# HEALTH INDUSTRY

## OPEN DATA IN SANITÀ E BANCHE DATI

OPEN GOVERNMENT E OPEN DATA

UNA CONVIVENZA POSSIBILE?

I LINKED DATA

Il data store del Ministero della Salute

Open data in Europa

KEKOA PUBLISHING

HEALTH INDUSTRY


**CARLOTTA GALEONE**

Dottorato in Statistica Metodologica e Applicata presso l'Università degli Studi di Milano Bicocca, si occupa da più di 15 anni di ricerca in ambito sanitario ed epidemiologico. Autrice di oltre 100 articoli scientifici su riviste indicizzate e vari capitoli di libro, ha sviluppato una grande esperienza nel campo nell'area salute con molti anni di collaborazione con il Dipartimento di Epidemiologia dell'Istituto Mario Negri di Milano. Ora ha una posizione Post-Doc presso l'Università degli Studi di Milano, dove si occupa di analisi di grandi banche dati. È Docente di Biostatistica ed Epidemiologia presso il corso di perfezionamento in Market Access in Life Science presso l'Università di Milano Bicocca.

**PAOLO MARIANI**

Paolo Mariani è Professore Ordinario di Statistica Economica presso l'Università degli Studi di Milano Bicocca. Insegna Statistica Aziendale e Analisi di Mercato. La sua attività di ricerca a fini applicativi di impresa è attualmente orientata ai Big Data e alle tecniche di accesso ai mercati, argomenti sui quali ha scritto articoli e monografie. È relatore in occasione d'incontri nazionali e internazionali.

**MARCO MORANDI**

Master al MIP del Politecnico di Milano, Marco Morandi si occupa da circa 20 anni di servizi per il settore farmaceutico e sanitario e ha sviluppato una grande esperienza nell'area comunicazione e marketing. Docente di Marketing Farmaceutico presso la Business School del Sole 24 Ore, appassionato d'innovazione, fa parte del Consiglio di Amministrazione di Innovits, no profit a supporto delle startup. Al momento ricopre il ruolo di BU Director per PKE, una business unit di DS Medigroup, dove si occupa di banche dati sanitarie, profilazione e business development.


# *Focus*

## • OPEN DATA IN SANITÀ E BANCHE DATI PROFESSIONALI

### Una convivenza possibile?

- Open. Un concetto multiforme
- Open Government e Open Data
- Il loro identikit
- I linked data
- Open data in Italia
- Open data in Europa
- Gli ambiti di applicazione in Sanità
- Il data store del Ministero della Salute
- Gli open data e le banche dati professionali

# OPEN DATA IN SANITÀ E BANCHE DATI PROFESSIONALI

## UNA CONVIVENZA POSSIBILE?

**Il concetto di open** appartiene ormai a molti campi della nostra realtà. Lo troviamo applicato nell'ambito dell'implementazione dei software, dove il concetto di open source - nato nel 1998 – ha dato modo agli sviluppatori di migliorare codici scritti da altri. Nella scienza e nella tecnica rappresenta uno dei concetti portanti relativi all'innovazione; non a caso si parla di open innovation. Open è un concetto fondamentale anche della ricerca medica, in particolare di quella virologica, e caratterizza molte fasi della ricerca farmaceutica, ambito nel quale anche grandi e consolidate aziende multinazionali sviluppano progetti di open innovation. Il "movimento degli open" si avvia a operare profonde trasformazioni, in particolare diffondendo e consolidando altri valori importanti delle nostre attività sociali, come la trasparenza, la condivisione e il lavoro in gruppo. In questo features magazine cercheremo di chiarire cosa siano gli open data e, dopo aver richiamato i principi dell'Open Goverment, l'attenzione si focalizzerà sugli open data nel contesto sanitario e farmaceutico, facendo il punto sullo stato dell'arte in Italia e ipotizzando una relazione rispetto all'attuale dotazione aziendale di banche dati professionali.

# Open Government e Open Data

**Il concetto di Open Government** è stato introdotto per la prima volta in modo strutturato dal governo Obama nel 2009 e, anno dopo anno, è diventato un principio cardine delle azioni di governo per oltre 70 Paesi nel mondo, Africa compresa. L'Open Government si basa sul principio che tutte le attività del Governo e delle Amministrazioni dello Stato devono essere aperte e disponibili *(concetto di trasparenza)* per favorire azioni efficaci e garantire un controllo diffuso sulla gestione del patrimonio pubblico da parte dei cittadini. In uno scenario a lungo termine, questo nuovo processo condurrà a un nuovo modello d'interazione e collaborazione fra Stato e cittadini, dove il singolo cittadino non sarà più solo un fruitore di prestazioni erogate da parte della Pubblica Amministrazione, ma parteciperà in modo attivo alle scelte del governo. Gli Open Data sono lo strumento chiave per tradurre il concetto di Open Government in un vero e proprio modello sostenibile, in cui il cittadino può valutare le decisioni prese dalla Pubblica Amministrazione. Esistono varie definizioni di Open Data. Una di quelle comunemente accettata è riportata dall'Open Data Manual (http://opendatamanual.org/ - Documento redatto dall'Open Knowledge Foundation), che li definisce come "dati che possono essere liberamente utilizzati, riutilizzati e redistribuiti, con la sola limitazione – al massimo – della richiesta di attribuzione dell'autore e della redistribuzione allo stesso modo (ossia senza che vengano effettuate modifiche)". Partendo da questa prima definizione, è possibile approfondire il tema affiancandolo a due altri concetti chiave; quelli di *linked data* e di *fonte*. Se un open data è un set di "dati che possono essere liberamente utilizzati, riutilizzati e distribuiti..." significa che devono essere assoggettati a un tipo di licenza d'uso che rispetti alcune caratteristiche chiave, quali *l'accessibilità* (il dato deve essere accessibile senza restrizioni legali o tecnologiche), *l'autorizzazione all'uso e riuso* e alla *ridistribuzione senza vincoli* (ad esempio la licenza non può obbligare i fruitori dei dati a generare solo contenuti open) e, infine, la possibilità di poter/dover citare la fonte di partenza. Sulla base di queste caratteristiche, s'intuisce facilmente come un open data sia assimilabile a un prodotto con caratteristiche definite, con una licenza d'uso che ne specifica l'ambito di utilizzo. In realtà, una corrente di pensiero più moderna (Gurstein, 2013) punta a trasformare il concetto di open data da prodotto a servizio, dove l'elemento chiave è l'esigenza del fruitore dei dati e non tanto il dato stesso; in altre parole l'open deve essere effettivamente fruibile e qualificabile come fonte dati.

Gli Open Data sono lo strumento chiave per tradurre il concetto di Open Government in un modello sostenibile

# Il loro identikit

**Aziende, istituzioni** e operatori pongono molta attenzione non solo agli open data come elementi d'informazione, ma anche al fatto che possano effettivamente divenire nel tempo delle fonti. Per questo è opportuno riflettere sulle caratteristiche fondamentali che gli open data, in quest'ottica, dovrebbero avere. Ecco l'elenco:

**PERTINENZA:** devono contenere misurazioni e prodotti statistici che riflettano le esigenze attuali e potenziali degli utenti.

**ACCESSIBILITÀ:** l'informazione deve poter essere ottenuta in modo semplice ed è determinata dalle condizioni attraverso cui gli utenti ottengono i dati.

**CHIAREZZA:** è la facilità con cui gli utenti vengono messi in grado di capire i dati. Essa è determinata dal contesto informativo in cui vengono presentati i dati, se sono accompagnati da metadati appropriati, se vengono utilizzate illustrazioni come grafici o mappe, se sono disponibili informazioni sulla loro accuratezza. Particolare importanza rivestono le eventuali limitazioni d'uso e l'assistenza aggiuntiva dal produttore del dato.

**ACCURATEZZA:** devono fornire informazioni precise e accurate, vale a dire che ci deve essere un alto grado di corrispondenza tra i dati contenuti in esse e la realtà.

**ATTENDIBILITÀ:** vicinanza del valore/stima iniziale ai valori/stime successive

**COMPARABILITÀ:** è una misura di quanto le differenze, temporali e geografiche, siano dovute a variazioni reali e non a differenze legate a concetti statistici, strumenti di misurazione e/o procedure

**INTEGRABILITÀ:** devono permettere di affiancare le informazioni integrandole con altre, analoghe, provenienti

da altre sorgenti.
**COMPLETEZZA**: riguarda non solo le omissioni o le lacune eventualmente presenti nei dati forniti, ma anche il fatto che molto spesso non siano disponibili i dati per alcune aree del territorio nazionale.
**TEMPESTIVITÀ:** gli open data non devono fornire dati riferiti a periodi troppo lontani nel tempo. In altre parole, il momento della rilevazione non deve essere troppo distante da quello in cui i dati vengono diffusi e resi disponibili.
**PERIODICITÀ:** devono aggiornare periodicamente le informazioni che mettono a disposizione dei diversi utenti.
**PUNTUALITÀ:** è il periodo di tempo tra la data del rilascio dei dati e quella pianificata da calendario, da regolamento o da accordo preventivo tra partner.
**COERENZA:** misura l'adeguatezza delle statistiche a essere combinate in modo diverso e per diversi usi.
**RICONCILIABILITÀ:** è la caratteristica che devono avere affinché le statistiche all'interno di una stessa fonte - relative a variabili diverse, calcolate su domini diversi, da fonti diverse o da processi con periodicità diverse - possano essere assimilabili tra loro.
Inoltre, gli open data devono essere corredati da una dotazione informativa, necessaria, in termini di documentazione, per valutare la qualità dei dati forniti - e di conseguenza migliorarla - e per un pieno e corretto sfruttamento di tutto il potenziale informativo.
Gli utenti devono essere messi in condizione di poter scegliere le fonti più adeguate per soddisfare i propri bisogni conoscitivi e comprendere fino a che punto i dati disponibili siano utilizzabili per ulteriori elaborazioni. Una descrizione dettagliata delle metodologie statistiche utilizzate, corredate da misurazioni degli errori campionari e non campionari, presenta un elevato contenuto informativo sia per gli utilizzatori dei dati stessi, sia per i produttori.

**Gli utenti devono poter scegliere le fonti piu' adeguate**

# I linked data

Una volta espressi i concetti di open data e di fonte, è opportuno introdurre un'ultima definizione, ricorrendo a Tim Berners-Lee, che nel 2009 ha introdotto il concetto di *linked data* (e di conseguenza di Linked Open Data). Secondo Tim Berners-Lee, i dati non solo devono essere accessibili (open) ma anche, "interpretabili da un PC", strutturati e collegabili tra di loro in modo da poter effettivamente trarre valore e significato dalle varie fonti. Berners-Lee dà indicazioni a chi progetta i sistemi software su come rendere i dati linked e fornisce anche una classificazione a stelle (https://www.w3.org/DesignIssues/LinkedData.html), dove i dati sono classificati con una stella se sono disponibili sul web con licenza open, e raggiungono le 5 stelle se, oltre ad essere disponibili in forma open, sono in un formato non proprietario, rispettano standard specifici (es. RDF) e sono interconnessi. Da notare che un linked data può essere o non essere open (Miller, 2010).

# Open data in Italia

In Italia, nell'ultimo quinquennio, le iniziative di apertura del patrimonio informatico da parte della Pubblica Amministrazione (PA) sono aumentate in maniera significativa. Nel 2010, la Regione Piemonte è stata la prima a mettere in rete un portale di open data (www.dati.piemonte.it), all'interno del quale, a oggi, sono contenuti oltre 800 dataset, provenienti anche dai singoli comuni e province. L'anno successivo, la Regione Emilia-Romagna ha reso disponibile un catalogo di open data (www.dati.emilia-romagna.it) che contiene 390 dataset. Al fine di rendere strutturato il processo di condivisione dei dati da parte delle PA, il 18 ottobre 2011 è stato inaugurato il portale nazionale (www.dati.gov.it), che ha dato una propulsione verso l'innovazione e la trasparenza delle PA. Allo stato attuale nel data store sono presenti oltre 10.300 dataset, provenienti da 76 amministrazioni. I dataset sono organizzati e consultabili secondo temi, generi, territorio di riferimento e amministrazioni che li hanno rilasciati. Questo portale in realtà non contiene tutti gli open data italiani disponibili, poiché, per esempio, sono presenti 8 dataset in ambito sanitario della Regione Lombardia, mentre sul portale degli open data lombardo (www.dati.lombardia.it) sono presenti 52 dataset. A oggi i data store regionali con

il maggior numero di dataset scaricabili sono quelli della Regione Toscana (oltre 1.400) e della Regione Lombardia (oltre 1.200). Nell'ottica della creazione e del mantenimento del portale nazionale www.dati.gov.it, nel 2011 l'Istat ha pubblicato il Vademecum sugli open data, una linea guida per i siti web delle PA sulla tematica dei dati aperti, al fine di armonizzare le modalità di pubblicazione dei dati delle diverse amministrazione. Il documento è strutturato in due macro aree:

• **Parte I, PA e dati aperti**, che introduce i concetti di Open Goverment e Open data, con approfondimenti sul quadro normativo

• **Parte II, Come procedere per aprire i dati della PA**, che descrive gli aspetti pratici, organizzativi e giuridici da considerare prima di rendere disponibili i dati dell'amministrazione. Le PA, nel rispetto dei principi del nuovo codice dell'amministrazione digitale, hanno la responsabilità nel rendere disponibili i propri dati in modalità open e digitale. Per poter diffondere gli open data, è stato studiato un apposito contratto di licenza, denominato *Italian Open Data License* (*IODL*) *v* 2.0. In questo modo si intende rendere facile e immediato il riutilizzo delle informazioni pubbliche nel contesto dello sviluppo della società dell'informazione (http://www.dati.gov.it/iodl/2.0/).

Il portale italiano degli open data è stato messo a punto dopo una serie di esperienze da parte di Governi di tutto il mondo. Il portale più famoso e meglio strutturato è quello americano (www.data.gov) nato a seguito della riforma Obama nel 2009. Successivamente è stato inaugurato il portale inglese (www.data.gov.uk) fortemente spinto da Sir Berners-Lee, padre del World Wide Web. Negli anni successivi si è assistito al fiorire di molti data-store governativi in tutto il mondo.

Degno di nota è il risultato del processo di digitalizzazione del Kenya, Paese pioniere su queste tematiche in Africa. Nel 2011 ha lanciato l'Open Data Initiative, un'operazione di trasparenza delle politiche governative, mediante un progetto finanziato dal Ministero dell'Informazione e della Comunicazione. A oggi il portale (https://opendata.go.ke/) contiene oltre 900 dataset, ben catalogati e di facile interrogazione.

# Open data in Europa

A fine 2012 è stata lanciata una versione beta del portale open data della Comunità Europea (https://data.europa.eu/euodp/it/) che negli anni successivi è entrata a regime e ad oggi contiene oltre 9000 dataset. Il sito descrive molte applicazioni di riutilizzo dei dati open provenienti da una singola fonte o da fonti multiple in ogni campo, sviluppate dalle istituzioni, agenzie e altri organi dell'UE o da terzi. Solo a titolo esemplificativo, si segnalano la mappatura delle attività di pesca, l'Osservatorio della Siccità, le statistiche sulla mobilità degli studenti e l'Atlante degli Incidenti stradali. Molto interessante è l'applicazione infografica "Il mio Paese in una bolla" che permette di confrontare in maniera immediata i dati dei Paesi europei attraverso circa 50 indicatori. Infine si segnala il motore di ricerca Drugle, che si concentra sulle informazioni riguardanti la farmacologia e i medicinali, elencando e analizzando le informazioni sui medicinali disponibili su Internet mediante una semplice interrogazione. La potenzialità di tale motore di ricerca risiede nella possibilità di combinarlo anche con altre applicazioni e sistemi di assistenza sanitaria.

**Il Kenya ha lanciato nel 2011 l'Open Data Initiative. Il portale contiene oltre 900 dataset**

# Nella Sanità gli open data cambieranno lo scenario di mercato

# Il data store del Ministero della Salute

**Il Ministero della Salute** ha lanciato un data-store dedicato (http://www.dati.salute.gov.it/) che a oggi conta 31 dataset nazionali, contro i 259 nella sezione "Sanità" del portale nazionale di open data (che comprendono anche quelli a livello regionale, comunale e di AIFA). I dataset più scaricati sono l'elenco completo delle farmacie aperte al pubblico (inclusi anche succursali, dispensari e dispensari stagionali), l'elenco completo dei dispositivi medici registrati presso la banca dati e il Repertorio del Ministero della Salute, l'elenco completo degli esercizi commerciali, diversi dalle farmacie, autorizzati alla vendita al pubblico di farmaci (art. 5 del D.L. 223/2006 - esercizi commerciali) e l'elenco completo dei Siti Logistici autorizzati alla distribuzione di medicinali a uso umano sul territorio nazionale (art.100 D.Lgs. 219/2006, già Decreto Legislativo 538/92). Per curiosità, si citano altri due esempi utili: l'elenco dei 50 medicinali senza obbligo di prescrizione più venduti, nell'ultimo semestre disponibile e i dati relativi ai posti letto accreditati e a carico del SSN, in ciascuna struttura ospedaliera pubblica, equiparata o casa di cura privata accreditata.

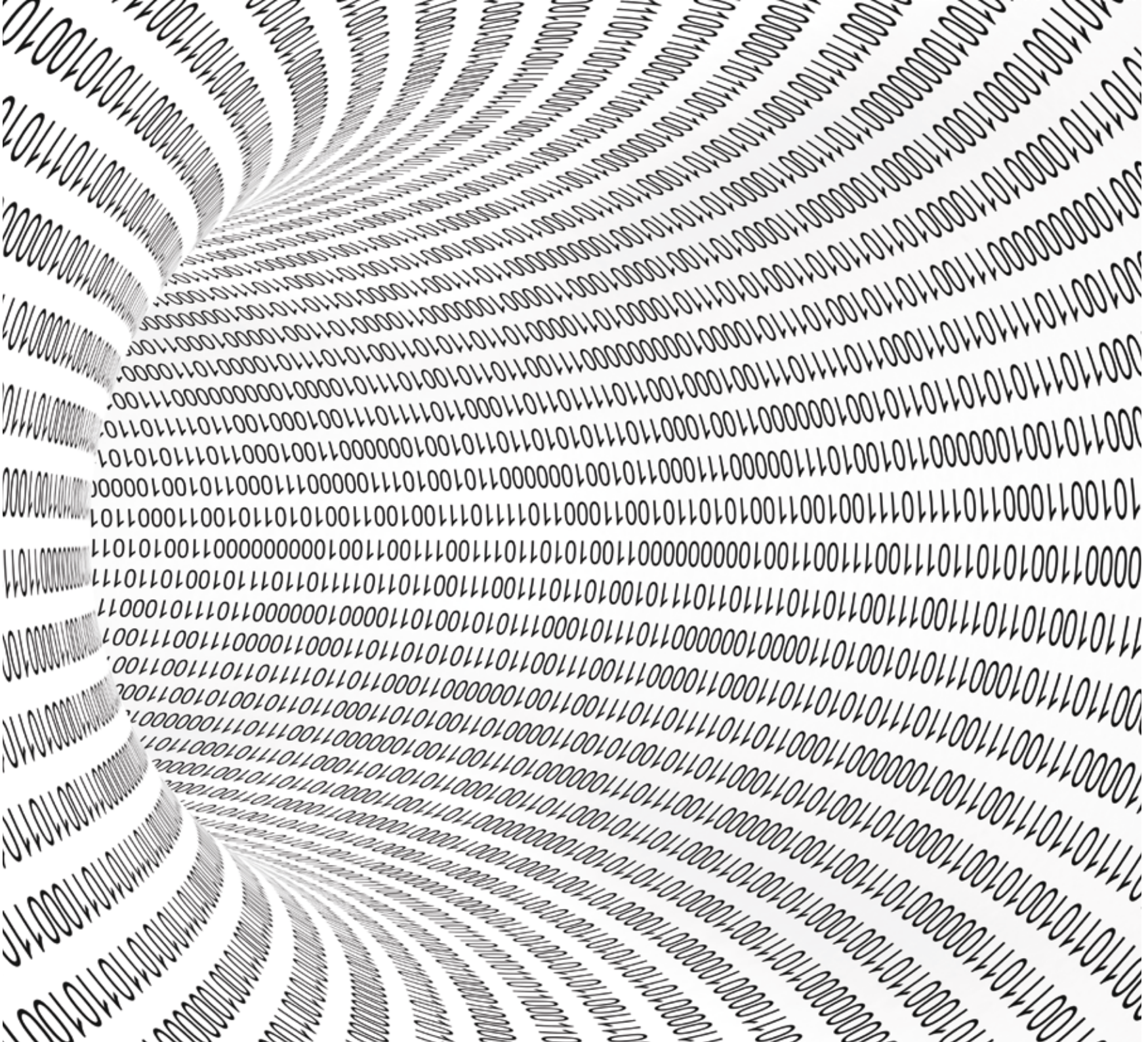
Dal 18 giugno 2013, l'Italia, insieme agli altri Stati membri del G8, aderisce alla "Carta dei Dati Aperti del G8" (http://opendatacharter.net/). I Governi di questi Paesi si sono impegnati ad adottare politiche di apertura del patrimonio informativo delle PA. I 5 principi strategici che tutti i Paesi membri intendono adottare sono: 1) Open Data by Default (tutti i dati amministrativi devono essere pubblicati on line e in modo automatico); 2) Qualità e quantità degli open data (i dati devono essere di alta qualità - metadati, tempestivi ed esaurienti; 3) Accessibilità universale (in formato aperto ed elettronico); 4) Utilizzo degli open data per migliorare la Governance (condivisione delle competenze tecniche e delle esperienze fra Paesi; trasparenza sulle modalità di raccolta, normalizzazione e procedimenti di pubblicazione dei dati); 5) Utilizzo degli open data per promuovere l'innovazione (incoraggiare il riutilizzo a fini commerciali e non). Sono state individuate 14 aree d'intervento condivise da tutti i Paesi. In particolare, in ambito della salute, si richiede la disponibilità a condividere i dati nazionali, come, ad esempio, quelli relativi alla prescrizione e i dati

Il Ministero della Salute ha lanciato un data-store dedicato che conta 31 dataset nazionali

di performance dei sistemi sanitari (http://eventipa.formez.it/sites/default/files/allegati_eventi/OpenData_politica_di_sviluppo_Bonelli.pdf). Per rimanere aggiornati sugli sviluppi di questo tema così interessante e dinamico, oltre a far riferimento a una continua attività di data surf in Rete, l'iscrizione alla mailing list di "Spaghetti Open Data" - la community italiana più longeva e attiva sull'argomento - potrebbe risultare utile.

# Gli open data e le banche dati professionali

**Il numero crescente in Italia** di open data e di portali delle PA è la cartina di tornasole della scelta dell'innovazione – seppure non priva di ostacoli - nell'azione sociale del cittadino italiano. La vera sfida è quella di trasformare l'informazione in conoscenza, altrimenti detta *Evidenza*, in campo sanitario. Il momento di svolta si verificherà quando il maggior numero di data open provenienti da varie fonti saranno linkabili fra loro, andando così a creare nuove sintesi informative. Queste nuove evidenze che si creeranno, non potranno mai sostituire quelle derivanti da studi in ambito sanitario con raccolta ad hoc delle informazioni, in quanto non avranno mai lo stesso livello di accuratezza e precisione dell'informazione. Tuttavia costituiranno evidenze complementari, sicuramente utili a completare un tema così complesso e dalle molte sfaccettature come quello della salute pubblica. Appare chiaro che l'avvento massivo degli open data contribuirà a un miglioramento delle banche dati professionali attraverso l'integrazione dei dati (anche ai soli fini di valida-



zione e di completamento) ma non le sostituirà. La distinzione rimarrà per almeno tre motivi:

• Gli open data sono, nella grande maggioranza, dati raccolti a fini amministrativi e non a fini di ricerca;

• La produzione degli open data non è asservita alle logiche professionali e di business;

• Non viene rispettata una buona parte delle caratteristiche che potrebbero fare di alcuni open data una fonte.

Di contro, alcune banche dati professionali che a oggi fanno della difficoltà di accesso al dato amministrativo il loro punto di forza, saranno destinate a scomparire o a trasformarsi attraverso opportune integrazioni con dati originali e di ricerca. In ultima sintesi, da questa breve disamina si può evincere la crescente importanza e utilità degli open data e lo scenario futuro che dipinge un crescente interesse verso di essi da parte di aziende, enti e cittadini. Sono risultati però evidenti i limiti delle "fonti" open. Ecco alcuni esempi di questi limiti: la possibilità che la struttura dati cambi ogni volta che venga pubblicato un aggiornamento; una frequenza di aggiornamento assolutamente non associabile a un utilizzo aziendale del dato; un obiettivo di raccolta delle informazioni difficilmente coincidente con le esigenze di utilizzo. Un altro evidente motivo che permette di intuire le difficoltà di affidarsi esclusivamente agli open data è legato alle modalità di creazione dei collegamenti tra i diversi dati: difficilmente si troveranno codifiche univoche e comuni tra open data diversi, obbligando così gli utilizzatori a un'attività di abbinamento costosa e, a volte, impossibile. Di contro, risulta altrettanto evidente che la disponibilità di fonti open ha iniziato a rivoluzionare il mondo dei dati, della loro analisi e della loro interpretazione: basti pensare che secondo la "Carta dei dati aperti", alla quale l'Italia ha aderito, anche i dati di prescrizione medica potrebbero/dovrebbero diventare open, andando così a modificare uno scenario di mercato e di ricerca più che consolidato all'interno del panorama farmaceutico nazionale.

## BIBLIOGRAFIA ESSENZIALE


**Istat (2016),** *Linee guida per la qualità dei processi statistici che utilizzano dati amministrativi Versione 1.1.*

**Christian Bizer, Tom Heath e Tim Berners-Lee,** *Linked Data -The Story So Far (PDF), in International Journal on Semantic Web and Information Systems, vol. 5, nº 3, 2009, pp. 1–22, DOI:10.4018/jswis.2009081901, ISSN 15526283*

**Rob Kitchin,** *Open and linked data - SAGE publication, 2014*

# POPULAR SCIENCE

KEKOA PUBLISHING