

# 1. BIG DATA E SOCIAL MINING: I DATI, A SAPERLI ASCOLTARE, RACCONTANO STORIE.

*In questo articolo ci soffermiamo alcune alcune importanti sorgenti di Big data e le opportunità che ne derivano per nuovi processi analitici. Presentiamo poi una selezione di esperienze pilota di analisi di datamining su Big data finalizzate alla comprensione della complessità di fenomeni sociali. Discuteremo inoltre gli aspetti etici più critici che emergono dall'analisi di Big data.*

## 1. Opportunità e rischi dei Big data

Nel mondo che abitiamo, oggi abbiamo l'opportunità di osservare da vicino e misurare il funzionamento della società attraverso i "Big data", le briciole digitali che le nostre attività quotidiane lasciano per effetto del nostro uso dei sistemi ICT. Briciole che registrano la nuda verità sui comportamenti individuali e collettivi con una precisione senza precedenti, in modo che le diverse dimensioni della nostra vita sociale trovano un'immagine riflessa nello specchio digitale: desideri, opinioni, stili di vita, movimenti, relazioni.

I nostri desideri, opinioni, sentimenti lasciano traccia nei social media a cui partecipiamo, nelle domande che facciamo ai motori di ricerca, nei tweet che inviamo e riceviamo, così come i nostri stili di vita lasciano traccia nei record dei nostri acquisti. I nostri movimenti lasciano traccia nelle traiettorie disegnate dai nostri smart-phone e dai sistemi di navigazione delle nostre auto.

Anche le nostre relazioni sociali lasciano traccia nella rete dei nostri contatti telefonici e delle email e nei link di amicizia del nostro social network preferi-



to. Possiamo cominciare a esplorare la rete di relazioni che costituisce la nostra società, il tessuto sociale e la sua robustezza o debolezza.

I Big data sono il nuovo microscopio che rende “misurabile” la società. Come la scoperta di ogni nuovo microscopio o telescopio nel passato, i Big data stanno spingendo verso una nuova scienza dei dati: il *social mining*, in grado di misurare e, in prospettiva, prevedere crisi economiche, epidemie e pandemie, diffusione di opinioni, distribuzione delle risorse economiche o energetiche, bisogni di mobilità.

Certo, bisogna tenere conto della qualità dei dati e della loro rappresentatività. Certo, bisogna essere consapevoli delle grandi opportunità così come dei nuovi rischi: occorrono tecnologie a sostegno della privacy, occorre un “new deal” sui temi della privacy, della trasparenza e della fiducia per far sì che l’accesso alla conoscenza dei Big data sia bene pubblico per tutti. Certo, bisogna superare la fase attuale, in cui la maggior parte dei Big data interessanti sono tutt’altro che “open” e se ne stanno chiusi nei database delle web corporations e degli operatori telecom. Certo, questi problemi – qualità, privacy e proprietà dei Big data – sono decisivi.

Sicuramente questo nuovo percorso ha forti ripercussioni su molti aspetti etici: privacy e protezione dei dati personali (chi può accedere ai miei dati?), proprietà e sfruttamento dei dati (Di chi sono i miei dati? Per quali scopi vengono usati?), trasparenza (Chi può fare cosa con i miei dati?), consapevolezza e conoscenza di sé (Come posso accedere alla conoscenza nascosta nei miei dati), monopoli ed asimmetrie (come controbilanciare il potere delle grandi corporation della conoscenza). Nel nostro mondo interconnesso non possiamo permetterci di perdere l’opportunità offerta dai Big data, ma dobbiamo trovare un nuovo ecosistema socio-tecno-legale in cui la conoscenza sia un bene comune sicuro .

## **2. Le sorgenti di Big data: gli effetti laterali dei servizi ICT che usiamo**

Il *social mining* si basa su collezioni massive di tracce digitali di attività umane mediate dalla tecnologia ICT. Spesso tali tracce sono le semplici memorizzazioni delle richieste di accesso a servizi: i cosiddetti “log” di tali servizi. Sono quindi registrazioni che memorizzano il contenuto minimale per attivare il servizio. Spesso tali registrazioni consistono di un identificativo del richiedente, di un time-stamp temporale, talvolta anche spaziale, talvolta un piccolo contenuto testuale o multimediale. In altri casi i dati sono i contenuti

veri e propri dei servizi. La sfida sta nell'utilizzare questi dati per scopi diversi dal servizio come specchio di un qualche fenomeno sociale da osservare e capire.

Molti degli esperimenti in letteratura si basano sull'utilizzo di dataset messi a disposizione della comunità scientifica con diverse modalità di accesso:

- *accesso aperto*: ad esempio i log di accesso ad un servizio "open" su internet da parte di un identificativo di utente (es. Europea), le foto geolocalizzate condivise sui social network su internet (es. Flickr), i log delle reti sociali (es. Foursquare, Twitter, etc.) e ovviamente gli Open Data disponibili sia nelle pubbliche amministrazioni che nel Linking Open Data cloud. Buoni strumenti di "crawling" su rete possono realizzare velocemente raccolte massive di questi dati, oppure i servizi stessi mettono a disposizione delle funzionalità (API) per scaricarli.

- *accesso ristretto*: sono il risultato di accordi tra ricercatori ed aziende per condurre sperimentazioni con un limitato diffusione dei risultati e non condivisione con la comunità scientifica dei dati. Esempi sono i record di utilizzo dei telefoni cellulari (GSM CDR – Call Detail Records datasets) messi a disposizione da alcuni operatori telefonici, le tracce dei dispositivi GPS sia su veicoli che su smart phone. Esempio del secondo tipo sono i dispositivi di tracking forniti da compagnie assicurative (in Italia l'azienda Octotelematics, che offre il servizio per diverse compagnie serve dal 3% to 7% di veicoli private registrati). Esistono esperimenti con datasets della compagnia TomTom su flotte di taxi (un dataset a NewYork ed uno a Shangay), datasets di query di motori di ricerca, comunità di Facebook che danno l'accesso ai loro profile ed alla rete delle loro amicizie, raccolte di e-mail da volontari, accessi ai servizi di Skype.

- *Participatory crowd sensing*: una modalità nuova di realizzare raccolta dati ed indagini coinvolgendo attivamente gruppi di utenti volontari nell'esprimere fatti, opinioni usando sistemi di gioco con scopo ed attività collaborative immerse in realtà più o meno virtuali.

### **3. Esperienze di Big data analytics per analisi sociali**

L'uso di Big data per misurare e comprendere fenomeni sociali è un'area recente, ma molto vivace. Illustriamo qui un elenco non esaustivo di prime esperienze.

### 3.1 Nowcasting con le tracce delle ricerche su Web

Vari studi hanno esaminato la capacità predittiva dei record di interrogazioni sui motori di ricerca web (query logs), come quelli messi a disposizione dal servizio di Google Trends, che fornisce il volume giornaliero e settimanale dei termini di ricerca più usati dagli utenti di Google <sup>1</sup>. I ricercatori Choi e Varian [Choi 2009] di Google hanno introdotto il termine "predire il presente", o "nowcasting", in contrapposizione a "forecasting", per indicare la correlazione che spesso esiste tra ricerche su web e vari indicatori socio-economici, correlazione che può rivelarsi utile per previsioni a breve termine. Il rilascio di indicatori di attività socio-economiche avviene in genere con un ritardo di diverse settimane o mesi, ed è quindi chiaramente utile avere stime più tempestive di questi indicatori.

Un esempio popolare in ambito epidemiologico è lo studio in [Ginsberg 2008] che ha dimostrato come i query logs possano predire l'incidenza delle malattie influenzali. Intuitivamente, si osserva una stretta correlazione tra il numero di persone che cercano in rete argomenti riguardanti l'influenza e il numero delle persone che in realtà hanno sintomi influenzali. Gli autori hanno confrontato i conteggi delle query con sistemi di sorveglianza dell'influenza tradizionali ed hanno scoperto che molte di queste query tendono ad essere molto popolari in coincidenza con la stagione influenzale. E' quindi possibile stimare l'incidenza del fenomeno influenzale nei diversi paesi e regioni di tutto il mondo in tempo quasi reale. Questa scoperta è importante perché le agenzie di sorveglianza sanitaria elaborano le loro stime con un ritardo di settimane, mentre la tempestività è fondamentale per consentire agli operatori sanitari di rispondere meglio alle epidemie stagionali e alle pandemie. Altri studi di nowcasting basati sui dati di ricerca web hanno interessato la stima di disoccupazione, inflazione, domanda commerciale etc.

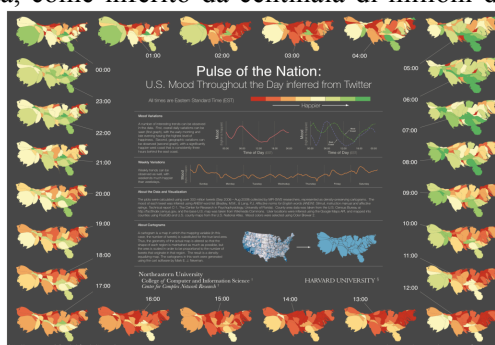


Attività influenza in US. Blu: Stima di Google Flu Trends; Arancio: Dati ufficiali Influenza-like illness data (ILI) forniti da U.S. Centers for Disease Control.

<sup>1</sup> <http://www.google.com/trends/>

### 3.2 *Misurare la felicità con le tracce nei Social Media*

Una direzione popolare di ricerca focalizza sull'analisi del contenuto di Social Media, in particolare Twitter per misurare e monitorare lo stato emotivo di una popolazione a varie scale geografiche e temporali. Diversi autori hanno proposto di utilizzare i tweets come proxy per la felicità, utilizzando metodi per mappare i 140 caratteri di testo in ogni tweet in uno stato emotivo, che viene poi aggregato utilizzando gli ancoraggi geografici e temporali. Un primo esempio è il progetto Twittermood<sup>2</sup>, volto a mappare lo stato d'animo negli Stati Uniti su tutta la giornata, come inferito da centinaia di milioni di tweets. Il contenuto di ogni tweet è mappato in uno stato emotivo usando il metodo ANEW (Affective Norms for English Words) che fornisce rating emotivo per un gran numero di parole in lingua inglese. Le valutazioni sono aggregate per stato degli Stati Uniti e ora del giorno, in modo che la semplice infografica riveli pattern evidenti.

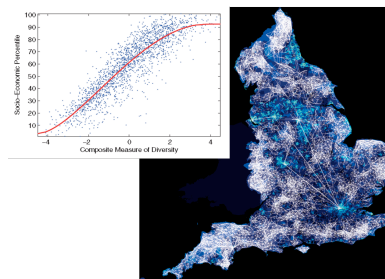


### 3.3 *Misurare il benessere con le tracce delle Reti Sociali*

Una direzione di ricerca emergente è quella di utilizzare concetti dall'analisi di reti sociali e relative misure inferite da record di attività telefonica rilevati a livello nazionale, al fine di stimare indicatori socio-economici di povertà, benessere e progresso sociale. La logica è che le reti sociali modellano il tessuto della società e costituiscono la spina dorsale della vita sociale ed economica: i dati di telefonia espongono la struttura macroscopica della rete sociale di una intera nazione e permette quindi di studiare a società con un approccio quantitativo. Uno studio pionieristico in questa linea è quello in [Eagle 2010], dove gli autori combinano una registrazione completa di una rete di comunicazione nazionale con dati nazionali di censimento sul benessere delle comunità. Gli autori mostrano che la varietà delle relazioni individuali è fortemente correlata con lo sviluppo economico delle comunità locali.

<sup>2</sup> <http://www.ccs.neu.edu/home/amislove/twittermood/> Alan Mislove, Sune Lehmann, Yong-yeol Ahn, Jukka-Pekka Onnela, J. Niels Rosenquist. Northeastern University, 2010.

L'ipotesi è che la varietà dei propri contatti sociali sia proporzionale all'accesso alle opportunità. La figura mostra la forte correlazione misurata, tra la varietà dei contatti sociali della rete e l'Indice di deprivazione multipla del governo Inglese, una misura composta di prosperità, in base al reddito, occupazione, istruzione, salute, criminalità, alloggi e la qualità ambientale di ogni regione.



#### 4. Comprendere la mobilità umana con i Big data.

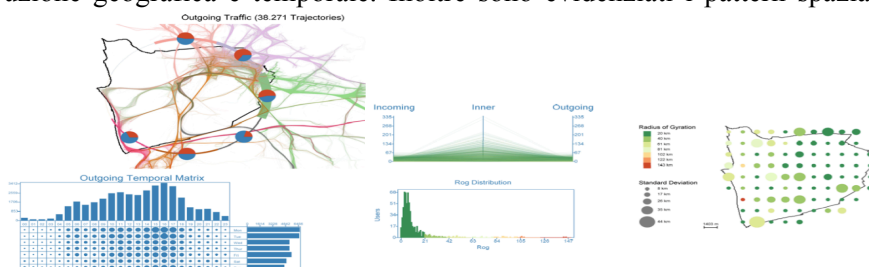
L'ingrediente chiave per il rinnovato interesse nell'analisi del funzionamento delle città emerso durante l'ultimo decennio è la disponibilità di Big data relativi alla mobilità umana, favorito dalla diffusione capillare delle tecnologie wireless, come il Global Positioning System (GPS) abilitato dai satelliti e le reti di telefonia mobile. Queste infrastrutture di rete, come sottoprodotto del loro normale esercizio, consentono il rilevamento e la raccolta di enormi archivi di dati spatio-temporali, quali i record delle chiamate provenienti da telefoni cellulari e le tracce GPS dei dispositivi di navigazione, che rappresentano i proxy a livello globale per le attività mobili umane. Questi dati possono aiutarci a capire i pattern e modelli che caratterizzano le traiettorie seguite dalle persone durante la loro attività quotidiana. Su questa base, è fiorita una vivace linea di ricerca che ha attratto scienziati di diverse discipline, non solo per la grande sfida intellettuale ma anche per l'importanza in settori applicativi quali la pianificazione urbana, la mobilità sostenibile, l'ingegneria dei trasporti, la salute pubblica, e la previsione economica [Giannotti 2008, Batty 2012].

Gli analisti ragionano su concetti di alto livello, come sistematicità o occasionalità del movimento, scopo di un viaggio, e caratteristiche di pendolari, residenti, visitatori. Di conseguenza, gli strumenti analitici tradizionali di ingegneria dei trasporti, come ad esempio le matrici origine / destinazione, si basano su dati semanticamente ricchi raccolti attraverso indagini sul campo e interviste. Non è quindi ovvio capire come usare Big data per superare i limiti delle indagini tradizionali, ossia il loro costo elevato, la periodicità infrequente, la rapida obsolescenza, l'incompletezza e l'imprecisione. D'altra parte, i dati di mobilità rilevati automaticamente fotografano la nuda verità: le attività mobili reali sono direttamente e continuamente campionate senza avere però

alcuna annotazione semantica o di contesto. Molti ricercatori hanno cominciato ad investigare come colmare questo divario. Esperimenti su larga scala stanno dimostrando come sia possibile rispondere a molte domande analitiche complesse sul comportamento in mobilità sulla base di Big data.

#### 4.1 Costruire l'Atlante della mobilità Urbana dai dati GPS

I dati, a saperli ascoltare, raccontano storie. Prendiamo ad esempio un territorio come la Toscana, osserviamo per qualche tempo i viaggi compiuti da un campione di automobilisti registrandone i viaggi e le soste. In uno dei nostri esperimenti abbiamo analizzato le tracce digitali lasciate da auto equipaggiate con una “scatola nera” che, per conto della compagnia di assicurazione dell’auto, registra la traiettoria di ogni viaggio attraverso il segnale di localizzazione via satellite (GPS, Global Positioning System): in 5 settimane, osservando il territorio intorno a Pisa, Lucca e Livorno, abbiamo contato circa 40.000 auto che hanno compiuto un milione e mezzo di viaggi, percorrendo oltre 16 milioni e mezzo di chilometri (più di 400 volte la lunghezza dell’equatore!) per una durata complessiva di oltre 60 anni. Il sistema di *mobility data mining* in [Giannotti 2011] è stato utilizzato su questi dati per creare un atlante<sup>3</sup> della mobilità urbana, cioè un catalogo completo dei comportamenti di mobilità in una città, sviluppata per tutte le città della Toscana, sulla base di dati GPS installati su veicoli privati che copre circa il 7 % del autovetture private registrate, le cui tracce GPS sono state osservate per un lungo periodo. Ogni città è rappresentata attraverso una infografica che illustra nuovi indicatori di mobilità, come il raggio di girazione dei residenti (la distanza media quadratica di ogni persona dalla sua posizione preferita) e la sua distribuzione geografica e temporale. Inoltre sono evidenziati i pattern spaziali e



L'Atlante della Mobilità Urbana

<sup>3</sup> <http://kdd.isti.cnr.it/uma/>

temporali dei viaggi dei pendolari (in ingresso ed in uscita), insieme con l'impatto del pendolarismo e della mobilità non sistematica sulla città (vedi figura). L'atlante della mobilità urbano è un punto di partenza per caratterizzare diverse città rispetto ai loro profili di mobilità.

#### 4.2 *Stimare i flussi di movimento dai dati della telefonia mobile*

I dati, a saperli ascoltare, raccontano storie. I record delle comunicazioni della telefonia cellulare forniscono un'interessante immagine sia di traiettorie individuali che di relazioni sociali, infatti mantengono la traccia di ogni telefonata tra due utenti e la localizzazione nello spazio e nel tempo dell'utente che avvia la chiamata.

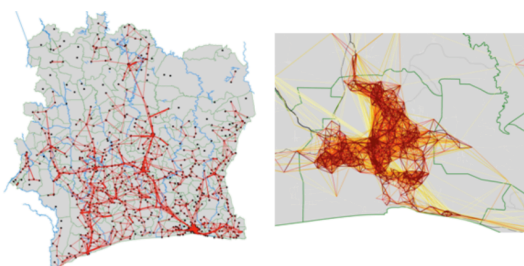


Figure 12: Mobile phone movements in Ivory Coast and Abidjan.

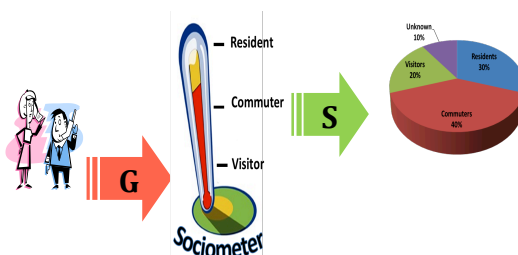
L'alta penetrazione dei telefoni cellulari fa sì che tali dati catturino una grande frazione della popolazione di un intero paese. La disponibilità di enormi quantità di CDR (Call Detail Records) ha attirato l'interesse di molti ricercatori in molti campi con una varietà di risultati interessanti. Recentemente, Orange France Telecom ha reso disponibile alla comunità scientifica un anno di dati di telefonia mobile di tutti gli abitanti della Costa d'Avorio in Africa. Lo studio condotto in [Nanni 2013] ha dimostrato che con questi dati è possibile stimare con precisione i flussi di mobilità e sostenere la creazione di strumenti di pianificazione sostenibili, anche in un paese in via di sviluppo, senza una infrastruttura di sensori sulle strade. I risultati empirici di questo studio mostrano come creare un modello della domanda di trasporto per la Costa d'Avorio e le sue principali città, stimando una matrice origine-destinazione dei flussi di mobilità dai dati di telefonia mobile, in combinazione con l'uso di dati della rete stradale accessibili al pubblico e un software di modellazione di trasporto standard. I risultati indicano che una stima continua ed affidabile dei flussi di mobilità dai dati GSM è a portata di mano .



### 4.3 Quantificare la popolazione urbana dai dati della telefonia mobile

I dati, a saperli ascoltare, raccontano storie. Un altro esempio di come utilizzare i dati GSM per il monitoraggio continuo di indicatori demografici è il *sociometer* della popolazione urbana proposto (Furletti et al 2012.). Il *sociometer* è finalizzato a stimare la percentuale di city users che rientrano in tre categorie: residenti, pendolari, e visitatori.

In questo studio, un dataset di dati di chiamata di telefonia mobile (CDR) è utilizzato per caratterizzare i profili di chiamata delle persone osservate in un'area urbana per un arco



temporale di cinque settimane. Il profilo di chiamata di un utente cattura le sue abitudini di chiamata durante i giorni feriali o nei fine settimana nell'intero periodo di osservazione: quanto appare durante le prime ore del mattino, durante l'orario di lavoro, o le ore notturne. Per mezzo di una campagna di rilevamento mirato, alcuni dei profili di chiamata disponibili sono classificati secondo le tre classi; il classificatore è costruito su un sottoinsieme di dati annotati utilizzando tecniche di apprendimento automatico / data mining, poi viene applicato a tutta la popolazione di utenti di telefonia mobile disponibile nel dataset, al fine di monitorare continuamente la percentuale di residenti, pendolari e visitatori in città. Sorprendentemente, i profili di chiamata delle tre classi di utenti della città sono stabili nel tempo: i residenti chiamano essenzialmente in qualsiasi momento, i pendolari tendono a chiamare solo durante i giorni feriali e orari di lavoro, i visitatori chiamano sporadicamente. Al contrario, la proporzione tra le tre categorie tendono può variare in modo significativo nel corso del tempo, e monitorare queste variazioni è una informazione importante per la pianificazione di servizi urbani, come la gestione dei rifiuti, l'approvvigionamento energetico o con i mezzi pubblici.

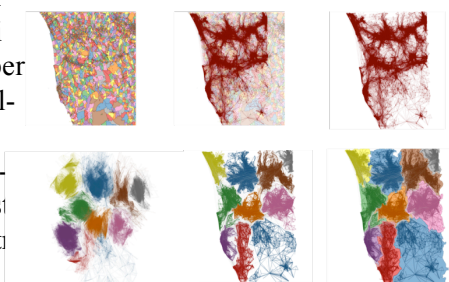
### 4.4 Scoprire i bordi geografici della mobilità umana

I dati, a saperli ascoltare, raccontano storie. Per riuscire a raccogliere una storia, però, bisogna porre le domande giuste. Con le persone, così come con i dati che registrano aspetti della vita delle persone. Bisogna fare in modo che

la conoscenza nascosta in questo apparente caos, nell'intricata trama dei nostri viaggi, venga fuori. Un esempio di domanda di questo tipo è: quali sono i confini geografici della mobilità umana? I nostri spostamenti possono suggerirci come delimitare le aree entro cui la nostra attività si svolge prevalentemente? In altre parole, è possibile far emergere gli "spartiacque" della mobilità dalle tracce digitali dei nostri viaggi?

Il problema di scoprire i confini geografici delle attività umane a bassa risoluzione spaziale dei comuni o delle province è un problema di vasta portata, motivato dalla necessità di fornire ai responsabili politici suggerimenti sui migliori partizioni amministrative per il governo del territorio.

In (Rinzivillo 2012), si utilizzano metodi analitici di rete applicati a dati di mobilità sfruttando le connessioni nascoste che la mobilità umana stabilisce tra luoghi diversi. La figura illustra il processo analitico. Partendo da una



suddivisione iniziale del territorio in zone censimento, gli autori costruiscono una rete i cui nodi sono le zone e gli archi pesati tra due zone rappresentano il numero di viaggi originari primo e terminando nel secondo. La fase di analisi consiste nella scoperta di aree densamente connesse per mezzo di un metodo di scoperta comunità, quindi si aggregano in comunità quelle zone che sono fortemente collegate da molti viaggi mentre rimane bassa la connettività tra diversi comunità / cluster evidenziando. Quando si torna alla geografia, queste comunità suggeriscono confini ben definiti, delimitano i bacini di mobilità dettate dalla vera motivazione della permanenza umana.

## 5. Conclusioni

Da un lato il pubblico chiede di veder più fatti e preme per la trasparenza dell'informazioni delle pubbliche amministrazioni; dall'altro lato i cittadini chiedono che la sfera individuale sia protetta. Che cosa deve rimanere un segreto e cosa no? Come garantire trasparenza e privacy? Un punto chiave è adottare l'approccio Privacy-by design per far coesistere l'analitica con la protezione delle informazioni personali nei Big data. C'è ormai una tecnologia solida che consente di costruire servizi analitici, per esempio di info-mobilità, a partire da dati trasformati in modo che non rilevare alcuna informazione

sensibile degli utenti a cui i dati si riferiscono. Ma più in generale è necessario un cambio di paradigma nel concetto di proprietà del dato personale, che ponga il singolo individuo al centro del gioco restituendogli trasparenza e diritti. In una parola democratizzare i Big data.

Questo è un vero e proprio new deal dei dati personali che realizza alcuni principi di base:

- *Self-awareness*: l'utente ha pieno controllo e consapevolezza dei propri dati personali e della conoscenza che può essere estratta da questi. Quindi una evoluzione da "consenso informato" a "consapevolezza".
- *Data liberation*: la possibilità di fare dei propri dati ciò che vogliamo, ad esempio ritirarli da un qualunque operatore. (data portability).
- *Oblivo*: il diritto di richiedere che i dati vengano dimenticati
- *Public good*: il diritto di avere accesso alla conoscenza collettiva come bene comune.

La grande sfida è come progettare ecosistemi per i dati personali che diano a tutti la possibilità di gestire la propria informazione personale e l'interscambio con le entità esterne, persone ed istituzioni, promuovendo self-awareness, trasparenza e trust.

Una ricca sorgente di articoli, news, e opinioni legali su questi temi si trova sul Privacy Observatory Magazine realizzato nell'ambito del progetto Europeo MODAP: Mobility, Data Mining and Privacy: [www.modap.eu](http://www.modap.eu). Il magazine è disponibile online a <http://www.privacyobservatory.org> ed ha un board scientifico che integra competenze tecnologiche, sociali e legali sui temi della data privacy e data protection, e più in generale di aspetti etici dei Big data.

Ultima considerazione ma non meno importante è il fatto che sta emergendo una nuova figura professionale: il "data scientist" capace di combinare competenze di programmatore software, competenze di statistico e capacità giornalistiche da "story teller" per estrarre pepite d'oro da montagne di dati.

Hal Varian, un economista di Google predice che questa figura professionale diventerà la più affascinante nei prossimi anni: i dati saranno sempre più largamente disponibili mentre è ancora molto scarsa la capacità di estrarre "saggezza" da essi. L'Economist nel 2010 ha dedicato un numero speciale ai dati intitolato: "Data, data everywhere".

E' quindi dovere della comunità scientifica ed accademica di promuovere iniziative formative di carattere interdisciplinari tra ICT, statistica, scienze umane e scienze economiche e sociali: la Laurea magistrale (MSc) in Business Informatics, dell'Università di Pisa va in questa direzione.

## 6. Bibliografia richiamata nel testo (esempio)

[Rinzivillo 2012] Salvatore Rinzivillo, Simone Mainardi, Fabio Pezzoni, Michele Coscia, Dino Pedreschi, Fosca Giannotti: Discovering the Geographical Borders of Human Mobility. *Künstliche Intelligenz* 26(3): 253-260 (2012)

[Giannotti 2008] Fosca Giannotti, Dino Pedreschi. *Mobility, data mining and privacy*. Springer (2008)

[Giannotti 2011] F Giannotti, M Nanni, D Pedreschi, F Pinelli, C Renso, S Rinzivillo, R Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal* 20(5), 695-719 (2011)

[Eagle 2010] Nathan Eagle, Michael Macy, and Rob Claxton. Network Diversity and Economic Development. *Science* 328 (5981), 1029-1031 (2010)

[Choi 2012] Choi, H. and Varian, H., Predicting the Present with Google Trends. *Economic Record*, 88: 2–9. doi: 10.1111/j.1475-4932.2012.00809.x (2012)

[Ginsberg 2008] Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, et al. Detecting influenza epidemics using search engine query data. *Nature* 457: 1012–1015. (2008)

[Batty 2012] M. Batty, K.W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Baz-zani, M. Wachowicz, G. Ouzounis and Y. Portugali. Smart cities of the future. *Eur. Phys. J. Special Topics* 214, 481-518 (2012) <http://dx.doi.org/10.1140/epjst/e2012-01703-3> (Open access)

[Nanni 2013] Mirco Nanni, Roberto Trasarti, Barbara Furletti, Lorenzo Gabrielli, Peter Van Der Mede, Joost De Bruijn, Erik De Romph, Gerard Bruil. MP4-A Project: Mobility Planning For Africa. *NetMob 2013, Third conference on the Analysis of Mobile Phone Datasets* (2013)

[Furletti 2012] B. Furletti, L. Gabrielli, C. Renso, and S. Rinzivillo. Identifying users profiles from mobile calls habits. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing, UrbComp '12*, pages 17–24, New York, NY, USA. ACM Press (2012).

## NOTE SUGLI AUTORI



Fosca Giannotti è dirigente di ricerca all'Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" del CNR a Pisa dove coordina il laboratorio Knowledge Discovery & Data Mining – KDD Lab.". I suoi interessi di ricerca includono mobility data mining, privacy-preserving data mining e social network analisi. Ha coordinato alcuni progetti Europei incluso il progetto FP6 GeoPKDD premiato al parlamento Europeo nel 2010. E' autrice di oltre 150 lavori scientifici e contribuisce attivamente all'indirizzo scientifico delle maggiori conferenze internazionali nel settore data mining e machine learning.