

A new dimension in urban planning: the Big Data as a source for shared indicators of discomfort.

Paolo Scattoni

Associate Professor

Dipartimento di Pianificazione, Design, Tecnologia dell'Architettura - *Sapienza* Università di Roma
Via Flaminia, 72 - 00196 - Rome, Italy – Paolo.Scattoni@uniroma1.it

Roberta Lazzarotti

Master ACT Lecturer

Dipartimento di Pianificazione, Design, Tecnologia dell'Architettura - *Sapienza* Università di Roma

Marco Lombardi

MA Student

Sapienza Università di Roma

Andrea Raffaele Neri

Lecturer in Urban Planning and Management

Ethiopian Institute of Technology, Mekelle University, Department of Architecture and Urban Planning

Roberto Turi

Research Fellow

Italia Lavoro

Jesus A. Zambrano Verratti

MSc Student

Dipartimento di Pianificazione, Design, Tecnologia dell'Architettura - *Sapienza* Università di Roma

ABSTRACT

The web has been used for years as a means of expression for the local communities in highlighting their problems, needs and hopes, often in the form of organized group discussions and fora. The enormous amount of information currently available, Big Data, is already used for business purposes in the private sector, but has never been truly available to decision makers who operate in urban planning and would represent an invaluable help for those communities that undertake the path of self-construction of their Community Strategic Frameworks. This paper elaborates methodological and operational proposals to identify sequences of words and common occurrences in sets of documents that would help understanding the problems of the communities on a geographically-located basis, creating the search engine “Social Debate” and devising new indicators for indices of disadvantage. Such tool could drastically change the perspective of public participation and planning practice and improve the quality of local public policies and decision making processes.

1. INTRODUCTION

This article will consider the potentials for the use of Big Data in urban planning in Italy looking at the measurement of social deprivation, disadvantage and discomfort from data derived mainly from social media.

In this paper for Big Data we intend large unstructured data set derived from internet space that cannot be processed through traditional methods.

The work presented here is the further development of a research conducted for a contest on “Producing official statistics with Big Data” launched jointly by Google and ISTAT (Italian National Institute of Statistics - Istituto Nazionale di Statistica) in December 2013¹. Therefore main references for the proposal presented in this paper are to the two promoting institutions; anyway the same proposal is valid for any official statistics agency collaborating with any Big Data provider.

1 In: <http://www.istat.it/en/archive/107117> (last retrieved: 16/09/2014).

While the phenomenon of Big Data has found several interesting applications in the private sectors in the last few years it is still almost absent from central and local governments. In particular urban and environmental planning seem to be unaffected. The ISTAT-Google contest itself is a sign of interest for the use of Big Data in government (Giovannini, 2014). As far as planning is concerned the paper looks at big data analysis as a tool to verify the relevance of the problems taken into consideration within a participatory planning process.

The growth of the information in the World Wide Web is impressive (Cosenza, 2012). The data produced in the web in 2010 were 800 exabyte (billions of megabytes) and in 2011 1.8 zettabyte (trillions of gigabytes). In the last few years a specific economic activity (market research) based on the use of Big Data has been developed. It is generally accepted that Big Data will change drastically the economy and the ways of production (Mayer-Schönberger and Cukier, 2013).

Our research proposition is that spatially distributed data, measuring the subjective perception of the social “discomfort” can be very beneficial for a specific planning practice where the local community itself has set up a Community Strategic Framework (CSF), that is a planning tool built according the rules of the Strategic Choice Approach (Friend, Jessop; 1969) and maintained independently from local governments, but to be taken into consideration when urban plans and projects are required.

For such a purpose techniques of “data mining” and “text mining” (see par 2.3) will be taken into consideration and be compared and the most appropriate forms of output for use in planning will be considered.

There are two possible outputs. On one side a set of possible fixed indicators of the perceived discomfort might be produced at regular times, at different scales (National, Regional, Municipal and local. There is then the possibility for ad hoc exploration using existing tools or even better using a specific search engine for the social debate data to be produced by one of the main www database owners (e.g. Google).

The paper is structured as follows. Section 2 will present a background to the state of the art on the deprivation analysis and how it has been used in Italian planning. In the same section, the nature and potential of textual analysis are presented, in particular the textual analysis of social media is

seen as the key factor for providing information on perceived social deprivation. Section 3 outlines a proposal for a search engine on the social debate, including the possibility to produce official statistics derived from Big Data; this could have a huge impact on urban planning. The new perspective is that local communities in Italy, as well as in most representative democracies, could have the ability to produce their own planning tools (e.g. a Community Strategic Framework) cheaply and effectively. Such tools then could be used to be related to plans, programmes and projects proposed by Local Governments with better awareness and consequently the whole planning process would be more effective.

The final proposed strategy is organised in three different steps. The first step considers the possibility of spatially referenced data derived from social media. The second step implies the development of a search engine specialized in social media contents, in order to detect word patterns related to specific geographic area problems. The third step concerns the setting up of an observatory of urban discomfort where a set of predetermined indicators are continuously monitored at different spatial levels (census areas, municipalities, regions, etc).

2. BACKGROUND

2.1 Measuring deprivation: the state of the art

The debate on how to enhance the process of assessing the deprivation of a community, a city, a region or a nation as a whole is not new. In many countries the traditional measures solely based on quantitative data, have been questioned in the last decades, giving way to alternative indices that have been developed and implemented with positive social and political effects on the countries involved. Even in a country such as Italy, which lagged behind in the international debate and practice on the subject, a number of interesting local experiments have been developed, but have had difficulty in reaching the social and the political debates, not having been sponsored by the decision makers or effectively used as decision-making tools (e.g. the Index of Deprivation for Geographical Analysis of inequalities

in Mortality, or Index Cadum, 1999; the Index of Deprivation, or Index Caranci, 2008; Material Indices of Deprivation, Social Deprivation and Disadvantaged Areas of the Sardinia Region, 2006; Indices of Deprivation and Disadvantaged Areas of the Abruzzo Region, 2001; the Index of Multiple Deprivation of the Savona Province, 2005; the Index of Deprivation for the City of Genoa (Ivaldi and Busi, 2004).

In 2013 ISTAT, with the collaboration of the Consiglio Nazionale dell'Economia (CNEL, 2012), published its measure of equitable and sustainable well-being, (the BES: Benessere Equo e Sostenibile), an ambitious and highly comprehensive index of well-being aiming to summarise and expand the huge amount of data available concerning the degree of well-being (and conversely deprivation) of communities. The well-being of the country is described by means of 12 domains (Health, Education and Training, Employment and time for well-being, Financial well-being, Social relationships, Politics and Institutions, Security, Subjective well-being, Landscape and cultural heritage, Environment, Research and innovation, Services quality), investigated by means of 134 indicators. However BES has important limitations of scale, describing well-being in Italy only at the regional level even though, as demonstrated by the British Indices and policies based on them, the most appropriate scale for assessing it should be that of the neighbourhood, easily identifiable by census output areas (Neri, 2012).

The traditional method of data acquisition for intrinsically unquantifiable information such as the degree of well-being or deprivation puts in place the difficult, uncertain and costly practice of assuming that certain quantifiable data have a positive relationship with the well-being of a community (for example, that a greater income, combined with a low incidence of admissions to hospitals and a high level of education necessarily leads to greater well-being). This operation does not provide certainty about the results, but was implemented in British countries at the census scale unit, giving rise to the implementation of Indices of Multiple Deprivation, which represented an important tool available to decision makers in the local government processes, for example with regard to the distribution of funds by public bodies.

Our proposal is based on the following consideration: in a country where access to the internet is now virtually guaranteed to the entire population, in

addition to traditional methods of data acquisition, ISTAT, with the support of Google, may obtain a important amount of information of a qualitative nature, not sufficiently included at this point in time. In other words, different types of behaviour existing within communities, representative of the perceptions, expectations and fears of the population, have been not sufficiently or promptly identified and studied by traditional statistical means, but they can now be understood and mapped thanks to the relevance and quantity of data available from the internet.

Therefore, the potential to use the information arising from finding and grouping sequences of words that connect documents available on the web becomes vitally important for the production of modern indices of deprivation. Decision makers in urban planning would have the opportunity to take advantage of tools that existed in the private sector for years, allowing companies to retrieve strategic information on customers and competitors.

In the near future, in addition to being able to use census data and traditional indices of deprivation based on census data to identify patterns of socio-economic deprivation at a local scale will thus exist rankings of local-based 'hot topics' concerning the discussions on the web in specific geographical areas. This innovative element would guarantee that the identification of qualitative data which are derived not from hypothetical assumptions (for example that high-income areas have less problems than low-income ones), but from a comprehensive analysis of people's actual perception of well-being or deprivation. The following section analyses the means by which this analysis would be conducted.

2.2 Index of deprivation

Instead of the traditional quantitative approach, the research team proposes to use the approach named "textual statistics", which combines both qualitative and quantitative data analysis. As a first step, it will use the specific search engine "Google Social Debate" to find a large amount of information related to the issue of crime, or some form of social discomfort. The texts found into the comments posted on blogs will be classified by some key words (like "diffusion type" or "date" or "geo-referencing"). The aim is to obtain a corpus of data that will be analysed through a lexicon metric approach (Bolasco, 1999). The hypothesis being tested is that

qualitative material found on several blogs posted by people living in different places means that the texts are “geo-referenced”, which means that the use of a large amount of textual data allows analysis to deal with a completely new connection between territory and needs in order to define priorities for action.

Then, as a second step, the researchers move forward with the creation of a "deprivation index", combining the keywords that determine a scale of seriousness of the problems. How to build indexes in the case of Big Data (Mayer-Schönberger, Cukier, 2013), we proceed in the following way: it makes the analysis of portions of the text as phrases ("lack of respect",...), in case multiple expressions of actions that address the crime ("make a pickpocketing", "have robbed a bank", "drug trafficking") or individual words ("steal", "sell", "criminal",...); second, we proceed to extract the positive and negative mood from messages (technically these documents) and analysed by making a classification of documents according to a polarity positive, negative or mixed, to calculate the polarity of the semantic associates of the motor (high, medium, low). Finally, the process of automatic analysis of portions of the text is in two main linguistic resources: the sentiment lexicon (Feldman, 2013) which is lexicons coverage level of single words and multi-word units enriched with information about their positive or negative valence and emotion that transmit and syntactic rules, for the semantic composition of the expressions of sentiment.

2.3 Textual Statistics. Knowledge Discovery in Databases – KDD

For the construction of a deprivation perception index, there are several steps to be considered. Firstly a choice of sources must be made. Secondly the documents have to be filtered through a set of analyses according to the nature of the considered data. Finally an appropriate analysis of the occurrences will produce an index.

The sources of Big Data that will affect our deprivation perception index mostly are those contained in the so- called "Human-sourced information" (Rayson et al, 2000) and then everything that relates to the concept of "Social Media", such as Facebook, blogs and fora.

The information derived from sets of specific analysis of Big Data is known as "Knowledge Discovery and Data Mining - KDD" (Fayyad, Piatetsky-Shapiro,

Smyth, 1996), which is reflected in both computer science and statistics. So the construction of an index of perception, as previously described, it is now possible through the combined use of different strategies that take the name of "Information Mining".

In recent years, the inexhaustible scientific development in the field of computer science has seen the great development of two different methods of unstructured data storages and collections: structured and unstructured.

Such differences in the methods and forms of data storage have produced a significant crossroads in Information Mining, giving rise to two methods of analysis clearly distinguished: "Data Mining" and "Text Mining". The latter is the method that will be used for the construction of indicators of perception of discomfort.

Text Mining (Nisbet, Elder IV, Miner, 2009) has as its goal the pursuit of knowledge from large collections of documents and then from textual sources. This practice allows the identification of sequences of words that are common occurrences and characterize a set of documents, enabling the group to identify principal matters of debate.

This type of approach is therefore particularly useful when you want to analyse the contents of a collection of documents, even if they come from heterogeneous sources. This technique also allows the identification of groups of subjects. It thus allows the creation of structured archives where the information can continue to be treated by iterative cycles of analysis and with textual approaches appropriate to the level of knowledge required.

2.4 The potential role of indicators of perceived discomfort in urban planning

In which way can a general index of perception of discomfort in general and a set of indicators of such perception obtained from the web have an impact on planning? The literature does not provide relevant information on this. A quite isolated contribution has tried to open up the debate, revealing the potential of the deployment of Big Data for territorial policies (Roccasalva, 2012).

In the first half of the past century the traditional approach of a survey as the basis for the construction of planning tools has gradually evolved from being tools for the management of physical transformations of the built environment into complex tools for the management of a growing multiplicity of variables, from environmental to those of local development.

In this paper we discuss the potential of Big Data for planning that looks at perceptions of current problems and treats them in a strategic dimension this interacts with a multiplicity of stakeholders. It will then reference the approach of strategic choice as originally outlined by Friend and Jessop (1969) and its subsequent theoretical (Faludi, 1989) and practical (Friend and Hickling, 2005) developments. In Italian planning practice a relevant implementation of the approach can be found in the Grosseto structure plan (Scattoni, 2007).

Thus has emerged a planning practice that begins with identifying and understanding the problems it faces in a dimension of the "here and now" of shrinking time spans, giving rise to a dimension of continuity. In this dimension, in practice the analysis activity commensurate with the problems turned out to be more costly for the development of ad hoc surveys and at the same time less and less effective.

Most often the recognition of the critical issues to be addressed through planning is delayed with respect to their first occurrence, that is exactly when public action would be most likely to have the desired effects. From this point of view the diagnosis that emerges from both the practice of politics and that of the current processes of public administration are most often partial and incomplete. Therefore in such a context the availability of a stream of data coming from what we call "social debate" would give rise to a profound transformation of current practices .

Indeed, this information flow would provide not only an important support for the interpretative analysis of an area and its problems, but also to the construction of participatory processes. It will help in active involvement of local communities that now permeate every level of planning by allowing the focus of the discussion on the areas "hottest topics" and pre-identify the prevailing stakeholders' positions in the debate itself.

2.5 Big Data for community planning and the Community Strategic Framework

In this section the potential of Big Data for community planning will be discussed. In a research work still in progress a reversal of present processes of land use planning and management has been suggested. No longer would there citizen participation in the formation of planning tools but rather there is the self- construction by the citizens, of a Community Strategic Framework (CSF): this is a general framework in which the public can recognize and

document the problems, needs and aspirations that a local community shares.

It is also assumed that such a framework a system which is 'homebuilt' by the local community and in which the role of technical support has to remain complementary, avoiding invasions and overlapping roles. In addition, the Community Strategic Framework must always be changed and updated in relation to the requirements as these mature over time. Finally, the CSF must be organized in a form that allows the possibility of being commented on and discussed to facilitate its updating.

The reversal of the planner-community relationship is obvious: planning must necessarily confront the perceived problems at the local level in the context set by the local community, rather than on themes offered by the "expert knowledge" of planners.

It is a process that not only covers land use planning, but a varied range of tools, including most of the projects and policies derived from European funding, for which the activation of processes of participation is an essential element.

For the construction of the CSF open source software was developed that facilitates the setting up of the framework in a self-managed activity, costing practically nothing.

The CSF is formed and managed over time through forms of participation which are now widely used and well known like Local Agenda 21, public meetings, etc.

Among available tools there is also a software to record the evolution of the CSF over time so as to ensure its traceability and be accessible on the web (Scattoni, Tomassoni, 2007).

In such a context, access to data on perceived discomfort becomes a fundamental element of importance, above all, to see if and how the problems identified by a limited number of citizens are perceived by a much wider audience, represented by local users of the network.

In other words it is assumed that the Strategic Choice Approach as originally conceived (Friend, Jessop, 1969) and later developed by Openshaw and Whitehead (Openshaw, Whitehead; 1985) can be built independently by local communities without planners and validated by Big Data analysis. The assumption has not been tested yet through empirical work.

3. THREE STEPS FOR BIG DATA IN PLANNING

The operational perspectives of Big Data for planning in general and the Community Strategic Framework in particular can be approached in three different steps. The first step is activating text mining and data mining techniques already available and usable. For this purpose, a simple application is presented below. It provides origin/destination data based on the flows obtained from Twitter. The second step assumes the construction of a search engine oriented to social debate; this would be similar to specialized search engines in other sectors (e.g. Scholar for academic publications). Such search engine would allow the exploration of databases from a variety of social media. Finally, the third step involves the ability to create a set of indicators of social problems collected by a certified body with a spatially detailed reference (e.g. census areas).

3.1 Spatially distributed data

There is an incredible amount of data obtainable from the Internet, nevertheless access to this information is not always granted. One example of simple access to a database with very few limitations is Twitter. A previous work developed the possible usage of “data mining” techniques, specifically on Twitter data, for the making of origin/destination maps with no cost and little effort (Zambrano Verratti, 2014). This research focused only on “quantitative data”, especially on the spatial coordinates associated with each message posted on Twitter. In this exercise the potential of the developed technique relies on the spatially distributed data that is implicit in the origin/destination maps produced.

The problem of the “data mining” approach is the wide variety of databases from which to obtain information not only from Twitter, but from the vast World Wide Web as a whole.

In fact it is extremely difficult to use information derived from so many sources and make it easily readable for both decision makers and for the population at large in a particular geographical context. For such a purpose it would be necessary to spend a considerable amount of time and efforts to compare data from heterogeneous sources, not to mention the great amount of time and resources that would be required to carry out the mining process

multiple times or even continuously.

The hypothesis to overcome these limits is to create a unified database that filters information from many sources and centralises it in a tool that should be as easy to use as a search engine. Due to the fact that this was a research was carried out for a contest that included Google among its organizers, the proposal is based on a Google search engine.

3.2 Unified and easy to use database.

Based on the specialized search engine called Google Scholar, a new engine called Google Social Debate was proposed that would be focused on Human Generated Data (Miller, 2014) obtained from the following Social Media categories (Kaplan and Haenlein, 2010):

1. Collaborative projects (e.g. Wikipedia)
2. Blogs and microblogs (e.g. Blogspot, Twitter)
3. Content communities (e.g. Youtube, Flickr)
4. Social networking sites (e.g. Facebook)
5. Virtual game worlds (e.g. World of Warcraft)
6. Virtual social worlds (e.g. Second Life)

In order to connect some of these databases Google has already made an experiment with the “Google Search, Plus your world” (2012) and its predecessor “Google Social Search” (2009) with the search option “include my circles”, which is now an official part of “Google Plus”. Nowadays with Google Dashboard there are already many databases linked which are Google property. But also other services (e.g. Facebook) can be directly connected to the Google Account, which grants access to external databases at least partially. So gradually a unified and easy to access database is widely possible if proper permissions are granted.

Over the last few years, the support of Open Data from public administrations has increased notably. Once again, many datasets and databases have little or no linkage. The proposed Google Social Debate could include specifically this information in its search engine in order to make Open Data more reliable. Image 1 is a photomontage made to graphically illustrate what has been proposed until now.



Figure 1 – Photomontage representing the possible interface proposed for Google Social Debate.

Google Social Debate would use text mining techniques (as previously described) in order to “measure” words written by users and calculate indicators of discomfort related to pre-established categories. Even if there is no reference to take from Google Scholar for spatially distributed data, the “metrics” used on this platform (and other academic databases) give a hint on how the resulting indicators could be expressed.

In order to achieve accurate geographical representation at local level the Google Location Service could be used, which cross references publicly broadcast Wi-Fi data from wireless access points, as well as GPS and cell tower data, in order to determine the position of a user.

As far as privacy is concern, the crucial issue is the protection of personal data. The big data collectors should guarantee that the provided data remain anonymous in order to avoid the identification of a specific user. The connection between privacy and spatially distributed data, as well as the law frame in Europe², has been addressed in the research made with cell-phone data from Telecom (Calabrese, Colonna, Lovisolo, Parata, Ratti; 2010). Analogous argumentation could be used for the development of the search engine hypothesized.

2 Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

The relation between the words written by anonymous users (processed with the techniques described in paragraph 2.3) and their specific upload location would allow the calculation of the Perceived Urban Discomfort Index directly associated with a community. The resulting measurement would be not only geographically accurate but also constantly updated, given the non-stop production of data which can be collected. Image 2 shows the proposed interface for the resulting indicators, specifically for the Italian context, located at different scales: neighbourhood, zone (municipio), municipality (comune), province, region and nation. Alternative graphic representations on maps would highly improve the legibility of the information, and the scale could be much more detailed (e.g. census areas).

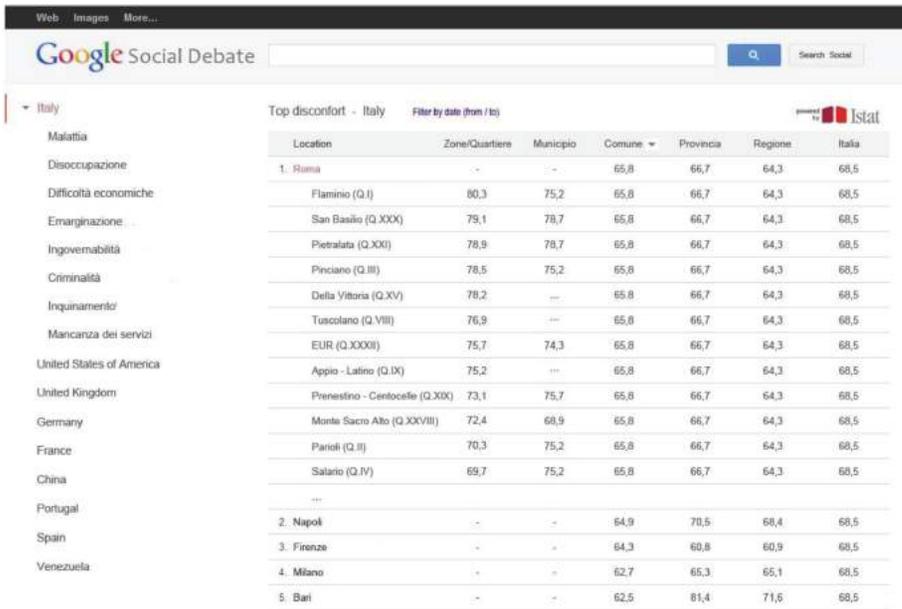


Figure 2 – Proposed interface (photomontage) for the “metrics” of Google Social Debate.

It is important to clarify that “Google Social Debate” is not a place of interaction between users, but a search engine that links to the sites where social debate is already happening. The user of this search engine will be able to have a wide view of the issues that interests him by introducing a simple query, so the participation in the discussion will be made after the redirection made by the search engine. In such context the continuous monitoring of

keywords recurrence can play an essential role for the CSF updating.

3.3. Big data and official statistics: an Observatory of Urban Discomfort

The information deriving from possibilities of interrogating social media can be helpfully utilized, among others by ISTAT, the National Institute of Statistics, either in ongoing activities or in the ISTAT BES project, which can usefully draw data and indicators from the proposed search engine; in fact, BES represents already a dataset on equitable and sustainable wellness, using 'new' sources.

A system detecting the perception of deprivation permits the extrapolation of useful indications for all BES thematic fields, but particularly for *social relationships, safety, environment and quality of public services*. It seems important to realise that this system could offer the possibility of organizing the information:

- at *spatial level*, for example extrapolating and mapping data relating to urban areas, particularly relevant for the definition of *ad hoc* policies;
- and for *specific periods*: while being aware of the changes in data bases and definitions over time, it is possible to compare the relevant phenomena at different points of time.

Immigrants and new citizens is another field of statistics that can be considered proper to the research potentialities of a social media interrogation system, especially considering the relationship between integration and life quality in settlements. Urban quality, particularly of public spaces, mobility and transport, equipment accessibility, participation in building policies and decision making are all topics variously recognized as key factors of integration in society and in towns. The possibility of measuring multidimensional social discomfort can represent a valuable tool for diagnosis and intervention.

It is possible to imagine a new activity field that can be informed by a regular measure of the online debate on the urban condition: this could be an *Observatory on Urban Discomfort*, which could enable researchers and planners to monitor one of the key topics on scientific, cultural and politic themes. The Observatory could open a new front not only in terms of data manipulation and indicator building, but mainly in the *interpretation and accumulation* of observations, thanks to the realization of a regular

reporting activity.

Finally, the Observatory could work also as a tool for measuring the impact of sectoral policies, an *in itinere* and *ex post* evaluation through the perceptions and opinions of local communities, which, crucially, have been expressed in a spontaneous and un-filtered way. Because of all previously said, it should be properly created and managed by official research institutions (ISTAT, Universities, sectoral research centres, ...).

4. CONCLUSIONS

The paper has focused on the way a general index of perception of discomfort in general and a set of indicators of such perception obtained from the web can have an impact on urban planning.

Techniques of “data mining” and “text mining” have been used for years in the private sector to investigate the Big Data for the purposes of classifying the internet users for business purposes. Contrarily the public sector largely ignores the potential of such tools, which would be easily used with little if none expense for the improvement and upgrade of the indices of deprivation or the measures of well-being (BES) already in place or proposed, which are based on processing traditional quantitative census data.

Identifying methodological sequences of words that are common occurrences and characterizing sets of documents would lead to a better understanding of the problems of the communities on a geographically-located basis, since Facebook, Twitter or fora discussions are informally, but timely used to express concerns of public relevance. This would lead to ranking the different forms of perceived discomfort into structured information for improving the local public policies and decision making for urban planning.

The paper has discussed the advantages of a proposed search engine able to focus on the “Social Debate” that could be of great importance both for improving the planning practice in the Public Administration and enabling the community to independently build its own tools like the Community Strategic Framework.

Therefore in such a context a low cost flow of data coming from the web thanks to the search engine “Social Debate” could drastically change the perspective of public participation and planning practice.

REFERENCES

- AGENZIA SANITARIA REGIONALE ABRUZZO (2011), *L'epidemiologia geografica comunale - territoriale: ambiente, qualità della vita, salute e sanità federale*.
Available at:
<http://sanitab.regione.abruzzo.it/osservatorio/l'epidemiologia%20territoriale%20d'abruzzo.pdf> [24 April 2014].
- BOLASCO S. (1999), *L'analisi multidimensionale dei dati*, Roma, Carocci.
- CADUM E. et al (1999), "Deprivazione e mortalità: un indice di deprivazione per l'analisi delle disuguaglianze su base geografica", in *Epidemiologia e Prevenzione*, 23, pp. 175-87.
- CALABRESE F., COLONNA M., LOVISOLO P., PARATA D. and RATTI C. (2011), "Real-time urban monitoring using cell phones: A case study in Rome", in *Intelligent Transportation Systems, IEEE Transactions on*, 12(1), pp. 141-151.
- CARANCI N. et al (2008), "Verso un Indice di Deprivazione a livello aggregato da utilizzare su scala nazionale: giustificazione e composizione dell'Indice 2001", *Convegno AIE Metodi e Strumenti per la Misura delle Disuguaglianze*, Roma 15-16 maggio 2008.
- CONSIGLIO NAZIONALE PER L'ECONOMIA E IL LAVORO (CNEL) e ISTITUTO NAZIONALE DI STATISTICA (ISTAT) (2012), *The BES Project*. Available at: *Misure del Benessere*, <http://www.misuredelbenessere.it/> [24 April 2014].
- COSENZA V. (2012), *La società dei dati [Kindle Version]*. Milano, MI: 40k.
Available at: *Amazon.it*.
- FALUDI A. (1987), *A Decision-centred View of Environmental Planning*, Oxford, Pergamon.
- FAYYAD U., PIATETSKY-SHAPIRO G. and SMYTH P. (1996), "From data mining to discovery in databases", in *AI magazine*, 17(3).
- FELDMAN R. (2013), *Techniques and Applications for Sentiment Analysis*. Communications of the ACM 56(4):82-89.
- FRIEND J., JESSOP N. (1969), *Local Government and Strategic Choice*, London, Tavistock.

- FRIEND J., HICKLING A. (2005), *Planning under Pressure: the strategic choice approach*, Abingdon, Routledge.
- GIOVANNINI E. (2014), *Scegliere il futuro. Conoscenza e politica al tempo dei Big Data*, Bologna, Il Mulino.
- IVALDI E., BUSI A. (2004), *An Index of Geographical Deprivation for Geographical Areas*.
Available at: *Diem Unige*, <http://www.diem.unige.it/23.pdf> [24 April 2014].
- KAPLAN A. M., HAENLEIN M. (2010), "Users of the world, unite! The challenges and opportunities of Social Media", in *Business Horizons*, 53(1), pp. 59-68.
- LILLINI R. et al (2005), "Costruzione di un indice di deprivazione socio-economica per la provincia di Savona, Istituto Nazionale di Ricerca sul Cancro".
Available at: *Ambiente Liguria*,
http://www.ambienteinliguria.it/eco3/DTS_GENERALE/20080805/4_stato%20di%20salute%20e%20deprivazione%20in%20provincia%20di%20savona.pdf [24 April 2014].
- MAYER-SHOMBERGER V., CUKIER K., (2013), *Big Data: A Revolution that Will Transform how We Live, Work, and Think*, Boston, Houghton Mifflin.
- MILLER P. (2014), *Applying big data analytics to human-generated data. Report from GIGAOM Research*.
Available at: <http://research.gigaom.com/report/applying-big-data-analytics-to-human-generated-data/> [24 July 2014].
- MINERBA L. e VACCA D. (2006), *Gli indici di deprivazione per l'analisi delle disuguaglianze tra i comuni della Sardegna*, Istituto Nazionale di Statistica.
- NERI A. R. (2012), The Importance of Indices of Multiple Deprivation for Spatial Planning and Community Regeneration, in *Italian Journal of Planning Practice*, ISSN: 2239/267X.
- NISBET R., ELDER IV J., MINER G. (2009), *Handbook of Statistical Analysis and Data Mining Applications*, Elsevier Inc., Amsterdam.
- OPENSHAW S., WHITEHEAD P. (1985), "A Monte Carlo simulation approach to solving multicriteria optimisation problems related to planmaking, evaluation, and monitoring in local planning", in *Environment and Planning B: Planning and Design*, 12(3), pp. 321-334.
- RAYSON P., EMMET L., GARSIDE R. and SAWYER P. (2000), *The REVERE*

project: experiments with the application of probabilistic NLP to systems engineering, Lancaster University, p. 3.

ROCCASALVA G. (2012), "How big data might induce learning with interactive visualization tools", in *Territorio Italia*, Agenzia del Territorio, pp. 22, ISSN: 2240-7707.

SCATTONI P. (2007), "Il piano strutturale di Grosseto e la memoria della pianificazione", in *Urbanistica*, 133, pp.63-82.

SCATTONI P. e TOMASSONI G. (2007), "Un sistema informativo per i processi decisionali della pianificazione", in *Urbanistica*, 133, pp. 68.

ZAMBRANO VERRATTI J. A. (2014), "Mapping urban flows of Buenos Aires, Mar del Plata and Caracas through Twitter location data", in *Urbanistica PVS on-line*, 2 (1).
Forthcoming: <http://urbanistica-pvs.arc.uniroma1.it/index.php/pvs-rivista>.