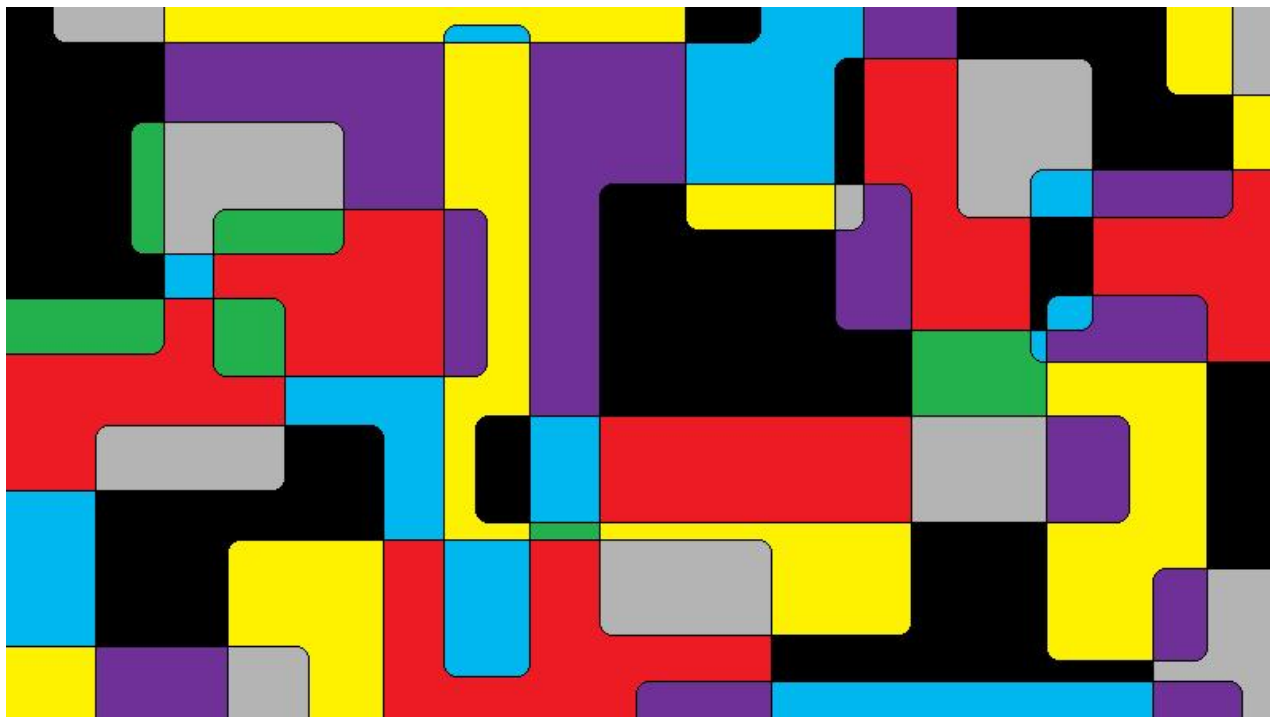


Big Data e Scienze Regionali: una relazione inesplorata

eyesreg.it/2016/big-data-e-scienze-regionali-una-relazione-inesplorata/



di: Barbara Martini

EyesReg, Vol.6, N.6, Novembre 2016

Il termine *Big Data* sta riscuotendo un notevole interesse in numerosi campi. Esso fu utilizzato per la prima volta nel 2011 dal McKinsey Global Institute e, allo stato attuale, non esiste una definizione univoca (Floridi 2012, Loverace et al., 2016; Kitchin, 2014). Le diverse definizioni oltre a identificare i Big Data con le usuali 3V: velocità (dati che possono essere ottenuti in tempo reale), volume (grande mole di dati) e varietà (dati provenienti da fonti diverse), ne sottolineano da un lato la complessità e dall'altro l'eterogeneità delle fonti. Ne consegue che i Big Data non sono "grandi" soltanto per le loro dimensioni ma anche per la possibilità di essere messi in relazione con altri dati spesso provenienti da fonti diverse (Boyd & Crawford, 2012). Nonostante il crescente interesse da parte di molte discipline le scienze sociali, ed in particolare l'economia regionale, sembrano non aver colto le potenzialità e le problematiche ad essi connesse rimanendo ancora ai margini del dibattito.

Obiettivo di questo lavoro è quello di esplorare lo stato dell'arte dell'utilizzo dei Big Data nell'ambito delle scienze sociali, e di trarre alcune considerazioni in merito alle potenzialità e ai rischi derivanti dal loro utilizzo.

Per un lungo periodo di tempo le scienze sociali hanno operato in un ambiente povero di dati. Come conseguenza le più comuni tecniche di analisi si basano sull'utilizzo di un numero di dati limitato provenienti da campionamento e insiemi di informazioni raggruppati in forma di *panel data* o di serie storiche costruite per rispondere ad una specifica domanda di ricerca. La possibilità di accedere a una grande mole di dati da un lato aumenta l'esattezza e la completezza delle misurazioni, ponendo lo studioso in grado di catturare fenomeni che prima erano difficili da indagare ma, contemporaneamente, dall'altro lato ne incrementa il livello di complessità. Questa crescente complessità richiede un cambio di paradigma che le tecniche di analisi comunemente usate nelle scienze sociali e nelle scienze regionali non hanno ancora affrontato (Einav and Levin, 2013). I *Big Data* provengono da fonti diverse, sono archiviati in basi dati differenti, non sono privi di rumore (*noise*) e sono spesso incompleti; sono pertanto dati non strutturati. In presenza di dati non strutturati il primo problema diviene quello di organizzare e strutturare i dati stessi, nonché ridurre la mole cercando di mantenere l'informazione (segnale) ed eliminando tutto ciò che non è necessario per l'analisi (rumore). Uno dei metodi più comunemente utilizzati è rappresentato dal così detto *Knowledge Discovery from Database* (KDD). Il KDD è un processo volto a scoprire informazioni utili da una grande mole di dati, e si basa sull'ipotesi che l'informazione possa essere trovata attraverso la scoperta di sentieri relazionali nascosti in una grande mole di dati, in cui la conoscenza è la conclusione di un processo *data driven*.

L'utilizzo dei Big Data e delle sue tecniche di analisi muta, in modo più profondo, l'approccio epistemologico ed il modo di fare scienza (Kitchin, 2014), passando da un modello *science driven* ad un modello *data driven*, in cui il rischio è che "i dati parlino da soli" senza che vi sia una vera teoria di supporto: la catena causale che fino ad ora ha guidato la ricerca scientifica potrebbe essere invertita. In altre parole, anziché partire dalla formulazione di ipotesi ed utilizzare i dati per sottoporle a verifica, l'utilizzo dei Big Data e le tecniche di analisi ad essi associati, potrebbero portare alla scoperta di significative associazioni tra i dati senza essere guidati da una ipotesi (Kitchin, 2014). Generalmente, l'approccio seguito nelle scienze regionali è quello di lavorare con campioni rappresentativi, costruiti per rispondere ad una specifica domanda di ricerca, che vengono generalizzati utilizzando tecniche quali l'inferenza. Nel caso di utilizzo dei *Big Data*, al contrario, si utilizza spesso la dizione "*correlation supersedes causation*" (Anderson's, 2008). L'idea sottostante questa affermazione è che i possibili *pattern* trovati all'interno di una base dati siano significativi. Questa osservazione, oltre ad essere falsa, è anche pericolosa: la correlazione tra due variabili all'interno di una base dati può essere casuale o spuria. Gli approcci basati sul KDD sono ben lungi dal proporre soluzioni automatiche. Il loro obiettivo è quello di descrivere soluzioni e scenari che potrebbero non essere catturati o percepiti. La scoperta di eventuali correlazioni, non precedentemente ipotizzate, può divenire oggetto di ulteriori analisi (Kitchin, 2014). Ne consegue che i *Big Data* costituiscono una opportunità per la disciplina che non si sostituisce, bensì si affianca, alle tradizionali tecniche di analisi.

Le nuove fonti di dati aprono la possibilità, soprattutto nel campo delle scienze sociali, di studiare fenomeni fino ad ora difficili da catturare. Come è stato sottolineato da Capello (2016), le scienze regionali focalizzano la loro attenzione sull'aspetto spaziale nella

analisi del funzionamento del mercato. Tutti i dati concernenti le città, i comportamenti sociali, i *network* di imprese e sociali che possono concorrere a determinare l'attrattività delle città, la concentrazione spaziale delle imprese e gli effetti sulle politiche economiche e sociali, sono stati fino ad ora collezionati attraverso l'utilizzo di *survey* ed interviste. I Big Data in questo campo possono rappresentare una opportunità per sviluppare nuovi tipi di analisi più accurate e geo-referenziate. Nonostante queste opportunità i Big Data non sembrano riscuotere interesse nelle scienze regionali.

Analizzando le riviste più importanti di scienze regionali (<http://www.regionalscience.org>) emerge che in solo 7 di esse la parola *Big Data* appare nel titolo, nelle *key words* o nell'*abstract*, come riportato nella tabella I in appendice. La tabella IA riassume le caratteristiche dei contributi che riportano nel titolo, nelle *keywords* o nell'*abstract*, la parola Big Data a cui sono state aggiunte due pubblicazioni apparse su NBER. Come si osserva dalle fonti dei dati non è ancora possibile fare emergere una applicazione organica dei *Big Data* nell'ambito della disciplina delle Scienze Sociali. Un copioso numero di contributi (ben 8, dal 2013 al 2016) è stato pubblicato in *GeoJournal* che, nell'agosto 2015, ha dedicato una sezione speciale al tema dal titolo "What's So Big about Big Data? Finding the Spaces and Perils of Big Data". Le Tabelle II e IIA sintetizzano i contributi ed i relativi obiettivi. Dall'analisi emerge come le scienze regionali possano beneficiare dei *Big Data* utilizzando informazioni di tipo spaziale in essi contenute, evidenziando le criticità connesse al loro utilizzo.

Nonostante questo scarso interesse da parte delle scienze regionali nei confronti del tema dei Big Data, è possibile affermare che essi possono rappresentare un valore aggiunto sotto una serie di profili. Utilizzando i Big Data è possibile avere migliori informazioni circa i comportamenti sociali, le opinioni politiche, i *network* relazionali. Queste informazioni risultano essere molto utili per analizzare tutta una serie di comportamenti sociali che spesso non potevano essere studiati con i dati a nostra disposizione. Inoltre, sotto il profilo delle *policies*, è possibile ottenere, in fase decisionale, significativi contributi circa le azioni da intraprendere attraverso l'analisi dei bisogni e dei desiderata dei cittadini, e, in fase operativa, dei *feedback* circa i risultati e il gradimento delle *policies* adottate. È inoltre possibile concorrere alla creazione e al miglioramento di *smart cities* attraverso un aggiornamento continuo sul modo con cui le città sono organizzate e regolamentate. Infine è possibile ottimizzare o migliorare l'uso di alcuni beni pubblici ottenendo maggiori informazioni in termini di domanda e di efficienza degli stessi. Per quanto i *Big Data* possano rappresentare una preziosa fonte di informazioni, esse devono essere utilizzate con le opportune cautele. Esiste infatti un problema connesso al campionamento. Nell'ambito degli *small data* il campione, oltre ad essere scelto per rispondere ad una specifica domanda di ricerca, è rappresentativo. Nel caso dei *Big Data* il campione, per quanto ampio, potrebbe essere non rappresentativo (*self selected*). Se si prendono tutti gli utenti che utilizzano Facebook per misurare qualche attributo della popolazione italiana, si è di fronte ad un campione *non rappresentativo*. Inoltre, i dati e le informazioni che le persone decidono di rilasciare attraverso i *social* sono filtrati dagli stessi utenti: le informazioni condivise sono una parziale rappresentazione dell'individuo e dei suoi comportamenti. La narrazione e l'immagine che gli utenti vogliono dare di loro stessi potrebbe non corrispondere alla realtà. Un

utente potrebbe decidere di postare una foto su Facebook mentre si trova ad un concerto lirico, ma potrebbe decidere di non postare una foto di se stesso in palestra alle 10 del mattino. Mentre la prima dà un'immagine colta e raffinata, la seconda potrebbe fornire una immagine negativa.

A conclusione, è necessario evidenziare un elemento importante nell'ambito della analisi. I *Big Data* generalmente provengono da imprese private e come tali sono di proprietà delle stesse. Essi rappresentano spesso, per le imprese detentrici, la possibilità di avere informazioni che conferiscono loro un vantaggio competitivo. L'accesso ai dati da parte della comunità scientifica è spesso inibito. Occorre pertanto che gli Atenei concorrano a stipulare accordi con le imprese detentrici dei dati affinché gli studiosi possano, nel rispetto della normativa vigente, accedervi ed utilizzarli nei loro lavori.

Barbara Martini, Università di Roma "Tor Vergata"

Riferimenti bibliografici

Anderson C. (2008), The end of theory: the data deluge makes the scientific method obsolete, *Wired*, 23.

Body D., & Crawford K. (2012), Critical Questions for Big Data, *Information Communication and Society*, 15,5: 662-679.

Capello R. (2016), *Regional Economics*, NY: Routledge.

Einav L., & Levin J. D. (2013), The Data Revolution and Economic Analysis, NBER *Working Paper*, n. 19035.

Floridi L. (2012), Big Data and their Epistemological Challenger, *Philosophy and Technology*, 25: 435-437.

Kitchin R. (2014), *The Data Revolution*, London: Sage.

Loverace L., Birkin M., Cross P., & Clarke M. (2016), From Big Noise to Big data: towards the verification of large data set for understanding regional retail flows, *Geographical Analysis*, 48, 59-81

Appendice

Tabella I: Riviste e contributi sui Big Data				
		Titolo	Parole chiave	Abstract
1	Applied Spatial Analysis and Policy	0	0	0
2	Cities	0	3	3
3	Computers, Environment and Urban Systems	1	7	7
4	Economic Geography	0	0	0
5	Economic System Research	0	0	0
6	Geographical Analysis	1	2	2
7	Growth and Change	0	0	0
8	International Journal of Urban and Regional Research	0	0	0
9	International Regional Science Review	0	0	0
10	Italian Journal of Regional Science	0	0	0
11	Journal of Economic Geography	0	0	0
12	Journal of Geographical Systems	0	2	2
13	Journal of Geography and Planning	0	0	0
14	Journal of Regional Analysis and Policy	0	0	0
15	Journal of Regional Science	0	0	0
16	Journal of Urban Economics	0	0	0
17	Letters in Spatial and Resource Sciences	0	0	0
18	Location Science	0	0	0
19	Networks and Spatial Economics	0	0	0
20	Papers in Regional Science	0	1	1
21	Region	0	0	0
22	Regional Science and Urban Economics	0	0	0
23	Regional Science Policy & Practice	0	0	0
24	Regional Studies	1	0	0
25	Regional Studies, Regional Science	1	1	1
26	Review of Urban & Regional Development Studies	0	0	0
27	Socio-Economic Planning Sciences	0	0	0
28	Spatial Economic Analysis	0	0	0
29	Studies in Regional Science	0	0	0
30	The Annals of Regional Science	0	0	0
31	The Review of Regional Studies	0	0	0
32	Transportation Research Part A: Policy and Practice	0	0	0
33	Urban Studies	0	0	0

Tabella IA Obiettivi dei contributi

	Rivista	Titolo	Autori	Anno	Affiliazione	Obiettivi dei contributi	Fonte dei dati
1	Cities	Evaluating the effectiveness of urban growth boundaries using human mobility and activity records	Ying Long et al	2015	Cina	Evaluate Bejiing's urban growth boundaries; real time representation of urban dynamics and evolution over time and space	Mobile device;
2		Head/tail breaks for visualization of city structure and dynamics	Bin Jiang	2015	Svezia	Studying natural cities defined by a bottom up approach and from individual people and thier interractions	VGI (Volunteer Geographic Information); Social media; images
3		How website users segment a city: The geography of housing search in London	Alasdair Rae, Ebru Sener	2016	UK, Turchia	Spatial patterns of housing search in London	Most popular Real Estate portal UK
4	Computers, Environment and Urban Systems	MERRA Analytic Services: Meeting the Big Data challenges of climate science through cloud-enabled Climate Analytics-as-a-Service	John L. Schnasea, et al	2014	USA	Study climate change	Coming from Nasa center from Climate Simulation and Climate Science
5		Volunteered Geographic Information: Towards the establishment of a new paradigm	Bin Jiang, Jean-Claude Thill	2015	USA, Svezia	Contribution of VGI to social science and urban environmental	VGI (Volunteer Geographic Information)
6		Understanding U.S. regional linguistic variation with Twitter data analysis	Yuan Huang et al	2015	USA, UK	Regional linguistic variation in US	Tweeter
7		An efficient data processing framework for mining the massive trajectory of moving objects	Yuanchun Zhou, et al	2015	Cina, USA	Efficient processing of trajectory data	Theoretical
8		Constructing gazetteers from volunteered Big Geo-Data based on Hadoop	Song Gao et al	2014	USA	Construct gazzetter using a data driven approach	VGI
9		A scalable framework for spatiotemporal analysis of location-based social media data	Guofeng Cao et al	2015	USA	Harm massive location based social media for spazio temporal analysis	Social media

Tabella IA Obiettivi dei contributi - segue

10	Computers, Environment and Urban Systems	Online interactive thematic mapping: Applications and techniques for socio-economic research	Duncan A. Smith	2016	UK	Investigate the use of interactive thematic mapping tools for visualizing, exploring and analyzing socio-economic data on line	US open data portal site; UK open data; Geospatial geometry data (free in US)
11	Geographical Analysis	From Big Noise to Big Data: Toward the Verification of Large Data sets for Understanding Regional Retail Flows	Robin Lovelace et al	2016	UK	Human movements from home location to shopping centers	(1) mobile telephone provider; commercial consumer survey; geotagged Twitter messages
12		Spatial Distribution of City Tweets and Their Densities	Bin Jiang et al	2016	USA, Svezia	Study natural cities and mapping them using tweeter; spatial distribution of tweets and density within the cities	Tweeter
13	Journal of Geographical Systems	Mobile phone usage in complex urban systems: a space-time, aggregated human activity study	Emmanouil Tranos, Peter Nijkamp	2015	UK, Paesi Bassi	To investigate the potential of the data generated from mobile phone usage in modelling the aggregate activity in space and time using Amsterdam as point of reference	Mobile phone operators
14		Optimizing distance-based methods for large data sets	Tobias Scholl, Thomas Brenner	2015	Germania	Improve the distance based methods for measuring spatial concentration of industries and adapt them, dealing with large data sets, bypassing limiting concerning memory requirements and running time	Theoretical
15	Papers in Regional Science	Uncovering regional characteristics from mobile phone data: A network science approach	Guanghua Chi et al	2014	Cina	Network science methods to cover inherent characteristics of functional regions to understand spatial interactions of regional structures	Mobile phone operators
16	Regional Studies, Regional Science	Putting big data in its place: a Regional Studies and Regional Science perspective	Alasdair Rae & Alex Singleton	2015	Irlanda	Position paper	
17	Regional Studies	The Data Revolution, Big Data, Open Data, Data Infrastructures and Their Consequences	Barbara Martini	2015		Book Review	
18	NBER	Big Data And Big Cities: The Promises And Limitations Of Improved Measures Of Urban Life	Edward L. Glaeser et al	2015	USA	Using big data in urban social science to answer the following questions: (a) how does urban development influence the economy? (b) how does the physical cities interact with social outcomes? (c) How much do people value urban amenities? (d) how can public policies improve the quality of physical space?	Street view
19		The Data Revolution And Economic Analysis	Liran Einav, Jonathan D. Levin	2013	USA	How the big data impact on economic policy and economic research	government administrative data

Tabella II: contributi in Geojournal

	Autore	Anno	Titolo
1	John Steenbruggen, Maria Teresa Borzacchiello, Peter Nijkamp, Henk Scholten	2013	Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities
2	Rob Kitchin	2014	The real-time city? Big data and smart urbanism
3	Ryan Burns, Jim Thatcher	2015	Guest Editorial: What's so big about Big Data? Finding the spaces and perils of Big Data
4	Ryan Burns	2015	Rethinking big data in digital humanitarianism: practices, epistemologies, and social relations
5	Linnet Taylor, Ralph Schroeder	2015	Is bigger better? The emergence of big data as a tool for international development policy
6	Rob Kitchin, Tracey P. Lauriault	2015	Small data in the era of big data
7	Jeremy W. Crampton	2015	Collect it all: national security, Big Data and governance
8	Harvey J. Miller, Michael F. Goodchild	2015	Data-driven geography

Tabella II A: obiettivi dei contributi

1	The aim of the paper is to provide a systematic overview of the main project addressing the use of data derived from mobile phone networks to obtain location and traffic estimations of individuals as starting point of further research on incident and traffic management. The results concern the use of the telco data to understand individual presence and mobility in regular situation and during non recurrence events.
2	The paper analyze the concept of smart cities underlying that these cities are being instrumented with digital device infrastructures that produce big data. These data are raw material for envisioning and enacting more efficient, sustainable, competitive, productive and transparent cities.
3	Introduction to special issue concerning big data; Geography as discipline can have a lots of benefits from the big data studies; The 60% of the data sets contain spatial information. Through an emphasis on space and spatial relations geographer have the opportunity to emerge as a central pillar of the big data research agenda. The aim of the special issue is to situate big data and its accompanying methodologies in their contingent social and historical contexts.
4	Big Data future is not inevitable, and would emerge embodying multiple forms of values and privileges. The attention is focalized on Big Data digital humanitarianism. Big Data is constituted by profound changes in how data is collected, processed, and visualized; and they can effect the political-economic and socio-political processes
5	The use of mobile phone give the opportunity to access a mobile phone location information and calling metadata. These data have the potential to fill some of the problematic gaps in data resources available to country policy makers. The paper considers the implication of these new types of data for development policy and planning. The analysis indicates that these new data sources represent an important complement to country level statistics
6	Small data are been used for long time in academic knowledge buildings. The small data are generated to answer to specific questions. The paper analyze the logic and the value of small data studies, their relationship to emerging bid data and data science and the implication of scaling small data into data infrastructures with the focus on spatial data
7	The paper focused on the contradictions and the complications of dealing with geospatial data. The geopolitical assemblage of the data includes and incorporates different institutions: military, state, legislature and knowledge process. None of these are easily separated from each other. Whit in this staring point in mind the paper highlight the importance of the concept of geopolitical assemblage.
8	Big Data can change the future of geography because now, more than ever, it is possible to capture georeferenced data coming from different sources. The beliefs that spatial context matters is an important theme in geography. There is a potential to use big data to inform both geographical knowledge discovery and spatial modelling. There are also some challengers such us how to formalize geographic knowledge to clean data and to ignore spurious patterns and how to build data-driven models that are true and understandable.