used to foresee possible exploitations with resects of the users demand. For e.g. all things been equal, the increasing of BD can:

- generate confusion in identifying data that are really useful. In fact, a too large set of data makes their selection, learning and dissemination very difficult, decreasing their aptness for communities; large data sets can even weight as waste to be contained or to be transferred in order to make room for new information and communication needs.
- cause problems in the community with respect to the common levels of "privacy";
- disable the current strategies that rely on the use of digital information.

Finally, if the graphs are looked together, it is possible to see some harmony and critical correlation. Moore's Law, in fact, defines the speed at which technology increases and highlighted a natural limit while, in theory, data expansion is not limited.

VISUALIZING FOR LEARNING

This article focuses on the analytical support of visualizing digital information for urban planning, however, research groups interested in BD can be found in different branches of Marketing, Business Intelligence. "Discover the world through our eyes," wrote Few[8], one of the most famous scientists in the field of Visual Communication and Data Visualization. Visual communication has evolved as a discipline, becoming more sophisticated. For example, the forms of representation for graphs have been multiplied, because graphs are usually more useful to understand trends and exceptions, unlike tables that only show lists of specific values. The graphs can also create relationships between values, assigning size, shape, color, and now also dynamism. All this affects the way people perceive

8 He is director of *Data Visualization Research Lab* at New Hampshire University.

and understand information. Nearly 70% of human body receptors reside in the eyes. The visual perception brings information directly to our brain. Colin Ware[9] explains that "*The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers …the perception and cognition are so closely interrelated that words such as 'know' and 'see' are synonymous*". The visual representation of data can make them clear or even invisible. It is necessary to understand how perception works and structuring proper display, this enable us to enhance communication and learning. Data Communication and Visualization, or better perception studies are becoming increasingly popular with resects to the tangible support to all sectors of the economy. Data visualization is a research topic in all disciplines as long as we all need to understand information and data. Above all, the recent augmented knowledge provided by Data Visualization tools is remarkable. Tools allow us to visually represent data and information, but also to interact with different forms of display, to change the format of their display, to enter into lower levels of data details, to filter what it is not relevant, to highlight subsets of data through multiple and simultaneous charts.

Initially, while geographers focused on the description of complex processes, today, the visualization is keen to support analysis and to induce a better understanding of complex systems (inductive learning[10]), offering sophisticated tools that allow to explore current and past patterns in real-time (e.g. traffic data in real time allow to trace patterns of congestion, route preferences during the weekdays or holidays, etc. …)

MOST EFFECTIVE VISUALIZATION TECHNIQUES

Several tools and techniques for viewing complex data were evaluated. The research work carried out with collaboration of CSI Piemonte aimed at writing a white book for the exploitation of local BD in planning practices. In particular, three tools were empirically studied:

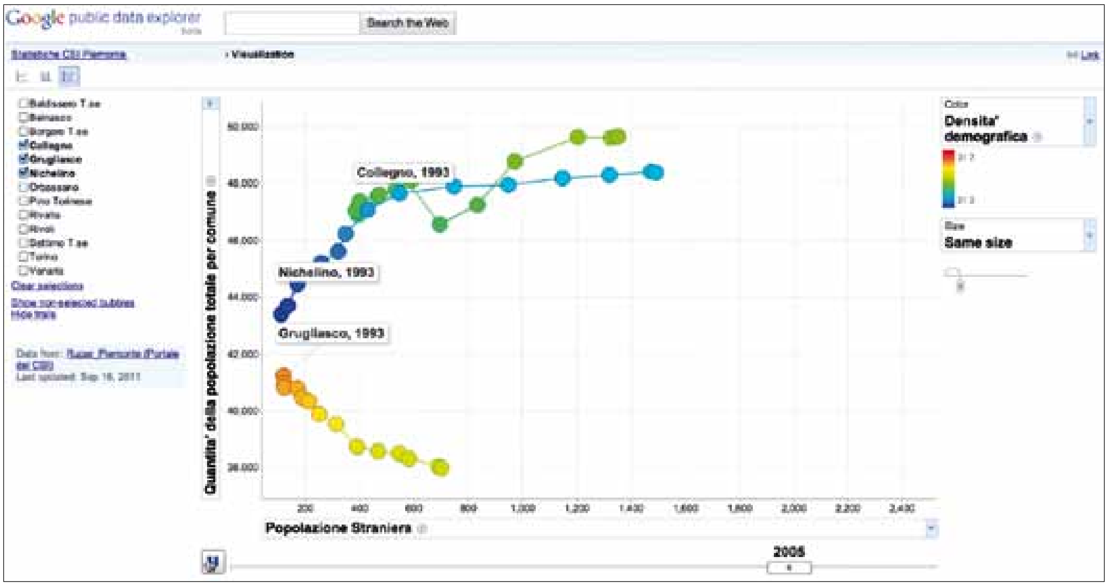
- Google public data Explorer

It is a tool for viewing large datasets, designed by Google. It has three views: a line chart, an animated bar chart and an animated bubble chart. The display format lets the user select one of the database indicators (knowledge of DSPL is required) and choose one of the available chart (line graph, dynamic histogram and animated bubble chart). The time series allow the comparison between the indicators and colors can be customized according to the needs of the users.

9 He is author of three books on data visualization, consultant for sounding companies as Apple, Microsoft, Siemens, Tableau Software and eBay. He has been lecturer at Harvard, Stanford and Oxford.

10 Roccasalva G., et al. "Visu-an-alyse indicators of urban quality"

Figure 5 View of Foreign citizens within the Metropolitan Area of Turin.



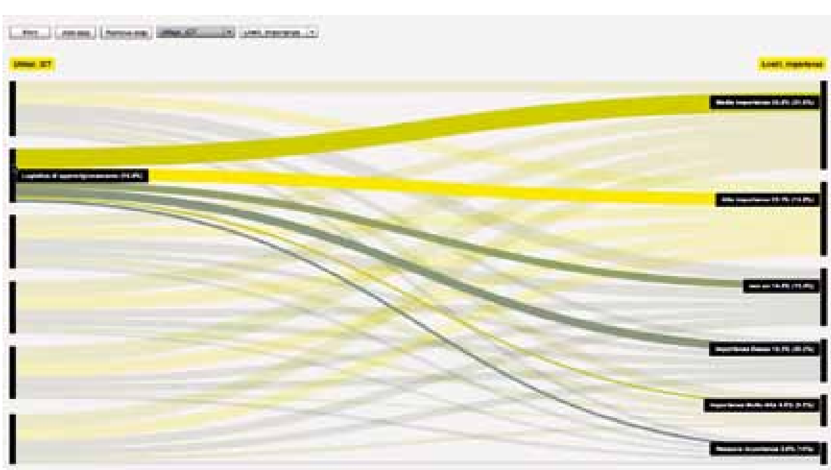
- ManyEyes

This is a viewer that crosses pairs of data and allows to create and share graphs of data. It was designed by IBM. This is a useful tool for experiments on large and even unstructured amounts of data. This tool is able to handle large amounts of data in real time with high-quality graphics. In addition it is possible to produce graphs with drill down query of the taxonomy. Above all, a third dimension can be achieved by choosing colors.

- Fineo

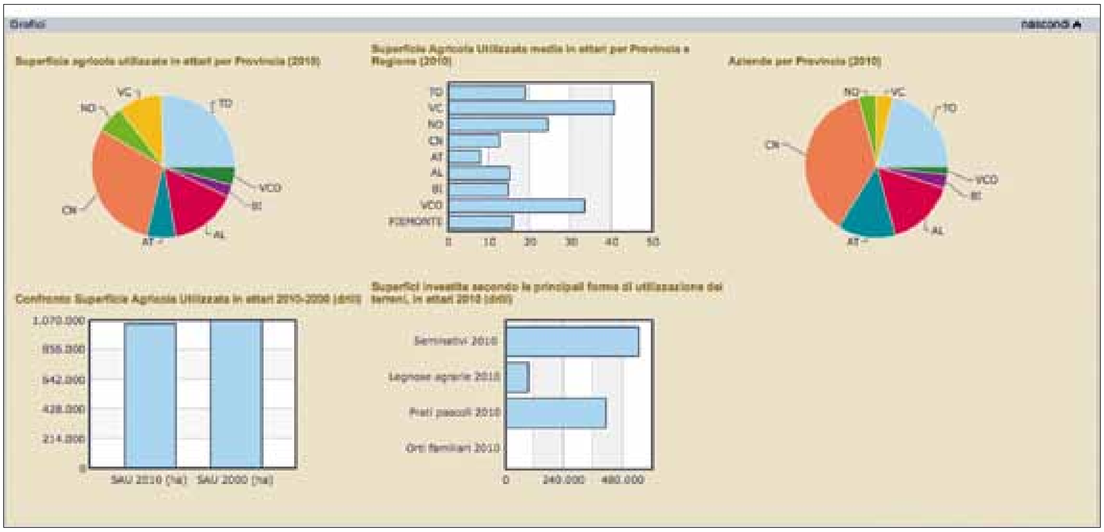
It is a web application that displays continuous streams of data (energy, money, traffic) in order to represent the relationship of the taxonomy in terms of size. This innovative application was produced by Density Design, a research center at the Polytechnic of Milan. The aim was to boost the knowledge with new forms of visualization. The percentages are self-distributed with respect of the implemented dataset.

Figure 6 Use flows of ICT within different economic sectors in Piemonte Region.



Among the most common visualization techniques, the dashboard combines on a single screen different information which are useful to monitor an aspect. If it is properly used, the dashboard offers a very effective picture of what happens in a monitored process. Rather than traditional visulization techniques, it gives a more awared information. The dashboard is a visualization system designed to be easily understandable for a management system. Dashboards make reports more effective for business programs and can be valuable tools for managers and decision-makers.

Figure 7 A dashboard example which was produced by CSI Piemonte.



The failure-success of a dashboard could be summed up by few basic rules[11]. The dashboards are useless when:

- Leave space to interpretation or display too aggregated data (if we provide index rather than an indicator);
- Let the user read too many data at the same time;
- Are produced by analysts who read the data but who do not live-work in the context in which the data are produced;
- Are developed by people who have no experience of the company.

On the contrary, a good dashboard must:

- Have at least one chart that shows the trend of the metric;
- Interpret trends and provide context to the studied dynamic;
- Explain the impacts to induce choices;
- Recommend actions to be taken and the steps to be followed.

OLAP (Online Analytical Processing)[12] is another interesting analytic tool. It is a system that collects a series of techniques for fast and interactive analysis of large amounts of data. Companies use this tool for studing the results of sales or the performance of purchase costs of goods, for marketing, for measuring the success of an advertising campaign or for testing a survey's output. An OLAP database is a photograph of information (such as a community database) at a given time, it works by transforming these individual pieces of information into multidimensional data. For example, a

11 Source is liberally excerpted from web communication studies of Prof. Leonardo Bellini, La Sapienza University, Rome.
12 Source: "The OLAP report" or "The BI verdict", edited by Nigel Pendse

customer's database of waste collection can be grouped by city, province, region; these customers can be classified by needs, area or district and every request (even exceptional) may be grouped by category. OLAP is the combination of all possible aggregations of data, this could potentially answer to every single work scheme related to waste collection services. An OLAP system allows the study of large amounts of data in order to see them from different perspectives and to help decision making processes. In particular an analyst can:

- Analyze the total of a dataset according to different dimensions and cross combinations (this function is called "slicing");
- Focus and select a sector of data which are of particular interest (this function is called "dicing");
- Break the structure of a dataset by its determinants within the same hierarchy (this function is called "drill-down");
- Break the structure of a dataset by its determinants through a hierarchy (this function is called "drill-across").

Another communication technique is the mostly fashion geo-spatial visualization. Today, companies need to monitor and understand information which are related to geographical locations (logistic, transport, environmental and so on). Often, services, sales or events can only be understood if we can see where they actually occur or where there is a demand. For example, the ability to see the trend of sales revenues on maps can support interpretation with a critical dimension.

Finally, among the analyzed tools, the animations function (the movement of part of the graphs) is commonly used to show the changes of a data over time or to compare two different variables. Some of the best works on dynamic charts have been developed by Rosling and his foundation, Gapminder. In 2006, the software Trendalyzer was produced, an application that lets you create interactive and dynamic charts; it neglected the static components which are in all other types of graphics. In March 2007, Google bought Trendalyzer from Gapminder, recognizing the value and efficiency of the work of the researchers who have contributed to the construction of such software.

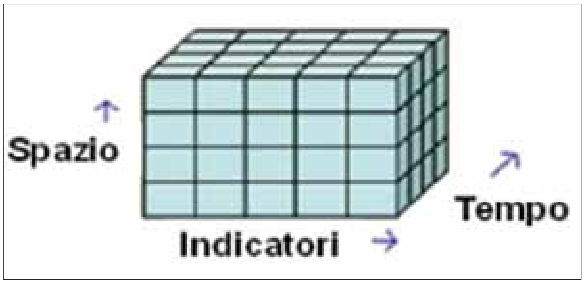
THE OPERATED TOOL

Among the applications experienced in this research, Open Explorer is one of the most complete and developed tool that includes multi-level scientific visualization, whose uses range from urban planning to the physics or medicine. It was developed at Norrkoping University (SE) by the team of Mikael Jern. This tool meets the needs of inductive learning (Roccasalva 2005) which is also the key for selecting the applications described in this article. In addition, this tool symplifies the implementation and customization of its analyses, thanks to an on-line translator of data (vislet) which is automatic managed by the group NcomVA (Norrkoping communication and Visual analytics). This makes the entered data ready to be used. Jern's studies have been recognized by international awards and by the sounding clients who adopted his patents (ISTAT, EUROSTAT, Worldbank, European Commission, Province of Kensington and Chelsea, The Economist).

Open Explorer is primarily a display of geographic data and metadata that can be put in or come from PSI in one of the most common formats (PC-Axis SDMX eXplorer Unicode). Data file can be made using M. Excel or similar, with a coherent structured relative to the position and associated metadata. The animations are available with the use of Adobe Flash Player. The layout is customizable and it is possible to have either one graph per page or up to three linked views. From a perceptual point of view, the main user interface is divided into four boxes which display different visualization techniques to communicate the data. The interface consists of a thematic map, a scatter plot, an histogram (or graph with parallel coordinates) and the area of content browser within which there is room for the so-called story-telling (or for users'opinions and ideas).

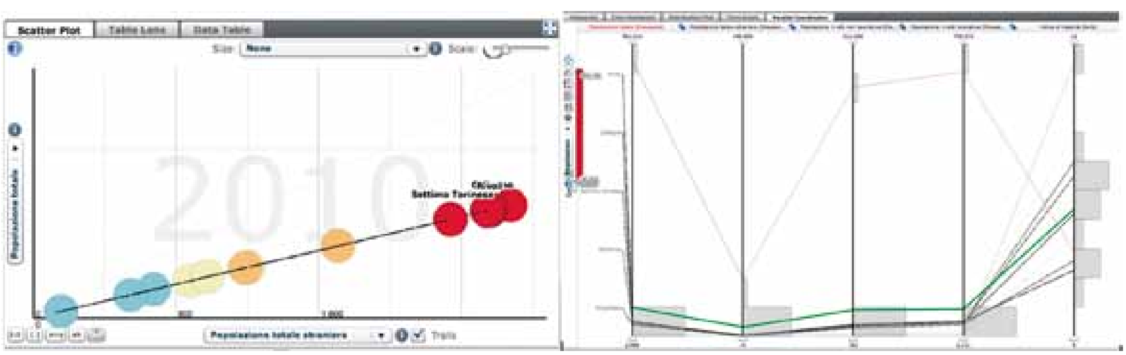
Open Explorer is based on GAV Flash, which is a conceptual model of data management optimized for space-time and multivariate indicators. This model can be represented by a data cube with the following three dimensions: space (city-regions), time (time series) and indicators.

Figure 8 Conceptual model of the data cube applied to Open eXplorer.



Open Explorer allows to view simultaneously data both in spatial and temporal dimensions thanks to a multiple linked views. The “play” bar, below the four boxes, starts the time component and it allows a dynamic display of the selected indicators simultaneously on the tematic map box, the histogram box and the scatter plot box.The navigation tools are also tools for the analysis. The Scatter Plot is a graph that allows the simultaneous display of four dimensions on the same territory; one dimension is the color of the scatterplot forms (boubble), two axes (X-axis, Y-axis) and a relative size of boubble.

Figure 9 Scatter plot example with evidence of traces about the selected data; coordinate parallel graph (PC).



The Parallel Coordinates graph (PC) displays different dimensions with lines and histograms producing an alternative view to what is shown on thematic map and scatter plot. PC can show the data evolution over time in a dynamic way, drawing attention to the average of a phenomenon by each selected indicator. PC is a geo-visualization technique used to identify trends, clusters and outliers; this supports in-depth analysis of the existing relations between the context environment and non-geographical information.

Thematic map allows the user to create effective visualizations of spatial data and to compare historical data and geographic information dynamically. It is possible to assign a desired color scale to represent the indicator on the map by selecting one of the eight custom combinations. The scale can also be dynamically adjusted by moving the cursor within the limit of the interval or by clicking on a single label, and then defining the desired value.

It is clear that this tool allows multiple views of the same data distribution; as written above, this makes visual perception critically important as long as the same sequence of information could become more or less understandable with respect to the communication structure that was chosen.

A CASE STUDY

In order to get a picture of the tools aptness, some experimental cases were studied, which have used interactive methods of display such as those described before. For example, in Italy, the National Institute for Statistical Survey (ISTAT) in collaboration with the Department for Development and Economic Cohesion (DPS) has made available large data for visual analytic experiments with Open Exporer. According to Jern, the project has allowed to communicate the main statistical data on a regional basis throughout the Italian territory. Moreover, a consulting company publically owned has bought some of the visualization tools described above for studying a new communication structure of economic data. This project aims at supporting the work of the Italian Ministry of Economic Development.

Next a case study will be introduced where these tools were applied to a new consortium of public administrations in the northern side area of Turin (Union of Municipalities to the north-east from Turin). The latter is a group of six Municipalities which have got about 120.000 inhabitants in an area of 140 km^2. The municipalities signed a protocol as it is suggested by Italian legislation (Art. 32, D. Lgs. 18/08/2000, N. 267), in order to provide more efficient service management and more effective choices concerning their policies.

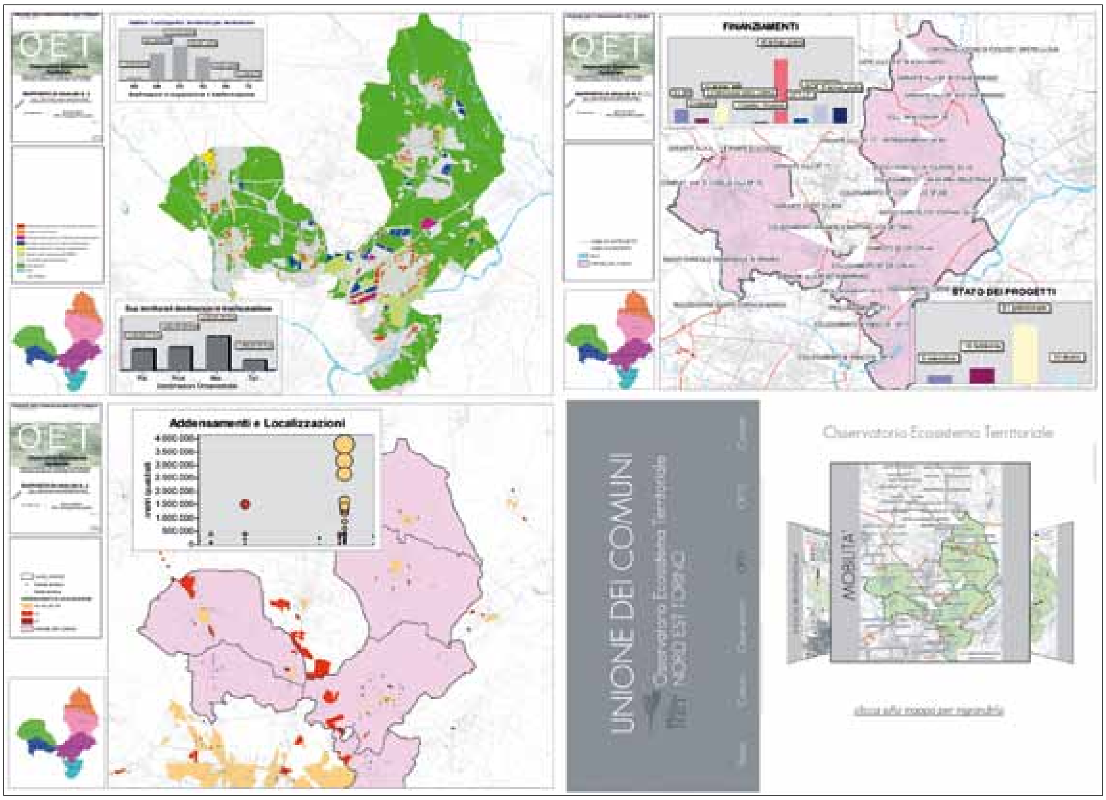
In March 2012, one of the consortium members commissioned a study about the different sources of available data. The study aimed at building a common knowledge of information for the new territorial boundaries. The study lasted six months, it worked on analytic readings and visual applications. The main objective was to build a digital communication system in order to know, understand and disseminate BD geographical information in a simple and intuitive way, to induce economies of scale for the community and for public services. The community has a good demand for information and internet communication by having a free Wifi system which covers the core areas of the six municipalities. Therefore, the demand curve is potentially wide and varied, ranging from a demand of generic information for citizens, a more specific demand from public activities and an under-exploited demand for private initiatives. However, the public offices do not have consistent and widespread knowledge about the use of the data, the management and maintenance of structured levels of geographic information. In addition, the different quantity and quality of data were hard to filter and be made coherently similar for the geographic area of the experiment. Therefore, initially, less detailed and more coherent data sources were selected, such as those which are monitored by comprehensive public bodies. The CSI Piemonte has joined the research process providing the initial data regarding the demographic composition of the main cities around Turin. The CSI Piemonte owns and collects a huge amount of data on behalf of its public clients (1300 alphanumeric databases, 1400 spatial databases and 160 databases supporting local Administrations).

Firstly, it was decided to structure specific maps that already contain processed information, and then it was chosen which simple data can be useful for the display and to study the demand coming from regular public practices (authorization work, analysis and planning studies). The first products of this research were so- called “expert maps”; these were produced on a GIS basis and were aimed at making order, assembling geographic information for all the municipalities. Common spatial analysis techniques were used to aggregate indices to discriminate the present situation from the future transformation choices, to notice relevant information that could induce the official political debates. Three main topics were investigated (see Figure n. 10):

- the master plan choices, distinguishing between consolidated and transformation areas and functions
- the infrastructure, distinguishing between completed and planned roads
- the amount of commercial areas currently available and licensed.

In particular, in the urban transformation topic, the authorized use and the level of transformation (consolidated, moderate and consistent transformation) were distinguished. About the infrastructure topic, the state of the art and the financing status of road in transformation were discriminated. As for the third topic, commercial activities were outlined by position, quantity of square meters, according to clustering categories and between the already existing ones and those under construction.

Figure 10 Geo-analytic maps (Expert maps) with data about three topics which were initially requested by the study. Clock work order, analysis of masterplan choices (surface per function), analysis about licensed commercial areas (clusters, surface per commercial category), analysis of infrastructures (divided by type of road, level of design process, type of funding). The image on the right low side is a screen shot of the web site which was published for disseminating expert maps.



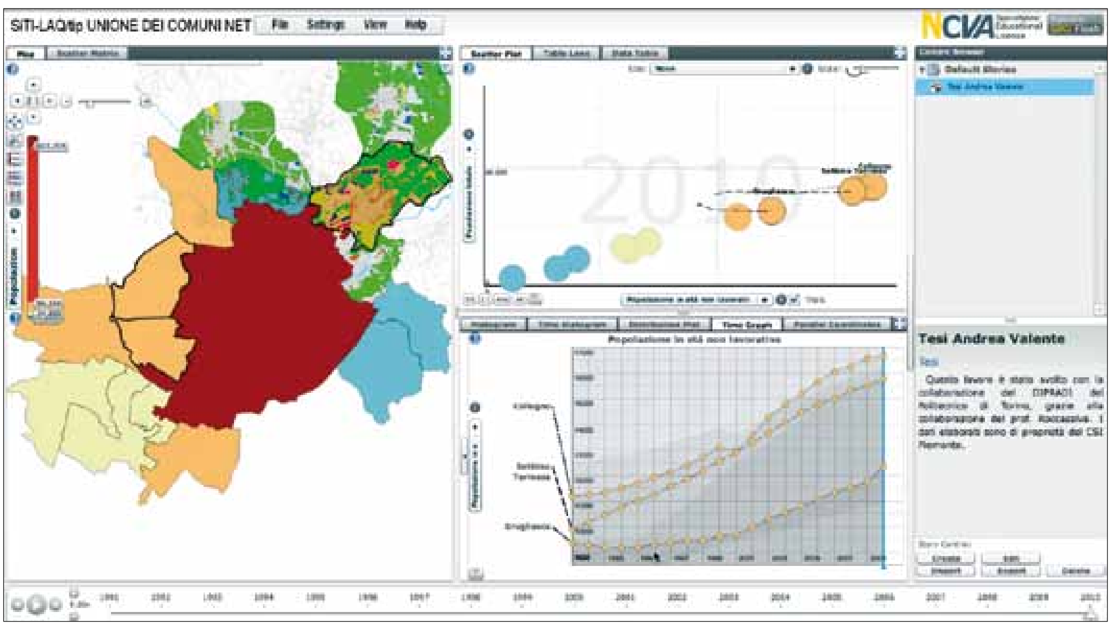
The expert maps were published on a web page that allows the visualization and dissemination of contents, while maintaining some of the features for analytic visualization (see Figure n.10). At the same time, the municipalities experts were provided with a first draft of data visualization system made with the Open Explorer tool (see Figure n.11). The observation made about the way the tool was used led the study to simplify and adjust the data visualized, in order to support actions and to draw some attention about future choices. The main decisions which were taken by this study are summarized as follows:

- A limited number of indicators out of those open information and available data has been offered, because it was initially difficult to interactive visually with more than seven categories of data. In this regard, statistical data, although well-known, were useful to easily trigger a dialogue between technical and social issues, widening the levels of induced knowledge with simple mix of geographical basic data. For example, the rate of unemployment, ethnicity or education are cross-cutting themes which induce better or implicit deductions if they are placed in a common and specific neighborhood.
- Interactive and searchable data were limited to statistical information, leaving the analyses made

with the expert maps as the background of the display. In this light, the processed geographic data were taken as given condition and the users could focus on eventual local mistakes with the information. In fact, the common user does not interact with the status of infrastructure (partially finished or still under construction), or with new supermarkets and new residential areas, but he rather can propose accuracy of information.

- Observing the way the tool was used, a small tutorial was produced that explains the basic forms of interaction. This choice was necessary due to consistent levels of initial dived with the tool, especially with those important features as the views of the synchronic data. It was noted that users limited their observation to the expert maps and automatically represented indicators without using the basic functions of the display, or proposing interpolations, or simply watching a time series of an indicator in a specific portion of the territory.
- The user's comments and requests were gathered. The tool is not self-explaining for either entering data or managing data remotely. The user must be initially supported with both direct or virtual aid. Some of these support functions could be automated because of their recurrence.

Figure 11 Screen shot of geo-visual and analytic tools used for the case study



It was observed that, among the users behaviors, there was not a genuine search for hidden meanings in the data but rather a will to classify. The common user was keen to distinguish his local area of work or living by combining information in coherent visual way. Although the users had initially a limited number of indicators, a large amount of customization of views were produced. In addition, the "story-telling" was among the most appreciated and utilized functions. It is important to let the user discuss the displayed data because it tells the level of acquaintance with the topic and with the display method. The user needs to put into context what he understood or discovered in order to make robust opinion also by writing this to other people. The longer stories are written by analysts who induce new and critical thoughts. It is our belief that story-telling function is useful not only to strengthen users understanding, but also for a better dissemination of statistical information.

The next goal will be to test the tool with a wider audience through more sensitive themes and indicators. A more interactive way of entering and correcting the initial data will be encouraged, by widening and improving geographical BD. Moreover, it will be important to study the effects of an increased level of accessibility to technical information, in order to assess profitability with a curve that links the available data to the recurrent demand of information.

DISCUSSION AND CONCLUSION

The empirical case, although producing modest results with respect to the effect on local choices, gave several hints on how to use visual analytics tools between the data and the request for information. The case study can be valuable for all those consortia (public and private) that want to try to benefit from their available information sources. In this article, it was chosen not to discuss the contents of maps and the related technical opinions, in order to focus on the scientific role of the user interaction with these analytical tools. The activities and choices made by the municipalities as a result of interactive views of data are not as important as the way the tool was adjusted and implemented, which were the alternative geo-spatial visualization, how to gather the open knowledge of the available data, and the use of interactive story-telling.

In this regard, the scholars who have discussed this research process suggested to proceed in the future with a special focus on:

Expanding the role of visualization:

- It is important to investigate and design innovative methods and tools to visualize, explore, encourage, submit, publish and share data (open data, public or expert made data for dissemination). In the future, the BD display will no longer have the task of describing information (it will not be possible or necessary) but it will have to study the complexity and/or to accept uncertainty[13];
- It will be useful to adapt visual analytic tools to be used on-line in order to reach more users and gain a greater amount of data.

Improving the analytic work of social phenomena:

- It may be important to increase the chances of achieving information and feedback in real time, in order to produce choices which are effective, shared and stimulating. For example, municipalities are more frequently having analytical views of public transport data where users and analysts help to reduce congestion, to understand and plan routes and to link routes to preference;
- It is crucial to enhance the story-telling function. Visual analytics researches pay a great attention to exploration of data rather than on developing tools which support the presentation of the acquired knowledge. Jern's studies suggested a way to record and store user's digital interactions during the analytical visualization process. It is proposed to store interactive views during the user's activities in order to facilitate the process.

Today, the intelligent use of BD is a research topic mostly influential for single person or communities in terms of economic growth and territorial competitiveness. Public Administrations (PA) are the largest generators, collectors and users of data and digital information. The PA possess data on different aspects of the economy, monitoring statistics on business licenses, on population trend over time (birth, death, marriage).

Since 2009, citizens and non-governmental organizations around the world are increasingly asking

13 Roccasalva G., et al (2012), The Future of Cities and Regions: simulation, scenario, Governance and Scale, Springer Geography.

both at national and local level for a free access to information and data (Open Data). Initially, the President of the United States, Barack Obama, started answering this question by putting on the internet all public government data. In Europe, countries such as Sweden have worked for turning from a service economy to a digital economy. In the law prospective, the Directive 2003/98/EC of the European Parliament has defined standard and minimum rules for Open Data (OD) owned by PA, with respect to conditions of transparency, re-use and non-discrimination among the users. In Italy the implementation of the European Directive was slow, but in the end it definitely started with the national law 96/2010 and a first operative regional law which was published together with the first portal of Open Data (OD)[14]. In the last years, OD have developed applications that brought benefits to the community through costs reductions in public services, new business opportunities for the private sector, but also enhancing consumers' choices by making them more aware of competitive prices (petrol prices, insurance etc ...). The OD are part of the "open knowledge" that we all are contributing to build daily. It is likely that "digital natives" will no longer have obstacles in getting information, just because they will be more harmonized with technology than current public and private decision-makers. However, the way to critically understand information is depending on how to access the data. The data accessibility is the new frontier of multidisciplinary work of social sciences and informatics scholars, territorial and economic analysts but also administrators and citizens. Good accessibility allows us to build tools, reports, programs, but also to express choices and make changes. The accessibility of public sector OD is operatively difficult, as the experiments described in this article have demonstrated, especially with regards to the various forms of demand and present equipment of interactive visualization tools. In order to define information really "open" it is not enough to just put data on internet; Berners-Lee[15] writes that OD must be structured to be able to be updated, exportable, and above all it must be recognized by other data through an appropriate link with other sources. The tools which reach the level of ability described in this article can generate a significant amount of direct and indirect benefits to the economy, industry and territory.

14 May 2010, opening of www.dati.piemonte.it

15 He coined the term "World Wide Web" and invented HTML hyper language.

Acknowledgements

I want to highlight that the initial collaboration on the visualization tools was made with Giuliana Bonello from CSI Piemonte through the first level thesis work of Andrea Valente. Most of all, I want to thank Professor Mikael for his suggestions and for collaborating in the debate about this research process.

References

Baack, S.,(2011) *A new style of News Reporting: Wikileaks and Data-driven Journalism*, Year 2011.

Batty M.,(2012) *Editorial, Environment and Planning B: Planning and Design*, volume 39, pages 413 – 415

Bellini L., *Le guide di DML: marketing automation,* Digital Marketing Lab, Milano

Corsico, F. Roccasalva, G., (2005) *Visu-an-alyse indicators of urban quality. The crucial role of forecasting Scenarios in sustainable decision making processes,* International Conference for Integrating Knowledge and Practice, Life in The urban Landscape, Gothenburg

Few, S.,(2007) *Data visualization, past, present and future.,* Cognos Innovation Center, January 2007.

Few, S.,(2010), *Visual communication,* IBM Cognos Innovation Center, September 2006.

Jern M., et al, (2011), *A web-enabled visualization toolkit for geovisual analytics,* in *Information Visualization,* pag. 1-21 SAGE

Jossa, B., (2005) *Macroeconomia*, Padova, Cedam

O'Sullivan A., (2011) *Urban Economics* Kindle Edition

Pendse N., (2001) The BI Verdict o OLAP report, Business Application Research Center

Roccasalva G., et al (2012), *The Future of Cities and Regions: simulation, scenario, Governance and Scale,* Springer Geography

Schilling M. A. (2009), *Gestione dell'innovazione*, McGraw-Hill, Milano

Ware, C., (2004) *Information visualization: perception for desing*, Morgan Kaufmann Publishers, Second Edition

Ware, C. (2008), *Visual Thinking for Design*, Morgan Kaufman.

MAGAZINE AND REPORTS

CSI Piemonte (2009), *Bilancio Sociale 2009*.

Il Sole 24ore, (maggio 2011), *Il dato è tratto*

Il Sole 24ore, (maggio 2011) *Il Big Bang dei dati*.

McKinsey Global Institute (Report 2011) Big data: The next frontier for innovation, competition, and productivity

The Economist (2010) *Data, data everywhere*. (A special report on managing information).

WEB BIBLIOGRAPHY

Information site on theGoogle Explorer application URL:http://www.google.com/publicdata/home .

First visualization experiment for Italian databases on the basis of InfoVis technology. URL: http://www.ncomva.se/flash/projects/italy/

Information site on manyEyes application, URL: http://www-958.ibm.com/software/data/cognos/manyeys/

Information site on Fineo application, URL: http://www.densitydesign.org/

Site for general statistical data of CSI Piemonte URL: http://www.densitydesign.org/

Site of the company producing Open eXplorer URL:http://www.ncomva.org

Site of McKinsley, URL:http://www.mckinsey.com/