

MIRELLA LODA - MARIO TARTAGLIA

DALL'INFORMAZIONE ALLA CONOSCENZA

LA GEOGRAFIA UMANA NELL'ERA DELLA PROLIFERAZIONE DEI DATI (*)

Premessa. – L'incremento iperbolico nella produzione di dati, seguito al diffondersi di dispositivi digitali e tecnologie informatiche, ha innescato profonde trasformazioni nei processi attraverso i quali si producono e si diffondono informazioni e conoscenza. La riflessione su tali cambiamenti, intensificatasi negli ultimissimi anni, ne sta mettendo in luce le importanti ricadute in molti campi della vita sociale, da quelli più propriamente socio-politici (*marketing* invasivo e minacce alla *privacy*, controllo dei movimenti di protesta ed erosione di libertà democratiche, diseguaglianze nell'accesso ai dati), a quelli istituzionali (maggiore trasparenza della pubblica amministrazione, ridefinizione del ruolo degli organismi pubblici di ricerca), sino al terreno della ricerca scientifica e accademica (rischio di marginalizzazione) (1).

In queste brevi note esporremo alcune osservazioni sulle opportunità che i nuovi scenari aprono alla ricerca geografica, ma anche sui problemi che la proliferazione di dati geo-spaziali comporta e sulle implicazioni per la collocazione della disciplina nel più ampio contesto del pensiero scientifico.

Rileggendo Lucio Gambi. – Questa recente produzione di dati geo-spaziali conferisce nuova attualità a una riflessione sulla natura e sui metodi della conoscenza geografica, già oggetto in passato di ampio dibattito e di controversie fra i cultori della disciplina.

Se infatti da un lato la sovrabbondanza di tali dati geo-spaziali, per lo più prodotti digitali collaterali a operazioni routinarie come la telefonia mobile, non-

(*) Questo articolo è stato sviluppato congiuntamente dai due autori; i paragrafi primo, secondo, quinto e sesto sono stati redatti da M. Loda; il terzo e il quarto da M. Tartaglia.

(1) Si pensi in particolare alle acute ma profondamente scettiche considerazioni espresse da Jay David Bolter (2010) in occasione della *lectio magistralis* tenuta presso l'associazione Vivere di Cultura (Roma).

ché la diffusione di strumenti tecnici relativamente economici per elaborarli, sembrerebbero ricollocare la geografia al centro della produzione del sapere scientifico, dall'altro la combinazione dei due fattori agevola l'inserimento della dimensione spaziale nell'orizzonte teorico e nella pratica di ricerca di molte altre discipline sociali (*spatial turn*), costringendo di nuovo la geografia a ridiscutere il proprio statuto epistemologico e a legittimare il proprio ruolo.

Dalla seduzione del primo scenario (quello che immagina la geografia come lo snodo del «teatro dei saperi») ci deve mettere in guardia la storia della disciplina, che ha ampiamente dimostrato l'inconsistenza scientifica di una geografia ancorata alla rilevazione/lettura del puro dato locazionale: proprio visioni di questo tipo sono in ampia misura responsabili della progressiva marginalizzazione che la geografia ha subito negli scorsi decenni in ambito scientifico. È il caso della «geografia degli oggetti» praticata nei primi lustri del Novecento da Otto Schlüter, cui spesso si riduceva la sedicente geografia dei fenomeni. Un destino non dissimile è toccato alla «geografia delle interrelazioni spaziali» di Richard Hartshorne, impossibilitata a cogliere – almeno al di fuori del dominio della geografia fisica – la natura profonda delle interrelazioni stesse, in virtù del divieto che essa stessa si autoinfliggeva di occuparsi delle «proprietà sostanziali dei fenomeni» (Harvey, 1969, pp. 70-77). È infine il caso della cosiddetta «geografia integrale», attardatasi sin dentro gli anni Settanta del secolo scorso nell'illusione di poter traversare tutti i campi dello scibile grazie all'attenzione (costante, ma – ahinoi – spesso anche di corto respiro) alla componente locazionale.

La stessa geografia quantitativa ⁽²⁾ – laddove ha trovato terreno culturalmente fertile per radicarsi, come nel mondo anglosassone – è riuscita a sviluppare tecniche per arginare i limiti di un'analisi basata interamente su dati spaziali ⁽³⁾, ma al prezzo di miniaturizzare gli areali di osservazione, quindi di ridurre fortemente l'interesse dei propri risultati ⁽⁴⁾, motivo non ultimo del rallentare dello slancio quantitativista anche oltreoceano.

Un punto fermo e di straordinaria attualità in questo delicato dibattito era stato fissato da Lucio Gambi già attorno agli anni Settanta del Novecento. Gambi ci ricordava innanzitutto la necessità di distinguere fra informazione e conoscenza, fra ricognizione di dati ed elaborazione scientifica degli stessi. Argomentando contro l'illusione di una «geografizzazione del sapere» ⁽⁵⁾, egli precisava che «l'uso via via più frequente [...] nelle discipline più diverse di carte, di planimetrie, di fotografie panoramiche [...] non forma "geografia": sono unicamente dei mez-

(2) Per una definizione di quell'universo complesso di pratiche che va sotto il nome di geografia quantitativa, si veda Murray (2010).

(3) Con un percorso aperto dalla riflessione pionieristica di Stan Openshaw e Paul Taylor (1979) sul modo in cui la scelta degli areali di riferimento condiziona i dati, sino a quella di Luc Anselin (1988) sul tema dell'autocorrelazione spaziale.

(4) È il caso dei cosiddetti metodi di analisi locale (Anselin, 1995).

(5) L'espressione era stata utilizzata con ingiustificato ottimismo da Dino Griboaudi (1963, p. 267).

zi per documentare in modo più efficace una trattazione» (Gambi, 1973, p. 95). Come vedremo nei prossimi paragrafi, questa posizione è sorprendentemente vicina a quella formulata da alcuni osservatori a proposito dei *big data*, e ci avverte che la quantità di dati e i più sofisticati assemblaggi degli stessi non producono necessariamente conoscenza geografica (6), né sono appannaggio di coloro che si professano geografi.

In linea generale Gambi indica come oggetto di studio della geografia umana il modo in cui le collettività umane si interfacciano con le vocazioni ambientali, e le modalità con cui le vocazioni vengono organizzate per l'azione di forze di fondo che muovono la società (Gambi, 1973, p. 93). Il richiamo a tali forze di fondo, alle «strutture sociali» intese in una prospettiva vicina al materialismo storico, ha contribuito grandemente a riorientare lo sguardo di generazioni di geografi, distogliendolo da letture superficiali delle forme visibili dell'organizzazione territoriale (denominate allora specialmente in termini di «paesaggio») e indirizzandolo verso una più profonda attenzione ai fattori che tali forme determinano. Trasposta nell'oggi, la visione gambiana, come approfondiremo nel quarto paragrafo, induce a interrogarsi a fondo sul rapporto fra proprietà geometriche o semplicemente locazionali dei fenomeni (quelle ad esempio che gestiamo e analizziamo attraverso i GIS) e natura sostanziale degli stessi, nonché fra i fenomeni e il contesto in cui essi si esplicano, per evitare il rischio di apofenia alimentato dal superpotere di calcolo conferitoci dalle tecniche informatiche (7).

L'opera di Lucio Gambi ha apportato un avanzamento fondamentale nel paradigma scientifico della geografia italiana, ma rappresenta al tempo stesso un modello di sapere geografico pressoché irraggiungibile: narrazioni suggestive e profonde come *I valori storici dei quadri ambientali* sembrano insomma corroborare la validità dell'aforisma proposto da Paul Claval (1972, p. 120) secondo cui «non si può essere geografo e mediocre».

A nostro parere, tuttavia, la lezione di Lucio Gambi dà indicazioni preziose anche per una pratica di ricerca geografica di portata più quotidiana, come quella con la quale si misura di fatto la maggior parte di noi, e soprattutto traccia a tutt'oggi la strada maestra per orientarsi negli scenari aperti dalla diffusione di dispositivi digitali e dall'iperbolica produzione di dati geo-spaziali. La lezione gambiana ha insegnato che nell'ambito delle scienze sociali non ha (più) senso una definizione disciplinare per oggetto o per metodo, e che la conoscenza dev'essere

(6) Boyd e Crawford (2012, p. 670) affermano: «The size of data should fit the research question being asked», come dire che la disponibilità di un *set* di dati non definisce in sé un problema di ricerca, né un problema è meglio definito quanti più dati si utilizzano per trattarlo.

(7) Con riferimento ai *big data*, Bollier formula la questione nei termini seguenti: «The existence of Big Data intensifies the search for interesting correlations. But correlations, as any first-year statistics student learns, does not establish causality. Causality requires models and theory» (2010, p. 16). Un gustoso esempio di «immotivata visione di connessioni» era stato fornito nel 1987 da Gerard Hard (1987-1988), nel pieno della «rivoluzione quantitativa» in geografia, dimostrando come con applicazione meccanica del calcolo di correlazione si potesse dimostrare che le cicogne portano i bambini.

organizzata per problemi: una constatazione sulla quale successivamente hanno convenuto altri noti geografi in Italia (ad esempio Turco, 1987) e all'estero (ad esempio Taylor, 2009), e dalla quale discendono conseguenze rilevanti nei nuovi scenari di cui stiamo discutendo. Ne consegue in particolare che la disponibilità di un determinato *set* di dati «geografici», per quanto ampio, è un elemento puramente informativo, e non identifica di per sé un problema di ricerca; la definizione di quest'ultimo può invece risultare soltanto da considerazioni di ordine superiore e di natura teorica o storico-politico-sociale: è il progetto conoscitivo a dettare la selezione dei dati, non viceversa. Anche in questo caso va sottolineata la convergenza con la riflessione dei più attenti commentatori contemporanei del cosiddetto «computational turn»: «The real challenge is to understand what kind of data points you need in order to form a theory or make decisions» (Bollier, 2010, p. 14) ⁽⁸⁾.

Anche il geografo deve quindi orientare di volta in volta il proprio lavoro alla trattazione di un nodo specifico di problemi (Gambi, 1973, p. 96), la produzione di conoscenza sui processi territoriali essendo esito di un approccio transdisciplinare e necessariamente collettivo alla ricerca, con l'apporto di diverse tecniche analitiche qualitative e quantitative e di sofisticati strumenti digitali di analisi e rappresentazione spaziale. Se tuttavia queste considerazioni debbono farci diffidare dell'illusione che dalla pura disponibilità di dati geo-spaziali sgorga la geografia, esse non possono certo risolversi nella preclusione verso i nuovi percorsi di ricerca che il progresso tecnologico apre alla geografia come alle altre scienze sociali. Anche le discipline si muovono infatti con la storia, ed è innegabile che a «radically new kind of “knowledge infrastructure” is materializing» (Bollier, 2010, p. 1) ⁽⁹⁾.

Si tratta allora di capire, con orientamento vigile e critico, ma aperto, quali prospettive si aprono alla ricerca socio-territoriale con il massivo incremento nella disponibilità di dati e quali siano le specifiche problematiche connesse al loro utilizzo ⁽¹⁰⁾, possibilmente prima che il «computational turn» ⁽¹¹⁾ si cristallizzi in una nuova ortodossia (Boyd e Crawford, 2012, p. 666).

L'inflazione delle informazioni spaziali. – Lo sviluppo dell'informatica, dalla seconda metà del secolo scorso, dall'invenzione dei primi circuiti integrati alla creazione di Internet, è stato tanto impetuoso da portare a riconoscere l'inizio di un'era della tecnologia informatica da contrapporre a quella industriale (Castells,

(8) Bollier riprende qui un'espressione formulata da Stefaan Verhulst, capo del centro studi della Markle Foundation.

(9) Una radicale formulazione del cambio di paradigma viene proposta da David Berry, che propone «to think critically about how knowledge in the 21st century is transformed into information through computational techniques, particularly within software» (2011, p. 12).

(10) Interpretano a mio parere pienamente questo spirito le recenti osservazioni di Mauceri (2015).

(11) Per una approfondita rassegna delle problematiche connesse al «computational turn» da diversi punti di vista disciplinari, si veda IACAP (2011).

2010). L'inizio del terzo millennio sta mostrando i frutti di questo cambiamento nella forma di un'enorme diffusione di informazioni, tant'è che mentre in passato uno dei maggiori problemi da affrontare nell'applicazione del metodo scientifico era il reperimento di dati, oggi le difficoltà maggiori sono rappresentate dalla loro gestione ⁽¹²⁾. Il presupposto di questo fenomeno è da rintracciarsi nella stretta interrelazione che si è innescata tra il progresso tecnologico e le tendenze economico-sociali, interrelazione che si sta manifestando nell'uso massivo delle tecnologie informatiche, nel crescente utilizzo delle apparecchiature e delle infrastrutture *smart*, nella crescente disponibilità della banda larga, nella grande diffusione dei *social networks* (OECD, 2013, p. 6).

In aggiunta, una nuova modalità di produzione di informazioni sta crescendo sensibilmente negli ultimi anni: quella dovuta alla crescente diffusione di sensori, dai piccoli rilevatori a radiofrequenza (RFID), alle telecamere, ad altri sensori di vario genere. La loro funzione è quella di trasporre i fenomeni reali in un insieme di dati quantitativi e processabili riguardanti i settori più disparati, da quello della salute a quello della mobilità, da quello della sicurezza a quello dell'ambiente. Si stima che già nel 2011 esistessero al mondo circa 30 milioni di sensori connessi a sistemi informatici, e che la loro crescita attuale sia circa del 30% all'anno (McKinsey Global Institute, 2011). È plausibile che questa interpretazione del mondo attraverso dati rilevati dai sensori, che alcuni autori ⁽¹³⁾ indicano con il termine di *datafication*, si evolverà attraverso un sempre più vasto sistema di comunicazione tra macchine interconnesse e reti di sensori (Gubbia e altri, 2013), ovvero attraverso la cosiddetta «Internet delle cose» ⁽¹⁴⁾ che diventerà sempre più indipendente dalla comunicazione umana producendo una quantità enorme di dati.

Una parte delle informazioni che viene oggi scambiata nel mondo utilizza la rete Internet. La quantità di dati che transita in Internet rappresenta tuttavia soltanto una piccola parte dell'intero universo digitale, che è stato recentemente stimato essere circa 4.400 miliardi di Gigabyte nel 2013 e si immagina crescerà di dieci volte entro il 2020, raggiungendo i 44.000 miliardi di Gigabyte, con un incremento annuo medio di oltre il 40% (IDC, 2014).

La disponibilità di una così grande quantità d'informazioni ha determinato l'esigenza di sviluppare un certo numero di classificazioni dei dati disponibili.

Innanzitutto, appare molto significativa la distinzione tra *public data* e *private data*. Mentre la prima categoria è utilizzabile da chiunque, ad esempio per attività di analisi e ricerca, la seconda rimane disponibile soltanto a un numero ristretto di utilizzatori (per un'articolata definizione: Wyrwoll, 2014, pp. 12-15).

(12) Un'acuta polemica contro la «information society» è stata sviluppata da Richard Lanham (2006, pp. 1-41).

(13) Si veda ad esempio Mayer-Schönberger e Cukier (2013).

(14) Il concetto di «Internet delle cose» è comunemente indicato con la nomenclatura inglese, «Internet of Things» (IoT). Per una illustrazione introduttiva dell'Internet of Things e una discussione sulle sue possibili definizioni, si veda ad esempio Perera e altri (2014).

Una particolare categoria di dati liberamente disponibili è quella degli *open data*. Anche se in linea teorica la definizione di *open data* si applica a tutti i dati e contenuti che possono essere liberamente usati, modificati e condivisi da chiunque per qualsivoglia proposito (Open Knowledge Foundation Network, 2015), nella pratica una caratteristica fondamentale degli *open data* è quella di essere digitalizzati in formati standard, in modo da poter essere fruiti con estrema facilità (Attard e altri, 2015). Questo aspetto indica che gli *open data* sono in realtà un sottoinsieme del più generale insieme dei *public data*.

Una parte importante degli *open data* oggi disponibili è fornita dal settore delle amministrazioni pubbliche (*open government data*: Kučera, Chlapek e Nečáský, 2013). L'importanza di quest'ultima categoria di dati sta nel fatto che essi dovrebbero essere dotati di ufficialità, oltre che di elevati livelli di qualità, cosa che allo stato attuale non sempre avviene a causa delle non uniformi politiche di produzione e rilascio degli *open data* nei diversi paesi (Zuiderwijk e Janssen, 2014).

Un'altra importante tipologia di dati – oggi largamente disponibile benché non sempre esente da vincoli di riservatezza – è quella dei cosiddetti *user-generated contents* (UGC), che si può definire come l'insieme dei contenuti pubblicati su una piattaforma *online* dagli utenti della piattaforma stessa (Wyrwoll, 2014, p. 15). Come fanno notare Kaplan e Haenlein (2010), gli *user-generated contents* sono generalmente identificati con i dati contenuti nei *social media*, tanto che gli stessi autori definiscono questi ultimi proprio come quella categoria di applicazioni basate su Internet e concepite nell'ottica del Web 2.0, che consente la creazione e lo scambio di *user-generated contents*. Questi dati sono prodotti attraverso l'uso di numerose applicazioni o siti Internet, ad esempio Facebook, Twitter, Youtube, Instagram, Google, Flickr, LinkedIn o *blog* di diverso genere. I dati possono poi riguardare le generiche interazioni sociali degli individui, oppure aree tematiche specializzate come il tempo libero, il turismo, il commercio, la formazione, la salute. Un'ulteriore categoria di dati, trasversale rispetto a quelle menzionate finora, è quella dei dati geo-spaziali, ovvero quei dati che contengono in sé informazioni sulla propria localizzazione rispetto allo spazio geografico. Grandi quantità di dati geo-spaziali concernenti lo spazio fisico sono oggi disponibili in Internet attraverso servizi cartografici (come Google Maps, Bing, OpenStreetMap) che forniscono cartografie e foto satellitari prodotte con sistemi tradizionali di telerilevamento, ma anche cartografie generate da sistemi di collaborazione volontaria per la creazione di dati geografici (per una panoramica al riguardo: Goodchild, 2007). Ma la vera innovazione emersa negli ultimi anni è l'enorme diffusione di dati geo-spaziali continuamente prodotti dalle più moderne tecnologie di comunicazione, che consentono di localizzare eventi e attività delle persone pur non essendo questo il loro scopo primario. Esempi di questo tipo di tecnologie sono le apparecchiature elettroniche di uso quotidiano dotate di GPS, i telefoni cellulari e le altre *smart devices*, le infrastrutture che ospitano *social media*, i servizi di diverso genere che utilizzano le reti informatiche. Benché non siano disponibili valutazioni esatte su quanta parte

dei dati oggi disponibile sia di tipo geo-spaziale, si ritiene che essa sia una porzione rilevante di tutti i dati in circolazione (UN-GGIM, 2013, p. 12). Ad esempio, la quantità dei soli dati contenenti la posizione geografica delle persone è stata stimata essere di circa un milione di Gigabyte nel 2009, con un tasso di crescita pari al 20% all'anno (McKinsey Global Institute, 2011, p. 87).

A prescindere dalle tipologie fin qui richiamate, la caratteristica più evidente di questi insiemi di informazioni è il fatto di essere estremamente cospicui. Questo è il motivo per cui la più nota locuzione riguardante i dati di cui oggi disponiamo è «big data», che secondo Diebold (2012) nacque alla metà degli anni Novanta del Novecento nella comunità delle aziende tecnologiche statunitensi semplicemente per indicare la grande quantità di informazioni che cominciava allora a essere fruibile. Numerose definizioni di *big data* ⁽¹⁵⁾ sono ancora oggi basate sul loro grande volume (ovvero sulla grande quantità di memoria che essi occupano) e sui conseguenti problemi computazionali. La discussione intorno alla definizione dei *big data* non si è tuttavia limitata al loro volume, sin da quando, nel concettualizzare il suo cosiddetto «3V Model», Laney (2001) mise in evidenza l'importanza di tre caratteristiche delle grandi quantità di dati che rendono peculiari le problematiche della loro gestione: oltre al loro volume, anche le loro grandi velocità (di trasmissione) e varietà. Ad ogni modo, è evidente che la grande consistenza, mutevolezza e complessità dei *big data* oggi disponibili non rende agevole estrarre da essi elementi di conoscenza. A tale scopo si rende infatti necessario realizzare un processo piuttosto complesso (Labrinidis e Jagadish, 2012).

Ma più che soffermarsi sulle problematiche derivanti dalla complessità e dalla mole dei dati, dal punto di vista della geografia umana è utile riflettere sul livello di corrispondenza tra questa moltitudine di informazioni e le fenomenologie comportamentali dei singoli individui o dei gruppi sociali. In questo senso, è evidente che una parte delle informazioni di pubblico dominio costituisce la rappresentazione di entità puramente fisiche (si pensi ai dati cartografici più tradizionali) oppure di processi talmente aggregati da non consentirne una lettura in termini delle dinamiche sociali da cui in ultima analisi sono generati (si pensi ai dati macroeconomici di un paese e a quelli che rappresentano le *performances* delle aziende o le attività delle pubbliche amministrazioni). Altre informazioni, invece, sono più direttamente riconducibili alle attività degli individui, e normalmente sono talmente numerose da poter essere etichettate come *big data*. Una parte di questi dati corrisponde ai contenuti generati dagli utenti dei *social media* ed è pertanto identificata con il termine di *social big data* (Bello-Orgaz, Jung e Camacho, 2015). Un'altra porzione dei dati originati dalle dinamiche so-

(15) Ad esempio: «big data is when the size of the data itself becomes part of the problem» (Loukides, 2010); oppure: «Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze» (McKinsey Global Institute, 2011).

ciali dei singoli individui non transita necessariamente attraverso i *social media* e può essere disponibile per un numero ristretto di fruitori sotto particolari condizioni, ad esempio a pagamento, o limitato a certi ambiti aziendali o istituzionali. Si tratta comunque di dati prodotti da tecnologie di uso quotidiano: i sistemi GPS, le telecamere e le fotocamere, i telefoni cellulari, le apparecchiature o le applicazioni di navigazione satellitare, i sistemi di pagamento elettronico, i servizi commerciali espletati tramite Internet, i sistemi di localizzazione delle chiamate di emergenza, i sistemi informativi di *marketing*, i giochi elettronici. Molte di queste tecnologie – così come anche molti *social media* – non producono semplicemente *social big data* ma, essendo intrinsecamente dotate di metodi per il rilevamento della posizione geografica, sono una fonte molto prolifica di quelli che potremmo chiamare *social geo-spatial big data* e che senza dubbio rappresentano un'importante risorsa per le scienze geografiche e sociali.

Potenzialità e problemi dell'analisi spaziale. – Come è facile intuire, disporre di *big data* associati a una posizione geografica, ancor più se variabile nel tempo, amplia notevolmente tipi e potenzialità di analisi. Per motivi legati alle risorse economiche che è spesso necessario dedicare alla raccolta e all'elaborazione dei dati, molte applicazioni per l'analisi di questo tipo di dati sono sviluppate da parte di aziende per fini commerciali e da parte di grandi istituzioni per applicazioni legate al loro settore di attività. Tra i più promettenti campi di utilizzo dei *geo-spatial big data* in ambito commerciale troviamo, ad esempio, il *marketing* di vendita al dettaglio, la tariffazione delle polizze assicurative dei veicoli stradali, l'ottimizzazione del posizionamento delle coltivazioni per scopi agroalimentari e il loro monitoraggio, la salvaguardia delle *public utilities* dalle catastrofi naturali (Woodie, 2015). Dal punto di vista istituzionale, l'utilizzo di *big data* georeferenziati può invece essere visto come un'opportunità di sviluppo sociale (Eagle e Greene, 2014; Mayer-Schönberger e Cukier, 2013), tanto che molti paesi hanno attivato progetti per la creazione di piattaforme governative e per la loro gestione ⁽¹⁶⁾. Cionondimeno, sono sempre più numerosi gli studi, in diversi campi e discipline, che utilizzano grandi quantità di dati georeferenziati per attività di ricerca scientifica e metodologica.

Una tipica fonte di dati spaziali è in questo caso la telefonia mobile, dalla quale si possono estrarre diverse informazioni legate alle attività di conversazione. Generalmente, gli studi che utilizzano questo tipo di dati sono mirati a evidenziare relazioni tra contesti sociali e spaziali, estrapolare modelli tipici di mobilità, identificare luoghi rilevanti delle aree urbane, rilevare la densità delle persone nell'ambiente urbano al variare del tempo (Steenbruggen, Tranos e Nijkamp, 2015, p. 338). I risultati vengono normalmente utilizzati nell'ambito del *manage-*

(16) Si veda ad esempio il recente progetto del governo coreano descritto da Lee e Kang (2015).

ment urbano, ad esempio nei settori dei trasporti (Steenbruggen e altri, 2013), dell'economia urbana ⁽¹⁷⁾, del turismo ⁽¹⁸⁾, dell'uso del suolo ⁽¹⁹⁾.

Analoghe analisi possono essere condotte utilizzando *geo-spatial big data* estratti dai *social media*. A titolo di esempio, si possono citare studi finalizzati a caratterizzare le dinamiche sociali di un'area urbana attraverso i concetti di prossimità spaziale dei luoghi e di prossimità sociale degli individui (Cranshaw e altri, 2012; Spielman e Thill, 2008), evidenziare interazioni sociali attraverso la mobilità delle persone (Wu e Wang, 2015), individuare i legami degli individui con i luoghi (Wang, 2015). In alcuni casi, diverse fonti di *geo-spatial big data* possono essere utilizzate per l'analisi, come nell'applicazione esposta da Cho, Myers e Leskovec (2011), che mette in relazione dati provenienti da telefonia cellulare e da *social media* per la ricerca di modelli di mobilità tra persone che sono in mutuo contatto attraverso la rete.

In definitiva, si possono evidenziare alcuni aspetti importanti legati all'utilizzo di *geo-spatial big data* nelle scienze sociali. Il primo, valido anche in senso più generale, è dato dal fatto che avere a disposizione dati georeferenziati abilita la possibilità di riconoscere relazioni – non altrimenti desumibili – tra insiemi informativi eterogenei. Attraverso gli strumenti di analisi spaziale, i dati possono essere infatti legati fra loro sulla base della posizione, attualizzando di fatto la molto discussa cosiddetta «prima legge della geografia», secondo cui «everything is related to everything else, but near things are more related than distant things» (Tobler, 1970), e sulla cui base si è di fatto fondato lo sviluppo dei sistemi GIS (Sui, 2004, p. 269). È il caso ad esempio in cui si associano su base spaziale le traiettorie di spostamento degli individui, magari rilevate attraverso GPS o celle telefoniche, alle infrastrutture o ai servizi di trasporto da essi utilizzati, in modo da attribuire agli individui stessi la vocazione per determinati modi di trasporto. Oppure il caso in cui si mettono in relazione con metodi di analisi spaziale le posizioni delle persone con i luoghi che occupano più frequentemente e per maggior tempo al fine di caratterizzare lo stile di vita delle persone stesse in base alle proprietà degli spazi (ad esempio luoghi di attività culturale, sportiva ecc.). Al contrario, in altre circostanze, è possibile qualificare i luoghi in base all'intensità d'uso che gli individui ne fanno oppure alle caratteristiche sociali degli individui stessi, se si dispone di questo tipo di informazioni. Un altro aspetto rilevante è poi dato dal fatto che i *geo-spatial big data* possono sopperire alla limitatezza quantitativa dei metodi tradizionali come le osservazioni, le interviste e i diari, ma spesso non posseggono un adeguato valore informativo nella definizione sociale delle persone (vedi an-

(17) Ad esempio, De Jonge, Van Pelt e Roos (2012) individuano l'intensità delle chiamate telefoniche nelle aree commerciali come un indicatore dell'intensità delle attività economiche.

(18) Un esempio di utilizzo di dati estratti da telefonia cellulare per la stima della densità di visitatori in siti di interesse storico a Roma è quello di Calabrese e altri (2011).

(19) Ancora a titolo di esempio, Soto e Frías-Martínez (2011) hanno utilizzato i diversi tipi di comportamento emergenti dai dati di telefonia mobile per identificare diverse classi di uso del suolo a Madrid.

che quarto paragrafo). Per quest'ultimo aspetto, diventa cruciale la possibilità di integrare i *big data* con fonti informative adeguate alla definizione delle caratteristiche degli individui che – in ultima analisi – dei *big data* stessi sono i generatori.

In ogni caso, l'analisi dei *geo-spatial big data* necessita di potenti strumenti computazionali, in parte peculiari rispetto a quelli che tipicamente vengono adoperati per il trattamento dei generici *big data* (Shekar e altri, 2015). Tali strumenti, che rappresentano un'evoluzione dell'analisi spaziale tradizionale (Miller e Han, 2009), hanno recentemente fatto emergere un campo di ricerca specializzato che Mennis e Guo (2009) hanno indicato come «*spatial data mining and geographic knowledge discovery*». Gli strumenti impiegati nello *spatial data mining* sono in grado di trattare grandi quantità di dati e diversi formati combinando diverse tecniche di analisi spaziale (ad esempio, classificazioni, regressioni, ricerca di regole associative, *clustering*, analisi visiva). Si tratta tuttavia di tecniche che si potrebbero definire di tipo esplorativo, ovvero tendenti a scoprire correlazioni o regole nella grande quantità di dati a disposizione. Esse possono quindi portare a un approccio eccessivamente induttivo con il rischio di apofenia citato nel primo paragrafo. Questa preoccupazione, espressa da Kitchin (2013) a proposito dell'utilizzo dei *big data* nell'ambito della geografia sociale, è confermata da Miller e Goodchild (2015) che parlano dell'emergere di una «*data-driven geography*» la cui genesi risiede proprio nell'abbondanza di informazioni geo-spaziali. Si tratta qui del riflesso di una discussione di carattere epistemologico di ampia portata, che ha indotto alcuni autori a riconoscere la nascita di un nuovo paradigma del metodo scientifico basato esclusivamente sull'esplorazione dei dati senza più la necessità di supporti teorici (Anderson, 2008), erede della scienza sperimentale prerinascimentale, della scienza teoretica precedente alla nascita dei computer e della successiva scienza computazionale basata sulla simulazione (Hey, Tansley e Tolle, 2009). Il giudizio sulla validità di questa sorta di nuovo empirismo fondato sui dati è tuttavia fortemente dibattuto (Graham e Shelton, 2013; Kitchin, 2014; Miller e Goodchild, 2015) ed è indubbio che tutto questo richieda ulteriori riflessioni di metodo nell'ambito delle scienze geografiche e sociali.

Big data e ricerca accademica. – Se l'avvento dei *big data* ha modificato in profondità lo scenario entro cui si svolge la ricerca socio-territoriale, questo vale innanzitutto per la ricerca socio-territoriale finalizzata a supportare le strategie di *marketing* delle imprese. Il mondo delle imprese, come quello della politica, abbisogna continuamente di un supporto agile e duttile a fini decisionali, e i *big data* corrispondono a questa necessità in quanto consentono di elaborare in tempi rapidissimi imponenti quantità di informazioni sugli aspetti più diversi.

Processando dati derivati collateralmente dalla telefonia mobile, da GPS, dai *social networks* eccetera è possibile costruire algoritmi che descrivono comportamenti umani e rispondere in maniera quasi immediata a interrogativi molto specifici, come ad esempio il livello di apprezzamento per un prodotto, l'appetibilità

e la potenziale domanda per un prodotto da lanciare sul mercato, ma anche il gradimento per un programma, o il grado di fiducia in un personaggio politico.

Si sono così aperti spazi per professionalità (analisti) e per attività imprenditoriali (centri studi, società di consulenza) specializzate nel *data mining*, come ad esempio la Crimson Hexagon, specializzata in «providing social media analytics to leading brands and agencies», che ha recentemente messo a punto una tecnologia di successo nel campo della «sentiment analysis» (20).

L'impetuoso sviluppo di questo sapere direttamente finalizzato a obiettivi di mercato sta sostanzialmente avvenendo al di fuori dell'accademia. Le università, fino a tutti gli anni Novanta del secolo scorso luogo di elezione della ricerca socio-territoriale, stanno ora subendo una competizione che le vede in posizione di debolezza, dato che gli strumenti di ricerca tipici delle scienze sociali – segnatamente la ricerca empirica e l'indagine campionaria – presentano incontestabili svantaggi rispetto ai *big data* in termini di costo, di tempi di realizzazione e di volume quantitativo: diversamente da quelli necessari per un'indagine campionaria, i *social big data* non debbono essere generati *ad hoc*; essi sono immediatamente disponibili, e lo sono su volumi quantitativamente inimmaginabili fino a pochi anni fa, atti proprio per questo a essere modellizzati.

Nella valutazione circa la nuova condizione in cui viene a trovarsi la ricerca sociale, si profilano tre posizioni principali. Una prima posizione legge l'avvento dei *big data* come il passaggio a una fase in cui la ricerca accademica (nonché quella degli organismi a essa assimilabili) sarà sempre meno in grado di rispondere alla domanda di conoscenza espressa da gruppi di interessi forti, e sarà quindi destinata a un'inarrestabile, progressiva marginalizzazione (Savage e Burrows, 2007). Una seconda posizione stigmatizza come non scientifiche e culturalmente distorsive le correnti procedure di *data mining* e rivendica la superiorità della ricerca sociale tradizionale (21). Una terza posizione mira a evitare derive scissorie nella ricerca sociale e propone di recuperare i *big data* in una logica di triangolazione metodologica (Johnson e Smith, 2014; Laser e altri, 2014; Mauceri, 2015).

Concentrandoci su quest'ultima posizione, che ci sentiamo di condividere, è utile approfondire da un punto di vista strettamente metodologico in che cosa i *big data* si distinguono dai dati raccolti con le consuete pratiche di rilevazione socio-territoriale empirica (censimenti e, specialmente, indagini campionarie).

Al vantaggio di assicurare un imponente incremento di dati geo-spaziali, nei *big data* si associa lo svantaggio che le caratteristiche dei dati medesimi ne rendono complesso un utilizzo scientificamente corretto. Rispetto ai dati di una rilevazione campionaria, i *big data* appaiono carenti specialmente quanto ad affidabilità, rappresentatività, e soprattutto quanto a contenuto informativo. In estrema sintesi, la minore affidabilità deriva innanzitutto dall'impossibilità di associare con certezza

(20) La *sentiment analysis* consiste in tecniche finalizzate a estrarre da brani o testi di diverso tipo informazioni sull'orientamento del soggetto verso determinati temi.

(21) Su una posizione unilateralmente critica verso i *big data* ci sembra si attesti il CENSIS (2013).

il dato a una identità individuale. Nel caso dei *social networks*, in particolare, non è possibile stabilire un'equazione fra *account* e utenti, perché molti utenti sono presenti con falsa identità o con *account* multipli, ovvero lo stesso account può venir utilizzato da persone diverse. La minore rappresentatività è imputabile al fatto che i *big data* non sono inequivocabilmente rapportabili a un universo di riferimento: pensiamo ad esempio allo squilibrio anagrafico negli utenti dei *social network*. Non entrando in gioco criteri riconoscibili di campionamento, la quantità di dati è in sé priva di significato e non basta a rendere casuale o rappresentativo il campione; è quindi arduo applicarvi le consuete tecniche inferenziali di analisi statistica. Quanto al contenuto informativo, i *big data* offrono pochissime possibilità di contestualizzare l'informazione, mancando di norma un legame con le caratteristiche socio-biografiche o con altre dimensioni cognitive del soggetto. Le stesse possibilità di incrementare il contenuto informativo attraverso i filtri forniti dalle società proprietarie dei dati (ad es. le API – *Application Programming Interface* – per ricavare dati da Twitter) sono molto ridotte. La modellizzazione dei fenomeni socio-geografici operata attraverso i *big data* è quindi assai efficace dal punto di vista descrittivo, ma lo è molto meno dal punto di vista esplicativo. Al momento i *big data* non consentono infine di effettuare analisi longitudinali, in quanto le società proprietarie non sono interessate a procedure di archiviazione.

Il tentativo di ovviare a tali carenze incrementando il contenuto informativo del dato è sovente all'origine di forzature interpretative, fondate sulla crescente tendenza a considerare l'informazione circa la localizzazione spaziale del soggetto come una *proxy* importante di tutti gli altri tipi di informazione sociologica, che non abbisognerebbero così di essere misurati (Savage e Burrows, 2007, p. 892); oppure a caratterizzare i luoghi in base ai contenuti o ai *sentiment* dei *tweet*; o ancora a sovrastimare il significato dell'interazione virtuale. È questo il caso quando ad esempio l'intensità dei legami sociali viene interpretata come funzione del numero dei collegamenti telefonici⁽²²⁾, con buona pace di decenni di *social network analysis*. Come sottolineano Boyd e Crawford (2012, pp. 671-672), nessuno dei due tipi di rete sociale affermatasi negli studi sui *social big data* – «articulated network» e «behavioral network»⁽²³⁾ – esprime la forza di un legame sociale nel senso di Granovetter (1973).

Osservazioni conclusive. – Pur entro i limiti di questa breve rassegna, risulta evidente come il ricorso ai *big data* ponga problematiche metodologiche di non

(22) Un esempio di questo tipo è contenuto nell'altrimenti eccellente studio di Pappalardo e altri (2015) sui comportamenti di mobilità.

(23) Si definiscono «articulated» le reti di cui il soggetto compone un elenco di nominativi, indipendentemente dal tipo di legame, ad esempio liste di «amicizia» sui *social networks*, dove l'elenco può comprendere amici, colleghi, celebrità eccetera. Si definiscono «behavioral» le reti costruite a partire da comportamenti registrati indirettamente, ad esempio la compresenza di persone in uno stesso luogo rilevata dalla cella telefonica, oppure su una foto taggata in rete.

poco conto. Tuttavia, se le carenze sopra richiamate vengono tematizzate e consapevolmente affrontate, l'utilizzo dei *big data* da parte delle scienze sociali e specificamente della geografia umana può rivelarsi un'interessante integrazione dei dati raccolti con strumenti più tradizionali (indagini campionarie, ma anche ricerca qualitativa e studio di casi), entro un approccio di attenta triangolazione metodologica ⁽²⁴⁾; l'analisi dei *big data* consente inoltre di avventurarsi su terreni altrimenti non praticabili per la ricerca sociale ⁽²⁵⁾.

Un importante aspetto da considerare per un utilizzo metodologicamente corretto dei *big data* è rappresentato dall'urgenza di migliorare la qualità dei *big data* dal punto di vista delle modalità di gestione, archiviazione, filtraggio eccetera; al riguardo è stata suggerita l'opportunità di costruire modelli istituzionali che facilitino la collaborazione fra l'accademia e gli organismi esterni che acquisiscono e gestiscono *big data* a significativa valenza sociale (Laser e altri, 2014), come compagnie telefoniche, Google, *social networks* eccetera. Il percorso che in tal modo si delinea per gli anni a venire apre senz'altro scenari complessi per la ricerca accademica, che impongono di ragionare costantemente sulla misura in cui il cambio negli strumenti di indagine possa modificare l'oggetto dell'indagine stessa ⁽²⁶⁾. Non si intravedono tuttavia alternative praticabili, e nemmeno auspicabili, a meno di voler scorporare l'università dai processi di trasformazione sociale in corso, e farne un'isola di sapere autoreferenziale e in definitiva inutile.

RIFERIMENTI BIBLIOGRAFICI

- ANDERSON C., *The End of Theory: The Data Deluge makes the Scientific Method Obsolete*, in «Wired Magazine», 2008, 16, 7 (<https://www.wired.com/2008/06/pb-theory/>).
- ANSELIN L., *Spatial Econometrics: Methods and Models*, New York, Kluwer, 1988.
- ANSELIN L., *Local Indicators of Spatial Association-LISA*, in «Geographical Analysis», 1995, 2, pp. 93-115.
- ATTARD J. e altri, *A Systematic Review of Open Government Data Initiatives*, in «Government Information Quarterly», 2015, 4, pp. 399-418.
- BELLO-ORGAS G., J.J. JUNG e D. CAMACHO, *Social Big Data: Recent Achievements and New Challenges*, in «Information Fusion», 2015, 28, pp. 45-59.
- BERRY D.M., *The Computational Turn: Thinking about the Digital Humanities*, in «Culture Machine», 2011, 12.

(24) Per un esempio di questo tipo ci permettiamo di rimandare a un recente studio in ambito turistico condotto presso il Laboratorio di Geografia Sociale dell'Università di Firenze, in cui dati di rilevazione empirica sono stati confrontati ed integrati con una *sentiment analysis* (Loda e altri, 2016).

(25) Fra gli esempi più interessanti ricordiamo gli studi che analizzano l'impatto dei *social networks* sui meccanismi di funzionamento del giornalismo contemporaneo, ad esempio Lotan e altri (2011).

(26) Appare qui particolarmente appropriata la formulazione che Bruno Latour adotta nel riflettere – a proposito di Gabriel Tarde ed Émile Durkheim – su come il ricorso a metodi quantitativi piuttosto che qualitativi si rifletta nella definizione dell'oggetto di studio della sociologia: «Change the instruments, and you will change the entire social theory that goes with them» (2009, p. 155).

- BOLLIER D., *The Promise and Peril of Big Data*, Washington, The Aspen Institute, 2010.
- BOLTER J.D., *The Role of the Intellectual in the Era of Digital Plenitude*, 2010 (www.vive-rediscoltura/digital_plenitude).
- BOYD D. e K. CRAWFORD, *Critical Questions for Big Data*, in *Information*, in «Communication & Society», 2012, 5, pp. 662-679 (<http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>).
- CALABRESE F. e altri, *Real-time Urban Monitoring using Cell Phones: A Case Study in Rome*, in «IEEE Transactions on Intelligent Transportation Systems», 2011, 12, 1, pp. 141-151.
- CASTELLS M., *The Rise of the Network Society*, I, *The Information Age: Economy, Society, and Culture*, Londra, Wiley-Blackwell, 2010.
- CENSIS, *Un mare di numeri senza interpretazione*, in *La società impersonale*, in «Un mese di Sociale», 2013, 2, pp. 35-59.
- CHO E., S.A. MYERS e J. LESKOVEC, *Friendship and Mobility: User Movement in Location-based Social Networks*, in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (San Diego, 2011)*, New York, Association for Computer Machinery Press, 2011, pp. 1082-1090.
- CLAVAL P., *L'evoluzione storica della geografia umana*, Milano, F. Angeli, 1972 (ediz. orig., 1971).
- DE JONGE E., M. VAN PELT e M. ROOS, *Time Patterns, Geospatial Clustering and Mobility Statistics Based on Mobile Phone Network Data*, Washington, Federal Committee on Statistical Methodology, 2012.
- CRANSHAW J. e altri, *The Livehoods Project: Utilizing Social Media to understand the Dynamics of a City*, in *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media (Dublino, 2012)* (<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4682>).
- DIEBOLD F.X., *A Personal Perspective on the Origin(s) and Development of 'Big Data': The Phenomenon, the Term, and the Discipline*, in «Social Science Research Network Electronic Journal», 2012 (https://www.researchgate.net/publication/254946975_A_Personal_Perspective_on_the_Origins_and_Development_of_'Big_Data'_The_Phenomenon_the_Term_and_the_Discipline_Second_Version).
- EAGLE N. e K. GREENE, *Reality Mining: Using Big Data to engineer a Better World*, Boston, The MIT Press, 2014.
- GAMBI L., *Una geografia per la storia*, Torino, Einaudi, 1973.
- GOODCHILD M.F., *Citizens as Sensors: The World of Volunteered Geography*, in «Journal of Geography», 2007, 4, pp. 211-221.
- GRAHAM M. e T. SHELTON, *Geography and the Future of Big Data, Big Data and the Future of Geography*, in «Dialogues in Human Geography», 2013, 3, pp. 255-261.
- GRANOVETTER M.S., *The Strength of Weak Ties*, in «American Journal of Sociology», 1973, 6, pp. 1360-1380.
- GRIBAUDI D., *Contro una critica demolitrice della geografia*, in «Rivista Geografica Italiana», 1963, pp. 245-270.
- GUBBIA J. e altri, *Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions*, in «Future Generation Computer Systems», 2013, 29, pp. 1645-1660.
- HARD G., *Die Störche und die Kinder, die Orchideen und die Sonne*, Berlino-New York, Walter de Gruyter, 1987-1988.
- HARVEY D., *Explanation in Geography*, Londra, Edward Arnold, 1969.

- HEY T., S. TANSLEY e K. TOLLE, *Jim Grey on eScience: A Transformed Scientific Method*, in T. HEY, S. TANSLEY e K. TOLLE (a cura di), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Redmond, Microsoft Research, 2009, pp. XVII-XXXI.
- IACAP–INTERNATIONAL ASSOCIATION FOR COMPUTING AND PHILOSOPHY, *The Computational Turn: Past, Presence, Futures?*, Münster, Verlagshaus Monsenstein und Vannerdat OHG, 2011 (<https://coeckelbergh.files.wordpress.com/2015/03/55.pdf>).
- IDC–INTERNATIONAL DATA CORPORATION, *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*, Framingham, IDC, 2014.
- JOHNSON T.P. e T.W. SMITH, *Big Data and Survey Research: Supplement or Substitute?*, in *Proceedings of the Workshop on Big Data and Urban Informatics (Chicago, 2014)*, Chicago, University of Illinois, 2014, pp. 201-212 (ubdc.ac.uk/media/1286/finalproceedings.pdf).
- KAPLAN A. e M. HAENLEIN, *Users of the World, Unite! The Challenges and Opportunities of Social Media*, in «Business Horizons», 2010, 1, pp. 59-68.
- KITCHIN R., *Big Data and Human Geography: Opportunities, Challenges and Risks*, in «Dialogues in Human Geography», 2013, 3, pp. 262-267.
- KITCHIN R., *Big Data, New Epistemologies and Paradigm Shifts*, in «Big Data & Society», 2014, 1, pp. 1-12.
- KUČERA J., CHLAPEK D. e M. NEČASKÝ, *Open Government Data Catalogs: Current Approaches and Quality Perspective. Technology-Enabled Innovation for Democracy, Government and Governance*, in «Lecture Notes in Computer Science», 2013, 8061, pp. 152-166.
- LABRINIDIS A. e H.V. JAGADISH, *Challenges and Opportunities with Big Data*, in «Proceedings of the VLDB Endowment», 2012, 5, 12, pp. 2032-2033.
- LANEY D., *3-D Data Management: Controlling Data Volume, Velocity and Variety*, in «Application Delivery Strategies», 2001 (<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>).
- LANHAM R., *The Economics of Attention. Style and Substance in the Age of Information*, Chicago, Chicago University Press, 2006.
- LASER D. e altri, *The Parable of Google Flu: Traps in Big Data Analysis*, in «Science», 2014, 343, pp. 1203-1205 (<http://gking.harvard.edu/files/gking/files/0314policyforumff.pdf?m=1394735706>).
- LATOURE B., *Tarde's idea of quantification*, in M. CANDEA (a cura di), *The Social after Gabriel Tarde: Debates and Assessments*, Londra, Routledge, 2009, pp. 145-162 (<http://www.bruno-latour.fr/sites/default/files/116-CANDEA-TARDE-FR.pdf>).
- LEE J.G. e M. KANG, *Geospatial Big Data: Challenges and Opportunities*, in «Big Data Research», 2015, 2, 2, pp. 74-81.
- LODA M. e altri, *Sustainable Destination Plan for the Ancient Cities of Upper Myanmar*, Firenze, Polistampa, 2016.
- LOTAN G. e altri, *The Revolutions were tweeted: Information Flows during the 2011 Tunisian and Egyptian Revolutions*, in «International Journal of Communications», 2011, 5, pp. 1375-1405 (<http://ijoc.org/ojs/index.php/ijoc/article/view/1246>).
- LOUKIDES M., *What is Data Science? The Future belongs to the Companies and People that turn Data into Products*, in «O'Reilly Radar», 2010 (<https://www.oreilly.com/ideas/what-is-data-science>).

- MAUCERI S., *Contro il paradigma scissorio*, intervento al convegno *Sulle tracce dei Big Data. Questioni di metodo e percorsi di ricerca* (Roma, 2015), (<https://www.google.it/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiE8IC27bDJAbUI3g4KHR8qAawQFggbMAA&url=https://www.aismetodologia.it/2Fconv2015%2FMauceri.docx&usq=AFQjCNH8imynaC2rec024HaTgWyewRUPlw&sig2=Pq2AKR-2u7zHn561gC8gHQ/>).
- MAYER-SCHÖNBERGER V. e K. CUKIER, *Big Data: A Revolution That will transform how We live, work, and think*, Boston, Eamon Dolan/Houghton Mifflin Harcourt, 2013.
- McKINSEY GLOBAL INSTITUTE, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, New York, McKinsey & Company, 2011.
- MENNIS J. e D. GUO, *Spatial Data Mining and Geographic Knowledge Discovery. An Introduction*, in «Computers, Environment and Urban Systems», 2009, 33, pp. 403-408.
- MILLER H.J. e M.F. GOODCHILD, *Data-driven Geography*, in «GeoJournal», 2015, 80, pp. 449-461.
- MILLER H.J. e J. HAN, *Geographic Data Mining and Knowledge Discovery: An Overview*, in H.J. MILLER e J. HAN (a cura di), *Geographic Data Mining and Knowledge Discovery*, CRC Press, 2009, pp. 1-26.
- MURRAY A.T., *Quantitative Geography*, in «Journal of Regional Science», 2010, 1, pp. 143-163.
- OECD, *Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by «Big Data»*, in «OECD Digital Economy Papers», 2013, 222.
- OPEN KNOWLEDGE FOUNDATION NETWORK, *Open Definition 2.1*, Cambridge, 2015 (<http://opendefinition.org>).
- OPENSHAW S. e P.J. TAYLOR, *A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem*, in «Statistical Applications in the Spatial Sciences», 1979, 21, pp. 127-144.
- PAPPALARDO L. e altri, *Returners and Explorers Dichotomy in Human Mobility*, in «Nature Communications», 2015, 6 (<http://www.nature.com/ncomms/2015/150908/ncomms9166/full/ncomms9166.html>).
- PERERA C. e altri, *Computing for the Internet of Things: A Survey*, in «IEEE Communications Surveys & Tutorials», 2014, 1, pp. 414-447.
- SAVAGE M. e R. BURROWS, *The Coming Crisis of Empirical Sociology*, in «Sociology», 2007, 5, pp. 885-899.
- SHEKHAR S. e altri, *Spatiotemporal Data Mining: A Computational Perspective*, in «ISPRS International Journal of Geo-Information», 2015, 4, pp. 2306-2338.
- SOTO V. e E. FRÍAS-MARTÍNEZ, *Robust Land Use Characterization of Urban Landscapes using Cell Phone Data*, in *First Workshop on Pervasive Urban Applications (San Francisco, 2011)* (<http://www.cs.columbia.edu/~vsoto/>).
- SPIELMAN S.E. e J.C. THILL, *Social Area Analysis, Data Mining, and GIS*, in «Computers, Environment and Urban Systems», 2008, 32, pp. 110-122.
- STEENBRUGGEN J. e altri, *Mobile Phone Data from GSM Networks for Traffic Parameter and Urban Spatial Pattern Assessment: A Review of Applications and Opportunities*, in «GeoJournal», 2013, 2, pp. 223-243.
- STEENBRUGGEN J., E. TRANOS e P. NIJKAMP, *Data from Mobile Phone Operators: A Tool for Smarter Cities?*, in «Telecommunications Policy», 2015, 39, pp. 335-346.
- SUI D.Z., *Tobler's First Law of Geography: A Big Idea for a Small World?*, in «Annals of the Association of American Geographers», 2004, 94, 2, pp. 269-277.

- TAYLOR P.J., *(In)disciplina*, in E. DELL'AGNESE, *Geo-grafia. Strumenti e parole*, Milano, UNICOPLI, 2009, pp. 25-38.
- TOBLER W., *A Computer Movie simulating Urban Growth in the Detroit Region*, in «Economic Geography», 1970, 46, pp. 234-240.
- TURCO A., *Geografia e scienze umane*, in G. CORNA PELLEGRINI (a cura di), *Aspetti e problemi della geografia*, Settimo Milanese, Marzorati, 1987, II, pp. 85-130.
- UN-GGIM-UNITED NATIONS INITIATIVE ON GLOBAL GEOSPATIAL INFORMATION MANAGEMENT, *Future Trends in Geospatial Information Management: The Five to Ten Year Vision*, New York, UN-GGIM, 2013.
- WANG W., *Using Location-Based Social Media for Ranking Individual Familiarity with Places: A Case Study with Foursquare Check-in Data*, in G. GARTNER e H. HUANG (a cura di), *Progress in Location-Based Services 2014*, Cham, Springer International Publishing, 2015, pp. 171-183.
- WOODIE A., *5 Ways Big Geospatial Data is driving Analytics in the Real World*, 2015 (<http://www.datanami.com/2015/05/21/5-ways-big-geospatial-data-is-driving-analytics-in-the-real-world/>).
- WU W. e J. WANG, *Exploring City Social Interaction ties in the Big Data Era: Evidence Based on Location-based Social Media data from China*, in *55th Congress of the European Regional Science Association. «World Renaissance: Changing roles for people and places» (Lisbona, 2015)* (<https://www.econstor.eu/handle/10419/124698>).
- WYRWOLL C., *Social Media. Fundamentals, Models, and Ranking of User-Generated Content*, Wiesbaden, Springer Vieweg, 2014.
- ZUIDERWIJK A. e M. JANSSEN, *Open Data Policies, Their Implementation and Impact: A Framework for Comparison*, in «Government Information Quarterly», 2014, 31, pp. 17-29.

FROM INFORMATION TO KNOWLEDGE: GEOGRAPHY IN THE ERA OF DATA PLENITUDE. – The increasing use of digital equipment and new media is currently resulting in the generation of very big streams of data, most of them enriched by geospatial reference. This availability of large amounts of data can generate both opportunities and challenges for geographic research. In this paper, we examine the characteristics of such data proliferation and investigate its main methodological implications. After recalling some significant milestones of the social geography methodology and summarizing the issues linked to the data growth, we highlight some main benefits and weak points of using spatial big data for geographic and social research.

Università di Firenze, Dipartimento di Studi Storici Archeologici Geografici dell'Arte e dello Spettacolo – SAGAS

mirella.loda@unifi.it

Università di Firenze, Laboratorio di Geografia Sociale

mario.tartaglia@unifi.it