



Corso di Laurea Magistrale
Design della Comunicazione
Scuola del Design
Politecnico di Milano

vixpress

Visualizzazione di dati
per servizi di e-journalism

a.a 2010/2011
Tesi di Chiara Girardelli,
matricola 749789
Relatore Paolo Ciuccarelli

Politecnico di Milano, Scuola del Design
Corso di Laurea Magistrale in Design della Comunicazione
Tesi di Laurea di Chiara Girardelli,
Relatore Paolo Ciuccarelli,
a.a 2010/2011

Indice

7 Indice delle figure

11 Abstract

13 Capitolo primo *Il fenomeno degli Open Data*

- Conoscenza aperta
- Il concetto di open data: dati liberi e riutilizzabili
- La tecnologia come veicolo per gli open data
- Chi produce open data?
- L'ideologia Open
- La storia degli Open Data
- I principi degli Open Data
- I limiti degli Open Data

61 Capitolo secondo *L'e-journalism e il ruolo della visualizzazione di dati*

- Nuovi modelli di informazione
- Data Driven Journalism
- Il ruolo della visualizzazione
- L'information visualization nell'ambito dell'e-journalism e del data journalism
- Citizen Journalism e informazione grassroots
- Gli aspetti sociali ed antropologici del giornalismo su internet
- Il Crowdsourcing e i vantaggi della collaborazione di massa

- Georeferencing: la diffusione di informazioni georeferenziate
- L'utilizzo di dati in tempo reale
- La situazione attuale sul web

145

Capitolo terzo

viXpress: visualizzazione di dati per servizi di e-journalism

- Perché l'ambito dell'informazione?
- Idea di progetto e spunti di innovazione
- Il ruolo del Guardian
- Il target
- Obiettivi preposti
- Tecnologie utilizzate e librerie per la visualizzazione di dati
- Ricerche preliminari alla fase realizzativa
- La raccolta dei dati
- L'analisi dei dati e la scelta delle visualizzazioni
- I livelli di lettura
- Fase di test con utenti
- Obiettivi raggiunti e possibilità di sviluppo

211 Bibliografia

213 Sitografia

214 Ringraziamenti

Indice delle figure

- Fig.1 - Open Knowledge Index 2009-2010. - 17
- Fig.2 - Open Knowledge Index e reddito pro capite. - 18
- Fig.3 - “Open Data Movement”, visual.ly. - 28
- Fig.4 - data.gov. - 34
- Fig.5 - dati.gov.it. - 35
- Fig.6 - Timeline dell’apertura di siti ufficiali di Open Data. - 46
- Fig.7 - Linked Open Data September 2010. - 48
- Fig.8 - Linked Open Data October 2007. - 50
- Fig.9 - Social Media Landscape, Fred Cavazza. - 68
- Fig.10 - “Quantitative easing: How much money has the Bank of England pumped into the British economy?”, Mark McCormick. - 71
- Fig.11 - “Diagram of the causes of mortality in the army in the East”, Florence Nightingale. - 72
- Fig.12 - Mappa tematica estratta da “Essay on moral statistics of France”, André-Michel Guerry. - 74
- Fig.13 - “Statistique Comparée de l’Etat de l’Instruction et du Nombre des Crimes”, Guerry and Balbi. - 75
- Fig.14 - “Carte figurative des pertes successives en hommes de l’Armée Française dans la campagne de Russie 1812-1813”, Charl Joseph Minard. - 76
- Fig.15 - Special survey report published by the Detroit Free Press on August 20 1967, Philip Meyer. - 78
- Fig.16 - Guardian data workflow, Mark McCormick for the Guardian. Parte 1. - 82
- Fig.17 - Guardian data workflow, Mark McCormick for the Guardian. Parte 2. - 83
- Fig.18 - Schermata tratta da un’analisi e rappresentazione di dati realizzata con Spotfire. - 85

- Fig.19 - Anemone, Ben Fry. - 87
- Fig.20 - Passaggi per la creazione di una visualizzazione su ManyEyes. - 89
- Fig.21 - Classificazione dei tipi di visualizzazione su ManyEyes. - 90
- Fig.22 - visual.ly - 92
- Fig.23 - “Visualising traffic on img.ly”, 9elements. - 93
- Fig.24 - “The Path of Protest”, Garry Blight, Sheila Pulham and Paul Torpey. - 94
- Fig.25 - “How Twitter tracked the News of the World scandal”, guardian.co.uk - 103
- Fig.26 - Crowdsourcing Landscape, crowdsourcingresults.com - 116
- Fig.27 - WeePlaces - 121
- Fig.28 - Google News - 125
- Fig.29 - News Map - 126
- Fig.30 - Storify - 130
- Fig.31 - Storyful - 132
- Fig.32 - News Dots - 134
- Fig.33 - Scheramenta tratta da “The Opportunity Gap”, propublica.org - 136
- Fig.34 - Scheramenta tratta da “Dollars for Docs”, propublica.org - 137
- Fig.35 - Scheramenta tratta da “How the Heart Rythm Society sells Access”, propublica.org - 138
- Fig.36 - Infomous’ World News Cloud. - 142
- Fig.37 - Schema elementi del concept di progetto. - 151
- Fig.38 - Quotidiano The Guardian, prima pagina. - 153
- Fig.39 - Schermata tratta dalla homepage di guardian.co.uk - 154
- Fig.40 - Diagramma a corde generato da d3. - 163
- Fig.41 - 3D Data Visualization Via Processing & OpenGL. - 164
- Fig.42 - Bubble Chart generato da Protovis. - 165
- Fig.43 - “Flickr Shapes”, Aaron Straup Cope. - 166

Fig.44 - Articoli della sezione News di guardian.co.uk - 170

Fig.45 - Articoli al di fuori della sezione News di guardian.co.uk - 171

Fig.46 - Articoli per categoria della sezione News di guardian.co.uk - 172

Fig.47 - Schermata concept di progetto. Rete di relazioni. - 178

Fig.48 - Schermata concept di progetto. Mappa geografica. - 179

Fig.49 - Rete di relazioni del giorno 1/11/2011 - 182

Fig.50 - Rete di relazioni del giorno 2/11/2011 - 182

Fig.51 - Rete di relazioni del giorno 3/11/2011 - 183

Fig.52 - Rete di relazioni del giorno 4/11/2011 - 183

Fig.53 - Rete di relazioni del giorno 5/11/2011 - 184

Fig.54 - Rete di relazioni del giorno 6/11/2011 - 184

Fig.55 - Mappa del giorno 1/11/2011 - 186

Fig.56 - Mappa del giorno 2/11/2011 - 187

Fig.57 - Mappa del giorno 3/11/2011 - 187

Fig.58 - Mappa del giorno 4/11/2011 - 188

Fig.59 - Mappa del giorno 5/11/2011 - 188

Fig.60 - Mappa del giorno 6/11/2011 - 189

Fig.61 - Esempio di timeline annuali su diversi argomenti. - 191

Fig.62 - Schermata principale. - 197

Fig.63 - Schermata del primo livello di dettaglio. - 198

Fig.64 - Schermata del secondo livello di dettaglio. - 200

Fig.65 - Struttura del primo livello di approfondimento. - 201

Fig.66 - Schema di navigazione del primo utente. - 203

Fig.67 - Schema di navigazione del secondo utente. Prima parte. - 204

Fig.68 - Schema di navigazione del secondo utente. Seconda parte. - 205

Fig.69 - Schema di navigazione del terzo utente. Prima parte. - 206

Fig.70 - Schema di navigazione del terzo utente. Seconda parte. - 207

Abstract

Questa tesi si struttura in tre fasi principali. La prima è costituita da un'analisi sulla tematica dei dati aperti. Negli ultimi anni si è diffusa sempre di più l'esigenza di poter accedere in modo libero e diretto a dati provenienti da diversi ambiti, tra cui, in particolare, quelli governativi e relativi alle pubbliche amministrazioni e di poterne usufruire, analizzandoli, rappresentandoli e ridistribuendoli in vari modi. A partire dal 2009, i governi di varie nazioni, tra cui Stati Uniti, Gran Bretagna, Norvegia e infine Italia con il sito data.gov.it, hanno iniziato ad aprire portali web ufficiali dedicati alla pubblicazione di dati aperti. Sono sorte, inoltre, diverse associazioni e gruppi di lavoro impegnati nella ricerca, nella raccolta, nell'analisi e nella diffusione di dataset.

La seconda fase di questa tesi comprende un'analisi di come l'avvento dei nuovi media e, in particolare, la diffusione del web e delle piattaforme di condivisione online abbia influito sul giornalismo e sulla produzione di informazione. Vengono toccate tematiche tra cui quelle del citizen journalism, del crowdsourcing, della georeferenziazione delle informazioni e dell'utilizzo di dati in tempo reale,

fenomeni che sono arrivati ad influenzare il concetto stesso di giornalismo.

Questa parte, ricollegandosi alla prima fase, vuole inoltre trattare della crescente importanza che la visualizzazione di dati e informazioni ha assunto in ambito giornalistico e dei principi che regolano la realizzazione di questo tipo di artefatti comunicativi.

La terza e ultima fase tratta del progetto che è stato realizzato in merito a queste tematiche. Si tratta di uno strumento web-based per la visualizzazione e l'analisi di dati tratti da servizi di e-journalism. viXpress (questo è il nome dato all'applicazione, dalla commistione dei termini visual, press ed experience) consente di esplorare visivamente il panorama delle notizie pubblicate in rete e si basa su tre aspetti primari: quello temporale, quello geografico e quello delle relazioni tra articoli giornalistici differenti. viXpress mira a mettere in luce le reti di relazioni che intercorrono tra le varie tematiche trattate e l'evoluzione di queste relazioni nel tempo.

1. Il fenomeno degli Open Data

Conoscenza Aperta

“Open knowledge is any material — whether it is content, data or information-based — which anyone is free to use, re-use and redistribute without restriction, subject only, at most, to the requirement to attribute and share-alike.”¹

L’Open Knowledge Definition, sviluppata dalla Open Knowledge Foundation, organizzazione no-profit fondata nel 2004, definisce in generale il significato di “conoscenza aperta” e le sue caratteristiche, includendo in questa definizione contenuti come libri, film e musica, dati di ogni tipo e informazione del settore pubblico.

La definizione, che viene periodicamente aggiornata e rivista a partire dal 2005, si basa sui principi di libero accesso al materiale, libero riutilizzo e redistribuzione, assenza di restrizioni legate all’identità dell’utente o al suo

¹ Definizione di Open Knowledge, presentata sul sito web della Open Knowledge Foundation < <http://okfn.org> >

settore di attività.

In particolare, un'opera, o informazione, è aperta se soddisfa undici punti specifici, che includono indicazioni riguardo ad accesso, ridistribuzione, riutilizzo e in merito alle caratteristiche della licenza relativa. Questi punti stabiliscono che l'informazione l'opera "open" deve essere disponibile in forma completa, senza costi se non eventualmente quelli di riproduzione e in formati riutilizzabili e modificabili. La licenza può richiedere come condizione che sia dato credito agli autori originali, ma deve consentire la modifica e la ridistribuzione dell'opera e non deve imporre limiti discriminatori legati ai possibili usi e ai campi di applicazione. La licenza infine non deve limitare la distribuzione di altre opere.

La Open Knowledge Foundation ha definito un indice² per misurare i progressi dei diversi paesi nell'apertura dell'informazione, dei dati e della conoscenza al pubblico. Questo indice si basa su tre aspetti: l'abilità di accedere alla conoscenza, la disponibilità di conoscenza e dati e la capacità di utilizzare e riproporre i dati.

2

< <http://openeconomics.net/open-knowledge-indicator/> >

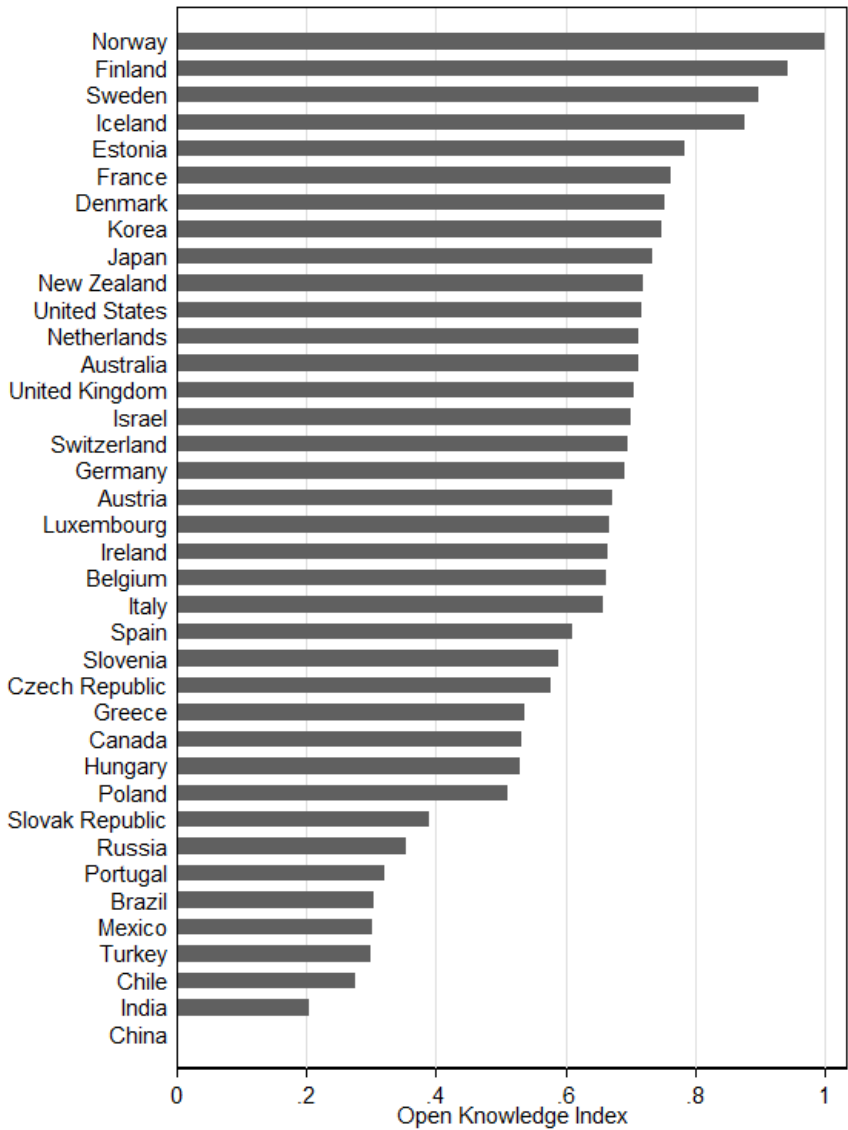


Fig. 1 - Classifica di un campione di 38 paesi in base all'Open Knowledge Index nel 2009-2010. Open Economics working Group della Open Knowledge Foundation. < openeconomics.net >

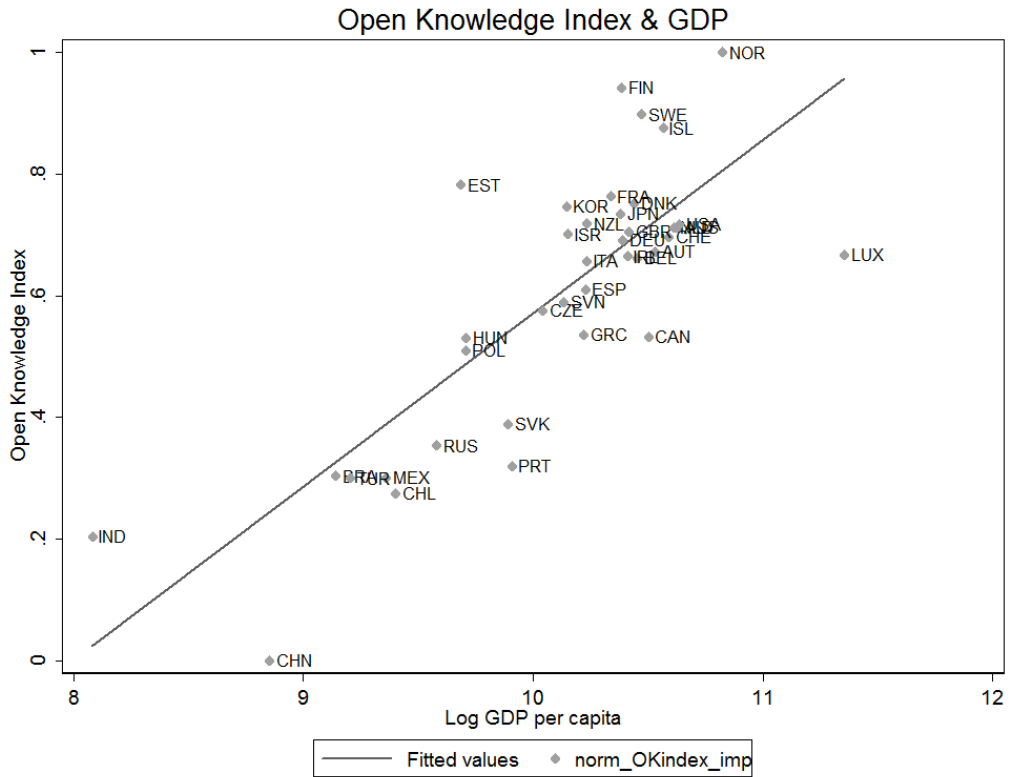


Fig. 2 - Open Knowledge Index messo in relazione con il reddito pro capite. Open Economics working Group della Open Knowledge Foundation. (openeconomics.net)

Il concetto di Open Data: dati liberi e riutilizzabili

Il concetto di dato aperto, o Open Data, fa parte della più ampia concezione di conoscenza aperta, ma assume anche delle sue proprie e specifiche caratteristiche.

Per Open Data si intendono tutte quelle pubblicazioni di dataset rese liberamente disponibili a chiunque. Qualsiasi dataset definito Open Data può essere ripubblicato dall'utente che vi accede senza alcuna limitazione di copyright, di digital patent o di qualsiasi altra forma di tutela e controllo dei contenuti digitali. L'obiettivo primario del movimento Open Data è simile a tutti gli altri movimenti definiti "Open" come nel caso dell'ideologia Open Source, che prevede il rilascio libero ed incondizionato dei codici sorgente dei software affinché la comunità degli utenti sia abilitata e, da un certo punto di vista, anche incoraggiata a sviluppare migliorie, aggiornamenti, versioni alternative, integrazioni o semplici modifiche personali.

Strettamente legato al concetto di OpenData vi è quello di “Open Access”, riferito al mondo delle pubblicazioni accademiche e basato sul principio della libera consultazione, e quindi del libero accesso, delle risorse bibliografiche.

“Open Access literature is digital, online, free of charge, and mostly free of copyright and licensing restrictions”³

Questo movimento si è rivelato particolarmente importante soprattutto per i paesi in via di sviluppo, dove i relativi istituti di ricerca ed accademie si sono trovati a dover affrontare costi troppo elevati per poter accedere a pubblicazioni e risorse aggiornate.

Un ulteriore aspetto legato all’ideologia Open che caratterizza il fenomeno dei dati aperti e a stretto contatto con l’Open Access è il concetto dell’Open Content, che prevede la possibilità di utilizzare e modificare il contenuto del documento digitale, dando così vita ad opere collaborative e non più identificate ed attribuite strettamente ad un’unica figura definita come l’autore e detentrica di tutti i diritti di proprietà intellettuale, di copia e di qualsiasi altra re-

³ Peter Suber, Open access overview. Focusing on open access to peer reviewed research articles and their preprints, 2004

strizione correlata alla risorsa.

La filosofia Open, nel suo complesso, a differenza di quanto si possa credere, non è del tutto recente ed è possibile trovare alcuni riferimenti a questa filosofia già in situazioni passate, come nel caso dell'approccio mertoniano. Robert Merton, sociologo statunitense, lavorò verso la metà del '900 ad un'analisi sui valori della comunità della ricerca e arrivò a definire la scienza come un'istituzione sociale separata dalla società. Egli studiò, tra le altre cose, le modalità di organizzazione della comunità della ricerca e le circostanze che possono agevolare o ostacolare la produzione di conoscenza scientifica.

Il termine "Open Data", tuttavia, è di più recente concezione. La sua popolarità è cresciuta parallelamente alla nascita ed alla crescente diffusione di Internet e del www. Il fenomeno degli Open Data, è stato poi ulteriormente accelerato dalle numerose iniziative istituzionali di diffusione di dati aperti, dando così vita all' Open Data Government, anche conosciuto come "OpenGov". Uno degli esempi più rilevanti e tradizionalmente più conosciuti è sicuramente data.gov, il portale statunitense mirato alla consultazione, alla diffusione ed al confronto di tutti i dataset prodotti e rilasciati dalle amministrazioni locali. Si tratta di uno dei primi esempi di comunicazione digitale Istituzione-Cittadino, diretto e trasparente. L'obiettivo di questa iniziativa

non è stato solo quello di trovare un metodo di diffusione di dati alla portata di tutti, ma anche quello di offrire ad ogni utente e cittadino le informazioni liberate da ogni tipo di suggestione prodotta dai media, anche se in forma di dati grezzi.

Se da un certo punto di vista questo tipo di approccio motiva l'interesse dell'utente nei confronti del dato, la totale libertà di interpretazione, priva di guide di lettura, può generare alcuni problemi. La comprensione dei dati risulta meno intuitiva per via del loro formato grezzo e non elaborato. La grande quantità di dataset e dati, inoltre, può portare ad un sovraffollamento di informazioni. E' per questo motivo che il ruolo della Data Visualization assume un'importanza centrale nei processi di diffusione ed interpretazione del dato grezzo.

Nonostante il termine "Open Data" venga frequentemente utilizzato, non è stata concordata una definizione ufficiale come è invece accaduto nel caso di "Open Access", la cui definizione è stata formalizzata e firmata da diversi enti. La produzione di open data si incentra soprattutto su materiale non testuale, come mappe, genomi, reti, composti chimici, formule matematiche e scientifiche, dati e pratiche mediche, scienze biologiche e biodiversità, oltre che su tutti quei dati di tipo ambientale, economico o sociale raccolti dalle pubbliche amministrazioni.

La maggior parte delle difficoltà nella diffusione e nell'utilizzo degli OpenData consiste nel fatto che in molti casi i dati prodotti rappresentano un valore commerciale o sono parte di opere commercialmente riconosciute. Purtroppo ancora oggi l'accesso o il riuso dei dati è spesso limitato dalle organizzazioni, pubbliche e private, che producono e distribuiscono i dataset. Il controllo sui dati viene esercitato attraverso le tipiche restrizioni proprietarie quali le licenze, i diritti di copia, i brevetti e le tassazioni.

I sostenitori dell'ideologia Open in merito ai dati affermano che tali restrizioni sono contro il bene comune e che risultano essere contraddittorie con la definizione stessa di Open Data. L'idea comune è quella di rendere questi dati disponibili senza alcuna restrizione o tassazione.

Molta importanza si pone anche sulla tematica del riutilizzo, ovvero la possibilità di rielaborare progetti sulla base dei dati prodotti per riproporre modifiche, miglioramenti, aggiunte o versioni alternative del progetto originale. Anche in questo caso è possibile scontrarsi con limitazioni che garantiscono la sola lettura, rendendo necessaria un'esplicita autorizzazione per qualsiasi forma di riutilizzo o redistribuzione.

Alcuni produttori di dati non considerano fondamentale il bisogno di esercitare il diritto d'autore o di definire i termini di utilizzo del proprio lavoro. Ad esempio, molti scienziati non considerano l'insieme dei dati che emerge dal loro lavoro come una loro proprietà da tutelare e da limitare con brevetti o altre restrizioni, ma si impegnano a diffondere i propri risultati in pubblicazioni con l'implicito intento di rilasciare i dati sottoforma di beni comuni (commons).

Tuttavia l'assenza di una licenza e la conseguente assenza di normative che possano portare alla definizione, in tempi relativamente brevi, di uno standard, potrebbe ostacolare la diffusione e l'utilizzo dei dataset, troppo eterogenei e di difficile aggregazione, limitando l'accesso e l'apertura a molti dei destinatari interessati al dato.

Alcune organizzazioni private e pubbliche, come il caso del "Institute of Electrical and Electronic Engineers" (in italiano: Istituto degli ingegneri elettrici ed elettronici), IEEE⁴, hanno intrapreso l'iniziativa di raccogliere, analizzare ed aggregare i dataset prodotti e di rivendere i dati coperti da copyright.

L'IEEE è un'organizzazione no profit che volge la sua attenzione allo sviluppo delle tecnologie elettroniche. Non-

4 < <http://www.ieee.org> >

ostante la protezione di questi dati sia esercitata attraverso l'utilizzo dei diritti di copia e l'accesso ne comporti necessariamente l'acquisto, i dati rilasciati risultano essere strutturati e di più facile comprensione.

La tecnologia come veicolo per gli Open Data

La natura stessa dei dati, i metodi di diffusione e di accesso alle informazioni rendono necessario ed imperativo l'utilizzo di dispositivi informatici. Come affermato in precedenza, il canale di diffusione principale dei dati aperti è Internet. E' proprio l'infrastruttura della rete stessa che garantisce all'utente di poter accedere alla pubblicazione di un dataset all'interno di un portale dedicato. In alcuni casi è anche possibile accedere ai dati in tempo reale ed osservarne l'evoluzione in diretta, ma questa possibilità dipende esclusivamente dalle modalità di rilascio del produttore.

Per quanto in teoria il concetto di open data implichi, tra le altre cose, l'assenza di limitazioni tecnologiche all'accesso, per l'utente finale si presenta una serie di requisiti generali minimali, ma comunque necessari per un corretto accesso. Il primo di questi, come accennato in precedenza, riguarda l'impiego di un computer che sia in pos-

sesso dei requisiti hardware minimi per poter supportare una corretta ed agile fruizione dei dati. Nei paesi più sviluppati al giorno d'oggi è pressoché impossibile trovare sul commercio dispositivi informatici che non soddisfino, nello specifico, una potenza di calcolo sufficiente all'esplorazione web (al di sotto del gigahertz) ed un'interfaccia di rete digitale ad alta velocità o a banda larga (ethernet, wi-fi, gprs). Alcuni problemi si possono riscontrare invece con un tipo di requisito hardware al di fuori del controllo diretto e delle competenze dell'utente, ovvero la disponibilità di banda offerta dall'infrastruttura di rete nella zona di riferimento del punto di accesso.

Come per i requisiti specifici della macchina in dotazione all'utente, i requisiti software necessari ad un corretto accesso, da 10 anni a questa parte risultano essere consolidati ed ampiamente coperti da tutte le dotazioni software commerciali (Microsoft Windows e Apple MacOS) e non commerciali (Linux). L'interfaccia primaria richiesta su una qualsiasi delle piattaforme citate è rappresentata interamente dal browser e da pochi altri software di supporto comunemente diffusi. Sulla rete sono disponibili diverse versioni di software open (privi di licenza e privi di limitazioni) e free (privi di qualsiasi forma di costo) in grado di dotare l'utente del giusto pacchetto operativo.

THE OPEN DATA MOVEMENT

THE EVOLUTION OF APIs

Increasingly, companies are making their data and inner workings publicly available through the release of APIs, which are used by developers in building new tools—like TweetDeck, based on Twitter's API. Since 2005, more than 3,700 APIs have been launched.

WH

An app
softwa
tracks

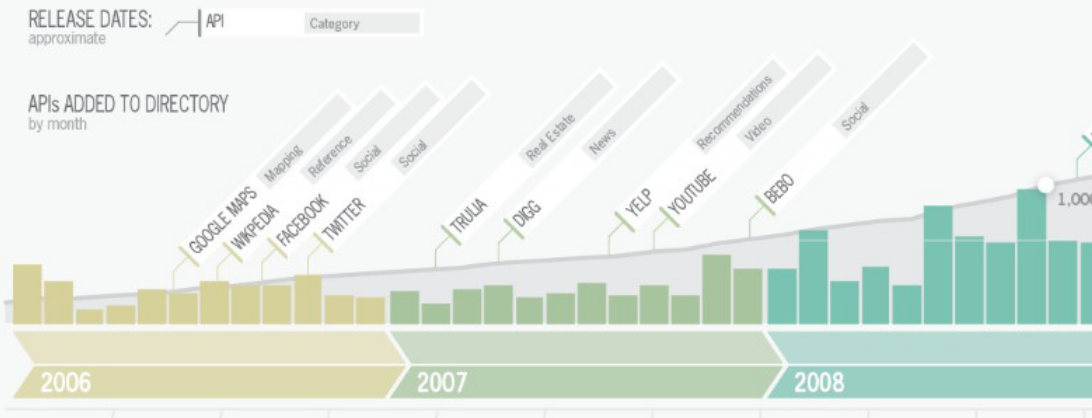
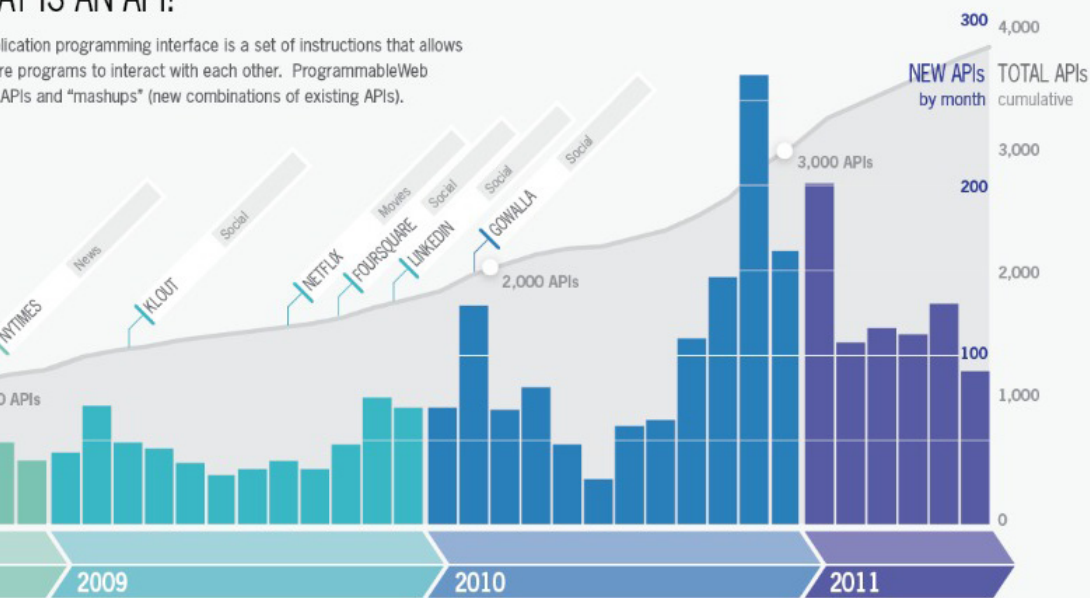


Fig. 3 - Estratto dell'infografica "Open Data Movement" realizzata da visual.ly. L'immagine mostra l'evoluzione nel tempo della disponibilità di API sul web.

WHAT IS AN API?

An application programming interface is a set of instructions that allows different programs to interact with each other. ProgrammableWeb tracks the growth of APIs and "mashups" (new combinations of existing APIs).



Source: ProgrammableWeb

Per quanto riguarda la fruizione dei dati sottoforma di visualizzazione, i requisiti necessari non si alzano a tal punto da giustificare una ridefinizione delle caratteristiche della macchina precedentemente definita adatta all'accesso. L'unica differenza significativa riguarda l'eventuale necessità di predisporre all'interno della dotazione software le librerie di visualizzazione richieste o alcuni plug-in comunemente diffusi (come Adobe Flash Player).

L'espansione della rete e la diffusione della banda larga, parallelamente alla crescente informatizzazione della popolazione mondiale negli ultimi 20 anni, hanno reso credibile l'idea di un accesso condiviso a risorse in costante modifica ed aggiornamento. La semplificazione delle procedure di accesso ai contenuti desiderati, l'indicizzazione ed il sempre più presente orientamento all'idea del Web Semantico, pongono le basi verso un nuovo approccio di accesso alle informazioni. All'interno di uno scenario caratterizzato da una produzione di dati sempre più vasta ed eterogenea, il concetto di semplice collegamento ipertestuale inizia a presentare problematiche legate all'accesso contestuale della risorsa. E' proprio per questo motivo che si presenta l'esigenza di un processo di accesso e relazione più elaborato, rappresentato appunto dal Web Semantico.

La costruzione di questo "ambiente" digitale è tuttora in corso ed il processo segue una metodica di costruzione a

strati, da quelli più basilari di identificazione di elementi semplici fino a livelli più esterni dove verranno considerate le relazioni ed i legami logici tra le varie risorse disponibili. I risultati ottenibili non dipenderanno più solamente da relazioni statiche e predefinite, ma sarà possibile ottenere aggregazioni di senso ora dipendenti anche da variabili di contesto aggiuntive.

Chi produce Open Data?

Considerando un'ipotetico scenario in cui ogni utente si sente coinvolto e risulta dotato dei mezzi necessari per effettuare rilevazioni ed elaborazioni significative per la costruzione di un dataset, si può affermare che tutti quanti potenzialmente sono in grado di produrre OpenData. E' chiaro come questa si tratti di un'affermazione puramente teorica che delinea il potenziale dell'effettiva produzione di dati, ma che non tiene conto delle competenze e degli strumenti necessari per la produzione di dati che presentino un adeguato livello di accuratezza e precisione per i molteplici contesti di applicazione.

Per questo motivo, si farà riferimento esclusivamente a tutte quelle fonti che risultino essere sufficientemente strutturate e costanti nel lavoro di produzione e pubblicazione dei dataset. Attualmente, gli enti che producono e diffondono le maggiori quantità di dataset sono le pubbliche amministrazioni, tra cui enti statali locali e regionali. Oltre a queste vi sono associazioni no profit e aziende favorevoli all'ideologia "Open".

A livello geografico, riferendoci ad una scala globale, i primi ed i maggiori produttori di OpenData sono stati gli Stati Uniti, con il noto portale Data.gov⁵, interamente dedicato ai propri cittadini e ricco di informazioni articolate riguardanti i diversi settori di competenza della PA locale.

Altri rilevanti esempi di produzione e diffusione di dati aperti sono alcuni portali web di stampo giornalistico, tra cui quello del Guardian⁶, nota testata giornalistica inglese, che attraverso la propria sezione DataStore rende disponibili ed accessibili numerosi dataset ai propri utenti. Il Guardian rappresenta uno dei più famosi casi di azienda privata, collocata nel settore giornalistico, impegnata nella diffusione e nella visualizzazione dei dati.

5 < data.gov > Sito del governo degli Stati Uniti, lanciato nel Maggio del 2009.

6 < www.guardian.co.uk >

DATA AND APPS

- 390,297 [law](#) and [geospatial](#) datasets
- 1,192 government [apps](#)
- 230 citizen-developed [apps](#)
- 95 [mobile apps](#)
- 172 [agencies and subagencies](#)
- [Suggest a dataset or app!](#)
- 2011 [Next Generation Data.gov](#) is interactive, explorable, and social.

COMMUNITIES

Come explore, discuss, meet others in the same field, and develop the data and apps in the community that you care about. Join in the discussions by going to communities below that interest you.

[OPENDATA](#) [SEMANTICWEB](#)

[HEALTH](#) [LAW](#)

[ENERGY](#) [OCEAN](#)

[EDUCATION](#)

OPEN GOVERNMENT

First open source code released for the Open Government Platform delivered by the governments of India and the U.S. [Find out more](#) and then [download the code](#).

What's coming up on Data.gov? Check out [our blog](#).

LEARN

Are you a teacher, professor, or student? Young people across America are learning about Data.gov and we want you to be part of getting data into the classroom.

See what teachers are doing, share a lesson plan, showcase an app. Learn what's going on in [classrooms](#) today!

SEMANTIC WEB

As the Web of linked documents evolves to include the Web of linked data, we're working to maximize the potential of Semantic Web technologies to realize the promise of Linked Open Government Data. We and our collaborators at the [Tethysness Working Constellation](#) at the [Bangalore Polytechnic Institute](#) are helping to lead the way in this exciting area!

DEVELOPER'S CORNER

Are you interested in sharing your mashups, apps, and ideas? Do you want to learn how to create app and mashups with some of the data hosted here on Data.gov? Whether you are here to share, learn, collaborate, or innovate —you've come to the right place.

Look at these great government [mobile apps](#) and help us to develop more!

Fig. 4 - data.gov.
Sito ufficiale di dati aperti del governo degli Stati Uniti.

Infografica: l'Open Data in Italia

Nonostante nel nostro paese non sia ancora fortemente radicato un modello di **società della conoscenza** - scenario nel quale l'accesso alle informazioni e il loro libero utilizzo diventa un diritto per il cittadino - sono comunque molteplici le iniziative d'apertura di dati pubblici già avviate.

più recenti **più lette**

Italian Open Data Licence 2.0: la nuova licenza italiana gli Open Data è ancora più semplice e aperta
In arrivo Italian Open Data Licence 2.0
Open data e i dati statistici delle province in un convegno dell'UPI
opendata.comune5.it, il nuovo portale dei dati aperti del Comune di Firenze
Quanti sono gli open data in Italia?

[altre notizie...](#)

4 di 5

<p>Cerco i dati</p> <p>Il catalogo degli open data contiene 181 dataset di 32 Amministrazioni</p> <div style="text-align: center; margin-top: 20px;"> </div>	<p>Voglio capire di più</p> <p>Come e perché fare open data:</p> <ul style="list-style-type: none"> • Definizione • Vademecum • Licenza italiana • Discussione online • Open data in Italia • Altri riferimenti utili 	<p>Condivido un dataset</p> <p>Segnalo un insieme di dati della pubblica amministrazione pubblicato in formato aperto</p> <div style="text-align: center; margin-top: 20px;"> </div>	<p>Le App della PA</p> <p>Le applicazioni già disponibili per accedere ai servizi della PA da uno smartphone, suddivise per</p> <ul style="list-style-type: none"> • Amministrazioni centrali • Regioni • Province • Comuni
---	--	---	--

In evidenza

<p>Funzione Pubblica</p> <p>Elenco delle partecipazioni a consorzi/società/fondazioni per comparto e regione.</p> <div style="text-align: center; margin-top: 10px;"> </div> <div style="display: flex; justify-content: space-between; align-items: center; margin-top: 10px;"> CSV </div>	<p>DigitPA</p> <p>SPC Data, i dati aperti dell'IPA (Indice delle Pubbliche Amministrazioni)</p> <div style="text-align: center; margin-top: 10px;"> </div> <div style="display: flex; justify-content: space-between; align-items: center; margin-top: 10px;"> Linked data </div>	<p>Regione Veneto</p> <p>Quadro d'unione degli elementi CTR a scala 1:5.000</p> <div style="text-align: center; margin-top: 10px;"> </div> <div style="display: flex; justify-content: space-between; align-items: center; margin-top: 10px;"> Shapefile </div>	<p>Provincia di Roma</p> <p>Elenco dei punti di accesso a Internet Wi-Fi della Provincia</p> <div style="text-align: center; margin-top: 10px;"> </div> <div style="display: flex; justify-content: space-between; align-items: center; margin-top: 10px;"> XML </div>
---	---	---	--

Fig. 5 - dati.gov.it.
Portale di dati aperti delle pubbliche amministrazioni aperto dal governo italiano.

L'ideologia “Open”

“Numerous scientists have pointed out the irony that right at the historical moment when we have the technologies to permit worldwide availability and distributed process of scientific data, broadening collaboration and accelerating the pace and depth of discovery.....we are busy locking up that data and preventing the use of correspondingly advanced technologies on knowledge”⁷

“Free software” is a matter of liberty, not price. To understand the concept, you should think of “free” as in “free speech,” not as in “free beer.”⁸

7 John Wilbanks, Executive Director, Science Commons

8 GNU, Free Software Foundation, Inc, 2006

Come accennato, l'orientamento verso l'informazione "Open" può essere considerato una vera e propria ideologia, il cui obiettivo è quello di portare il maggior numero di dati, software, pubblicazioni e documenti al più vasto numero di destinatari possibile. L'entusiasmo per questa filosofia di apertura e condivisione risulta essere fortemente incentivata dal progresso dell'Information Technology, la quale ha reso nulle le distanze fisiche in un vero e proprio ambiente virtuale caratterizzato da continui scambi di informazioni in tempo reale attraverso la vasta distesa dei nodi costitutivi della rete.

Secondo la Open Knowledge Foundation, la diffusione e l'applicazione dell'ideologia open potrebbe portare ad ampi benefici sociali, tra cui migliori forme di governo, maggiore diffusione della cultura e sviluppi positivi nell'ambito della ricerca e dell'economia. La vicinanza dei concetti costitutivi di OpenData, OpenSource ed OpenAccess permette di considerare le motivazioni motrici dell'ideologia Open come le stesse per ognuno degli ambiti sopracitati. Questi concetti vengono trattati dal manifesto GNU "Why software should not have owners"⁹ (Perchè il software non dovrebbe essere proprietario).

9 Testo completo consultabile a < <http://www.gnu.org/philosophy/why-free.en.html> >

Per quanto riguarda l'Italia, un documento intitolato "Italian Open Data Licence" (Licenza Italiana sugli Open-Data), in acronimo IODL¹⁰, inizia a definire meglio l'ambito degli OpenData sul territorio nazionale, indicando i diritti e le modalità di condivisione e riuso dei dati aperti. Nonostante si trovi ancora oggi alla versione 1.0, questo documento rappresenta un notevole passo avanti verso il mondo Open.

10 Pubblicata da Formez, centro servizi, assistenza, studi e formazione per l'ammodernamento delle P.A. <formez.it/iodyl/>

Italian Open Data License v1.0

Preambolo

La “Italian Open Data License” (IODL) è un contratto di licenza che ha lo scopo di consentire agli utenti di condividere, modificare, usare e riusare liberamente la banca di dati, i dati e le informazioni con essa rilasciati, garantendo al contempo la stessa libertà per altri. La presente licenza mira a facilitare il riutilizzo delle informazioni pubbliche nel contesto dello sviluppo della società dell’informazione.

Le banche di dati, i dati e le informazioni sono protetti dalle leggi applicabili in materia di diritto d’autore (incluso il diritto sui generis del costituente di banche di dati) e/o dalle altre leggi applicabili. I diritti concessi dalla presente licenza non implicano alcun trasferimento di diritto di titolarità sulle banche di dati, sui dati e sulle informazioni pubbliche.

Gli utenti possono utilizzare e riutilizzare la banca di dati ed i dati in essa contenuti in base ai termini della presente licenza; ogni utilizzazione che non sia espressamente autorizzata ai sensi della presente licenza o delle leggi vigenti è proibita.

Con l’esercizio di uno qualunque dei diritti qui previsti, accetti e ti obblighi a rispettare integralmente i termini della presente licenza.

1. Definizioni

Ai fini e per gli effetti della presenta licenza, si intende per:

- “Informazioni”, la banca di dati ed i dati (testi, data base e data set, immagini, video, ecc.) offerti per l’uso secondo i termini di questa licenza;
- “Lavoro derivato”, ogni traduzione, adattamento, trasformazione, diversa disposizione ed ogni altra modifica delle Informazioni o di una parte sostanziale della banca di dati;
- “Licenziante”, il soggetto (Ente, persona fisica o persona giuridica) che fornisce le Informazioni secondo i termini e le condizioni della presente licenza;
- “Tu/Te”, il soggetto (Ente, persona fisica o persona giuridica) che usa le Informazioni ed esercita i diritti derivanti dalla presente licenza, impegnandosi a rispettarne i termini;
- “Licenza compatibile”, una delle licenze elencate in Appendice, che sia stata riconosciuta da FormezPA come essenzialmente equivalente alla IODL, poichè le sue condizioni hanno lo stesso scopo, significato ed effetto delle condizioni della IODL.

2. Concessione della Licenza

Il Licenziante concede una licenza per tutto il mondo, gratuita, perpetua, non revocabile e non esclusiva alle condizioni di seguito indicate:

Sei libero di:

- riprodurre, distribuire al pubblico, concedere in locazione, presentare e dimostrare in pubblico, comunicare al pubblico, messa a disposizione del pubblico inclusa, trasmettere e ritr-

asmettere in qualunque modo, eseguire, recitare, rappresentare, includere in opere collettive e/o composte pubblicare, estrarre e reimpiegare le Informazioni;

- creare un Lavoro derivato ed esercitare sul Lavoro derivato i diritti di cui al punto precedente, per esempio attraverso la combinazione con altre informazioni (mashup).

A condizione di:

- indicare la fonte delle Informazioni e il nome del Licenziante, includendo, se possibile, una copia di questa licenza o un collegamento (link) ad essa;
- pubblicare e condividere gli eventuali Lavori derivati con la stessa licenza o con una Licenza compatibile;
- non riutilizzare le Informazioni in un modo che suggerisca che abbiano carattere di ufficialità o che il Licenziante approvi l'uso che fai delle Informazioni;
- prendere ogni misura ragionevole affinché gli usi innanzi consentiti non traggano in inganno altri soggetti e le Informazioni medesime non vengano travisate.

3. Usi consentiti

Tu puoi esercitare i diritti concessi con la presente licenza in modo libero e gratuito, anche qualora la finalità da Te perseguita sia di tipo commerciale.

La presente licenza non intende in alcun modo creare ulteriori diritti in capo al Licenziante rispetto a quelli previsti dalla legge sul diritto d'autore o ridurre, limitare o restringere alcun diritto di

libera utilizzazione o l'operare della regola dell'esaurimento del diritto od altre limitazioni dei diritti sulle Informazioni derivanti dalle leggi applicabili.

4. Garanzie ed esonero di responsabilità

Salvo che sia espressamente indicato diversamente, le informazioni concesse sotto la presente licenza sono rilasciate dal Licenziante "così come sono"; il Licenziante non fornisce alcuna garanzia di qualsiasi tipo con riguardo alle Informazioni, sia essa espressa o implicita, di fonte legale o di altro tipo, essendo quindi escluse - tra le altre - le garanzie relative all'idoneità per un fine specifico, alla non violazione di diritti di terzi (d'autore o d'altro tipo), alla mancanza di difetti latenti o di altro tipo, all'esattezza o alla presenza di errori.

Il Licenziante non è responsabile nei tuoi confronti a qualunque titolo per qualsiasi tipo di danni derivante dalla presente licenza o dall'uso delle Informazioni; nessuna clausola di questa licenza esclude o limita la responsabilità nel caso in cui questa dipenda da dolo o colpa grave.

5. Varie

La presente licenza è condizionata risolutivamente, e pertanto i diritti con essa concessi cesseranno automaticamente senza necessità di alcuna comunicazione in tal senso da parte del Licenziante, al rispetto delle condizioni previste all'art. 2 della presente licenza da parte tua.

Il Licenziante si riserva il diritto di rilasciare le Informazioni

sulla base dei termini di una differente licenza o di cessare la distribuzione delle Informazioni in qualsiasi momento.

La presente licenza deve ritenersi regolata dalla Legge italiana e deve, di conseguenza, essere interpretata applicando tale normativa.

Tu sei tenuto ad utilizzare le Informazioni licenziate nel rispetto della legge. In particolare, la presente licenza non costituisce un'autorizzazione a violare la legislazione italiana in materia di diritto d'autore (Legge n. 633/1941) e di protezione dei dati personali (D. Lgs. n. 196/2003).

Per ogni eventuale controversia legata alla presente licenza sarà competente il Foro di Roma.

Appendice

LICENZE COMPATIBILI

Ai sensi dell'art. 1 IODL, sono licenze compatibili:

- la licenza Creative Commons, Attribuzione Condividi allo stesso modo (CC-BY-SA), sia internazionale in versione 3.0 o successiva (disponibile all'URL <http://creativecommons.org/licenses/by-sa/3.0/>) che adattata a specifiche giurisdizioni, in versione 2.5 o successiva (in italiano disponibile all'URL <http://creativecommons.org/licenses/by-sa/2.5/deed.it>)
- la licenza Open Data Commons, Open Database License (ODbL), in versione 1.0 o successiva (disponibile all'URL: <http://www.opendatacommons.org/licenses/odbl/>)

La storia degli OpenData

Gli OpenData sono un fenomeno digitale piuttosto recente. Come affermato in precedenza, l'avvento di questa ideologia è stato possibile quasi esclusivamente in seguito alla larga diffusione delle tecnologie informatiche nell'ambito privato delle famiglie mondiali. E' stata, di conseguenza, riscontrata una più rapida adesione alla produzione di OpenData in tutti quei paesi dove la diffusione del personal computer è stata più rapida e dove è stata garantita la possibilità di connettersi ad una rete dati a banda larga.

Nel 2006 Tim Berners-Lee (Sir Timothy John Berners-Lee, nato a Londra l'8 giugno 1955, informatico britannico, co-inventore insieme a Robert Cailliau del World Wide Web) propose l'utilizzo di dati aperti sul web. Questa rivoluzionaria idea dovette aspettare 3 anni per poter essere applicata in un contesto reale. Solamente nel Maggio del 2009, negli Stati Uniti, durante l'amministrazione Obama, il governo americano ha aperto il primo portale di open data riguardanti la pubblica amministrazione, la spesa pubblica e

numerosi altri campi d'interesse utili per tutti i cittadini desiderosi di informarsi. Il portale, come già citato in precedenza, è stato reso raggiungibile all'indirizzo data.gov.

Appena quattro mesi dopo, nel Settembre 2009, anche la Gran Bretagna propone il suo portale di dati aperti per il cittadino. Pioniere europeo nel panorama degli OpenData, il Regno Unito, basandosi sull'esempio americano, dà vita a data.gov.uk.

Nell'Aprile 2010 anche la Norvegia pubblica online il proprio portale di dati aperti, data.norge.no.

A seguire, nel Marzo 2011 Australia e Canada pubblicano rispettivamente data.gov.au e data.gc.ca. Anche il Kenya, pioniere africano, apre il proprio portale all'indirizzo open-data.go.ke.

Finalmente, nell'Ottobre 2011, anche l'Italia dà vita al suo portale online per i dati aperti all'url data.gov.it.

Parallelamente alla diffusione di questi portali al servizio del cittadino, molti altri enti (privati, non governativi, no-profit) hanno iniziato a produrre dataset in quantità sempre maggiori.

2009



data.gov (Stati Uniti)



data.govt.nz (Nuova Zelanda)



data.gov.uk (Regno Unito)

2010



data.norge.no (Norvegia)



2011



opendata.go.ke (Kenya)



data.gov.it (Italia), datos.gob.es (Spagna)

data.gouv.fr (Francia), dados.gov.br (Brasile)



data.gov.au (Australia), data.gc.ca (Canada)



data.overheid.nl (Olanda), datos.gob.cl (Cile)

datos.gub.uy (Uruguay)



2012



Fig. 6 - Timeline dell'apertura di siti ufficiali di Open Data

Molte società di data visualization coinvolte nella ricerca e nello sviluppo, stimolate da questo crescente movimento si sono attivate per produrre strumenti e artefatti capaci di portare il significato dei dati sempre più vicini all'utente finale.

Sono sorte inoltre diverse associazioni e gruppi di lavoro impegnati nella ricerca, nella raccolta, nell'analisi e nella diffusione di dataset. Alcuni degli esempi più celebri sono l'inglese ckan.org e, citando qualche esempio italiano, spaghettiodpendata.org e linkedopendata.it. L'obiettivo è quello di accelerare i processi di selezione e ricerca dell'utente predisponendo un portale unitario capace di raccogliere le diverse pubblicazioni di dati e definire una breve descrizione tecnica e del contenuto di ogni risorsa resa disponibile.

Linked Data¹¹, in particolare, è un'associazione senza fini di lucro costituita da privati esperti nel settore delle tecnologie Web e attivamente coinvolti nella creazione del Semantic Web. L'obiettivo dell'associazione è quello di migliorare la qualità di vita del cittadino diffondendo dataset e servizi innovativi in modo libero e trasparente.

11 <linkeddata.org>

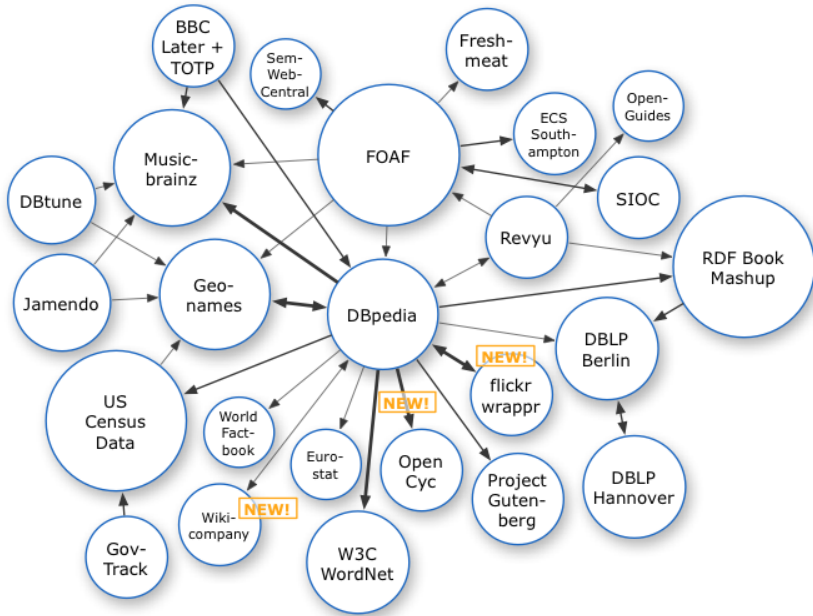


Fig. 8 - Linked Open Data October 2007. Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. < <http://lod-cloud.net> >

“Linked Data is about using the Web to connect related data that wasn’t previously linked, or using the Web to lower the barriers to linking data currently linked using other methods.”¹²

12 Citazione tratta dalla homepage di Linked Data.
< linkeddata.org >

Uno dei progetti di Linked Data è Linked Open Camera, progetto tuttora in evoluzione legato alla produzione di dati e strumenti riguardanti le collaborazioni e le consulenze della Camera dei Deputati, unificati all'elenco delle aziende con quali sono stati stipulati contratti di collaborazione. I dati forniti dal parlamento sono stati digitalizzati e resi disponibili agli utenti.

La breve storia degli OpenData prova come questo sia un fenomeno contemporaneo in continuo sviluppo ed evoluzione. Il numero di portali, ufficiali e non, dedicati ai dati aperti sta crescendo e, parallelamente, stanno aumentando la quantità e la dimensione dei dataset disponibili. In merito a questi fenomeni si intravedono possibilità di sviluppo di elevata importanza come la creazione e la diffusione della metodologia definita dal Semantic Web e la crescente e più tecnica necessità di definire uno standard.

I principi degli OpenData

L'assenza di uno standard preciso per quanto riguarda la produzione di OpenData ha reso necessario elaborare una serie di linee guida con l'obiettivo di produrre e rilasciare dati con modalità e caratteristiche almeno comuni tra loro. L'attenzione è stata rivolta principalmente alle modalità di produzione dal punto di vista della mentalità Open e meno a dettagliate specifiche tecniche. Il risultato è stato quello di porre sotto la definizione di "OpenData" un numero elevato di tipi eterogenei di dataset che presentassero caratteristiche comuni. L'elevata eterogeneità tuttavia rende inequivocabilmente difficile la possibilità di aggregare in modo strutturato informazioni e dataset prodotti da fonti differenti, senza una fase preliminare di riorganizzazione e filtraggio dei dataset. Ogni produttore infatti rilascia i dati autonomamente ed in completa libertà, senza la necessità di tenere conto delle modalità di rilascio di altri produttori.

Nel giugno del 2010 l'Ufficio del Gabinetto inglese ha invitato i propri cittadini ad esporre le proprie opinioni in merito alla bozza sulle linee guida di pubblicazione dei

dati sulla Pubblica Amministrazione online. Le linee guida proposte dal governo britannico nell'appena formato Comitato di Trasparenza del Settore Pubblico (Public Sector Transparency Board)¹³ mirano a definire strategie di diffusione il più possibile capillare ed omogenea, incentivando l'utilizzo e la ripubblicazione dei dati raccolti.

*“In just a few weeks this Government has published a whole range of data sets that have never been available to the public before. But we don't want this to be about a few releases, we want transparency to become an absolutely core part of every bit of government business. That is why we have asked some of the country's and the world's greatest experts in this field to help us take this work forward quickly here in central government and across the whole of the public sector.”*¹⁴

13 Articolo relativo pubblicato su < <http://data.gov.uk/blog/new-public-sector-transparency-board-and-public-data-transparency-principles> >

14 Francis Maude, Minister for the Cabinet Office, Giugno 2010

Il primo punto esposto riguarda le condizioni di licenza e l'utilizzo dei dati, definiti esplicitamente come controllati direttamente dai privati e dalle organizzazioni Business to Business e Business to Consumer che vogliono farne uso.

Il secondo punto esprime la necessità di produrre e pubblicare dati in forma riusabile e pronti per poter essere processati dagli elaboratori.

Il terzo punto chiarisce come tutti i dati utilizzati dovranno essere rilasciati con le stesse "Open Licence" in fase di acquisizione, includendo senza limitazioni il libero riutilizzo, anche commerciale.

Il quarto punto pianifica di rendere disponibili i dati attraverso un'unica piattaforma di accesso, che faciliti il reperimento dei dati.

Il quinto punto prevede la pubblicazione delle informazioni rispettando gli standard open, evitando quindi di utilizzare sistemi e software proprietari.

Il sesto punto afferma che tutti i dati pubblici appartenenti ai siti web del governo saranno pubblicati in modo da garantirne la riusabilità per gli utenti.

Il settimo punto garantisce che i dati saranno aggior-

nati, puntuali e conterranno un alto livello di dettaglio delle voci riportate.

L'ottavo punto esprime la necessità di pubblicare i dati in tempi brevi e di ripubblicarli successivamente sottoforma di dati strutturati ed aggregati.

Il nono punto garantisce che tutti i dati saranno liberamente disponibili per essere utilizzati in qualsiasi forma legale (ad esempio non saranno ammesse forme di riutilizzo volte al favoreggiamento del terrorismo).

Il decimo punto afferma che tutti gli enti pubblici dovranno attivamente incoraggiare il riutilizzo dei propri dati pubblici, in modo da garantire per il maggior numero di dataset un utilizzo concreto della risorsa.

L'undicesimo e ultimo punto della bozza presentata prevede che tutti gli enti pubblici produttori di dati redigano un inventario dei dati rilasciati, garantendone la tracciabilità temporale e di allocazione.

Questa iniziativa richiama in modo esplicito gli otto principi degli OpenData definiti nel 2007 da Tim O'Reilly¹⁵ e Carl Malamud¹⁶, affermando che, per potersi definire tali, gli OpenData prodotti nell'abito delle Pubbliche Amministrazioni debbano essere completi, primari, aggiornati, accessibili, processabili, non discriminatori, non proprietari e privi di licenza. Più precisamente, questi principi affermano che i dati pubblicati devono comprendere tutte le informazioni ricavate da un fenomeno e non devono essere parziali. Nella filosofia Open, infatti, non è consentito celare intenzionalmente e arbitrariamente alcune tipologie di dati. I dati devono inoltre provenire direttamente dalla fonte senza essere modificati, in modo che sia possibile lavorare direttamente sugli originali prodotti. Gli elementi pubblicati devono sempre rappresentare l'ultima versione prodotta di quel dato. A questo proposito è importante considerare l'ipotesi della creazione di un inventario di pubblicazione, consentendo di risolvere facilmente i vari casi di obsolescenza.

Una delle questioni fondamentali nel mondo degli OpenData riguarda poi gli utenti ed il fatto di consentire ad ognuno di questi di raggiungere i dati nel modo più facile ed

15 Fondatore della O'Reilly Media e sostenitore del software libero e dei movimenti Open Source.

16 Fondatore di public.resource.org, autore, ed ex chairman dell'Internet Software Consortium.

allargato possibile, senza limitazioni di accessibilità e uso dipendenti da fattori discriminanti quali sesso, età, etnia, religione, orientamento politico, ideologia, livello di benessere economico, stato o regione di appartenenza.

Il punto riguardante la processabilità dei dati esprime invece l'esigenza di produrre i dati in modo che risultino poter essere facilmente elaborabili dalle macchine. Questo principio incentiva il riuso e la ripubblicazione delle informazioni processate per poter poi essere ulteriormente rielaborate da altri soggetti.

I dati prodotti non possono essere identificati come proprietà di un ente o di un privato, ma devono essere rilasciati come “beni comuni” (commons) e non possono essere soggetti a nessuna forma di controllo o licenza (copyright, brevetto, etc...). Devono, infine, poter essere prelevabili senza costi diretti.

I limiti degli OpenData

Nonostante gli OpenData rappresentino un vero passo avanti verso i concetti di trasparenza ed informazione non suggestiva, le difficoltà che si incontrano nel processo di diffusione ed apertura a quest'ideologia sono rilevanti.

Il primo dei problemi riguarda l'utente finale, il quale, se lasciato sprovvisto di strumenti di lettura ed interpretazione adeguati, potrebbe semplicemente non comprendere i dati o, nella situazione peggiore, fraintendere il significato del dato. Da un certo punto di vista, questa libertà di interpretazione potrebbe essere considerata positiva se si accetta il principio (in qualche modo democratico) secondo cui ogni utente ha il diritto di formulare la propria idea sulla base di ciò che riesce a comprendere. Tuttavia una cattiva interpretazione rischia di portare a conclusioni necessariamente erronee ed in alcuni contesti questo tipo di incomprendimento potrebbe non essere di secondaria importanza.

La seconda riflessione riguarda l'estrema eterogeneità dei dati. Questo fatto è motivato dall'elevato numero di

soggetti produttori di dati che, operando autonomamente e svincolati per principio da normative e restrizioni, definiscono autonomamente le modalità e i formati, sempre scelti all'interno della gamma di possibilità "Open", con cui rilasciano il dataset. Questo processo da un lato incentiva la completa autonomia, dall'altro mina la possibilità di aggregare e strutturare con facilità e senza conversioni ad hoc dataset provenienti da più fonti.

L'interesse comune di tutti i soggetti che aderiscono all'ideologia Open è sì quella di promuovere la libera diffusione di dati, ma il crescente interesse verso la possibilità di poter usufruire di database associativi in grado di poter fare riferimento a diversi insiemi di dati richiederebbe, per quello che possono offrire le tecnologie digitali odierne, di uniformare a livello tecnico alcuni elementi identificativi dei vari campi dei dataset.

2. L'e-journalism e il ruolo della visualizzazione di dati

Nuovi modelli di informazione

“Dall’avvento di internet e dei media sociali, l’ecosistema dell’informazione ha conosciuto un’accelerazione dell’innovazione sotto tutti i profili. In particolare la caduta delle barriere tecnologiche ed economiche alla pubblicazione ha moltiplicato i partecipanti attivi alla produzione di conoscenze pubbliche: il che essenzialmente ha spostato il sistema strategico dei filtri all’informazione da prima a dopo la pubblicazione generando un nuovo paradigma dell’ecosistema dell’informazione.”¹

Nell’ultimo decennio si è diffuso il termine e-journalism, che costituisce un nuovo concetto nel mondo dell’informazione.

1

Fondazione <ahref, visione. < <http://www.ahref.eu> >

Si tratta comunque di giornalismo, il che significa saper produrre e riportare una storia che sia in grado di informare i lettori degli eventi che sono accaduti e degli attori che ne hanno preso parte. La differenza sostanziale dalla forma di giornalismo tradizione consiste nell'implementazione delle nuove tecnologie disponibili, quali computer, e l'utilizzo della rete come mezzo di comunicazione digitale "uno verso molti". Con questa nuova forma di giornalismo, è possibile veicolare contenuti multimediali la cui diffusione segue le dinamiche della rete.

La tecnologia è l'aspetto fondamentale dell'e-Journalism, ma questa nuova forma di informazione presenta implicazioni che vanno ad influire fortemente sulla concezione del giornalismo e sul ruolo del lettore. L'impiego della tecnologia e in particolare di Internet consente ad ogni utente di avere un'ampia libertà di scelta sull'informazione a cui attingere. E' inoltre possibile per chiunque diffondere informazioni o far conoscere il proprio pensiero tramite la rete. Proprio per questo motivo, assistiamo oggi ad un importante cambiamento nel mondo del giornalismo tradizionale, messo alla prova dalla produzione "dal basso" di notizie ed informazioni.

Per produzione "dal basso" si intendono tutti i contenuti ed i documenti prodotti direttamente dall'utente, prima relegato alle sole attività di lettura e confronto. Il ruolo

del giornalista, nel vecchio modello, era di fondamentale importanza e non sostituibile, in quanto il messaggio ed i collegamenti tra una notizia e l'altra venivano veicolati attraverso le sue parole e l'analisi del suo punto di vista, per quanto, professionalmente parlando, mirato ad essere il più imparziale possibile. L'utente si trovava dunque una versione dei fatti raffinata e mediata in diversi passaggi da diversi attori.

L'infrastruttura stessa della rete permette oggi di integrare il precedente modello di giornalismo con le emergenti realtà "2.0". Il valore aggiunto dell'esplorazione di risorse digitali permette di andare oltre le possibilità della carta stampata tradizionale e consiste principalmente nell'offerta di percorsi ipertestuali personalizzabili a discrezione dell'utente, integrati da risorse provenienti da più fonti che partecipano alla costruzione del dato giornalistico. All'interno del cyber-spazio è possibile attivare modelli 2.0 grazie a diverse piattaforme concepite e sviluppate proprio sulla base delle possibilità offerte dalla rete. Il modello di comunicazione che prevale su Internet è quello "tutti verso tutti", dove l'informazione non è più generata da un unico mittente e diffuso a tutti gli utenti, ma bensì risulta essere il frutto di una costante aggregazione collaborativa di sotto-unità informative in costante aggiornamento.

Questo modello di comunicazione, tuttavia, presenta al-

cune citicità legate principalmente alla verifica dei contenuti e delle fonti. Spetta solitamente ai gestori dei centri di aggregazione delle informazioni attuare delle policy di autenticazione del materiale raccolto.

Proprio riguardo a queste tematiche sono sorte alcune iniziative, tra cui quelle portate avanti dalla fondazione [Timu](#). Questa fondazione, istituita dalla Fondazione Bruno Kessler², con sede a Trento, propone come proprio obiettivo lo sviluppo di una ricerca sulla qualità dell'informazione che emerge da internet e dai media digitali. Considerata la natura del web, una definizione dall'altro al basso della qualità dell'informazione non sarebbe possibile e sarebbe, anzi, controproducente. Secondo la visione di [Timu](#), la concezione stessa di qualità dell'informazione dovrebbe essere rinnovata, in modo da poter giungere ad un insieme di pratiche condivise che la tutelino secondo una logica "bottom up".

La Fondazione [Timu](#) ha sviluppato, a questo proposito, un servizio chiamato Timu³, che propone una metodologia di lavoro per la produzione di informazione di qualità.

2 La Fondazione Bruno Kessler svolge attività di ricerca principalmente negli ambiti delle Tecnologie dell'Informazione, dei Materiali e Microsistemi, degli Studi Storici italo-germanici e delle Scienze Religiose.

3 Termine che in lingua swahili significa "squadra".
< www.timu.it >

Questo metodo si basa su quattro principi che sono una sintesi dei principi fondamentali del giornalismo: accuratezza, indipendenza, imparzialità e legalità.

Non esiste un solo tipo di giornalismo. Nel corso del tempo sono state coniate terminologie per indicare modelli di produzione di informazione che differiscono dalla classica stesura della notizia in forma testuale da parte di giornalisti di professione. Tra queste diverse forme vi sono, in particolare, data-driven journalism, visual journalism e citizen journalism.



Fig. 9 - Social Media Landscape. By Fred Cavazza.

Data driven Journalism

Un fenomeno di larga portata che risulta essere strettamente correlato a quello dell'e-journalism è, appunto, il data driven journalism o data journalism. Per data driven journalism si intende quel processo che, a partire da un insieme di dati oggettivi, porta alla costruzione di una narrazione del fenomeno descritto. Questa forma di giornalismo richiede un approccio che è sotto molti aspetti simile a quello della ricerca. Legato alla componente narrativa, infatti, nel data journalism vi è anche l'intento di far emergere elementi non immediatamente evidenti a partire dai dati, di produrre nuova informazione. Questo processo è spesso associato alla rappresentazione in forma grafica o visiva dei dati presi in considerazione.

“Data journalism is obtaining, reporting on, curating and publishing data in the public interest.”⁴

⁴ Jonathan Stray, professional journalist and a computer scientist, Gennaio 2011.

Da questo punto di vista, il Data Journalism, nonostante abbia visto una grande espansione negli ultimi anni e abbia assunto nuove caratteristiche legate alle possibilità offerte dalla tecnologia (gestione dinamica di grandi quantità di dati, interazione e animazione) e dal web (utilizzo di dati in tempo reale, utilizzo dei social media), non è un fenomeno nuovo.

Si possono trovare esempi di analisi e visualizzazioni di dati risalenti ai secoli scorsi, per quanto al tempo questi lavori non venissero direttamente associati all'ambito giornalistico.

Nell'800, Florence Nightingale, infermiera e statistica inglese, realizzò delle rappresentazioni grafiche di dati relativi ai tassi di mortalità nell'esercito britannico.

Il suo istogramma circolare è particolarmente noto e risulta essere un formato usato ancora oggi per la rappresentazione di dati.

How we got to £200bn in quantitative easing

Cumulative total of weekly purchases of gilts and other assets by the Bank of England's asset purchase facility
 Monetary value is proportional to the area of a complete circle of the diameter shown.
 Circles have been cut away to show the number of quantitative easing steps at the end of each month

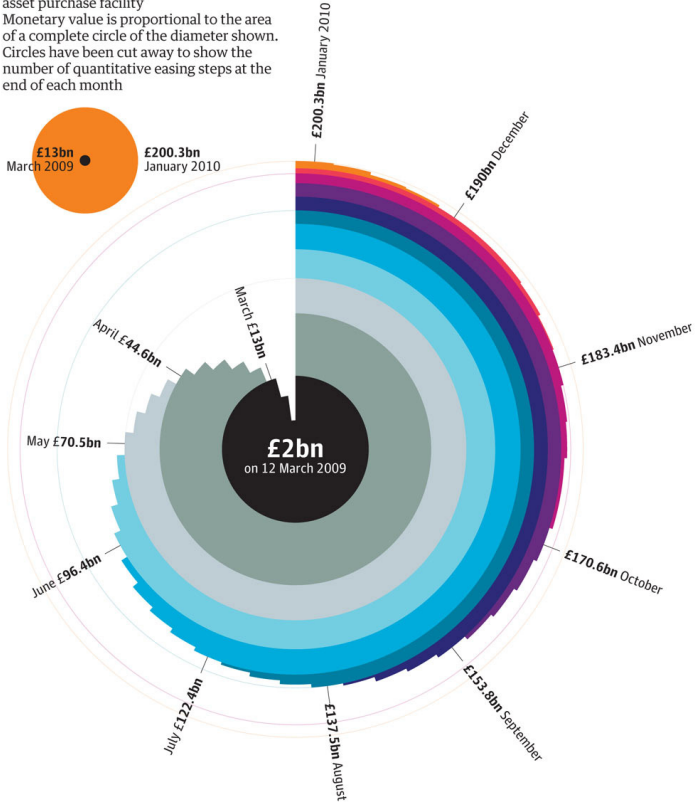
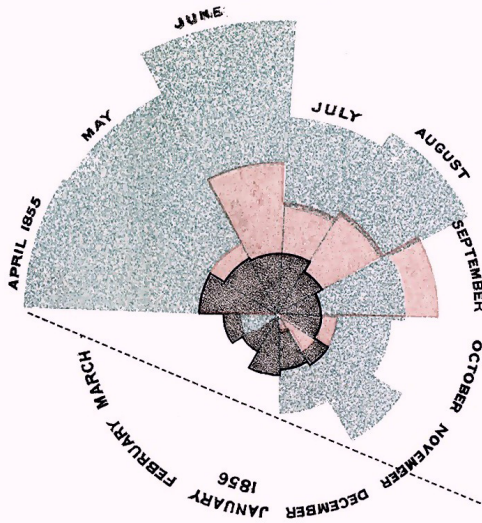


Fig. 10 - “Quantitative easing: How much money has the Bank of England pumped into the British economy?”
 Mark McCormick for the Guardian, 2011

DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY

2.
APRIL 1855 TO MARCH 1856.

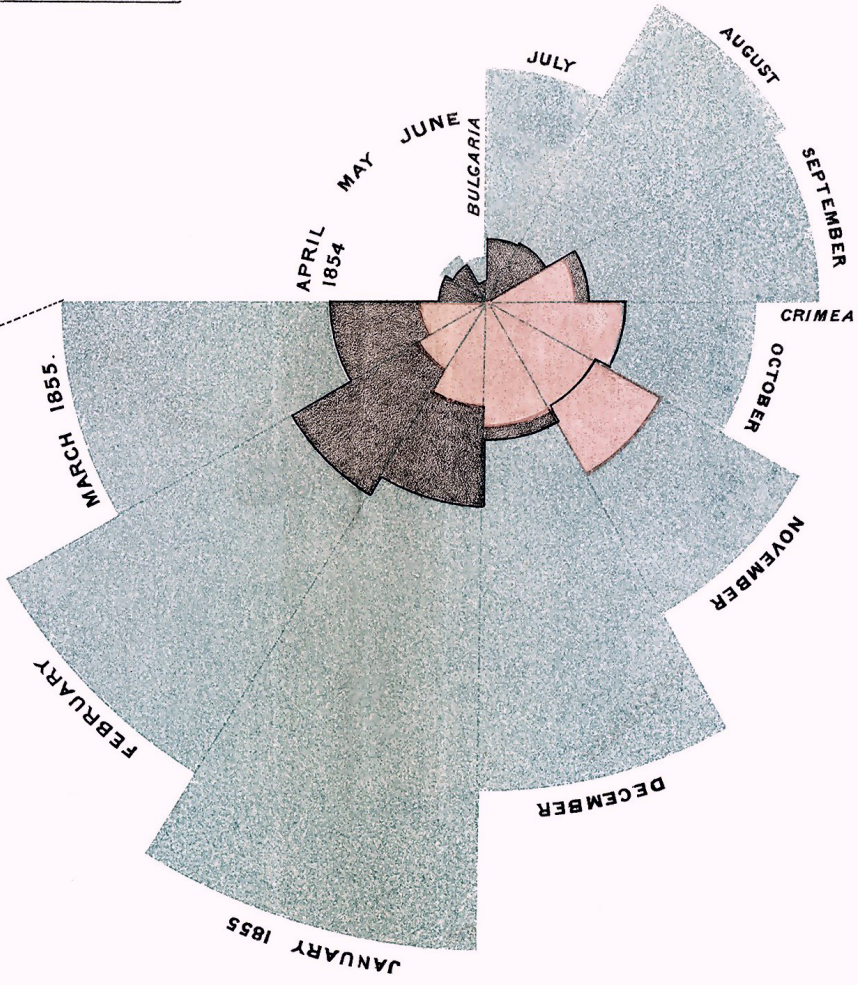


The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Fig. 11 - "Diagram of the causes of mortality in the army in the East" by Florence Nightingale, 1858

CAUSES OF MORTALITY IN THE EAST.

1.
APRIL 1854 TO MARCH 1855.



Nel 1832, André-Michel Guerry, avvocato e statista francese, presentò in un suo saggio (“Essay on moral statistics of France”) alcune mappe tematiche del territorio francese che confrontavano dati relativi a diversi fattori, tra cui tasso di criminalità, tasso di suicidi e livello di istruzione.

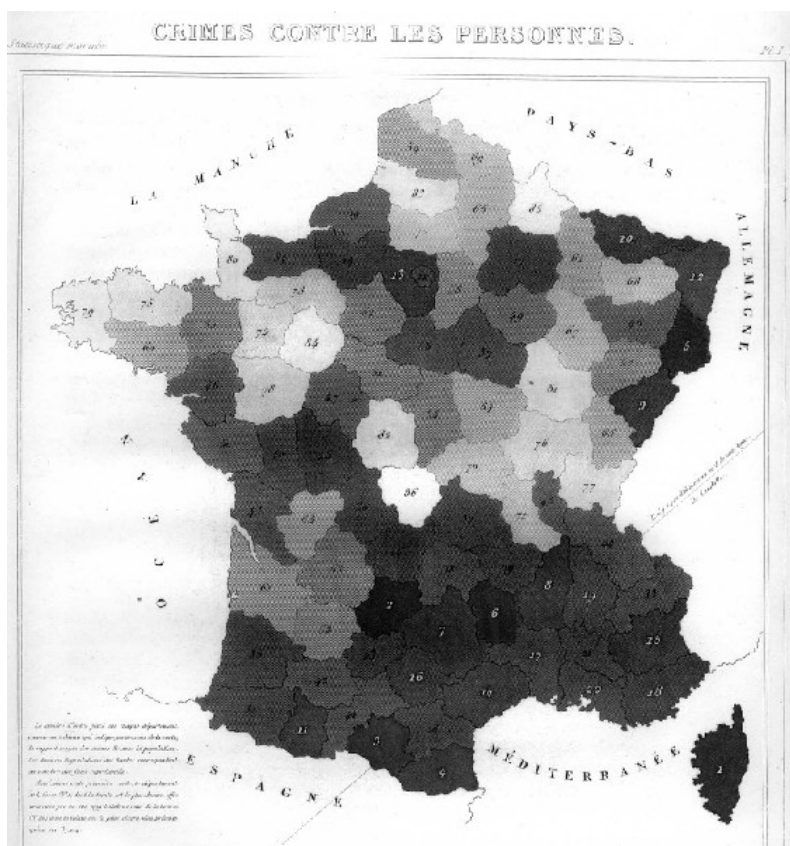


Fig. 12 - Mappa tematica estratta da “Essay on moral statistics of France”. André-Michel Guerry, 1832

Charles Joseph Minard, ingegnere civile francese, considerato uno dei principali pionieri dell'infografica e della data visualization, nel 1869 pubblicò una mappa relativa alla campagna di Napoleone in Russia del 1812 e 1823.

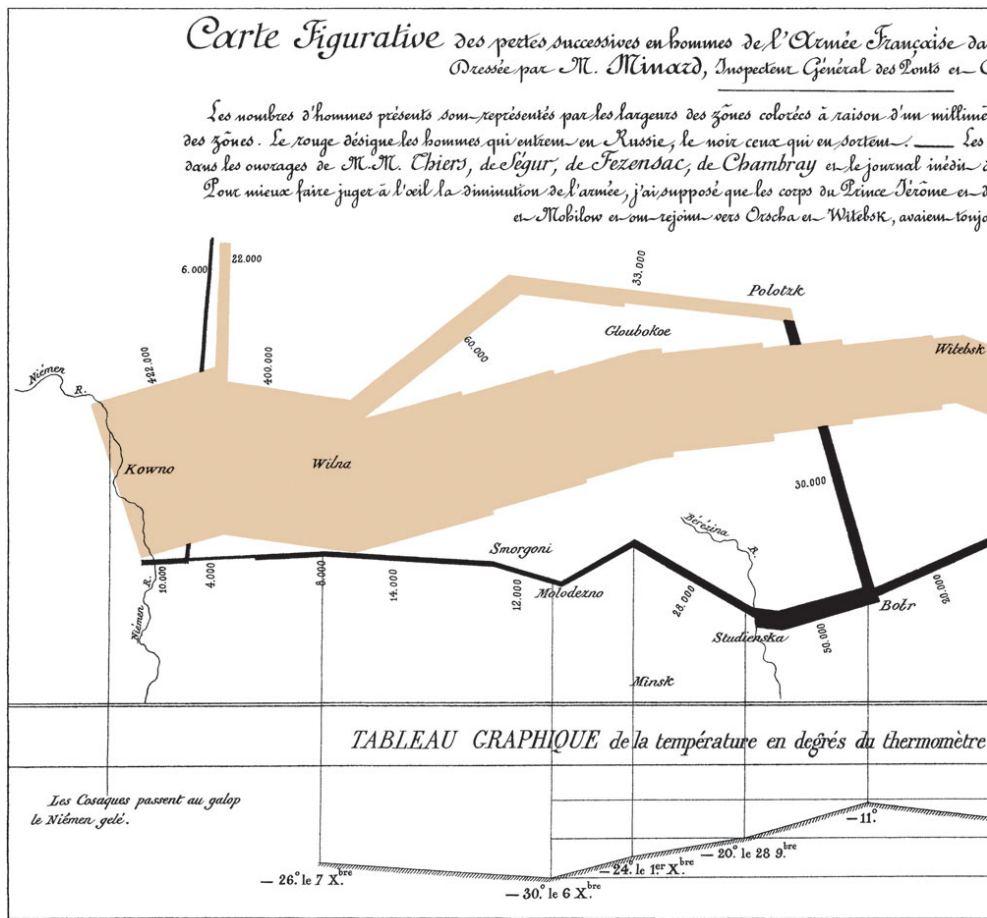


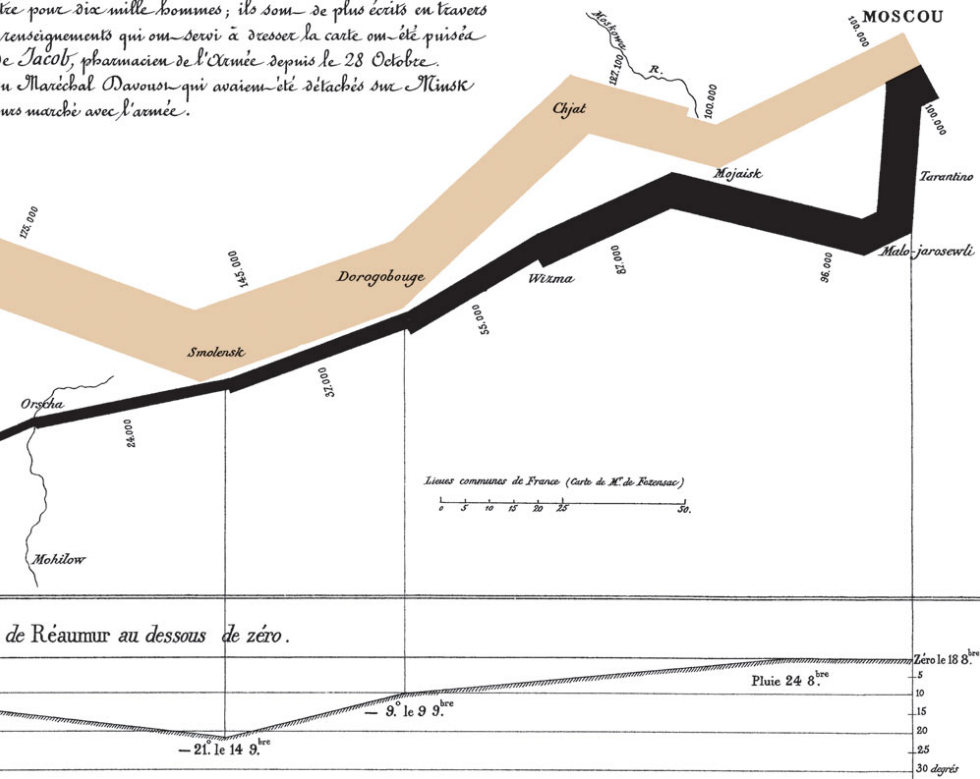
Fig. 14 - "Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813". Charl Joseph Minard, 1869

Questa rappresentazione metteva in relazione la dimensione dell'esercito di Napoleone, le coordinate geografiche relative agli spostamenti dell'esercito, le date e le temperature.

de la campagne de Russie 1812-1813.

Paris, le 20 Novembre 1869.

pour dix mille hommes; ils sont de plus écrits en travers des renseignements qui ont servi à dresser la carte ont été puisés dans le Journal de l'Armée depuis le 28 Octobre. Le Maréchal Davout qui avaient été détachés sur Minsk ont marché avec l'armée.



Imp. lith. Rognier et Dourlot.

Uno dei primi esempi di data journalism vero e proprio, poi, è il reportage sulla rivolta di Detroit del 1967, realizzato dal giornalista Philip Meyer, corrispondente nazionale per il Knight Newspapers. Per la realizzazione del reportage, venne steso un questionario con lo scopo di indagare l'identità e le abitudini degli insorti. Il questionario venne diffuso poi da 30 intervistatori appositamente reclutati ad un campione di oltre 400 individui all'interno delle aree interessate dagli eventi.

The Non-Rioters: A

BY PHILIP MEYER

Free Press Staff Writer

Despite the recent week of rioting, Detroit's Negro community is basically stable and committed to traditional American values.

The rioters were a small and deviant minority—protesting not only against society in general, but against prevailing Negro social standards.

These are the basic findings from a survey research project in the central riot areas of east and west Detroit. The study, based on interviews with a representative sample of 437 Negroes, was sponsored by the Urban League with assistance from the Free Press. Two University of Michigan faculty members acted as consultants.

The findings bear out, to a surprising degree, many of the views of Detroit race relations that were widely held before the riot: Detroit Negroes agree that they are as well off as or in better position than Negroes in other northern cities.

They have many problems but these are not, for the most part, overwhelming. And in seeking solutions, most Detroit Negroes look to the established social structure rather than the revolutionary ideology of black nationalism.

THE PEOPLE
BEYOND 12TH STREET



Then why the riot?

In the first place, it does not take many dissidents to create a riot. Historians have estimated that only a third of the colonists supported the American Revolution when it began.

Persons in the Detroit area were asked, with a guarantee that they would remain anonymous, to assess their own extent of participation in the riot. Twelve percent of the total sample (18 percent of those who replied to the question) admitted some degree of active participation.

Le interviste così completate ogni giorno furono trascritte su schede per computer e i dati raccolti furono analizzati per la stesura del reportage.

Da queste analisi emersero informazioni utili alla comprensione dei fenomeni legati alla rivolta e che contribuirono a sfatare alcuni pregiudizi legati all'identità dei partecipanti.

Fig. 15 - Special survey report. Published by the Detroit Free Press on August 20, 1967. Philip Meyer, 1967

A Hopeful Majority

rioters stand in sharp contrast to majority of Detroit Negroes. Page 1B.

Police brutality cited as the No. 1 cause of Negro discontent. Page 4B.

Many rioters were settling old grievances. Page 4B.

This is consistent with other estimates that most riots have the active support of only 10 to 15 percent of the Negro community.

It is possible that the true rate of riot involvement in Detroit was somewhat higher, due to a reluctance on the part of some to admit participation or to talk to interviewers.

However, the interviewers—all Negroes—reported that most of their subjects were co-operative. For purposes of comparison, the study took all self-descriptions of riot activity at face value.

There was not a basic or lasting breakdown of respect for law and order. The vast majority of Negroes interviewed think of looting, burning and shooting as crimes. They favor fines or jail for looters and jail for the more serious offenses.

Only a few Detroit Negroes see themselves as worse off in terms of education, income and jobs than Negroes in other northern cities.

Of those who see any difference at all, far more see Detroit Negroes as better off—25 percent of those surveyed think Detroit Negroes have more education compared to 10 percent who think they have less.

Forty-six percent think Detroit Negroes have more income, only eight percent think they have less. Thirty-nine percent think Detroit Negroes have better jobs, only eight percent think Detroit Negroes have worse jobs.

There are grievances, of course. And these grievances are related to rioting. But the pattern suggests that they did not constitute an explosive potential that demanded release. The riot could well have been an unlucky accident.

These points are explored more fully in the articles beginning on Page 1B of today's Free Press.

Al giorno d'oggi il numero e la dimensione dei dataset rilasciati e disponibili sta diventando decisamente notevole, anche grazie alla diffusione di open data.

Rendere i dati più accessibili e organizzati è dunque divenuto un aspetto rilevante del data journalism. In questo caso, il supporto delle tecnologie informatiche è alla base dei processi di analisi ed elaborazione dei dataset presi in considerazione. Per questo pare esserci un legame diretto tra e-journalism e data journalism.

Attraverso la rete, è possibile garantire l'accesso a banche dati (anche di considerevoli dimensioni grazie alla diffusione della banda larga) ad un elevato numero di utenti. In questo caso, gli strumenti di visualizzazione ed analisi dei dati ricoprono un ruolo fondamentale, in quanto consentono all'utente di poter trarre significato dal dato grezzo e, nei casi più evoluti, di praticare confronti o aggregazioni.

Il Data Driven Journalism consente di costruire contenuti informativi che si basano su dati oggettivi e convalidati alla fonte. Si tratta di una solida risorsa che oltre a corredare la descrizione del fenomeno con un dato oggettivo permette, nel caso di dati in formato digitale, l'aggiornamento e l'ulteriore elaborazione delle informazioni nel tempo.

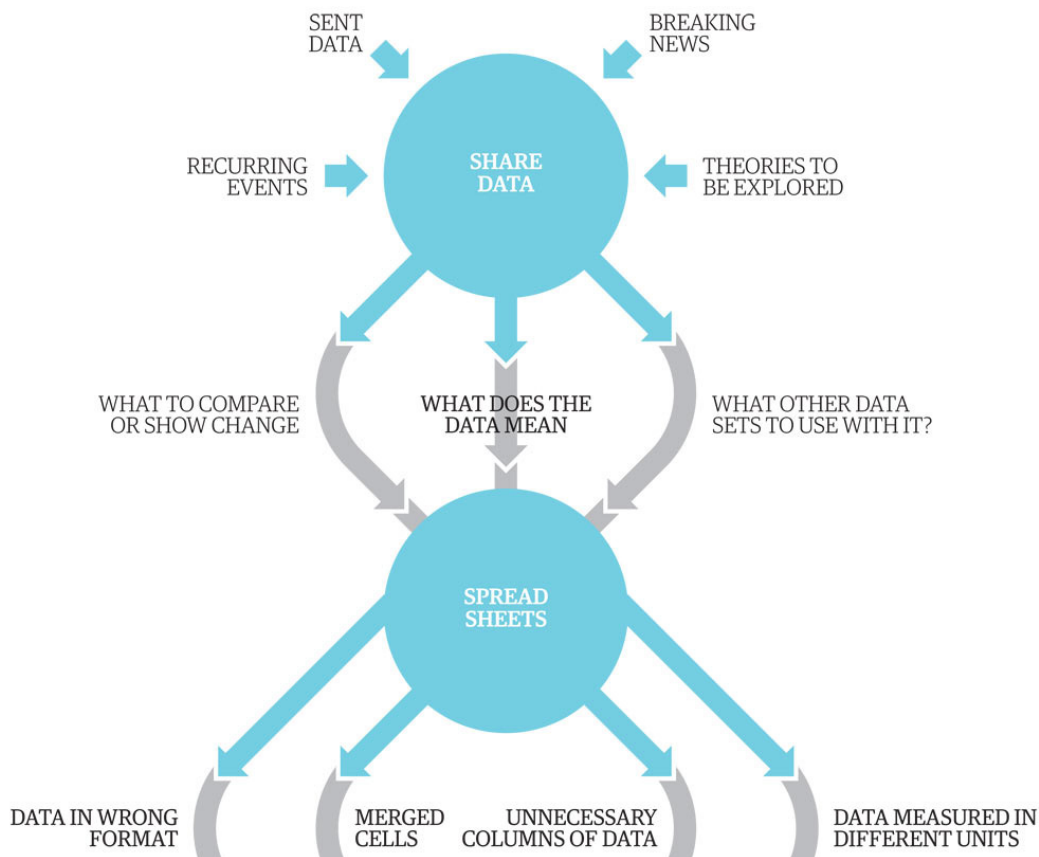
Uno degli esempi più celebri di Data Driven Journalism è quello rappresentato dalla testata giornalistica inglese “Guardian”, che mette a disposizione numerosi dataset fruibili online dall’utente grazie a strumenti di visualizzazione direttamente integrati nella pagina web.

Mark McCormick⁵ ha realizzato una rappresentazione grafica del processo di raccolta, elaborazione e pubblicazione dei dati che viene messo in atto dalla testata. Questo schema ben rappresenta i passaggi fondamentali del data-driven journalism.

5 Graphic artist del Guardian
< <http://www.guardian.co.uk/profile/markmccormick> >

In generale, si possono identificare diversi passaggi attraverso cui passa il processo del data-driven journalism a partire dalla raccolta dei dati fino alla pubblicazione.

Fig. 16 - Guardian data workflow. Parte 1.
Mark McCormick for the Guardian, 2011



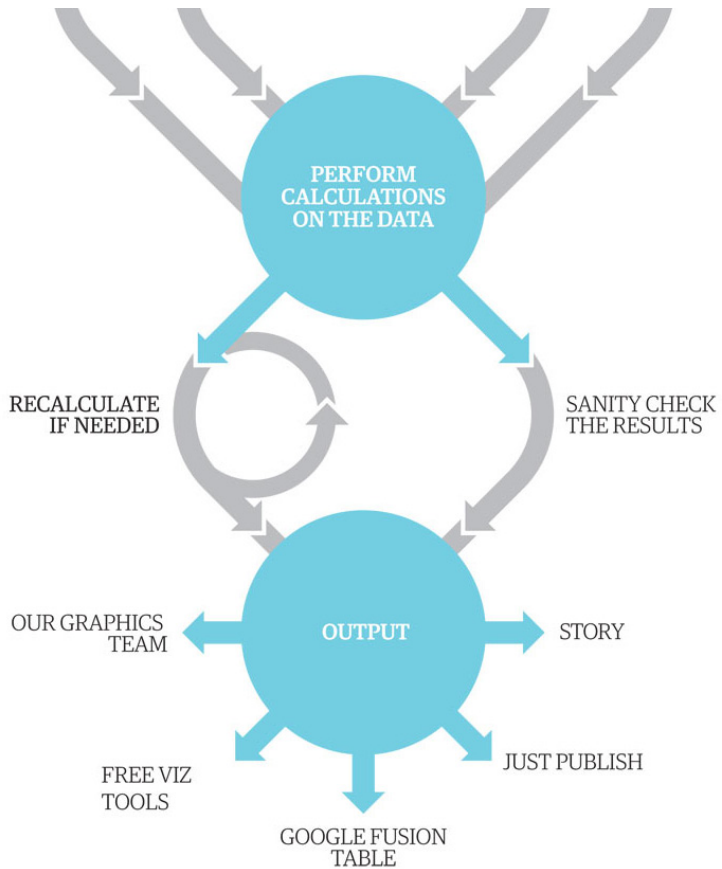


Fig. 17 - Guardian data workflow. Parte 2.
 Mark McCormick for the Guardian, 2011

La prima fase consiste nel reperimento dei dataset e della loro trasposizione in formati coerenti con i formati di utilizzo definiti, come ad esempio può essere la formattazione del database in uso.

La seconda fase prevede l'elaborazione dei dati raccolti e resi processabili attraverso diverse sotto-fasi di raffinamento ed analisi. Nello specifico, è in questa fase che la presenza di standard di formattazione ben definiti consente l'eventuale automatizzazione (attraverso script o semplici macro) dei processi di calcolo ed aggregazione.

La terza fase consiste nel rendere visive le informazioni processate. In questa fase risultano essere fondamentali conoscenze inerenti all'ambito del data visualization e degli strumenti di sviluppo necessari alla creazione di una visualizzazione web-based e, quando possibile e pertinente, interattiva.

La quarta ed ultima fase, strettamente legata alle precedenti e parallela ad esse, rientra a pieno titolo nell'ambito delle conoscenze giornalistiche e prevede la capacità di costruire una narrazione quanto più efficace al racconto dei dati visualizzati. Lo storytelling permette di produrre senso ed informazione utile a partire da un insieme di dati, riunendo i vari aspetti in un insieme organizzato ed omogeneo. Si può dunque affermare che il dato non svilisce la funzione narrativa e che, a sua volta, la narrazione non compromette necessariamente l'oggettività del dato. Entrambi gli aspetti traggono solidità e reciproco sostegno dall'altro, instaurando un rapporto sinergico finalizzato a proporre un'informazione più completa. Dovrebbe inoltre essere pre-

sente, a monte, un passaggio di verifica dell'autenticità dei dati, processo di fondamentale importanza nel mondo degli OpenData.

Per quanto riguarda il passaggio dell'analisi e dell'elaborazione dei dati, esistono diversi strumenti e servizi sviluppati per agevolare queste fasi. Esempi di questi strumenti possono essere Google Refine ⁶, improntato al trattamento di dataset grezzi, e Spotfire ⁷, strumento improntato all'analisi statistica basata su dati.

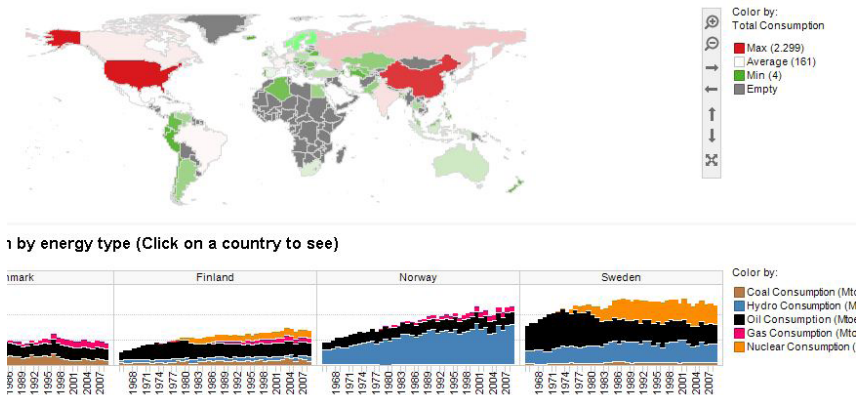


Fig. 18 - Schermata tratta da un'analisi e rappresentazione di dati realizzata con Spotfire.

6 Sviluppato e lanciato da Google nel 2009.

7 Distribuito da TIBCO Software Inc.

Il ruolo della visualizzazione

Perchè visualizzare graficamente i dati?

Nel corso del tempo sono stati effettuati molti studi sul ruolo dell'infografica e della visualizzazione di dati. La rappresentazione grafica di dati non è solo un modo per dare una forma estetica alle informazioni, ma è un aspetto che, oltre a rendere i dati più comprensibili, può a sua volta generare informazione, mettendo in luce aspetti e fenomeni che i semplici dati non elaborati non sarebbero in grado di comunicare. La realizzazione di visualizzazioni di dati richiede un insieme di diverse competenze, che vanno a coprire ambiti che vanno dal design della comunicazione all'informatica, fino alla psicologia cognitiva. Una rappresentazione grafica di dati dovrebbe innanzitutto avere un obiettivo e dei destinatari, deve rispondere ad uno scopo o a una domanda ed essere adatta, nella forma, al tipo di dati che si sta trattando. Dando una particolare attenzione al contesto d'uso, i principi da ricercare dovrebbero essere quelli di efficacia, informatività, novità ed estetica.



Fig. 19 - Anemone. “Using the process of organic information design to visualize the changing structure of a web site, juxtaposed with usage information” By Ben Fry.

L'ideazione e la creazione di un artefatto comunicativo basato su dati è un lavoro che presenta aspetti di complessità al suo interno e che passa attraverso diverse fasi, a partire dalla raccolta dei dati fino ad arrivare all'effettiva realizzazione.

Ben Fry nel libro “Visualizing Data”⁸ identifica a questo proposito 7 fasi: acquisizione, analisi, filtraggio, estrazione, rappresentazione, raffinamento, interazione.

Un elaborato di questo tipo può essere realizzato per comunicare ad altri delle informazioni di cui si è già a conoscenza, oppure al fine di scoprire qualcosa di nuovo a partire da un insieme di dati di cui si è in possesso. Questo secondo aspetto è di particolare interesse nell'ambito della ricerca e del data journalism. La rappresentazione grafica, infatti, può risultare essa stessa uno strumento utile per l'analisi dei dati e consentire, in diversi casi, di arricchire e approfondire la conoscenza in merito a determinati fenomeni.

Si può dire che la crescente disponibilità di dati aperti abbia dato un nuovo impulso all'information visualization. Da una parte si sono diffuse piattaforme che facilitano la realizzazione di rappresentazioni grafiche a partire

8 Scritto nel 2007 per O'Reilly. Il testo tratta dei processi per la visualizzazione di dati ed esamina in dettaglio una serie di casi studio.

da dataset, dall'altro sono state ideate e sviluppate nuove forme di visualizzazione.

Uno dei principali esempi del primo caso è ManyEyes, sviluppato da IBM Research nel 2007. Questo strumento consente agli utenti di caricare dei dati, produrre rappresentazioni grafiche di vario tipo a partire da essi e di visualizzare e commentare le rappresentazioni pubblicate da altri utenti.

Create a visualization in three easy steps



Start with a data set

You have two choices:

- use one of the **existing data sets**  on the site, or
- **Upload your own data set**


Fig. 20 - Passaggi per la creazione di una visualizzazione su ManyEyes.

< www-958.ibm.com/software/data/cognos/manyeyes/ >

See relationships among data points

-  Scatterplot
-  Matrix Chart
-  Network Diagram




Compare a set of values

-  Bar Chart
-  Block Histogram
-  Bubble Chart

Track rises and falls over time

-  Line Graph
-  Stack Graph
-  Stack Graph for Categories

See the parts of a whole

-  Pie Chart
-  Treemap
-  Treemap for Comparisons

Analyze a text

-  Word Tree
-  Tag Cloud
-  Phrase Net
-  Word Cloud Generator

See the world

-  Ottawa Neighbourhood Map
-  US County Map
-  World Map
-  Massachusetts Map
-  New Jersey Map

Fig. 21 - Classificazione dei tipi di visualizzazione su ManyEyes.
< www-958.ibm.com/software/data/cognos/manyeyes/ >

I tipi di visualizzazioni possibili sono piuttosto numerosi e comprendono bar chart, tag cloud, mappe geografiche di diverso tipo, scatterplot e altre tipologie comuni. Un aspetto interessante è il fatto che le modalità di rappresentazione siano suddivise e classificate in base alla funzione: analisi di un testo, comparazione di insiemi di valori, rappresentazione delle relazioni tra gruppi di dati, visualizzazione di frazioni di un insieme. Questo tipo di classificazione aiuta gli utenti a realizzare visualizzazioni di dati in modo corretto e coerente.

Oltre a ManyEyes sono sorti diversi servizi e comunità online con obiettivi e funzioni simili, come ad esempio visul.ly, servizio per la creazione e la condivisione di infografiche e rappresentazioni visive di dati.

L'avvento del web e delle moderne tecnologie informatiche, inoltre, ha aperto nuove possibilità alla visualizzazione, aggiungendo la dimensione dell'interazione e ampliandone altre, come quella dell'animazione.

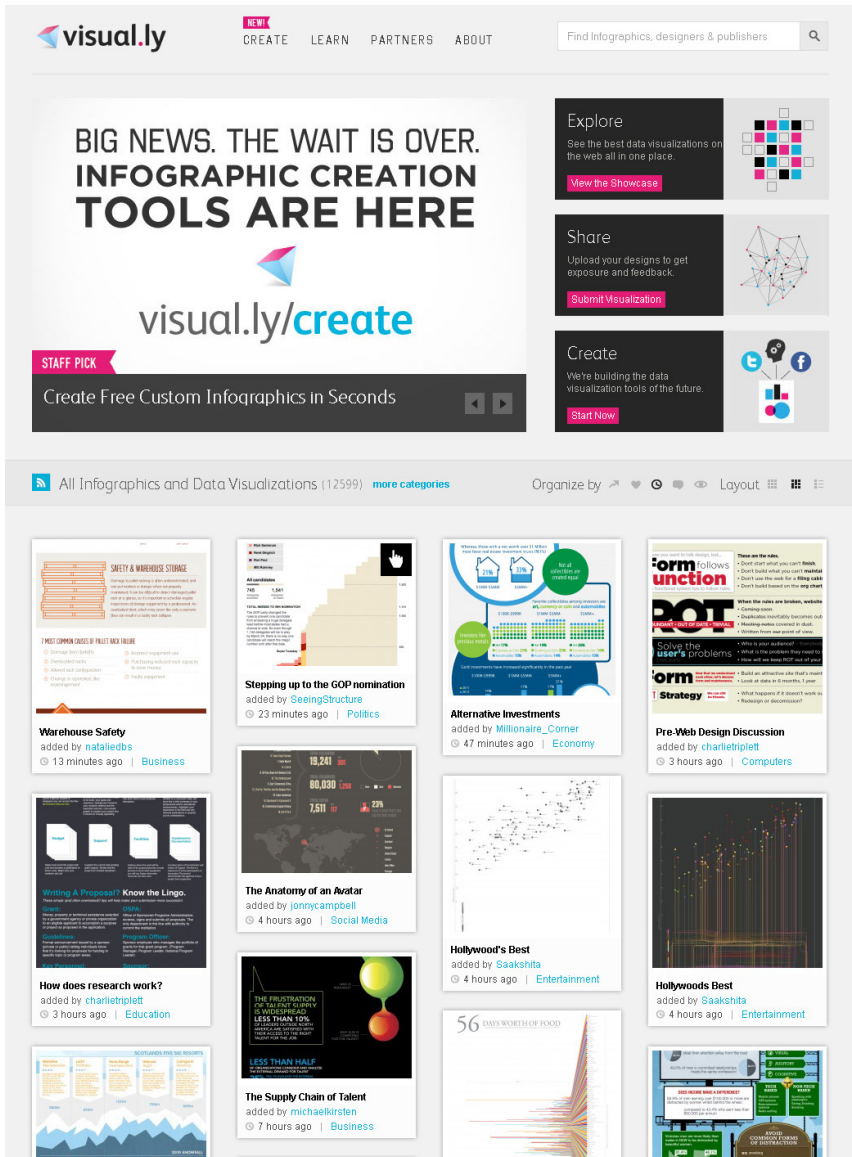


Fig. 22 - visual.ly

Per quanto riguarda invece la produzione di visualizzazioni innovative, è possibile trovare molti esempi di esperimenti o artefatti finiti che utilizzano metodi di rappresentazione non convenzionali o che mettono in relazione in modi particolari diversi fattori, come ad esempio quello temporale e quello geografico.

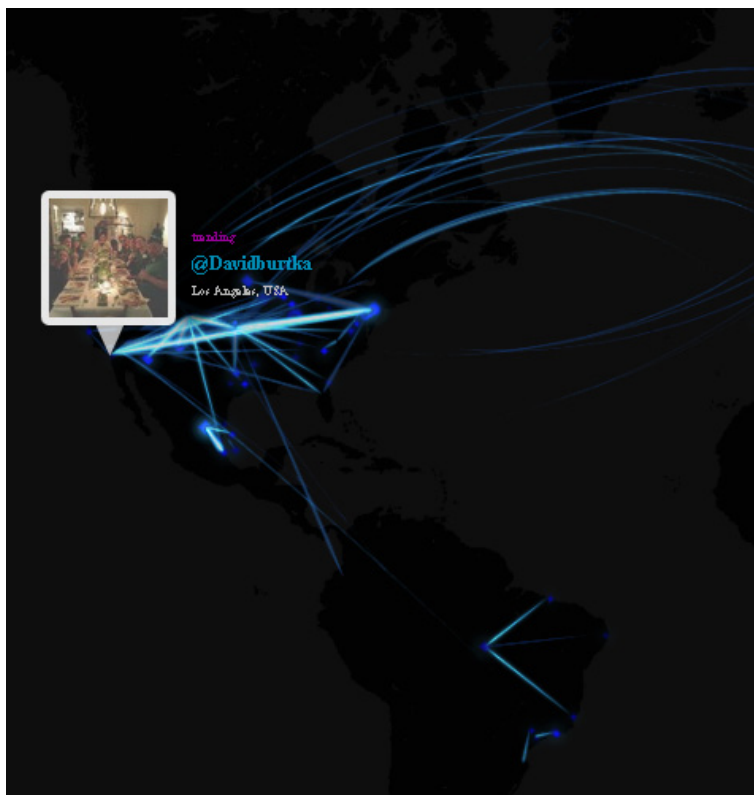


Fig. 23 - Visualising traffic on img.ly. Visualizzazione interattiva in tempo reale realizzata da 9elements nel 2011.

The path of protest

Drag the timeline slider above or use the navigation stick below to explore the road. Roll over the icons to find out more

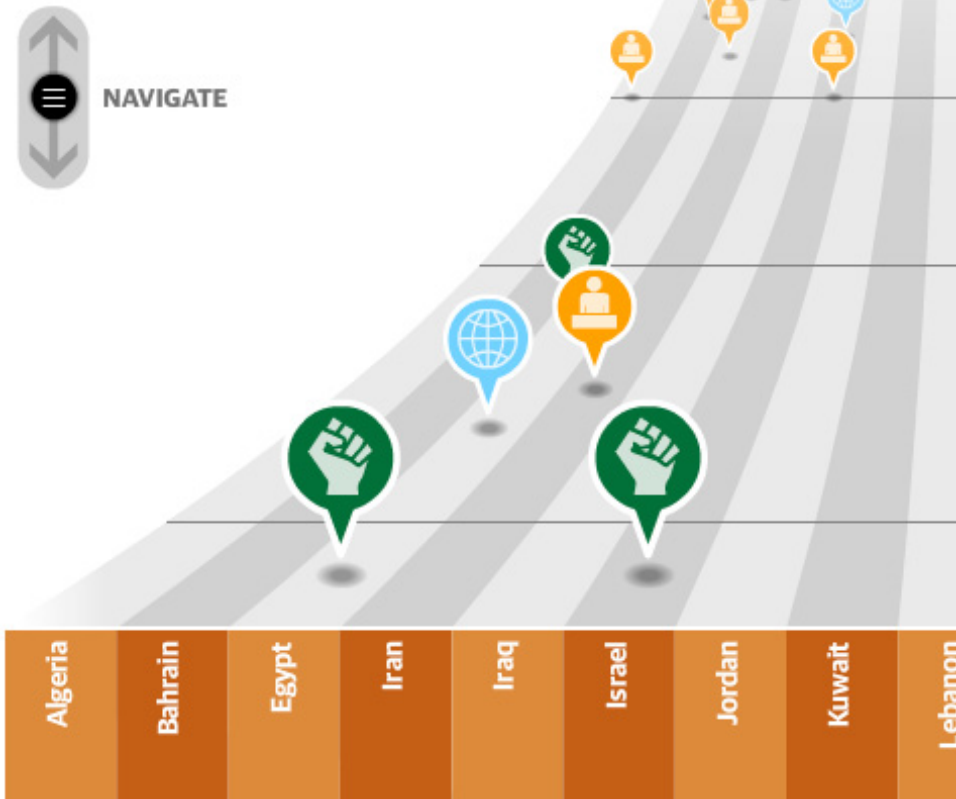
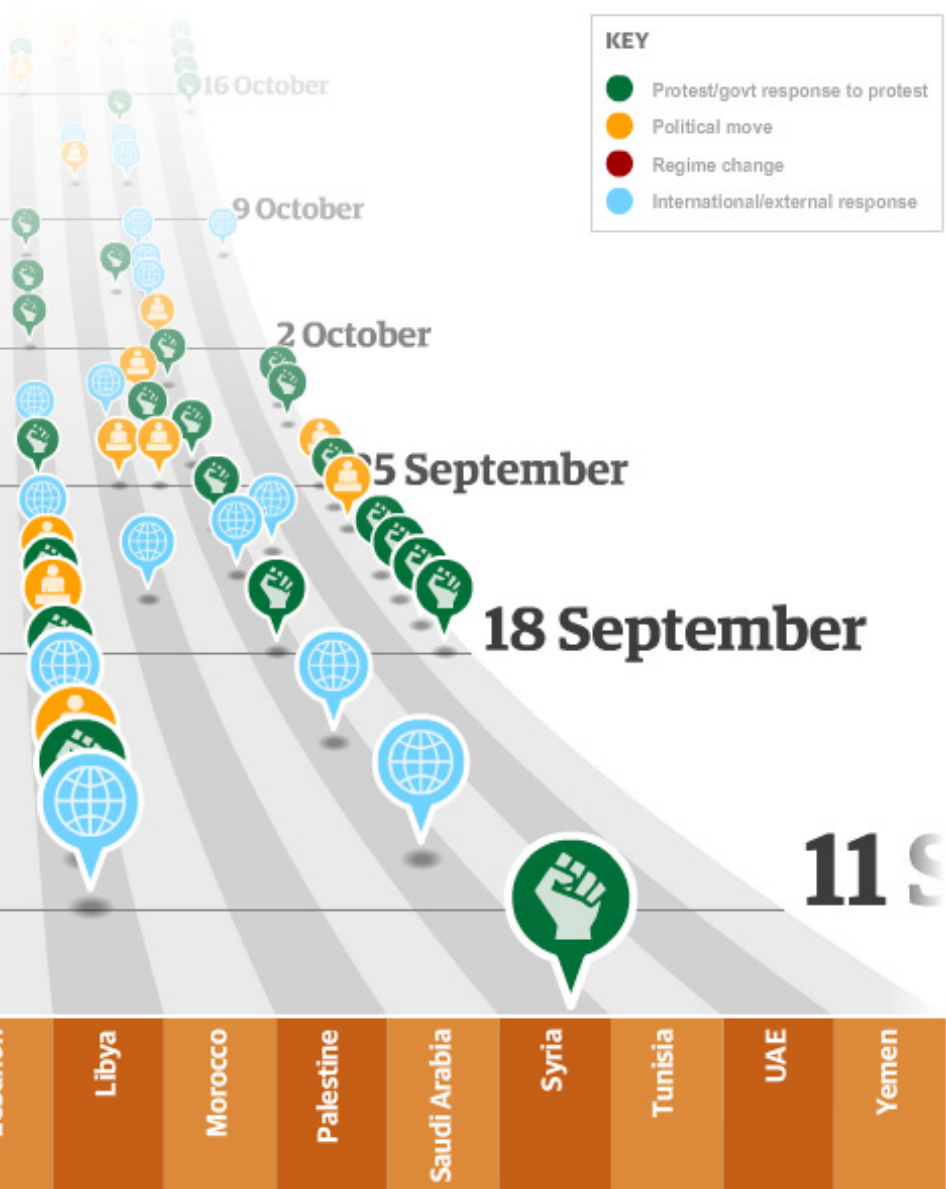


Fig. 24 - "The Path of Protest".

Garry Blight, Sheila Pulham and Paul Torpey, guardian.co.uk, 2012



L'Information Visualization nell'ambito dell' e-journalism e del data journalism

“Journalists are coping with the rising information flood by borrowing data visualization techniques from computer scientists, researchers and artists. Some newsrooms are already beginning to retool their staffs and systems to prepare for a future in which data becomes a medium.”⁹

Come conseguenza della diffusione e della condivisione di un vasto numero di dataset, comprensivi anche di dati capaci di raggiungere un alto livello di definizione e preci-

9 Geoff McGhee, giornalista specializzato in multimedia e information graphics. McGhee è autore del video report “Journalism in the Age of Data”, documentario prodotto nel 2009 che tratta visualizzazione di dati come mezzo di storytelling.

sione, si può affermare che l'Information Visualization assume ora un ruolo di primaria importanza, in quanto unico strumento al tempo stesso capace di essere sintetico ed esaustivo nel descrivere i vari e possibili livelli di lettura proposti dalle fonti.

Il valore aggiunto offerto da questo tipo di strumento risulta essere complementare e spesso sostitutivo alla tradizionale modalità narrativa lineare testuale proposta dal vecchio modello del mondo dell'informazione. Tra le potenzialità offerte da un artefatto digitale vi è senza dubbio quella relativa all'esplorazione a focalizzazione variabile, capace di offrire sempre all'interno della stessa interfaccia un taglio macro o micro a seconda delle esigenze e degli interessi del lettore/utente. E' inoltre presente la possibilità di ottenere aggiornamenti in tempo reale, consentendo all'utente di seguire l'evoluzione di un fenomeno attualmente in atto ed in costante evoluzione.

Un ulteriore aspetto che consente agli artefatti digitali di superare in diversi casi la carta stampata riguarda le possibilità offerte dall'interazione. Dalle più semplici operazioni di filtro e scrematura alle più complesse ed articolate fasi in cui si consente un approccio collaborativo da parte dell'utente (User Generated Content), l'interazione rappresenta il momento di contatto fondamentale tra il fruitore e l'artefatto proposto in cui l'esplorazione del contenuto ed

i percorsi narrativi si modellano a seconda delle esigenze dell'interesse dell'individuo. Ogni interfaccia diventa così uno strumento in grado di rispondere a domande precise e capace di offrire ad ogni visita tagli differenti, ammesso che siano disponibili i relativi dati.

Bisogna comunque tenere in considerazione che il data journalism, come ogni altro tipo di giornalismo, non è del tutto oggettivo. Le visualizzazioni di dati sono in grado di veicolare opinioni. Il modo in cui sono trattati e rappresentati i dati e le informazioni dipende dalle scelte e dai punti di vista del data journalist. La selezione dei dati e degli aspetti a cui dare risalto, la scelta del tipo di visualizzazione e delle scale di grandezza, tra le altre cose, sono tutti aspetti che vanno ad influenzare la percezione finale che i lettori, o utenti, avranno del fenomeno mostrato.

Citizen Journalism e informazione grassroots

Uno degli aspetti fondamentali del processo di digitalizzazione dei mezzi di informazione riguarda la possibilità degli utenti non solo di fruire liberamente dei contenuti, ma anche di partecipare attivamente all'arricchimento dei dati relativi ad un determinato fenomeno. Nuovi tipi di news media, gestiti e utilizzati da singoli utenti o da gruppi, hanno avuto un'ampia diffusione sul web. Tra questi si possono citare blog, siti di informazione, podcast e social network.

Si può dunque dire che un altro aspetto che si è diffuso anche grazie al Web 2.0 è quello del Citizen Journalism, detto anche giornalismo partecipativo o collaborativo. In realtà, il termine è nato precedentemente. Nel testo “We the media”¹⁰, Dan Gillmor, giornalista statunitense, fa risalire le origini del citizen journalism al diciottesimo secolo,

¹⁰ “We the media: Grassroots Journalism by the People, for the People”, scritto da Dan Gillmor e pubblicato da O'Reilly nel 2004

quando alcuni intellettuali iniziarono a diffondere le proprie pubblicazioni tramite pamphlet.

La definizione attuale di citizen journalism, tuttavia, si è diffusa molto più recentemente. Questo tipo di giornalismo, infatti, comporta oggi la partecipazione attiva dei cittadini o degli utenti del web alla produzione di informazione. Questo fenomeno è reso possibile soprattutto grazie alla diffusione di internet e di piattaforme di condivisione, quali social networks o reti di blog, che permettono lo scambio libero di notizie, idee e materiale multimediale. L'informazione viene generata o approfondita dagli utenti e diffusa attraverso questi tipi di siti web o altri servizi di condivisione. In questo modo, ogni singolo utente può contribuire all'incremento delle informazioni prodotte attraverso i mezzi con cui può interagire. Non è necessario che il cittadino che produce e diffonde notizie abbia competenze specialistiche, ma solo che segua alcuni principi che garantiscano un certo livello di qualità dell'informazione prodotta.

Secondo CitMedia¹¹, iniziativa no-profit affiliata all'università di stato dell'Arizona e all'università di Harvard e mirata a supportare e incoraggiare l'informazione "grassroot", i principi del citizen journalism comprendono

11 < <http://citmedia.org> >

accuratezza, capacità di approfondimento, trasparenza e indipendenza, concetti simili a quelli che regolano l'operato dei giornalisti di professione.

Un aspetto che rende importante il lavoro svolto dal lato utente è la rapidità di diffusione e la possibilità di condivisione dell'informazione.

Tra gli strumenti più utilizzati per questa forma di giornalismo vi è Twitter¹², noto strumento di microblogging creato nel 2006 da Jack Dorsey. Questa piattaforma gratuita ed open source, impostata e presentata come social network, consente di pubblicare messaggi, ipertestuali caratterizzati dalla peculiarità di essere limitati ad una lunghezza massima di 140 caratteri. Questo aspetto consente all'utente di inviare aggiornamenti anche attraverso dispositivi mobili, utilizzando un comune SMS.

La popolarità di questa piattaforma ha subito un notevole incremento nel 2007, al South by SouthWest festival, un evento pubblicitario di notevole rilevanza e visibilità durante il quale il numero di messaggi giornalieri inviati è aumentato da 20 mila a 60 mila.

12 < twitter.com >

I messaggi inviati a Twitter sono indicizzati in base a dati quali data, orario, località di provenienza e attraverso hashtag¹³. Ogni query lanciata dalla form di richiesta cercherà di rintracciare i messaggi scegliendo prioritariamente tutti quelli che presentano come hashtag la parola o le parole della ricerca, consentendo un buon livello di ricerca tematica.

Nel Luglio del 2011, il Guardian ha pubblicato una timeline interattiva ricavata dall'analisi di circa mezzo milione di tweet relativi alla decisione di Rupert Murdoch di chiudere il "News of the World"¹⁴. Questo è un esempio della crescente importanza che viene attribuita ai social network nell'ambito dell'informazione e del giornalismo. Quelli che una volta erano fruitori passivi delle notizie ora contribuiscono a diffondere l'informazione e la critica relativa ad essa.

13 Con questo termine si intendono particolari parole chiave evidenziate direttamente dall'autore all'interno del proprio messaggio, precedute dal carattere "#" e riconosciute da Twitter come tag di ricerca.

14 La rappresentazione, dal titolo "How Twitter tracked the News of the World scandal" è visualizzabile su < <http://www.guardian.co.uk/media/interactive/2011/jul/13/news-of-the-world-phone-hacking-twitter> >



7 events
20 PM



**nes Murdoch announces
 closure of the News of
 World**

chairman of News International says the final issue of the 168-year-old paper is to be published on 10 July – ‘out of respect’ for the press caused by the scandal. There is widespread disillusionment among News International staff. Ed Miliband, Labour leader, calls for Rakesh Brookes to resign.



Most retweeted



@stephenfoley
 Stephen Foley
 New York, NY

Nice one @burrell RT @S_R_Morrison Read last par from Independent piece on Wednesday: <http://t.co/bsBQRwR> #NOTW

Most frequent words

- sun
- out
- arrested
- tomorrow
- hacking
- phone
- staff
- over

Fig. 25 - Schermata tratta da “How Twitter tracked the News of the World scandal”. guardian.co.uk, 2011

Gli aspetti sociali ed antropologici del giornalismo su internet

Il cambiamento del tradizionale punto di vista del lettore passivo di notizie legate al mondo dei media, principalmente riferito a giornali e televisione, obbliga ad una rivalutazione del ruolo dell'utente ed alle modalità sia di diffusione che di fruizione dei contenuti informativi.

Una prima considerazione che si può fare riguarda il cambiamento del ruolo attivo dei media da singolo punto di riferimento per la diffusione di informazioni a figura condivisa all'interno dell'insieme di sorgenti informative. Nel primo caso il soggetto gode di un'aura di autorevolezza quasi totale, l'unico caso in cui si crei un confronto diretto prevede il ruolo attivo del fruitore che si attiva per attingere da più fonti le notizie relative ad un unico fenomeno. Nel secondo caso, invece, la moltitudine di fonti a contatto suggerisce di per sé la possibilità di effettuare facilmente confronti ed approfondimenti paralleli. La struttura stes-

sa della rete favorisce ed incentiva questo processo di esplorazione trasversale.

Questo nuovo contesto in cui sono entrati a far parte i media, garantisce un approccio maggiormente meritocratico, in quanto le discriminanti principali saranno relative alla pertinenza alla domanda dell'utente, la ricchezza di contenuti e al livello di approfondimenti trasversali proposti. Tuttavia proprio questo aspetto può essere problematico per alcune testate affermate nel panorama giornalistico, che possono essere messe in crisi dalla presenza di fonti meno autorevoli, le quali, limitatamente ad un particolare argomento, possono però risultare più specifiche e approfondite.

A loro volta, le fonti sono sottoposte al confronto ed alle opinioni degli utenti. Nonostante la maggioranza dei fruitori non sia dotata di competenze professionali, ogni utente può esprimere il proprio pensiero sulla qualità dell'informazione diffusa da un soggetto, o sito web, piuttosto che da un altro. La risultante di tutti i commenti e di tutte le opinioni scambiate può assumere una rilevanza tale da poter essere decisiva.

Anche in questo caso, la possibilità di garantire all'utente uno spazio di scambio direttamente sul portale di accesso dei contenuti informativi (da semplici commenti

postati, ad interventi puntuali e categorizzati come l'area "discussione" di ogni pagina creata sul modello wiki) garantisce grandi vantaggi per quanto riguarda l'interazione e la vicinanza dei lettori, ma espone necessariamente il portale al giudizio comune.

L'opportunità che si presenta a tutti i produttori di informazioni online o in formati digitali che hanno modo di attingere a diverse fonti è quella riguardante l'offerta informativa e il modo in cui questa viene proposta agli utenti. L'utilità, in questo caso, non consiste semplicemente di aggregare diversi contenuti informativi all'interno di un unico spazio visivo. Questa modalità di presentazione, purché più semplice ed immediata, viene sorpassata dai servizi che offrono la possibilità di aggregare diverse sorgenti di feed in un'unica pagina. Questi servizi e applicazioni sono detti aggregatori.

Attraverso l'uso di un aggregatore, è l'utente che sceglie come popolare la sua prima pagina e lo farà in base ai propri interessi o alle proprie necessità. E' proprio per questo motivo che una riorganizzazione meccanica di contenuti proposta a priori non potrà risultare così efficace come quella costruita personalmente dal lettore.

Anche in questo caso lo spunto ed il vantaggio per creare valore aggiunto risiede nella possibilità di costruire diver-

si percorsi tematici o narrativi in grado di riunire attorno ad un unico macro argomento tutta una serie di contenuti correlati, arricchiti, quando possibile, da materiale multimediale, sfruttando quindi l'aspetto crossmediale della piattaforma su internet.

Oltre a questo, diventa ora di cruciale importanza essere in grado di portare all'utente la notizia più fresca e recente possibile. Il tema dell'aggiornamento dei feed spesso diventa il parametro discriminante principale su cui si basa la scelta di una fonte piuttosto che un'altra. La ricerca dell'ultimo aggiornamento, tuttavia, richiede alla testata in questione un'attenta e tempestiva valutazione di autenticità e correttezza delle informazioni che stanno per essere pubblicate.

L'esempio più immediato riguarda l'utilizzo di Twitter come strumento di "live feed" in merito ad un determinato argomento: trattandosi di un canale fortemente orientato dall'opinione di utenti spesso non esperti del settore, con precisi orientamenti politici o, più banalmente, con una percezione del fenomeno sensibilmente distorta dal proprio punto di vista, l'impiego di questo mezzo come contenuto integrativo impone un'attenta valutazione di quale messaggio pubblicare.

Molto spesso, per far fronte a questo tipo di problema, lo spazio dedicato ai feed come possono essere quelli provenienti da Twitter è relegato ad un box all'interno del quale viene dichiarato espressamente che il contenuto è costituito da updates provenienti da non esperti del settore.

Si può dunque affermare che l'utente sia ora al centro del nuovo sistema informativo. Da lui dipendono tutte le scelte legate ai percorsi tematici intrapresi, non essendo più attore passivo che fruisce di un'impostazione narrativa decisa a priori dalla redazione. Data l'estrema libertà e facilità con cui lo strumento telematico connesso in rete fornisce alternative, non è più possibile evitare il confronto diretto con altri centri di informazione.

La differenza tra "locale" ed "internazionale" diventa ora un aspetto puramente tematico e non più un'impostazione legata a precise scelte politiche o redazionali. In rete infatti è molto difficile evitare il confronto con altre fonti di tipo internazionale o provenienti da realtà differenti.

Un esempio rilevante di come questo aspetto abbia avuto implicazioni notevoli è quello della censura di internet attuata dal governo cinese al fine di impedire questo tipo di confronto con informazioni ed opinioni fortemente divergenti dalla linea di pensiero ufficiale proposta dal governo.

In realtà, questo processo di confronto avveniva già in precedenza anche sui media tradizionali. In quel caso tuttavia il numero di fonti con cui confrontare i fatti e i dati di un certo fenomeno era fortemente limitato alla disponibilità locale; per i giornali, principalmente a quelli appartenenti alla propria nazione o addirittura regione, per radio e TV ai canali disponibili.

Anche per quanto riguarda questi media, comunque, un'informazione ambigua o parzialmente scorretta può mettere in dubbio la credibilità della testata, ma il processo è meno rapido e meno evidente. Inoltre, la condivisione di opinioni e la verifica collettiva di un determinato fatto descritto si limita ad un ristretto raggio di utenti, mentre su internet la scala è esponenzialmente più elevata e la rapidità di diffusione non ha precedenti.

Un'ulteriore differenza riscontrabile riguarda la trasformazione del concetto di prima pagina. Dapprima vetrina delle notizie e delle tematiche che avrebbero accompagnato le uscite all'interno della settimana, ora la "prima-pagina" su internet, o meglio, la home-page di un giornale introduce le novità del giorno, senza necessariamente decretare un filone tematico persistente per un determinato periodo.

L'approccio dell'utente, molto spesso, è quello di selezionare le categorie di notizie da cui costruire una propria prima pagina attraverso i più svariati strumenti di aggregazione.

Molto importante per le testate giornalistiche online è anche il tema della persistenza delle notizie in archivi digitali e la loro facile ed immediata reperibilità. Ogni notizia pubblicata risulta essere indicizzata ed archiviata secondo criteri informatici (tag e link) che consentono all'utente l'accesso a quella precisa risorsa quasi istantaneo e permettono di correlare due notizie pubblicate anche lunga distanza tra loro, consentendo all'utente di visionare entrambe le risorse come se si trattasse di un unico corposo argomento.

E' evidente come il web lasci ampio spazio non solo a tutto ciò che può essere definito "ufficiale" ed "istituzionale", ma soprattutto alla popolazione di utenti attiva, in grado di produrre contenuto e, potenzialmente, anche in grado di costituire una massa critica dotata in egual misura di forza e visibilità rispetto alla fonte ufficiale. Si tratta di un aspetto che difficilmente potrà essere ignorato dal lato "ufficiale", che dovrà essere in grado di promuovere un certo livello di attività lato utente e garantirne la totale libertà, pena l'abbandono o, perlomeno, una sensibile riduzione del proprio bacino d'utenza.

E' molto importante dunque attivare processi di osservazione ed analisi di tutto ciò che è ambito "social" sul web, per monitorare trend e opinioni, soprattutto se questi spazi di condivisione sono contenuti all'interno della popolarità infrastruttura digitale.

Esiste, infatti, una profonda differenza tra ciò che un individuo afferma se direttamente interrogato e ciò che sarebbe in grado di esprimere in piena autonomia di pensiero. E' sempre stato possibile raccogliere pareri ed informazioni dagli utenti, ma è sempre stato necessario impostare determinate situazioni e contesti di scambio dove difficilmente l'utente, o lettore, era pienamente incentivato a condividere la propria opinione, come nel caso di convegni, rassegne o altre modalità di coinvolgimento riguardo a determinati fenomeni. Ancora più difficile sarebbe stato trovare degli spazi da impiegare per avviare discussioni in merito alla testata stessa.

L'avvento della rete è dunque riuscito a dare all'utente quel livello di piena autonomia e libertà di espressione fino a quel momento impossibile da ottenere attraverso altri canali.

La sfida più grande oggi è dunque quella di non limitarsi semplicemente a quantificare il volume dei commenti, ma è riuscire a capirne il significato ed interpretarne la di-

reazione. E' molto difficile, soprattutto attraverso tecniche automatizzate, interpretare in termini assoluti se un commento sia "positivo" o "negativo". L'analisi semantica automatizzata, ad oggi, non è ancora in grado di percepire i significati intrinseci del linguaggio verbale, tanto meno di comprendere l'uso di alcune figure retoriche che negano il significato affermato, come ad esempio l'ironia.

La popolazione attiva sulla rete è in costante aumento e le tecnologie digitali sembrano muoversi a supporto dell'utente, cercando di facilitare la condivisione di qualsiasi forma di comunicazione prodotta. Grazie a questo fenomeno, è possibile disporre di dati in costante aumento e diffusione.

E' dunque necessario muoversi lungo due direzioni parallele. Una prevede il supporto verso l'utente di tutti i contenuti informativi disponibili, l'altra prevede il supporto di tutte le tecnologie disponibili per attivare uno spazio virtuale di scambio tra utenti.

Proprio per la natura dell'ambiente virtuale che va a costituirsi nel cyber spazio, è facile comprendere come i confini tra una fonte informativa ed un'altra siano possibili e definibili solo dal lato precedentemente indicato come "ufficiale", mentre per quanto riguarda il lato "social" di qualsiasi piattaforma, la commistione di molteplici pool di

dati ed informazioni non solo risulta inevitabile, ma spesso rappresenta una fonte di ricchezza e competenza della piattaforma su cui sta avvenendo questo tipo di scambio.

Legati a queste evoluzioni nell'ambito del giornalismo sul web, si sono sviluppati diversi fenomeni relativi al modo di produzione e organizzazione dell'informazione.

Il Crowdsourcing e i vantaggi della collaborazione di massa

Come già accennato, uno dei nuovi fenomeni, particolarmente in risalto e dotato di una certa rilevanza nel panorama dell'e-Journalism, è il crowdsourcing. Nello specifico con questo termine si identifica un sistema costituito da individui riuniti all'interno di uno spazio virtuale (la rete).

All'interno di questo gruppo prendono vita attività dedicate alla realizzazione di un progetto comune. In particolare modo, riferendosi al mondo del giornalismo, per crowdsourcing si intende l'intenzione di sfruttare le tecnologie offerte dal Web 2.0, ovvero la possibilità di trarre beneficio dalla collaborazione di massa, soprattutto per quanto riguarda i canali di microblogging, come nel caso di Twitter.

I vantaggi che si possono ottenere dall'ascolto e dall'utilizzo dei dati provenienti dalle masse riguardano

soprattutto il modesto costo, inteso in termini economici, di accesso alle informazioni. In quasi tutti i casi si assiste ad un processo di generazione collettiva e spontanea dei contenuti e questo, per ovvi motivi, non comporta un costo diretto. L'ascolto di queste informazioni consente all'ente di ricerca di conoscere a priori le opinioni o le reazioni che un determinato fenomeno suscita nelle masse.

Il crowdsourcing può assumere due forme distinte: la prima prevede un approccio collaborativo, dove ogni singolo utente è consapevole di contribuire al raggiungimento di un fine comune attraverso la condivisione diretta e conscia del proprio lavoro con altri utenti. La seconda, invece, prevede la raccolta del materiale individuale di ogni utente, aggregato poi in un secondo momento.

Il limite principale per questo modello democratico consiste, anche in questo caso, nel processo di autenticazione e di verifica delle informazioni raccolte. E' facile pensare che argomenti che trattino di temi sensibili all'opinione pubblica, come ad esempio l'ambito politico, possano presentare un alto tasso di faziosità, aspetto che, definito un limite di tolleranza oggettivo, può rendere inattendibile una determinata fonte.

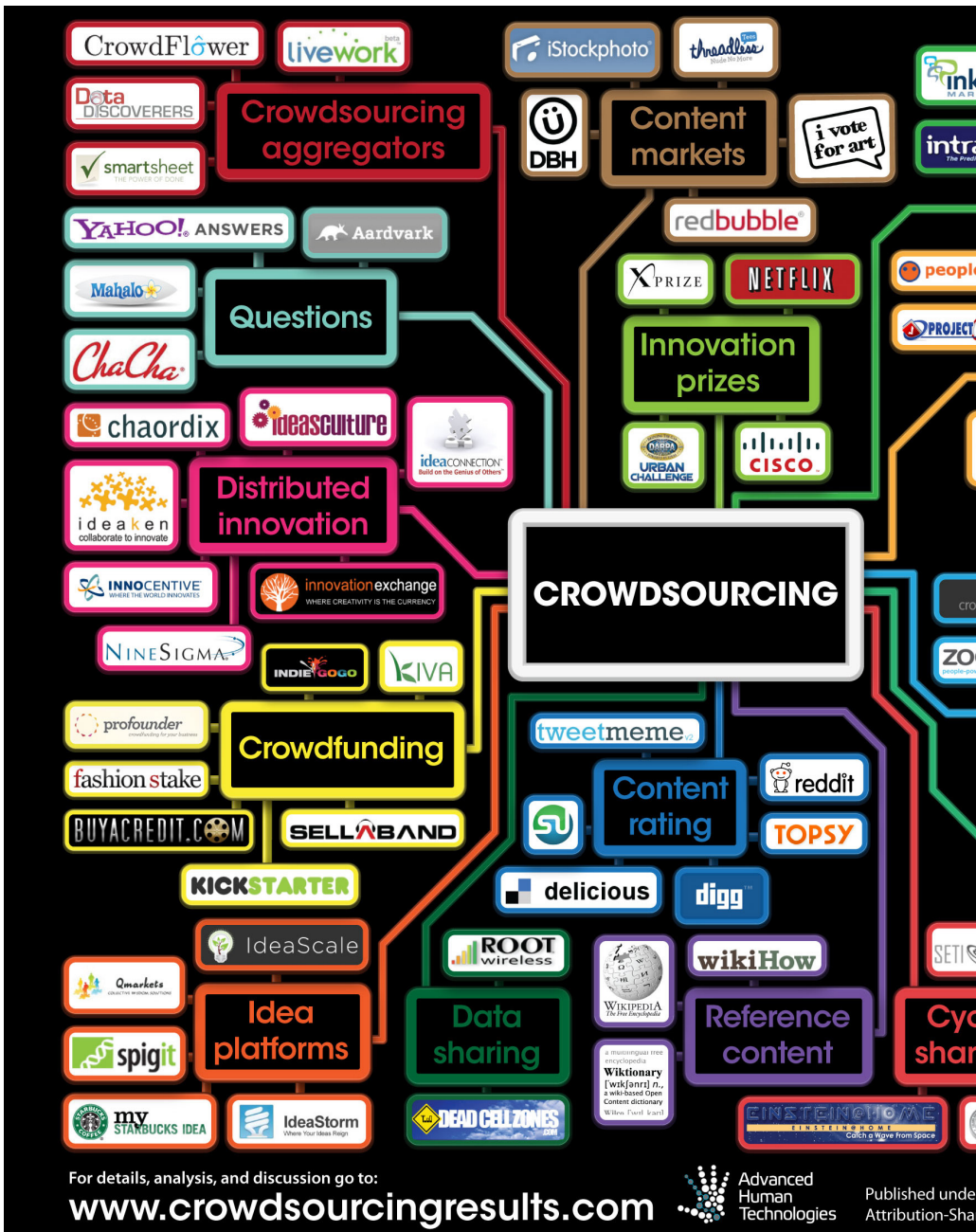


Fig. 26 - Crowdsourcing Landscape. crowdsourcingresults.com, 2010



Common Crowdsourcing Tasks and Examples

- 3D object design
- Advertising
- Business ideas
- Clothing
- Consumer research
- Crisis information
- Data analysis
- Fact checking
- Graphic design
- Human reading
- Investigative reporting
- Journalism
- Lending
- Mapping
- Movie reviews
- Music
- Observation
- Patent research
- Philanthropy
- Political activism
- Product design
- Proofreading
- Scientific problems
- Software
- Software development
- Software testing
- Stock picking
- Tagging
- Translation
- Trends
- TV programming
- Word of mouth
- Writing and editing

for a Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License

Created by Ross Dawson
Design by Daniil Alexandrov

Georeferencing: la diffusione di informazioni georeferenziate

Un ulteriore aspetto che pare opportuno trattare è quello della georeferenziazione. Georeferenziare un dato significa definire le caratteristiche che lo legano allo spazio fisico. Molto spesso per georeferenziazione si intende quel processo secondo il quale si attribuiscono coordinate spaziali di un sistema di riferimento topografico o geografico. Il primo requisito fondamentale per poter condividere informazioni all'interno di un sistema georeferenziato è assicurarsi di utilizzare un sistema di riferimento comune. Proprio per questo motivo, il sistema di riferimento utilizzato come standard è quello delle coordinate GPS, un sistema di posizionamento e navigazione satellitare che fornisce posizione ed orario in ogni condizione meteorologica ovunque sulla Terra, ammesso che vi sia un contatto privo di ostacoli con almeno quattro satelliti.

Il vantaggio principale ottenuto tramite l'uso di informazioni che sono riferite direttamente al territorio, consiste nell'aver la possibilità di identificare pattern o relazioni al di là della dichiarazione di senso oggettiva dei dati stessi. Una particolare zona geografica, ad esempio, potrebbe essere caratterizzata da un contesto o da un insieme di eventi che, se analizzati singolarmente ed indipendentemente, non sarebbero in grado di fornire un quadro completo su determinati fenomeni, mentre, se scollegati dalla propria collocazione spaziale, possono fornire informazioni più approfondite e far emergere relazioni che intercorrono con altri eventi.

Un esempio può essere l'analisi delle rivolte civili dal 2003 ad oggi, localizzate nell'area medio orientale e nord africana: è evidente come il contesto politico dell'area sia un dato di fondamentale importanza per poter comprendere i mutamenti di quei popoli. Più nello specifico, è possibile notare come dopo il conflitto in Iraq del 2003, i regimi di diversi stati confinanti abbiano iniziato ad essere destabilizzati da rivolte civili.

Questa correlazione di fatti legati all'area geografica permette di comprendere come gli effetti di un particolare fenomeno possano influenzare altri avvenimenti o aree, per motivazioni che possono essere riconducibili alla vicinanza dal fatto di riferimento oppure a fattori di diversa

natura. La disponibilità di informazioni georeferenziate consente di individuare ed analizzare queste particolari relazioni. Si nota, infatti, come negli ultimi anni l'utilizzo di informazioni georeferenziate sia cresciuto sempre di più, soprattutto negli ambiti "web 2.0".

I maggiori esempi dell'utilizzo di questo nuovo tipo di informazione sono sicuramente Twitter, il quale consente all'utente di allegare ad ogni tweet la propria posizione al momento del post e Facebook, che analogamente a Twitter, permette di integrare questo tipo di informazioni con ogni aggiornamento.

Molti servizi ed applicazioni fanno uso di questi dati, avvicinando in alcuni casi la distanza tra virtuale e reale ed offrendo all'utente nuove possibilità di relazione con il territorio, grazie alle informazioni tratte dalla rete. Esempi di questo tipo comprendono Google Maps e foursquare¹⁵.

15 Social network basato sulla geolocalizzazione. Foursquare è stato lanciato nel 2009 da Foursquare Inc.

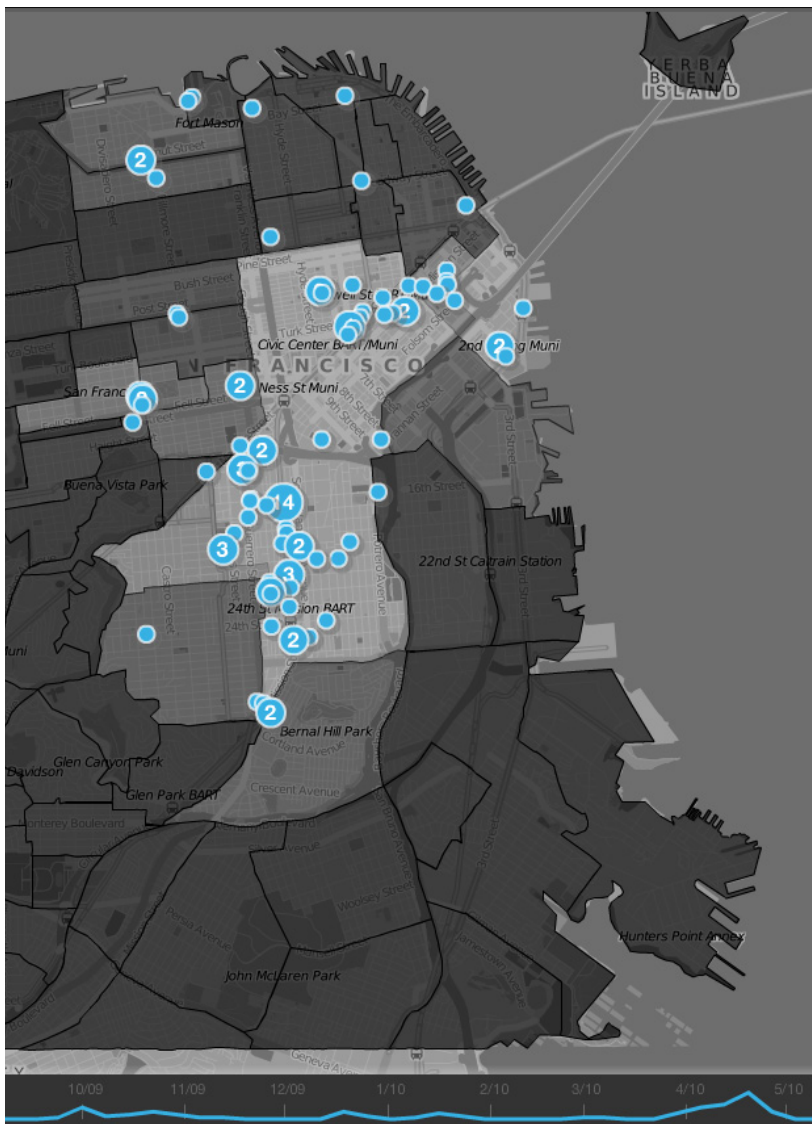


Fig. 27 - WeePlaces. Schermata tratta da WeePlaces, applicazione che visualizza i dati registrati tramite foursquare.

L'utilizzo di dati in tempo reale

Oltre al crowdsourcing e alla geolocalizzazione, una delle caratteristiche forse più importanti offerte dalle possibilità derivanti dall'utilizzo di tecnologie informatiche, supportate dalla connessione costante alla rete, è senza dubbio la possibilità di ricevere e inviare dati ed informazioni in tempo reale.

Per “dati in tempo reale” si intende un accesso istantaneo ed immediato all'informazione, virtualmente senza alcun ritardo e senza che sia necessario attendere che essa venga processata in un secondo momento, a posteriori dalla sua raccolta.

L'approccio dell'utente in questo caso cambia radicalmente: dal report di un fenomeno già accaduto in un passato più o meno recente, ora il lettore è partecipe e testimone dell'evoluzione degli eventi, avendo la possibilità di assistere ed osservare tutti i mutamenti che si susseguono istante per istante. Questo nuovo modo di confrontarsi con le informazioni, unito alla possibilità di condividere pareri

ed alla possibilità di intervenire direttamente con il proprio contributo all'interno delle public areas, pone l'utente all'interno di un nuovo concetto di giornalismo.

Il saper comprendere i dati, processarli, estrarre significato, visualizzarli e comunicarli sta diventando sempre più importante negli ultimi anni. Sono aspetti che stanno continuamente evolvendo e che probabilmente ricopriranno un ruolo sempre più rilevante anche negli anni a venire.

Nel mondo del giornalismo, questo processo è di fondamentale importanza. Come già citato diverse volte, un chiaro esempio dell'impiego dei dati applicati alla visualizzazione all'interno del mondo del giornalismo è il Guardian's Data Store. Un altro esempio molto importante di applicazione del data journalism è stato anche il fenomeno "Wikileaks", la cui forza mediatica era concentrata nel significato di quei dati, in grado di raccontare in modo oggettivo aspetti considerati come "non pubblicabili".

L'interesse mostrato dai giornalisti verso questo nuovo mondo dell'informazione porta una parte di questi ad investire sempre di più nella data visualization e nella creazione di una narrativa capace di far comprendere meglio un determinato fenomeno.

La Situazione attuale sul web

Alla luce delle considerazioni trattate fin'ora, è interessante andare ad osservare quali sono le piattaforme oggi attive sul web che presentano particolari aspetti di innovazione nell'ambito dell'e-journalism e della gestione e comunicazione di informazioni di tipo giornalistico.

Innanzitutto è utile distinguere tra alcuni diversi tipi di servizio. Come descritto in precedenza, sono definiti aggregatori di notizie quelle applicazioni o piattaforme che riuniscono un insieme di informazioni, spesso provenienti da diverse fonti, e le aggregano, proponendole all'utente in modo organizzato così da agevolarne la ricerca e la consultazione.

Google News e News Map

Uno dei principali esempi in questo senso è Google News, lanciato in versione beta nel 2002 da Google Inc. e reso definitivo nel 2006. Questo servizio raccoglie notizie giornalistiche da alcune migliaia di fonti ufficiali e le aggrega in base ad un algoritmo.

Prima pagina

- Fonsai
- Elsa Fornero
- Lucas Papademos
- Lin
- Coppa Italia
- Telecom Italia
- Gianni Alemanno
- Cesare Cremonini
- Tonino Guerra
- Vasco Errani
- Michigan, Stati Uniti
- Esteri
- Italia
- Economia
- Scienza e tecnologia
- Spettacoli
- Sport
- Salute
- In evidenza

Prima pagina



Torino, sparano contro il consigliere

il Giornale - 33 minuti fa 🔍 +

Il capogruppo del Terzo Polo al Comune è stato centrato un casco, che lo attendeva appena sotto casa, in via B all'ospedale Le Molinette, ...



Strage Tolosa, killer braccato "Mi arrendo"

il Giornale - 33 minuti fa

Il mostro è sotto assedio. Da stanotte la polizia ha circospezione l'uomo sospettato di essere il killer dei tre bambini e di ...



Lavoro: offensiva Cgil contro riforma "scio"

AGI - Agenzia Giornalistica Italia - 34 minuti fa

(AGI) - Roma, 21 mar. - Si inasprisce il clima tra il Governo e i sindacati. Il presidente della Cgil, ...



Del Piero: "Avanti, siamo la Juve"

Sport Mediaset - 56 minuti fa 🔍 +

Alessandro Del Piero è il trascinato della Juve in Coppa Italia. Anche all'Inter. "Nessun traguardo ci deve essere precluso. Rimanerci dentro un ...



Pd diviso sulla riforma del lavoro. E la rabbia

Il Sole 24 Ore - 1 ora fa

La rabbia degli elettori Pd monta su web. Dopo la rottura con il governo, una parte della base è sempre meno convinta dell'appoggio al governo. ...



Sim false, indagata Telecom per violazioni

Milano Finanza - 1 ora fa

Nell'inchiesta aperta dalla Procura di Milano sulle sim false, Telecom è sotto accusa. Nell'inchiesta aperta dalla Procura di Milano sulle sim false, Telecom è sotto accusa. ...

Fig. 28 - Google News. < news.google.com >

REGISTER LOGIN CUSTOMIZE + SELECT ALL

ARGENTINA AUSTRALIA AUSTRIA BRASIL CANADA FRANCE GERMANY

ES

Calderón pide respuesta internacional "abrumadora" para crisis en ...

Ministros de Salud y Transporte llegan a Coyhaique y se reúnen con ...

Los rectores defienden el programa de campus de excelencia

El PSC "suspende" sus relaciones con CiU tras el pacto ...

Jefe de Policía ve "proporcionada" la acción policial y dice que ...

Unos doscientos jóvenes expresan en Sol su solidaridad con el IES ...

La CEOE pide retirar la prestación a quien rechace un trabajo ...

España y otros once países europeos instan a Bruselas a adoptar ...

El BCE interrumpe la compra de deuda soberana por primera vez ...

"Eva Hache fue tan buena payasa como siempre, aunque se echó en ...

Una flamante Salma Hayek pisa la alfombra roja de los Premios Goya ...

Diego estará un mes de baja

US

Chris Wishe Happ 40s

AU

Red Negro Hum Fire

NL

Lawi Frisco grote geda

UK

Lloyds bank claws back £1.5m bonuses from directors

Mortgage lending rises 10% in January

Manchester United's Sir Alex Ferguson to retire 'in two or three years'

Six Nations 2012: Wales coach Rob Howley tells players to use the pain of 2010 ...

Hopes for gene analysis device lift IP Group

ECB bond buying halted after eurozone debt rally

OPF says it could go bust in weeks

SE

newsmap

Mon February 20, 2012 21:29:12

powered by Google™ hosting by htcml

NEWSMAP IS ©2004-2010 MARCOS WESKAMP | CONTACT | BLOG | ABOUT

BETA FEEDBACK

Fig. 29 - News Map. < <http://newsmap.jp/> >

Brown
es Rihanna
y Birthday

s and 50s:

Former Glock CEO's
racketeering trial begins
this week in Georgia

Spain sends planes
to US for shipwreck
treasure

Source: A's
agree to deal
with Manny

Ron Jaworski: Jets should
go after Colts' CB Peyton
Manning

Andre Villas-Boas: I would
not have accepted Chelsea
job if I could not handle ...

Rahul defies Kanpur
authority, booked

DMDK entry makes it a
four-cornered contest

Securing border is
top priority: Antony

MCP wants
Congress help for
Maharashtra
improvement

Cross
negotiating
humanitarian Cease-
fire in Syria

FR Le contrat ponctuel
avec Shell redonne
espoir aux salariés
de ...

Le Paris-SG est
encore en rodage

Après Royal, Lang
brigue le «perchoir»

La France
toujours
opposée aux
OGM

MEXICO México y EU
firman acuerdo
petrolero

Bella Villalón
sube en
mercado
turquesa, espere
relación Grecia

01:55PM
MARTÍN

Un minuto disputado elección
hablen subido: FE

Chavistas
medicinas
consiguen indio

Difícil
situación
no

Balcón de la
presidenta
muerte en Zoológico

Solicitud
Congreso
en la
SCA

Bajo cero perfume
con olor a "prostituta
caja"

Re
go
no

line die
trof
er dan
acht

IT Terza: nuovo
attacco Boko
Haram, morti

Napolitano ai
genitori di
Rossella Urru:
"Sta bene"

BR registra redução de
acidentes no Carnaval
em relação a 2011

De Si Simpson? chega a região
500 com contêiner político

DE Neuer Rekord an der
Zapfsäule: Benz ist so
teuer wie nie

BR Si Simpson? chega a região
500 com contêiner político

Sinque rebeldia
Berta-400000
da festa de jazz no
Sulão

Olga
Chavichid
nãofica de
Ustadhi na
Liga Europa

News Map, sviluppato dall'ingegnere statunitense Marcos Weskamp, fornisce una rappresentazione grafica, in forma di tree map, dei contenuti più recenti di Google News, suddividendo le informazioni in base alla categoria di riferimento.

News Map fornisce all'utente alcune opzioni per personalizzare la visualizzazione e i contenuti da mostrare. In particolare, è possibile scegliere la nazione di provenienza delle notizie e le tematiche, suddivise in categorie quali affari, tecnologia, sport, intrattenimento e salute. La luminosità dell'area che rappresenta ogni notizia, inoltre, fornisce un'indicazione su quanto questa sia recente.

Esistono poi piattaforme che fungono in parte da aggregatori, ma in cui il processo di raccolta e organizzazione delle informazioni non è eseguito in modo automatizzato dal sistema. In questi casi, questo passaggio è eseguito dagli utenti o dai gestori del servizio; vengono solitamente trattate quantità più ristrette di informazioni e fonti e vi è una particolare attenzione all'aspetto dello storytelling, della narrazione.

Storify

Storify¹⁶, piattaforma web ideata da Burt Herman e Xavier Damman, è uno strumento che consente di creare storie a partire dall'informazione e dai contenuti proposti da altre fonti o diffusi, spesso in modo frammentario, sui diversi social networks. Gli utenti, una volta registrati, possono iniziare a comporre liberamente le proprie storie anche grazie ad un sistema di ricerca implementato che consente di cercare direttamente i contenuti desiderati su Google e alcuni dei principali social network del momento e di trascinarli nell'area di lavoro di Storify.

16 Storify ha vinto il Knight-Batten Awards nel 2011 per l'innovazione nel giornalismo ed è stato nominato uno dei 50 migliori siti Web per il 2011 da Time Magazine. < storify.com >

Gli utenti hanno anche la possibilità di arricchire le composizioni con i propri contenuti e di condividere le storie così realizzate. Questo sistema di aggregazione e organizzazione permette di generare informazione e arricchire quella già presente online. Il servizio, lanciato nel 2010, è destinato sia a giornalisti professionisti che ai singoli utenti.

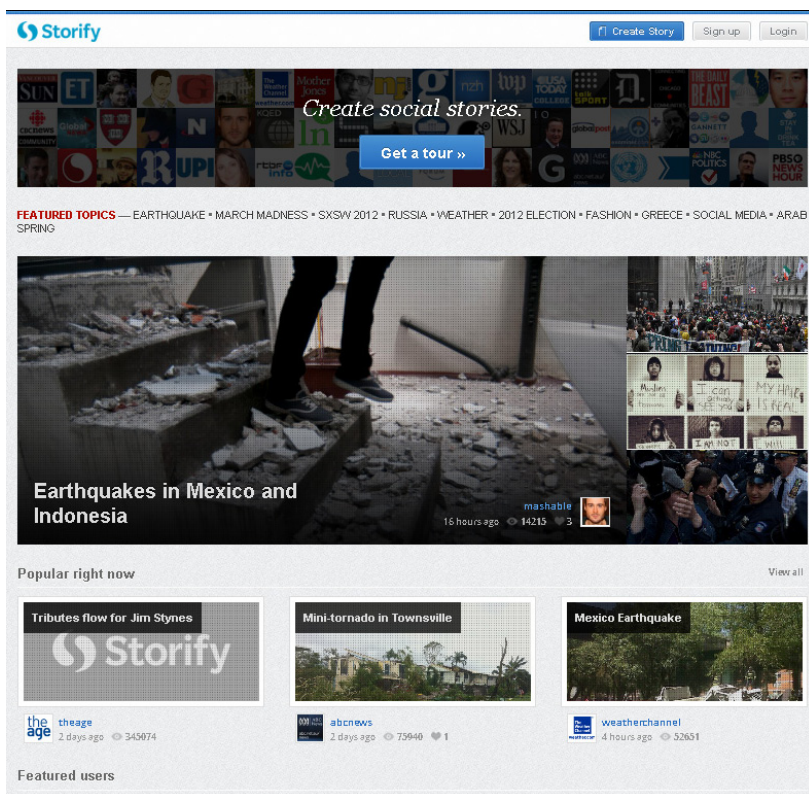


Fig. 30 - Storify.

Storyful

Storyful è stato creato da un gruppo di giornalisti, con l'intenzione di mettere in evidenza l'informazione di qualità che può essere estrapolata dal web e, in particolare, dai social networks. Uno degli obiettivi principali di Storyful è quello di far emergere la voce di persone in qualche modo legate o vicine alle notizie prese in considerazione, come giornalisti locali, fotografi amatoriali o abitanti del luogo.

“Storyful’s golden rule is there is always someone closer to the story. And in the last few months, we’ve worked with people at the heart of the action, capturing turning points in history in words, pictures and video.”¹⁷

Storyful si concentra principalmente sulle notizie che presentano una certa diffusione su Twitter e cita diverse fonti, sia testuali che multimediali, aggregandole in modo da facilitarne la consultazione.

17 < storyful.com >

I curatori del servizio selezionano il materiale e ne controllano i contenuti e le fonti, in modo da garantire la qualità dei contenuti.

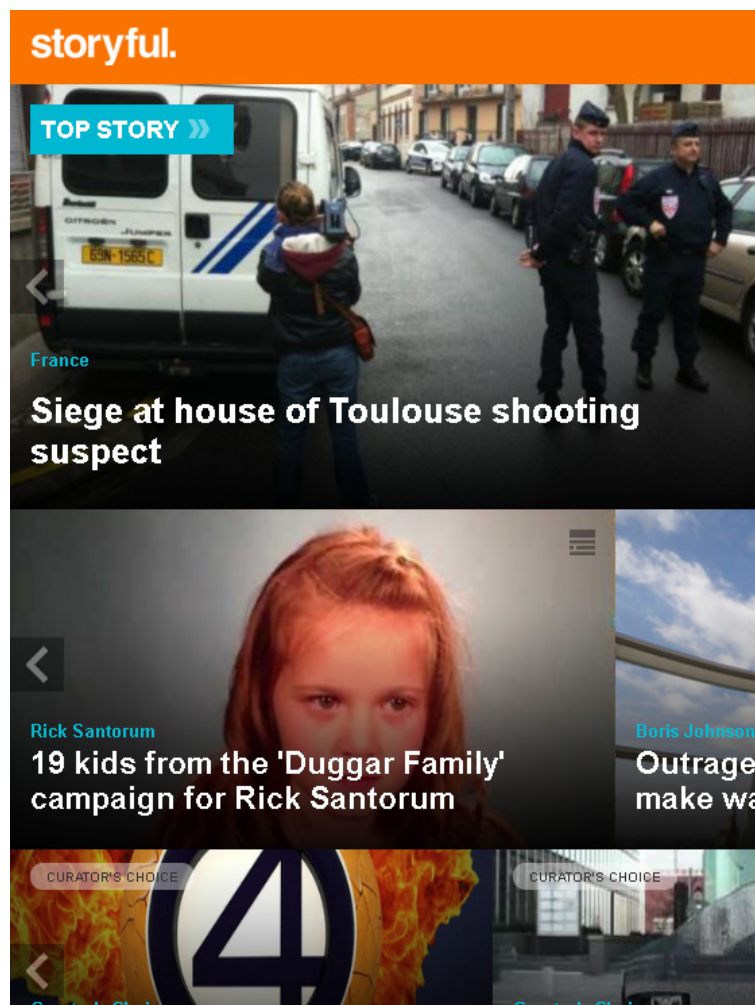


Fig. 31 - Storyful.
Schermata tratta dalla homepage di storyful.com

Vi sono poi elementi/serivzi di tipo differente, che, più che ad aggregare dati da diverse fonti, mirano a mostrare le informazioni in modo innovativo dal punto di vista visivo e comunicativo e ad arricchire l'informazione relativa ad un determinato insieme di dati.

Tra questi citiamo alcuni esempi, tra cui Newsdots, ProPublica.org e Infomous.

Newsdots

Newsdots¹⁸ è uno strumento interattivo che visualizza reti di relazione tra gli attori (o elementi chiave) di articoli giornalistici, generando una mappa tematica a partire dai tag dei singoli articoli. Lo strumento è stato sviluppato da Slate, testata giornalistica online fondata nel 1996.

Newsdots recupera e filtra gli articoli del giorno pubblicati online dalle principali testate giornalistiche. Tramite Calais, uno strumento di analisi semantica, ne estrapola gli elementi principali, raggruppati per tipologia (persone, enti, società, luoghi, argomenti).

18

< labs.slate.com/articles/slate-news-dots/ >

Questi elementi chiave vengono rappresentati all'interno di una rete e collegati tra loro a seconda degli articoli che li accomunano, mettendo così in luce il grado di separazione tra i vari elementi e le relazioni che intercorrono tra i diversi fenomeni del momento, seppur senza possibilità di approfondimento, se non quella di consultare gli articoli specifici.

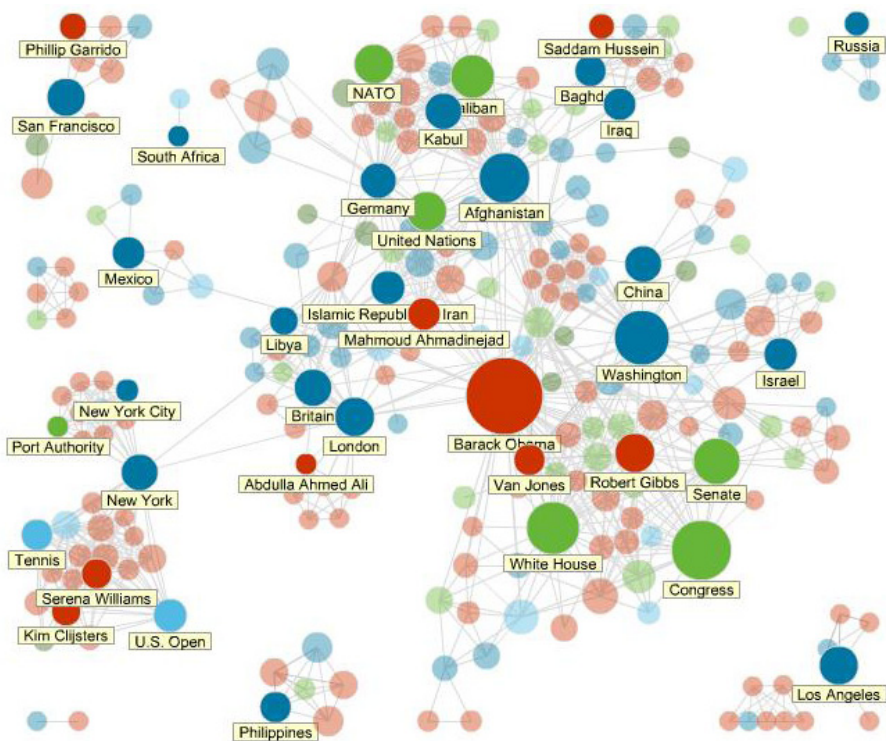


Fig. 32 - News Dots.

ProPublica

ProPublica¹⁹ è un sito statunitense di informazione che si occupa di giornalismo d'inchiesta d'interesse pubblico. Ha vinto il premio Pulitzer per i suoi lavori di giornalismo in ambito finanziario, tra cui "The Wall Street Money Machine".

ProPublica fa uso di dati e visualizzazioni che accompagnano e supportano articoli e inchieste e produce anche applicazioni basate sui dati, che permettono di analizzare fenomeni o comparare elementi sul territorio, relativamente a tematiche che possono essere di una certa utilità ai cittadini. Queste applicazioni presentano visualizzazioni interattive, mentre le visualizzazioni presentate all'interno di articoli e inchieste sono per lo più statiche.

Alcuni esempi di strumenti realizzati da ProPublica e basati su dataset sono "The Opportunity Gap" e "Dollars for Docs".

19 < propublica.org >

Il primo è basato su un database di scuole pubbliche statunitensi e permette di osservare e confrontare come i diversi stati gestiscono l'accesso paritario di scuole benestanti e povere a particolari corsi di istruzione avanzati e a programmi di educazione speciali.

How States Compare

...at providing high poverty and wealthier schools equal access to...

WORSE BETTER

AP CLASSES
GIFTED/TALENTED
ADVANCED MATH
PHYSICS
CHEMISTRY

Okla.	■	■	■	■	■	■	■	■	■
Kan.	■	■	■	■	■	■	■	■	■
Md.	■	■	■	■	■	■	■	■	■
Ala.	■	■	■	■	■	■	■	■	■
N.Y.	■	■	■	■	■	■	■	■	■
Va.	■	■	■	■	■	■	■	■	■
Ark.	■	■	■	■	■	■	■	■	■
S.C.	■	■	■	■	■	■	■	■	■
Colo.	■	■	■	■	■	■	■	■	■
Mich.	■	■	■	■	■	■	■	■	■
Ga.	■	■	■	■	■	■	■	■	■
N.M.	■	■	■	■	■	■	■	■	■
Neb.	■	■	■	■	■	■	■	■	■

Find a School

Your Address, ZIP, or school name

Submit

For example, 1605 E. 55th St. Chicago, IL or 77054 or Stuyvesant High

Connect to Facebook to quickly see your schools and your friends' schools.

[f Sign in with Facebook](#)

Latest Local Stories Based on This Project

Faribault Daily News: [Faribault School District needs to improve AP opportunities](#)

Shawnee Dispatch: [As student bodies diversify, school districts have trouble finding minority candidates](#)

California Watch: [Low-income students score lower on AP tests](#)

NBC New York: [AP Opportunity Gap: NY's Poor Students Enroll in Fewer College-Prep Courses](#)

StateImpact Ohio: [Why the Feds are Investigating the Toledo Public Schools](#)

Fig. 33 - Scheramenta tratta da "The Opportunity Gap".
< propublica.org >

Il secondo strumento utilizza un database di compagnie farmaceutiche e medici di professione e consente agli utenti di ottenere informazioni riguardo a quali medici abbiano ricevuto compensi da compagnie farmaceutiche per pubblicizzare i loro prodotti.

About this Database

\$761.3 million | **12**
in [Disclosed Payments](#) | [Disclosing Companies](#)

Our Stories



Piercing the Veil, More Drug Companies Reveal Payments to Doctors

by *Charles Ornstein, Tracy Weber and Dan Nguyen, ProPublica*

An update of ProPublica's Dollars for Docs database includes more than \$760 million in payments from 12 pharmaceutical companies to physicians and other health-care providers for consulting, speaking, research and expenses.

Also:

- [Doctors Dine on Drug Companies' Dime](#)
- **Full Coverage:** [Dollars for Docs](#)
- **Nerd Blog:** [Scraping for Journalism: A Guide for Collecting Data](#)

Payments in Your State

Click on a state name to see the disclosed payments made to healthcare practitioners in that state. Notes: Allergan is not included in the state or national totals because it reports payments in ranges. Also, Valeant did not report payment the fourth quarter of 2010.

Name	Total
Alabama	\$11,088,67
Alaska	\$263,33
Arizona	\$12,171,27
Arkansas	\$4,953,71
California	\$89,422,40
Colorado	\$13,383,82
Connecticut	\$10,835,25

Company Disclosures

The totals listed here range widely in scope at time period, and so aren't directly comparable. Some companies have added new data since the last downloaded their information.

Fig. 34 - Scheramenta tratta da "Dollars for Docs".

Un ulteriore esempio è quello di "How the Heart Rhythm Society Sells Access" in cui viene mostrato, tramite un'infografica interattiva, quali compagnie hanno finanziato vari aspetti della conferenza annuale della Heart Rhythm Society.

Spending: Sanofi-Aventis

Programs & guides	\$110,000
Educational support	\$110,000
Exhibit spaces/lounges	\$96,000
Turndown service	\$25,000
Bag inserts and cards	\$10,000
Total:	\$351,000

See table listing

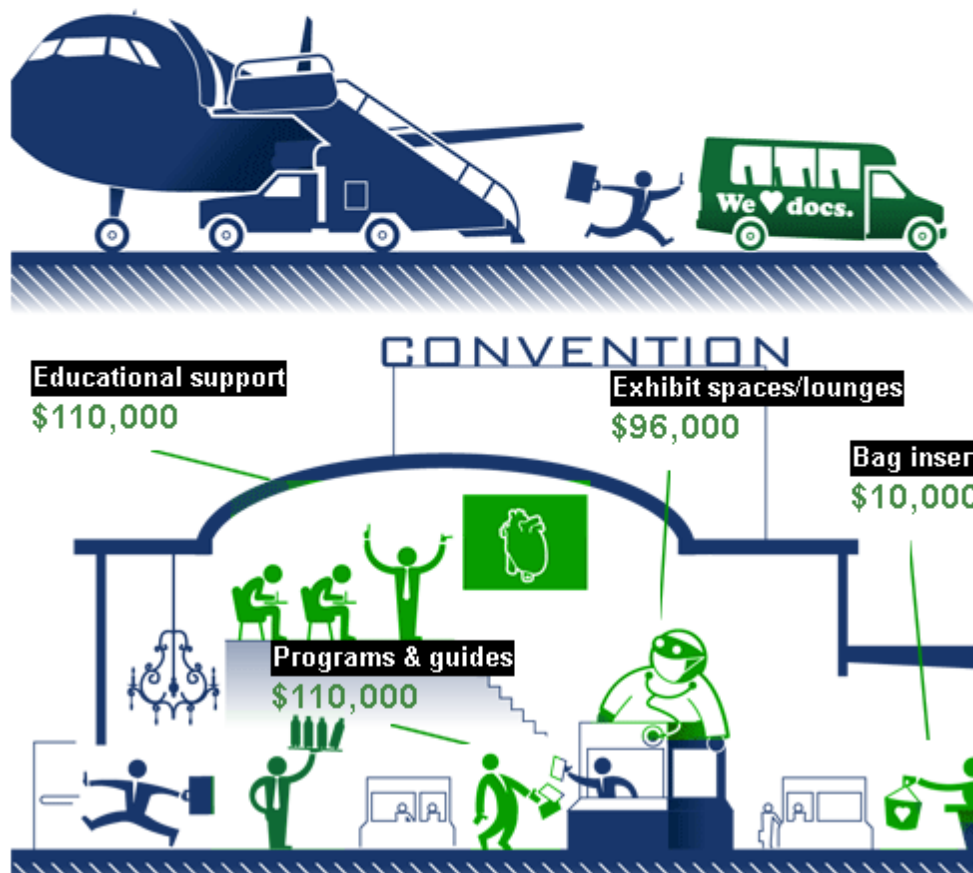
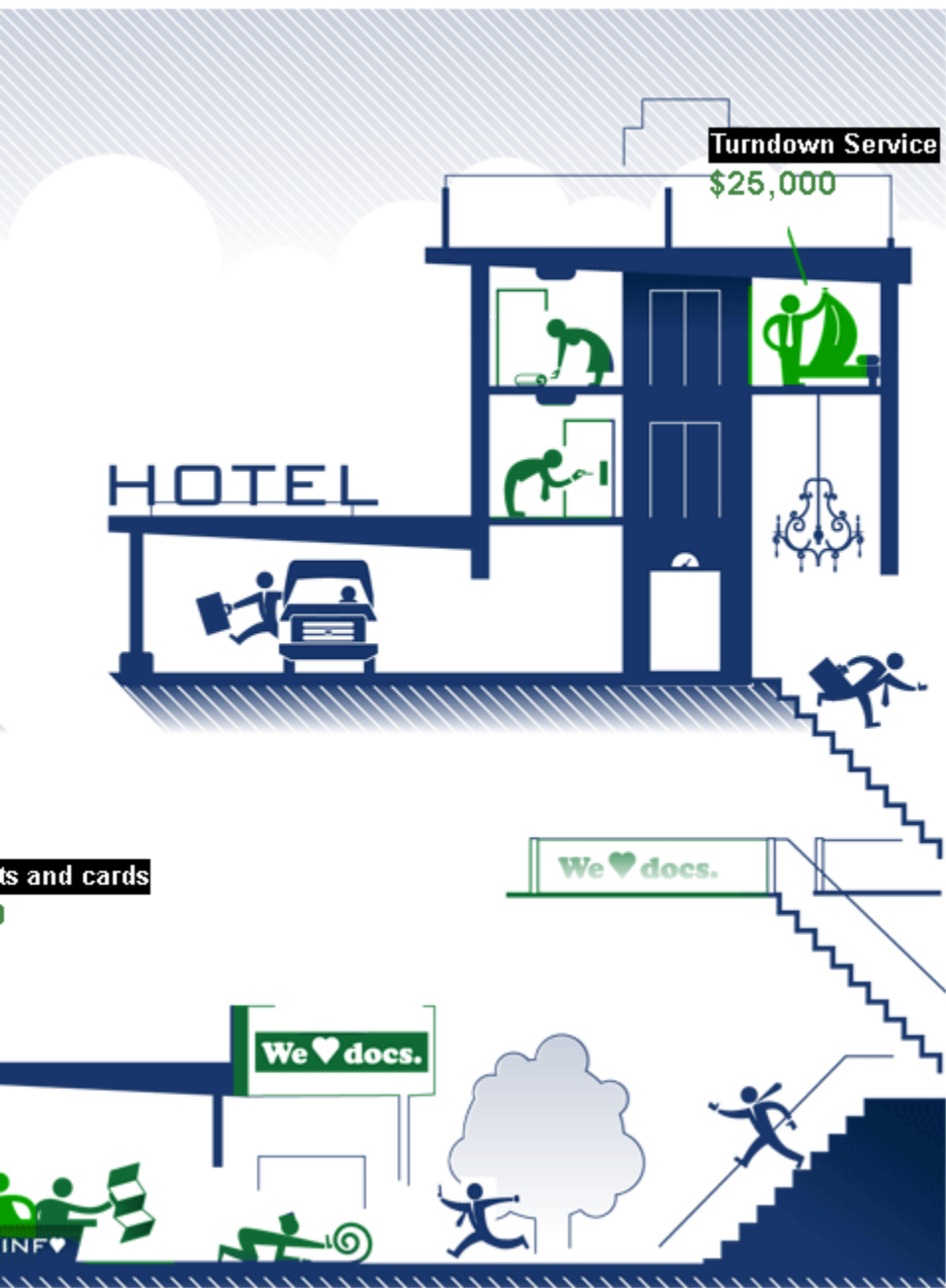


Fig. 35 - Scheramenta tratta da "How the Heart Rythm Society sells Access".



Infomous

Infomous²⁰, inizialmente sviluppato dalla Icosystem Corporation e successivamente divenuto un servizio indipendente, è uno strumento che permette di esplorare del contenuto web attraverso una visualizzazione interattiva in formato di rete di relazioni tra vocaboli. I termini che risultano maggiormente citati e che sono correlati tra loro all'interno delle fonti scelte vengono visualizzati all'interno della rete, collegati da linee e raggruppati in aree.

*“Our philosophy was to create a simple algorithm that does not try to ‘understand’ language, but simply provides a quick snapshot and allows the user to explore for further understanding.”*²¹

Selezionando una parola all'interno della rete, è possibile risalire alle fonti relative ad essa.

20 < www.infomous.com/ >

21 Paolo Guardiano, Infomous' president and CEO, 2011

Lo strumento permette inoltre di focalizzare la ricerca sugli elementi correlati ad un termine specifico e di raffinarla eliminando gli elementi non voluti.

Per gli utenti è possibile consultare le visualizzazioni già presenti o realizzarne di proprie, inserendo un certo numero di fonti scelte, tra cui feed rss, pagine web e Twitter feeds.

Questo sistema è utilizzato in vari ambiti, tra cui quello giornalistico. Citiamo a questo proposito la Comments Daily Chart di The Economist e la World News Cloud direttamente presente su Infomous.

La Comments Daily Chart sul sito di The Economist mostra gli argomenti più citati nei commenti degli utenti del sito, con collegamenti tra gli argomenti che vengono citati assieme. Questa rappresentazione raggruppa i topic tra loro correlati e offre alcune opzioni aggiuntive sulla visualizzazione, ma non permette ulteriori livelli di approfondimento ed è basata esclusivamente sul testo dei commenti e non sugli articoli pubblicati.

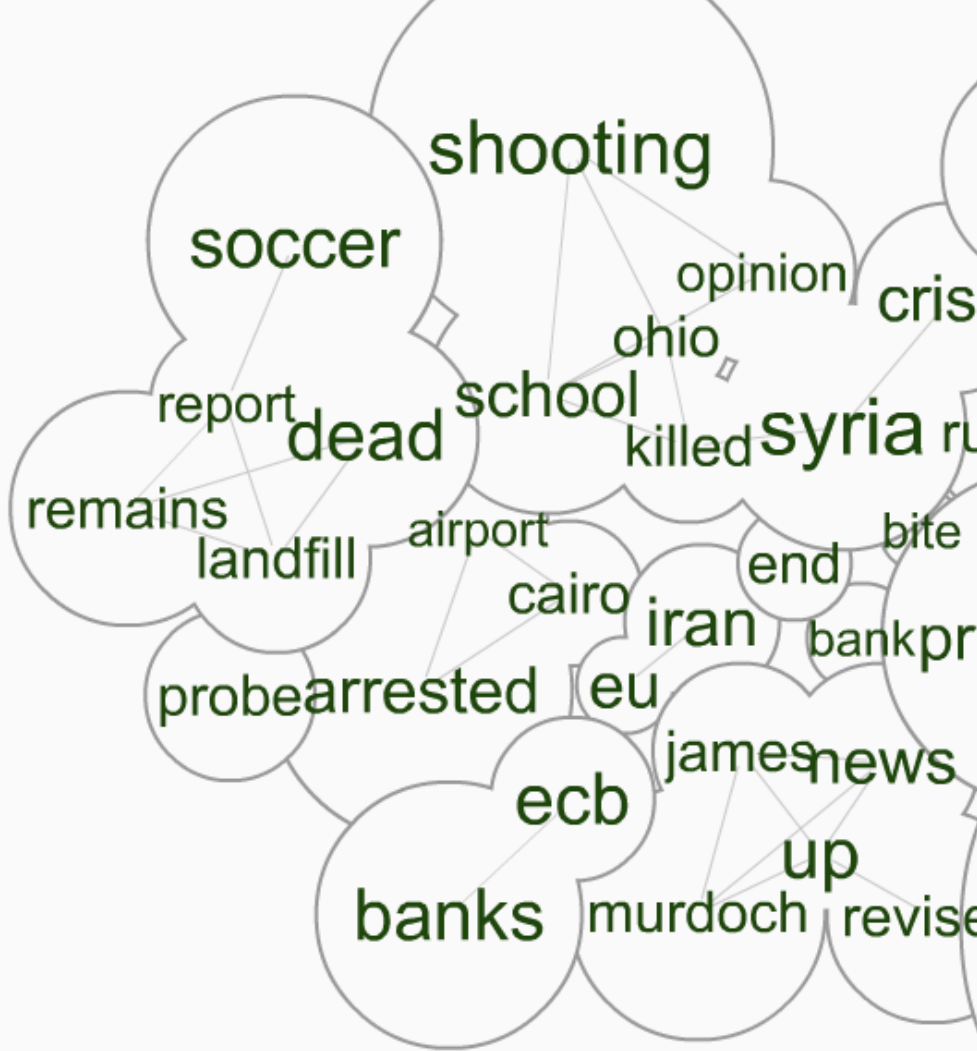


Fig. 36 - Infomous' World News Cloud

La World News Cloud, mostrata nella homepage di Infomous, utilizza come fonti i titoli degli articoli giornalistici tratti dai feed e dai siti web di diverse testate giornalistiche internazionali, fornendo così una rappresentazione in tempo reale delle tematiche più citate del momento, pur senza una vera analisi semantica.

3. viXpress: visualizzazione di dati per servizi di e-journalism

Perchè l'ambito dell'informazione?

L'ambito dell'informazione e del giornalismo costituisce un aspetto di grande rilievo per ogni individuo, oltre che per gli addetti del settore. Si tratta di un settore in costante evoluzione, che soprattutto negli ultimi anni ha subito molti cambiamenti e trasformazioni. Al giornalismo classico, infatti, si è andato ad accostare un aspetto più dinamico e mutevole, ovvero quello dell'e-journalism, il quale ha dato forti spinte a fenomeni quali il data journalism e il citizen journalism, e da questi è stato a sua volta alimentato.

Con l'avvento del Web 2.0, l'informazione di stampo giornalistico si è sviluppata ed evoluta sotto nuove forme, che hanno visto una diffusione su larga scala attraverso mezzi quali siti web, blog, social network e attraverso nuovi tipi di piattaforme di tipo collaborativo. Anche il modo di trattare i dati e il data driven journalism è cambiato. Con l'avvento del concetto di open data, con l'intensificarsi dell'attenzione verso la visualizzazione di dati e con la diffusione di strumenti volti ad agevolare la realizzazione e la condivisione di questi tipi di rappresentazioni, il data-

driven journalism è diventato oggi una componente fondamentale dell'informazione.

Alcune testate giornalistiche hanno seguito questi cambiamenti, espandendosi in diversi modi nell'ambito del web e iniziando a diffondere più apertamente e in diversi formati il materiale prima pubblicato solo in formato cartaceo.

Appare dunque interessante andare a trattare questo ambito e lavorare sugli aspetti di comunicazione e innovazione relativi alle nuove possibili forme di rappresentazione di dati e informazioni in ambito giornalistico.

Sono stati presi come punti di partenza gli aspetti descritti nei capitoli precedenti ed, in particolare, quelli riguardanti l'impiego e la visualizzazione dei dati, l'espansione delle testate giornalistiche sul web e le modalità di aggregazione e rappresentazione delle notizie.

Idea di progetto e spunti di innovazione

L'idea di progetto che è nata da queste ricerche e riflessioni consiste nella creazione di un'applicazione per il trattamento e la rappresentazione di dati provenienti da siti di testate giornalistiche. Più precisamente, l'idea è quella di un sistema che fornisca nuovi punti di osservazione e metodi di lettura sulle vicende trattate dalle testate e sull'evoluzione di esse, andando così ad arricchire l'informazione offerta dai siti di e-journalism e dai singoli articoli giornalistici.

viXpress è, sotto alcuni aspetti, uno strumento di aggregazione e confronto di notizie che mira ad andare oltre le tipiche funzionalità comunemente offerte da aggregatori e altri strumenti già tuttora diffusi.

Uno dei fattori che sono stati presi in considerazione, è il modo in cui viene proposto l'insieme delle notizie da parte di fonti giornalistiche autorevoli sul web, ancora

troppo simile alle modalità della prima pagina del formato cartaceo.

Nella maggior parte dei casi, infatti, nonostante le potenzialità offerte dal web e dalle applicazioni digitali, sui siti di testate giornalistiche le notizie vengono presentate in modo lineare: in formato testuale e accompagnate da immagini, similmente a come avviene sui quotidiani stampati.

A volte questi contenuti sono accompagnati anche da file multimediali. Gli articoli disponibili, poi, sono solitamente suddivisi in categorie e per data, e la navigazione tra essi avviene tramite semplici liste e link testuali.

Questo progetto si propone di realizzare un sistema di visualizzazioni che arricchisca l'informazione presentata da siti di e-journalism, offrendo nuove chiavi di lettura e un mezzo per esplorare le notizie in modo non convenzionale.

viXpress vuole mettere in luce aspetti quali la correlazione tra diverse notizie, l'evoluzione dell'aspetto mediatico di un fenomeno nel tempo e la relazione di particolari tematiche con il territorio geografico.

La creazione di questo sistema parte da dati tratti da siti di e-journalism. In particolare, per la realizzazione del prototipo, ci si è basati sulle informazioni provenienti dal sito web del Guardian.

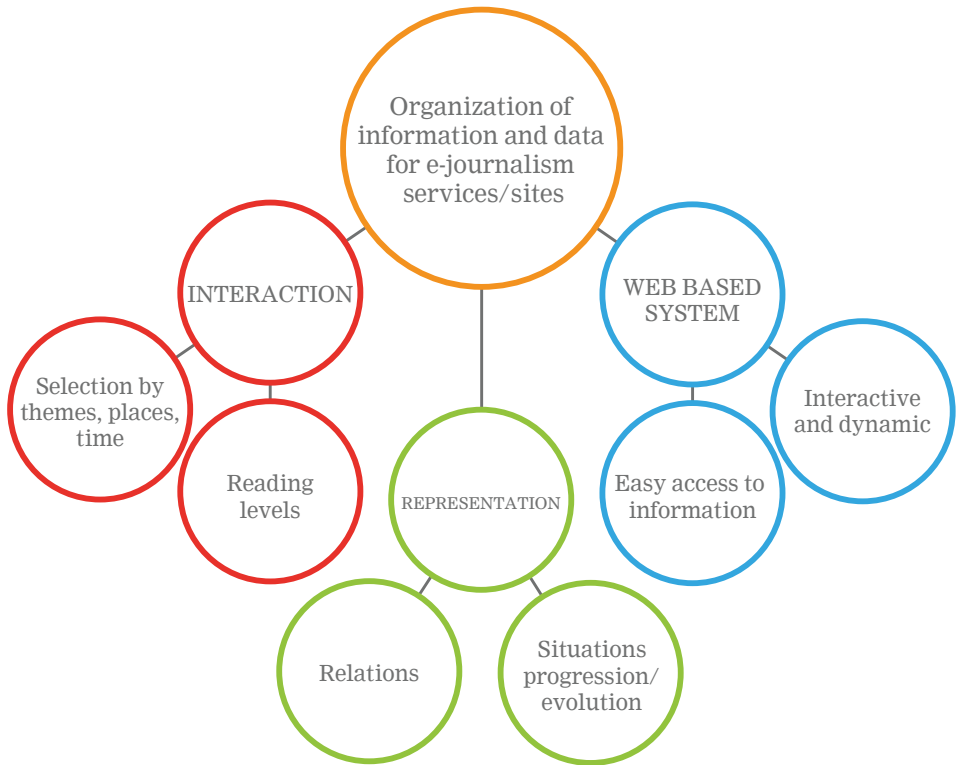


Fig. 37 - Schema elementi del concept di progetto.

Il ruolo del Guardian

Il sistema ideato è pensato per poter essere applicato a qualsiasi sito di e-journalism che pubblichi articoli corredati di una serie di informazioni di base tra cui tag, categoria e data di pubblicazione. E' stato necessario comunque scegliere una piattaforma su cui basarsi per la realizzazione del prototipo del progetto. Questa scelta è ricaduta sul Guardian, ritenuto l'opzione più adeguata per diversi motivi.

Il Guardian, fondato nel 1821, è una testata giornalistica britannica. Il suo formato cartaceo ha diffusione nazionale, ma la testata è particolarmente conosciuta per la sua presenza sul web. Il Guardian è considerato un punto di riferimento a livello mondiale per l'e-journalism e il data journalism.

Molte testate giornalistiche stanno puntando sempre di più sull'ambito digitale, ma il Guardian, in particolare, ha adottato una strategia "digital-first". Nel Giugno del 2011, infatti, attraverso un comunicato stampa, il Guardian ha affermato di voler incentrare le proprie attività sul web e sull'informazione digitale.

Toulouse gunman dead after jump from flat window

Last updated 14 minutes ago



LIVE Mohamed Merah found dead on ground after final stand-off with police, interior minister says

LATEST

"11.44am: The French socialist presidential candidate, Francois Hollande, has paid tribute to the "courage and determination" of the security forces..."
Toulouse siege – live updates

- [Toulouse shootings suspect dies](#)
- 📺 [Shootings suspect was planning more killings](#)
- [Tragedy may transform France's presidential race](#)
- 📷 [Toulouse siege – in pictures](#)
- 🗺️ [Interactive map of the Toulouse shootings](#)

The 'master of the whistle' Abraham Klein: the world's best referee



[Mad Men](#)

[More on the budget](#)

Fig. 39 - Schermata tratta dalla homepage di www.guardian.co.uk

“The Guardian has a history of innovation and openness with its digital business. This is the next step in that evolution. Our vision is to weave the Guardian into the fabric of the Internet, to become ‘of’ the web rather than ‘on’ the web.”¹

In particolare, il suo sito web presenta una sezione chiamata Datablog, incentrata sul trattamento e sulla rappresentazione di dati in ambito giornalistico. Quando è stato lanciato il Datablog, il data journalism non era ancora così affermato come parte integrante del giornalismo. Anche per questo il Guardian è considerato tra i pionieri di questo aspetto.

Scopo del Datablog era ed è quello di sviluppare storie e informazione a partire da dati. Il Guardian, da questo punto di vista, si propone di analizzare scrupolosamente i dati e di fornire gli elementi chiave di ogni storia. Nella maggior parte dei casi vengono realizzate rappresentazioni grafiche, in alcuni casi interattive.

1 < www.guardian.co.uk >

I dati utilizzati vengono al contempo resi disponibili in formato grezzo, consentendo così anche ai lettori di farne uso e sviluppare le proprie rappresentazioni a partire da essi. L'utente è anzi incentivato a riutilizzare, rielaborare e ripubblicare questi dati.

Nel 2010, inoltre, il Guardian ha dotato il proprio sito web di una sezione chiamata Open Platform, che comprende un insieme di servizi e strumenti per lo sviluppo di applicazioni e prodotti a partire dalle risorse messe a disposizione dalla testata stessa, aspetto che testimonia l'interesse del Guardian per la diffusione dei dati aperti e per l'innovazione nell'ambito della data visualization.

Per la realizzazione di viXpress è stata utilizzata la API presente nella sezione Open Platform.

Target

Il progetto è pensato per essere utilizzato sia da addetti del settore, che da un pubblico più ampio, come utenti finali. Gli appartenenti alla prima categoria potrebbero essere interessati, ad esempio, ad esaminare l'evoluzione di particolari fenomeni o delle relazioni che intercorrono tra diversi eventi, mentre un utente (occasionale/generico/non esperto) potrebbe essere interessato ad avere una visione d'insieme degli argomenti caldi del momento e a trovare in modo rapido e intuitivo gli articoli inerenti ad una particolare tematica.

E' necessario considerare, comunque, che nell'ambito dell'e-journalism, alla luce dei nuovi fenomeni che si sono sviluppati grazie al Web 2.0 e alla diffusione della definizione di citizen journalism, non è possibile e forse non è utile compiere una distinzione rigida e specifica tra giornalisti e lettori. Individui che non sono giornalisti di professione possono comunque essere considerati addetti del settore e utilizzare le proprie conoscenze per produrre e diffondere informazione.

Questo progetto vuole tenere conto di questi aspetti e fornire uno strumento che consenta ad ogni utente di avere una visione più approfondita del panorama delle notizie giornalistiche e dei fenomeni ad esse collegati.

Obiettivi preposti

Gli obiettivi che vogliono essere raggiunti attraverso questo progetto spaziano in vari ambiti, da quello comunicativo e informativo a quello tecnologico. Si è pensato innanzitutto all'importanza di arricchire l'informazione sulle notizie giornalistiche pubblicate online e a come può essere possibile fare questo a partire dai dati già disponibili.

Il principale obiettivo che questo progetto si pone riguarda dunque l'ambito della comunicazione e dell'informazione ed è quello di presentare in modo efficace e in qualche modo innovativo i dati offerti da siti di testate giornalistiche.

In particolare, l'intento è stato quello di sviluppare un sistema che possa mostrare le relazioni che intercorrono tra i diversi argomenti di attualità e come queste relazioni si evolvano nel tempo, cosa che difficilmente emerge dalla consultazione dei giornali o delle loro versioni online, ma che può avere una notevole rilevanza per l'analisi dell'andamento di particolari fenomeni o per mettere in luce correlazioni a prima vista non evidenti.

Un ulteriore scopo è quello di partecipare all'ambito "open", di contribuire a illustrare come il fenomeno dei dati aperti possa svolgere rilevanti funzioni di utilità pubblica. Questo intento è stato portato avanti lavorando sulle potenzialità dei dati messi a disposizione, aggiungendo senso e producendo nuova informazione a partire da essi.

Da un punto di vista tecnico, si è voluto inoltre lavorare sugli strumenti web-based per il trattamento di dati e per la realizzazione di visualizzazioni di dati, sfruttando ed esplorando le potenzialità di interazione offerte da alcuni linguaggi di scripting.

Tecnologie utilizzate e librerie per la visualizzazione di dati

La realizzazione del progetto ha richiesto l'utilizzo combinato di diversi strumenti e tecnologie. In generale, dal punto di vista tecnico, il risultato può essere descritto come un'applicazione web-based che presenta componenti sia lato client che lato server.

I linguaggi e i principali strumenti utilizzati comprendono HTML, servito a realizzare il layout dell'applicazione, e quindi come elemento di unione per i diversi elementi; JavaScript, impiegato per la realizzazione vera e propria delle visualizzazioni interattive, grazie a D3 e Protovis, librerie gratuite e open source per la manipolazione di documenti basati sui dati; PHP, utilizzato per creare gli script che costituiscono la parte "lato server" attiva dell'applicazione, svolgendo funzioni di recupero e trattamento dei dati, ed

infine MySQL, database management system impiegato per la gestione e la conservazione dei dati raccolti.

La libreria JavaScript D3, realizzata da Mike Bostock, merita, in questo ambito, un'attenzione ulteriore, trattandosi di uno strumento sviluppato allo scopo di facilitare la realizzazione di visualizzazioni grafiche interattive di dati sul web.

Strumenti di questo tipo, distribuiti secondo una modalità open source e forniti di documentazione ed esempi, presentano un approccio di tipo democratico. Da una parte, infatti, possono essere utilizzati anche dai non esperti del settore, nonostante alcune basi di conoscenza tecnica siano necessarie al fine di ottenere dei buoni risultati, e, d'altra parte, consentono agli individui più esperti di arricchire e migliorare lo strumento in questione.

Vi sono altri strumenti e librerie simili a D3, tra cui ricordiamo Protovis, sviluppato anch'esso da Mike Bostock, Processing.js, sviluppato da Ben Fry and Casey Reas, e Polymaps, realizzato da SimpleGeo e Stamen Design e incentrato sulla visualizzazione di dati su mappe geografiche.

La nascita e la diffusione di strumenti improntati alla realizzazione di visualizzazioni di dati è un'ulteriore testimonianza della crescente rilevanza che l'information visu-

alization (infografica) e il data journalism stanno assumendo nel mondo dei media e nell'ambito del web 2.0.

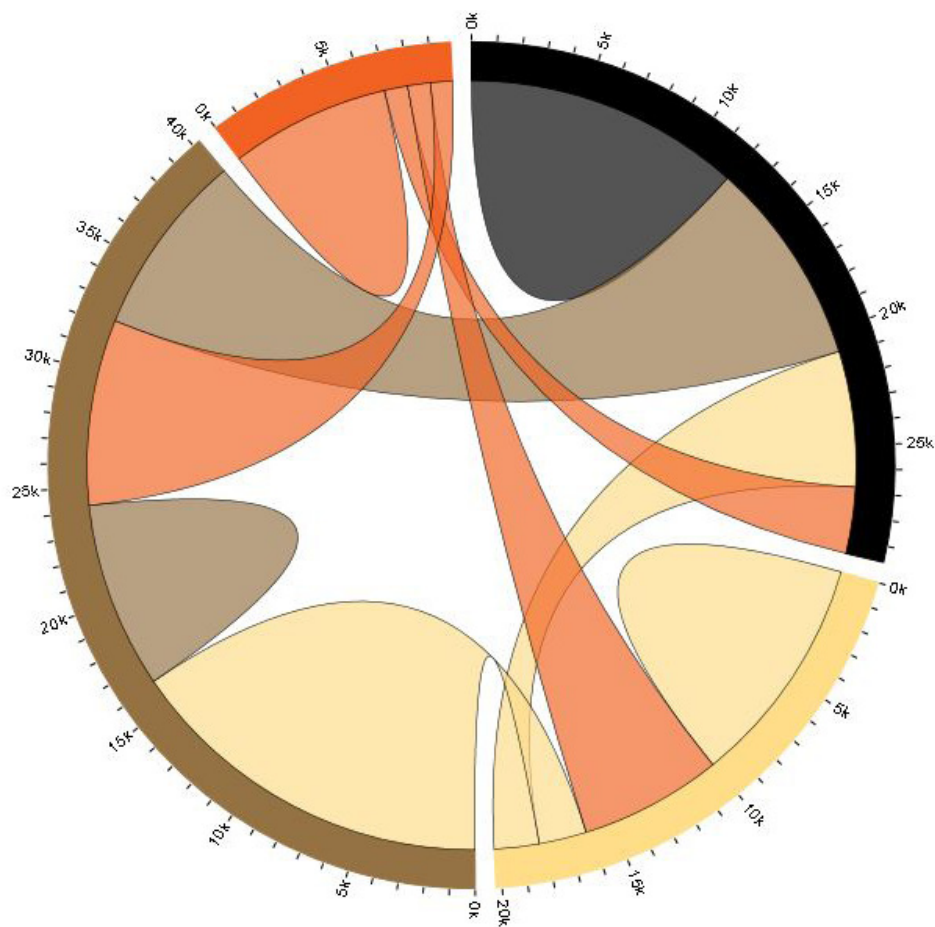


Fig. 40 - Diagramma a corde generato da d3.
[mbostock.github.com/d3/](https://github.com/mbostock/d3/)

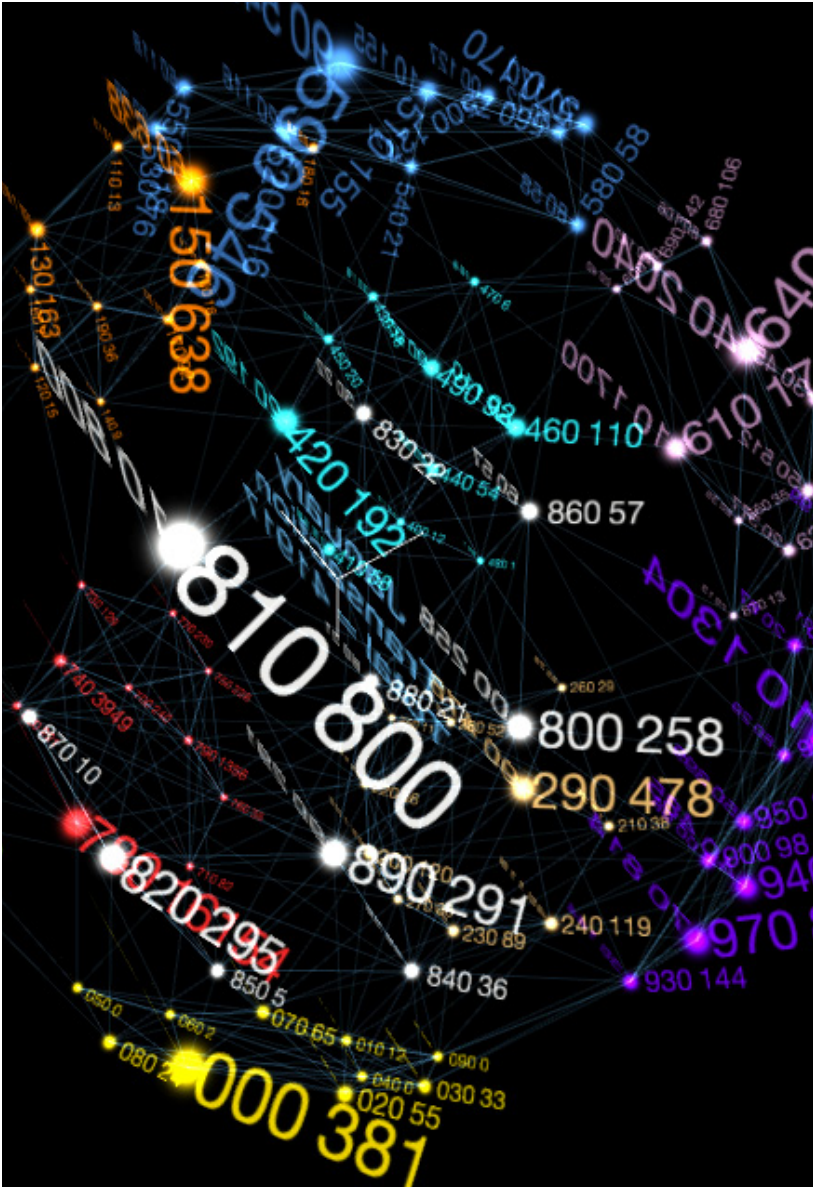


Fig. 41 - 3D Data Visualization Via Processing & OpenGL.
www.syedrezaali.com

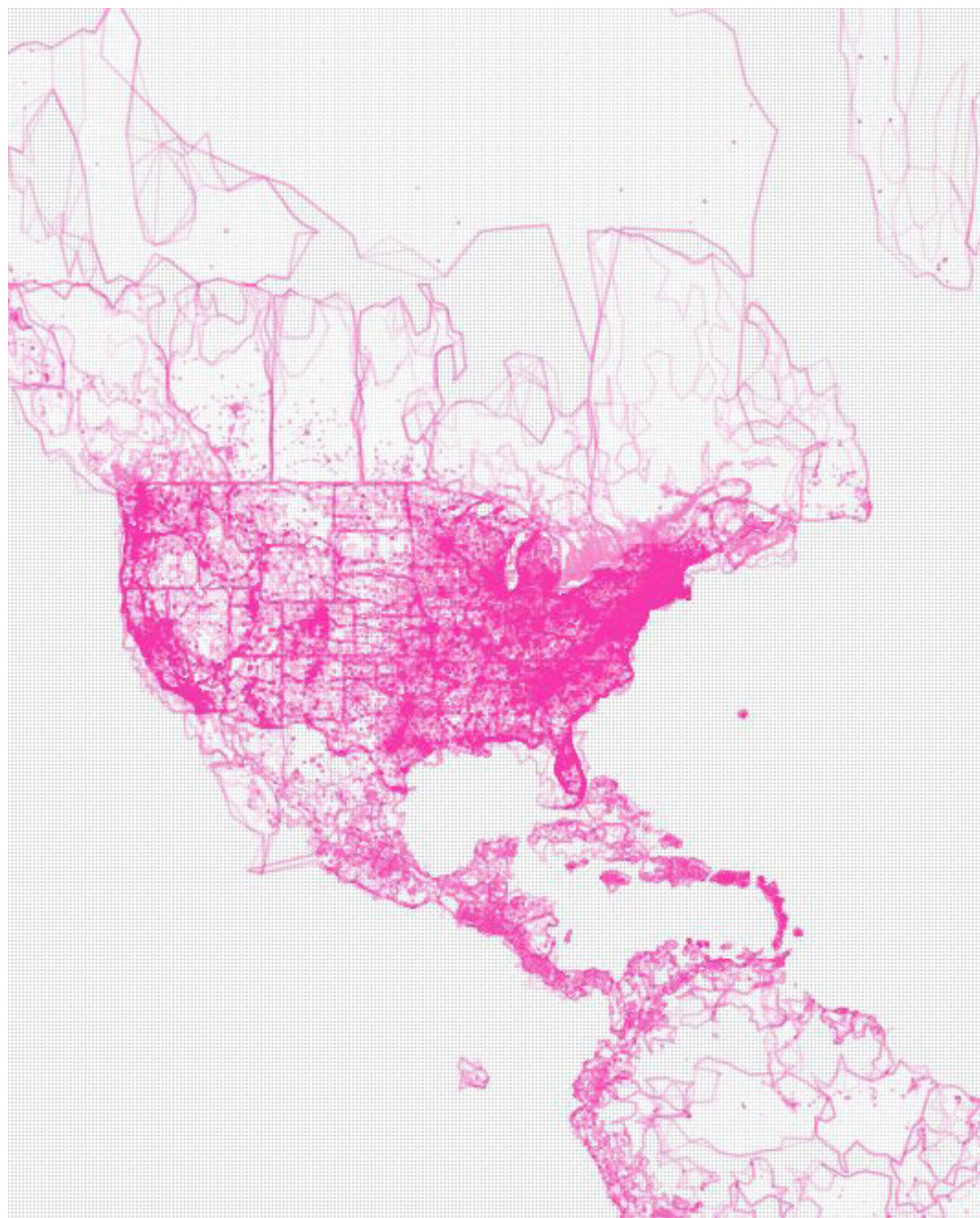
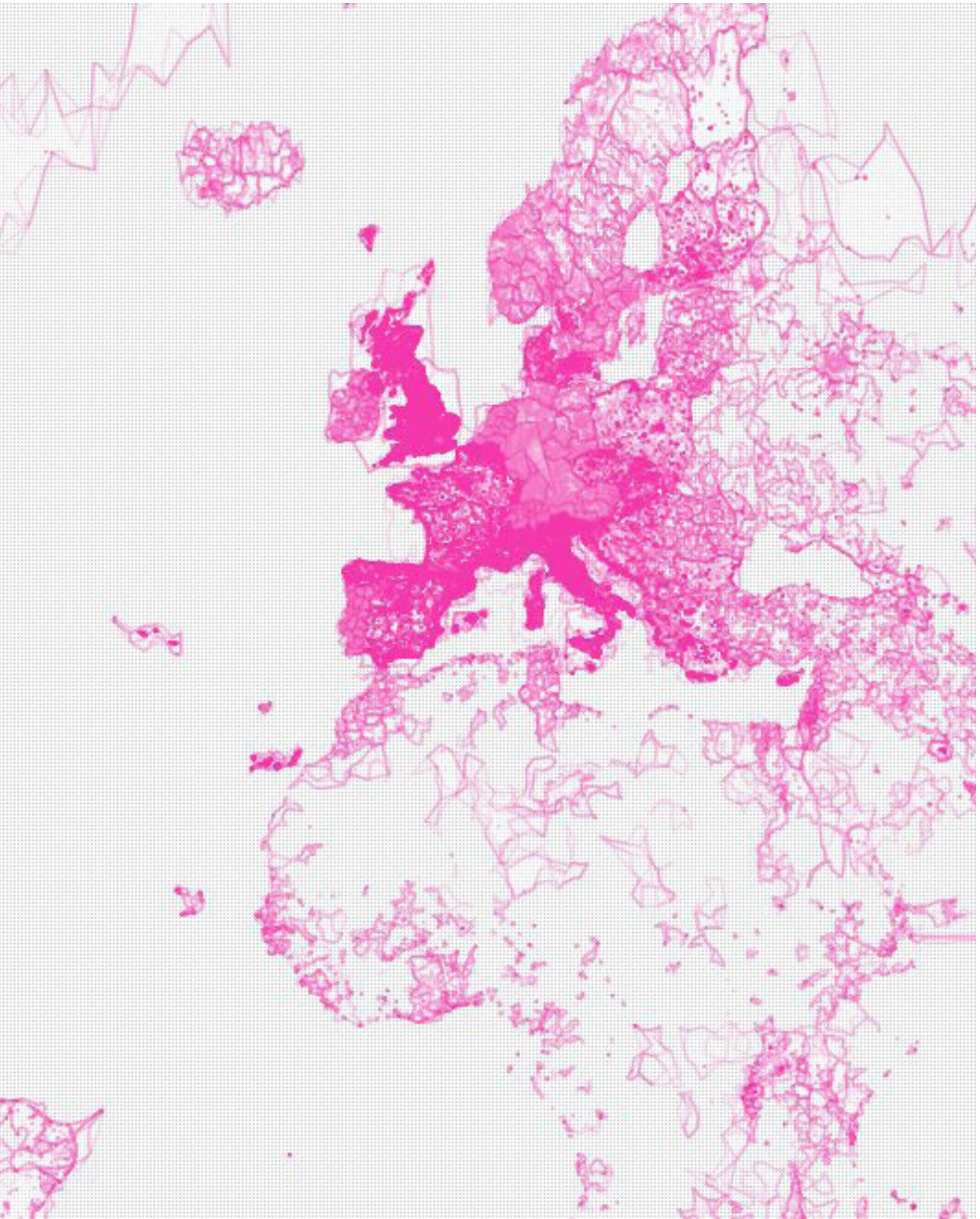


Fig. 43 - Flickr Shapes. Aaron Straup Cope.
<http://polymaps.org/ex/flickr.html>



Ricerche preliminari alla fase realizzativa

La prima fase effettiva del progetto è consistita in una serie di ricerche sui dati disponibili e sulla loro struttura, in modo da approfondire il contesto di riferimento e definire gli aspetti tecnici e le fasi successive a partire da questo e dagli scopi preposti. In particolare, questa fase di ricerca si è concentrata sul materiale messo a disposizione dal Guardian tramite il suo sito web.

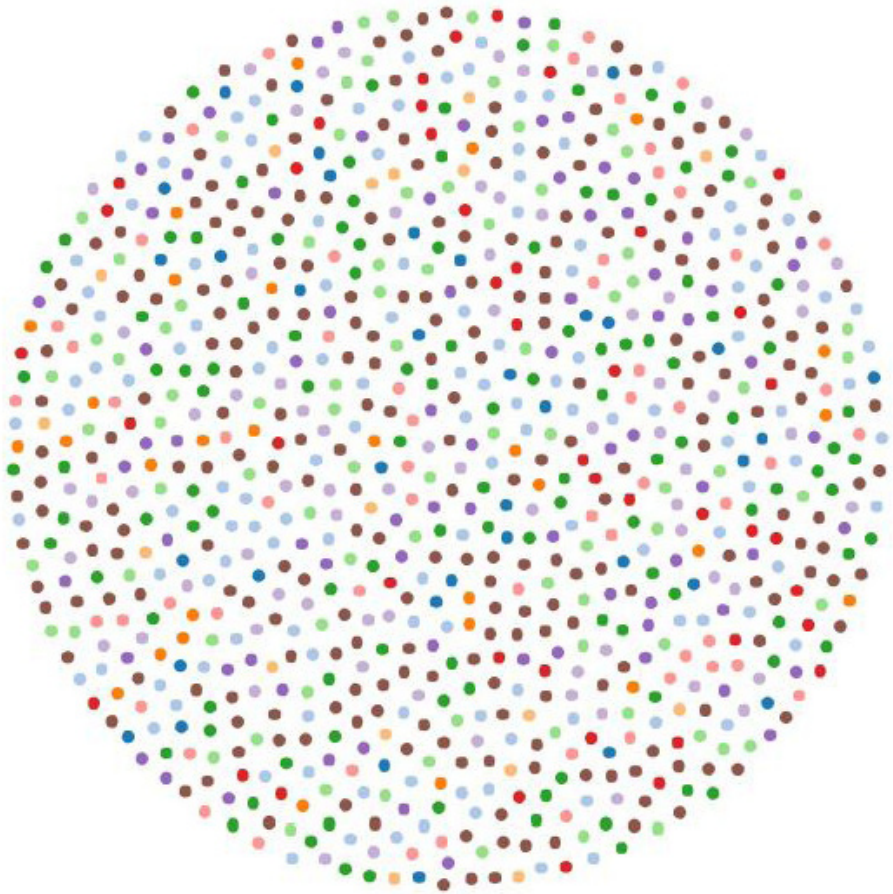
Innanzitutto si possono distinguere diverse tipologie di dati e materiale. Vi sono dataset veri e propri, messi a disposizione tramite la sezione Data Store del sito web e vi è poi l'insieme degli articoli che il Guardian pubblica quotidianamente e a cui permette di accedere direttamente tramite il sito web, oppure attraverso una API. I dataset riguardano diverse tematiche tra cui, in particolare, politica, economia, ambiente e sanità. Sono presentati in formato di tabella, accessibili e scaricabili tramite Google Documents, e sono, in diversi casi, accompagnati da una descrizione e da visualizzazioni e infografiche, per lo più statiche o parzialmente interattive. Gli articoli giornalistici veri e propri.

invece, sono suddivisi in categorie, le quali comprendono UK, World, Development, US, Politics, Media, Education, Society, Science, Tech, Law, Football.

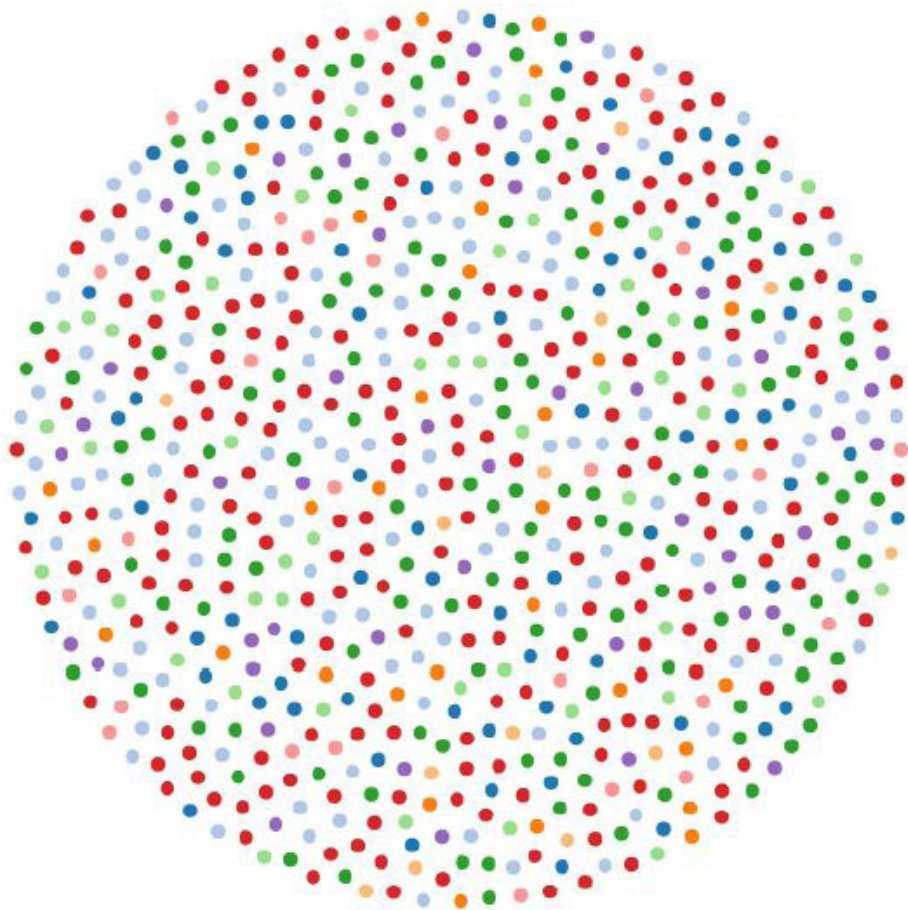
Il sito web permette agli utenti di consultare gli articoli per giorno e per categoria. Nella pagina principale di ogni categoria, gli articoli più recenti sono presentati tramite box testuali corredati in molti casi da un'immagine o da una fotografia. E' inoltre possibile consultare gli articoli tramite un elenco di tematiche specifiche, presentato in forma testuale e in ordine alfabetico.

Uno dei primi passaggi ha comportato il conteggio del numero di articoli pubblicati giornalmente e settimanalmente dal Guardian per ogni categoria principale, in modo da avere una stima della mole di dati che l'applicazione avrebbe poi dovuto gestire. Questo conteggio è stato fatto sul periodo che va dal 13 al 20 Giugno 2011 ed è stato eseguito in parte manualmente ed in parte tramite la API messa a disposizione da guardian.co.uk.

I dati così ottenuti sono stati utilizzati per realizzare delle rapide e schematiche visualizzazioni che mostrassero la ripartizione degli articoli presi in considerazione.



**Fig. 44 - Articoli della sezione News di guardian.co.uk
Periodo 13 - 20 Giugno 2011**



**Fig. 45 - Articoli al di fuori della sezione News di guardian.co.uk.
Periodo 13 - 20 Giugno 2011**

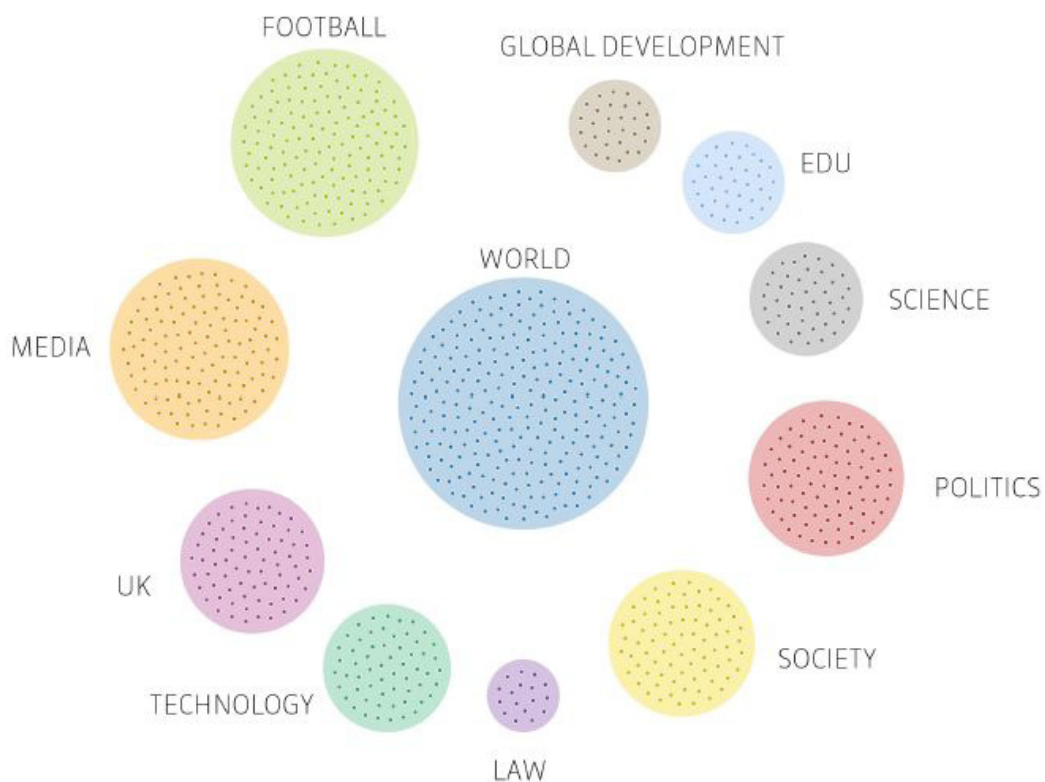


Fig. 46 - Articoli per categoria della sezione News di guardian.co.uk. Periodo 13 - 20 Giugno 2011

Successivamente è stata analizzata la struttura e la varietà dei dati ottenibili tramite la API. I criteri di ricerca consentiti comprendono categoria di appartenenza, data e tag o parole chiave. La API permette di ottenere un file xml o json in cui, per ogni articolo, sono presentate informazioni che includono

- id dell'articolo;
- url a cui reperire l'articolo;
- sezione di appartenenza;
- data di pubblicazione;
- titolo;
- sommario;
- tag associate all'articolo.

Il concept prevedeva come elementi principali l'aspetto temporale, quello geografico e le relazioni tra i diversi elementi. Osservando la struttura dei dati ottenuti tramite la API si è pensato di utilizzare le informazioni relative ai titoli e alle tag per sviluppare l'aspetto delle relazioni e quello geografico.

In base a questo, un ulteriore esperimento è stato quello di estrarre una lista di tutte le tag associate agli articoli del periodo .. - .. Giugno. Da questa lista sono stati isolati i termini relativi alle località geografiche, in particolare gli stati mondiali, e la ricorrenza di ognuno di questi all'interno

dell'elenco è stata conteggiata. Dal dataset ottenuto sono state realizzate alcune semplici visualizzazioni tramite ManyEyes che mostrano con che ricorrenza gli articoli pubblicati nel periodo preso in esame hanno trattato in modo rilevante dei diversi stati del mondo.

La raccolta dei dati

Dopo le prime fasi di ricerca e sperimentazione sui dati, è stata avviata una vera e propria raccolta di dati tramite la API del Guardian. A partire da questo momento, la gestione dei dati è stata in buona parte automatizzata. Sono stati realizzati diversi script in PHP che accedessero alla API per recuperare i file json necessari e li elaborassero in modo da estrarre le informazioni e salvarle su un database. Sono stati impostati diversi cron job in modo da richiamare questi script quotidianamente.

Per quanto riguarda i dati conservati su database, questi includono, in primo luogo, una serie di informazioni di base riguardo agli articoli (data di pubblicazione, categoria, autore, titolo, indirizzo url che rimanda all'articolo stesso...) e le tag che sono state associate ad ogni articolo. I titoli sono poi stati suddivisi in singole parole, in modo da poter eseguire un'analisi delle parole chiave anche a partire dal titolo, oltre che in base ai tag e al corpo dell'articolo.

Seguendo questi procedimenti, nel corso di diverse settimane sono stati recuperati i dati relativi agli articoli pubblicati a partire dal 1 Gennaio 2011 fino al termine dell'anno e oltre.

I dati raccolti comprendono circa 80.000 articoli e 600.000 parole chiave.

L'analisi dei dati e la scelta delle visualizzazioni

A seguito delle ricerche preliminari e in concomitanza con l'effettiva raccolta dei dati, sono stati definiti i principali aspetti su cui incentrare il lavoro ed il modo in cui sarebbero stati trattati. Questi elementi primari, interconnessi tra loro, comprendono relazioni, luoghi e tempo. Si tratta di fattori che hanno notevole spazio ed importanza nel mondo dell'informazione. L'analisi e la combinazione di essi può portare ad una più profonda e ricca comprensione degli eventi e dei fenomeni che scaturiscono da questi.

A livello di concept sono stati delineati alcuni modelli di rappresentazione che potessero ricoprire questi aspetti. L'aspetto e la struttura finali dello strumento sviluppato si sono differenziati in molte parti da queste immagini di concept, ma ne hanno mantenuto alcuni principi di base, tra cui l'idea di una rete di relazione per rappresentare le correlazioni tra gli elementi.

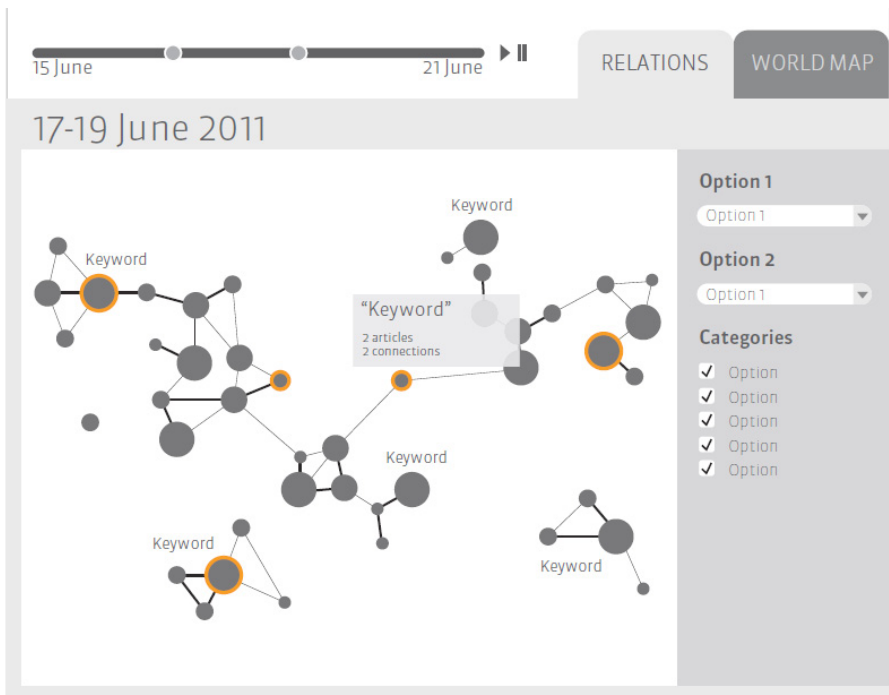


Fig. 47 - Schermata concept di progetto. Rete di relazioni.

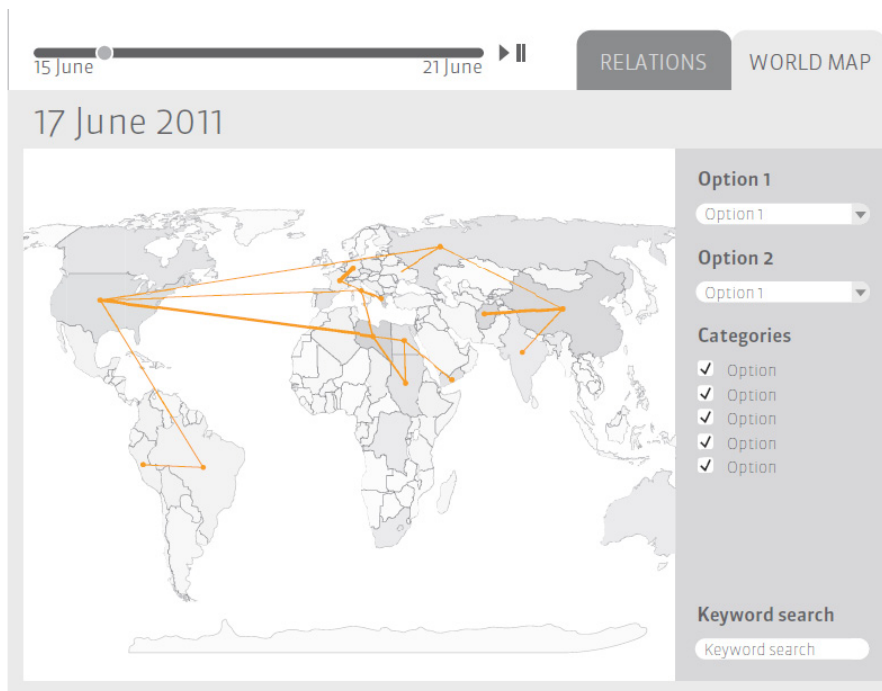


Fig. 48 - Schermata concept di progetto. Mappa geografica.

Aspetto delle relazioni

Si è pensato a quello delle relazioni come primo aspetto su cui lavorare. Si tratta di un aspetto piuttosto complesso e che presenta grandi potenzialità, soprattutto in un ambito come quello giornalistico. Le relazioni tra diversi elementi o fenomeni non sono facilmente individuabili da una normale lettura degli articoli, o da una consultazione per categoria, ma possono costituire una parte importante di informazione.. Dall'osservazione delle relazioni, infatti, è possibile individuare fenomeni non evidenti, o relazioni di secondo livello, ovvero legami tra elementi non direttamente collegati tra loro ma accomunati da un terzo elemento.

Il lavoro, da questo punto di vista, si è incentrato sul mettere in relazione articoli e parole chiave e sulla rappresentazione di questi legami.

Si è deciso di separare le relazioni tra i termini estrapolati dai tag da quelli tratti dai titoli degli articoli, in modo da non correlare tra loro elementi con pesi differenti, cosa che avrebbe potuto portare a rappresentazioni ingannevoli o comunque non corrette.

Per la rappresentazione delle reti di relazione è stato utilizzato un force-directed graph realizzato tramite la libreria javascript D3.

Ogni parola chiave (estratta dai tag o dai titoli) è rappresentata da un cerchio e costituisce un nodo della rete. La dimensione di questi elementi rappresenta il numero di volte in cui quel termine compare. Come e numero di citazioni dell'elemento sono indicate testualmente accanto al cerchio. Gli argomenti che sono trattati insieme all'interno di almeno un articolo sono collegati da una linea, il cui spessore indica il numero di articoli che li accomunano. I nodi vanno a disporsi in modo da mantenere una certa vicinanza tra gli elementi collegati tra loro.

Fig. 49 - Rete di relazioni del giorno 1/11/2011

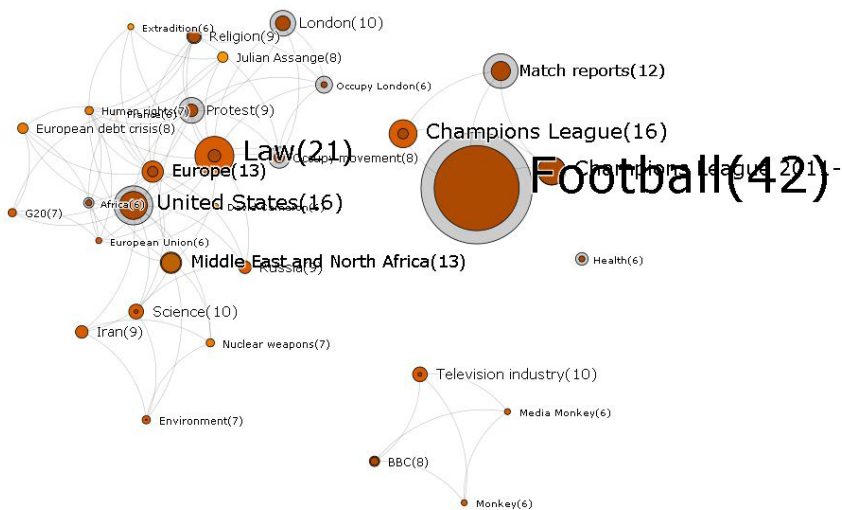
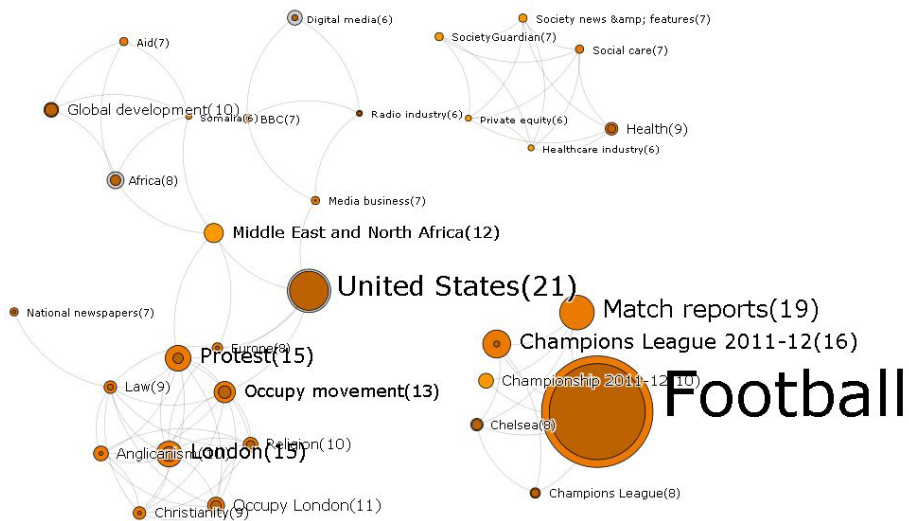


Fig. 50 - Rete di relazioni del giorno 2/11/2011

Fig. 51 - Rete di relazioni del giorno 3/11/2011

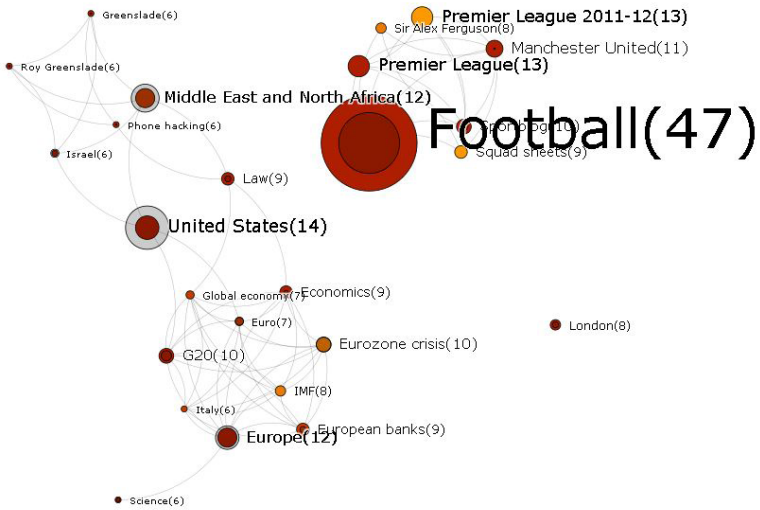
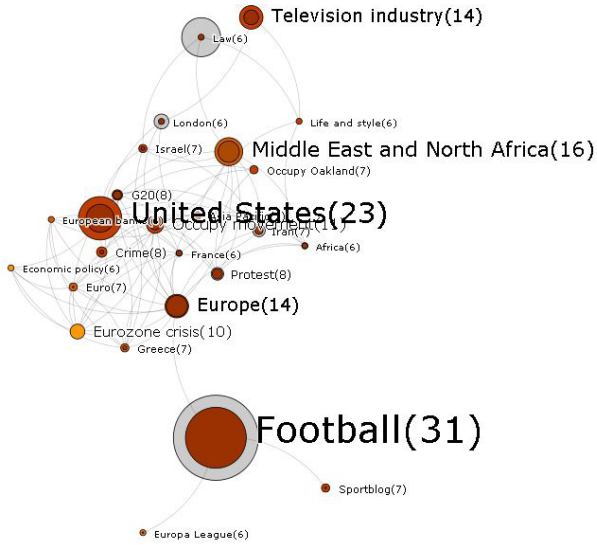


Fig. 52 - Rete di relazioni del giorno 4/11/2011

Fig. 53 - Rete di relazioni del giorno 5/11/2011



Fig. 54 - Rete di relazioni del giorno 6/11/2011

Aspetto geografico

I termini relativi alle località geografiche (stati e continenti) sono stati estratti, separatamente, dai tag e dai titoli dei singoli articoli.

Per visualizzare questi dati si è optato, in primo luogo, per una choropleth map, ovvero una mappa geografica in cui le diverse aree sono colorate in base al valore numerico di un certo dato. In questo caso, il dato è il numero di volte che uno stato compare tra i tag o nel titolo degli articoli di un giorno o di un dato periodo. Ad una maggiore saturazione del colore corrisponde un maggior numero di citazioni.

Si è poi fatto in modo che venissero richiamati dal database i termini che compaiono insieme al nome di ogni stato e che fossero inseriti all'interno del dataset utilizzato per la visualizzazione. Spostando il cursore su uno stato compariranno i termini relativi ad esso. E' dunque possibile così osservare, oltre che quali luoghi hanno interessato le vicende del momento, anche quali argomenti, persone o eventi sono associati ad un particolare luogo.

Per la realizzazione di queste mappa è stato utilizzato Protovis, una libreria Javascript per la visualizzazione di dati, simile a D3.

L'aspetto geografico è stato trattato non solo tramite una mappa, ma anche attraverso un force-directed graph, in modo da mettere in evidenza i collegamenti tra le vicende relative a diversi stati o continenti. Il funzionamento di questa visualizzazione è molto simile a quello descritto per le reti di relazione tra parole chiave. In questo caso, però, la colorazione dei nodi è stata utilizzata per distinguere gli stati appartenenti a continenti diversi.

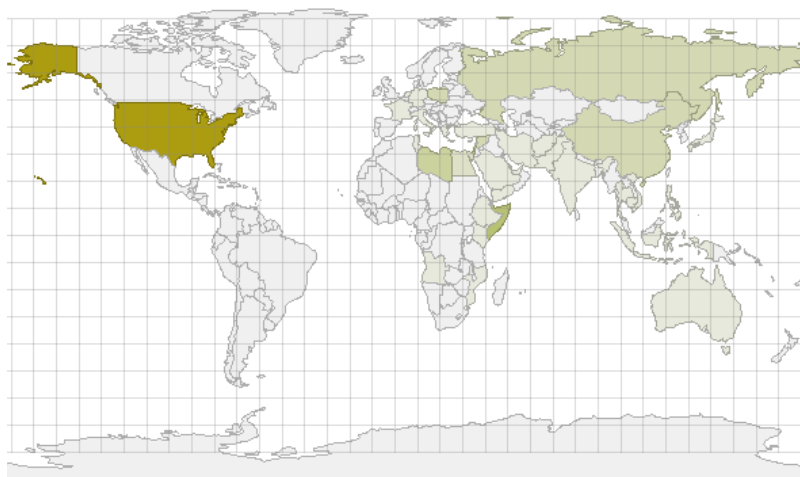


Fig. 55 - Mappa del giorno 1/11/2011

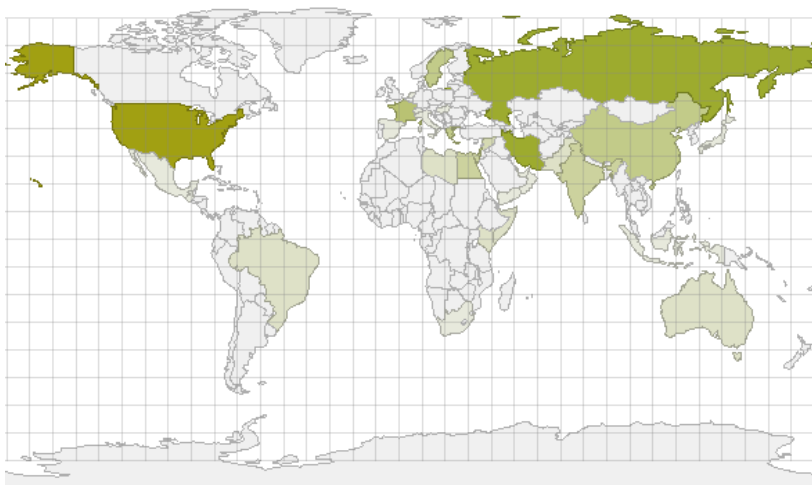


Fig. 56 - Mappa del giorno 2/11/2011

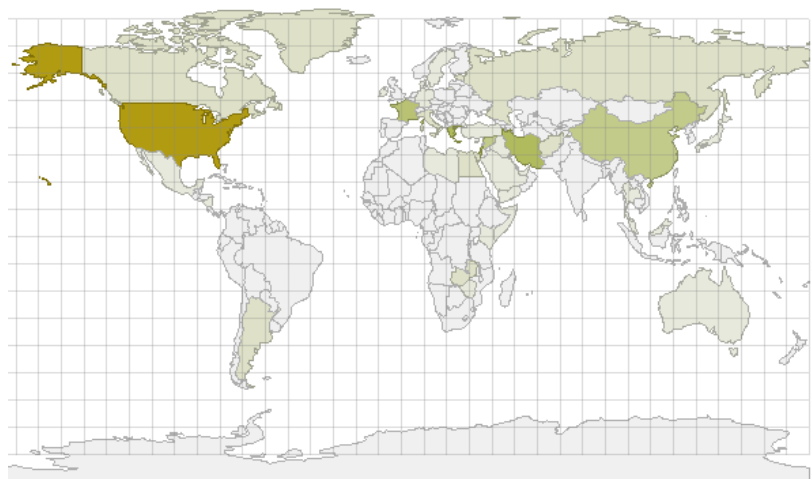


Fig. 57 - Mappa del giorno 3/11/2011

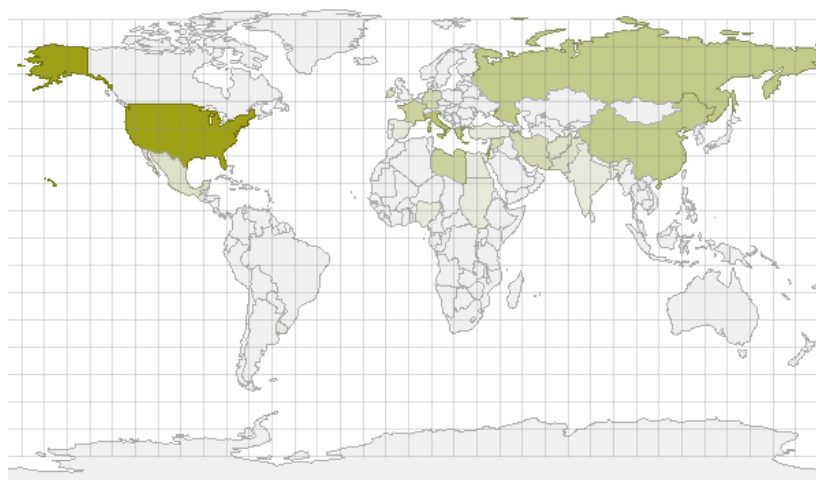


Fig. 58 - Mappa del giorno 4/11/2011

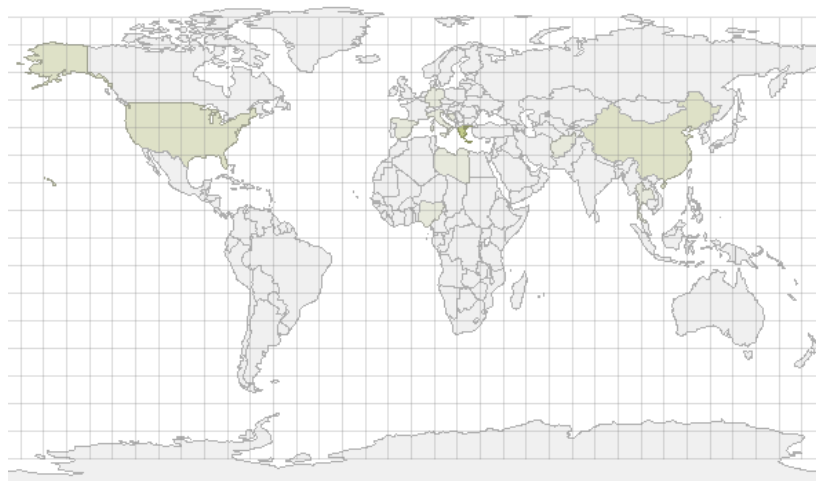


Fig. 59 - Mappa del giorno 5/11/2011

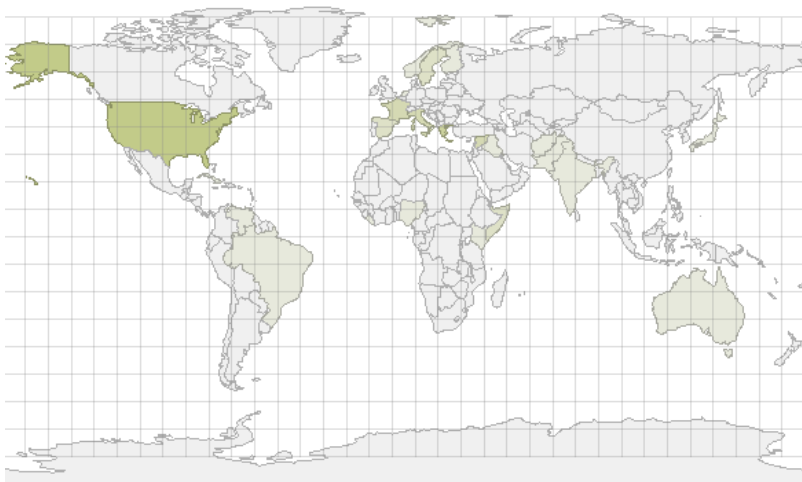


Fig. 60 - Mappa del giorno 6/11/2011

Aspetto temporale

Un altro fattore importante è quello temporale. Si è pensato che sarebbe stato importante lavorare su aspetti quali la persistenza e l'andamento nel tempo di fenomeni e tematiche.

La rete delle relazioni e gli altri dati rappresentati si basano su un singolo giorno o su un periodo di più giorni consecutivi. E' possibile per l'utente selezionare manualmente le date da visualizzare.

E' poi stata realizzata una timeline che mostra il numero di citazioni tra i tag degli articoli per ogni argomento selezionato, in ogni giorno dell'anno. Ogni giorno è rappresentato da una tacca e la quantità di citazioni dell'argomento è indicata dall'intensità cromatica di ogni tacca. Questo consente di osservare intensità e persistenza di un argomento del corso del tempo e di capire se ci sono stati periodi particolarmente "caldi" in relazione ad una determinata tematica. A partire da questa timeline, è poi possibile selezionare un giorno e visualizzare quali sono stati gli articoli che hanno trattato quell'argomento, proseguendo così la navigazione a partire dalle informazioni di quella data.

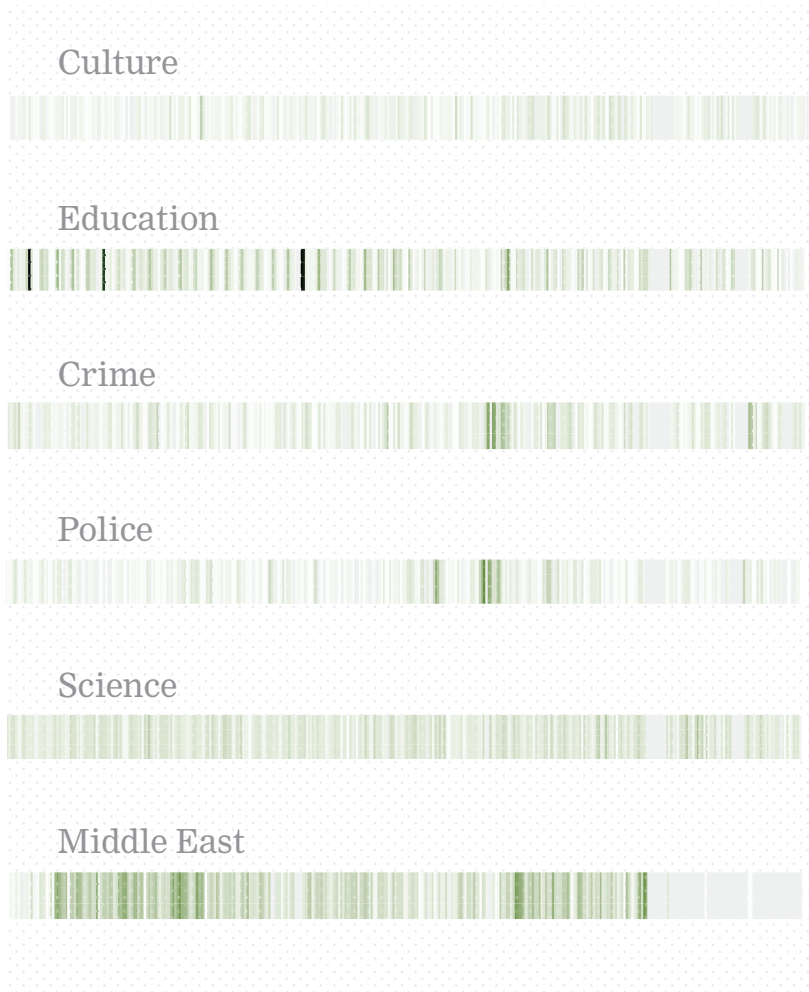


Fig. 61 - Esempio di timeline annuali su diversi argomenti.

Si è voluto, però, trovare un modo per rendere l'aspetto temporale anche all'interno della stessa rete delle relazioni, in modo da dare maggior peso a questo fattore che risulta essere di grande rilevanza nell'ambito giornalistico.

Si è dunque pensato, a questo proposito, di lavorare sull'aspetto cromatico e grafico dei nodi della rete. Per ogni elemento che andrebbe a costituire un nodo, lo script che genera il file json utilizzato dalla visualizzazione verifica la persistenza di quello stesso elemento nei sette giorni precedenti al momento che si è scelto di rappresentare con la rete.

Vengono così contati i giorni di persistenza di ogni argomento e, a seconda di questo valore, viene attribuita una colorazione ai cerchi che identificano i nodi, sulla base di una scala cromatica. Gli argomenti che non compaiono nei giorni immediatamente precedenti al momento considerato e che sono, quindi, argomenti "occasional" o di cui si è appena iniziato a trattare, sono rappresentati con un colore chiaro. Tonalità più scure identificano invece gli argomenti che persistono da un maggior numero di giorni, ovvero tematiche consolidate o "calde" del periodo.

Attorno al centro del nodo compare poi un ulteriore cerchio, semitrasparente, che mostra la portata di quell'argomento nel giorno precedente al momento consid-

erato, in modo che sia possibile fare un rapido confronto diretto e avere una prima idea di quale sia l'andamento, ovvero se si tratti di una tematica in espansione o in regressione.

Questo consente di poter osservare e conoscere con più chiarezza l'andamento di un fenomeno nel tempo anche visualizzando le reti di un singolo giorno.

L'analisi semantica

Un particolare strumento che è stato utilizzato per l'analisi degli articoli giornalistici e per l'estrazione di dati da essi è Alchemy API.

Questo strumento si basa sull'analisi semantica di contenuti testuali e presenta diverse funzionalità con svariati possibili usi. Tra queste vi è l'estrazione del testo da pagine web, l'identificazione del linguaggio, l'identificazione di relazioni soggetto-azione-oggetto e l'estrazione di parole chiave. In particolare, AlchemyAPI analizza il contenuto testuale ed estrae da esso informazioni semantiche, individuando persone, luoghi geografici, compagnie, organizzazioni e argomenti.

All'interno del progetto, AlchemyAPI è stato utilizzato per l'analisi del testo dei singoli articoli, in modo da ottenere gli elementi con cui costruire un ulteriore livello di lettura che potesse arricchire l'informazione presentata dalla rete delle relazioni tra parole chiave.

Ogni articolo estratto dal sito web del Guardian viene analizzato automaticamente da AlchemyAPI e sul database vengono salvate le informazioni relative all'articolo stesso, alle named entities individuate e agli indici che identificano la ricorrenza e predominanza di ognuna di queste named entity. Questi dati vengono poi richiamati e utilizzati dall'applicazione.

Così, selezionando una parola chiave all'interno della rete, verrà mostrato l'insieme degli articoli che ne trattano, organizzati in una treemap che li gerarchizza in base ad un indice che tiene conto del numero di citazioni e della alla posizione della parola chiave al loro interno.

Andando a selezionare uno degli articoli verranno invece mostrate tutte le named entities estratte tramite AlchemyAPI dal corpo dell'articolo, anch'esse gerarchizzate in base alla loro rilevanza all'interno del testo.

Questo sistema permette così di trovare articoli correlati ed esaminare i collegamenti tra i diversi articoli. Tra

articoli che trattano di argomenti simili, permette inoltre di osservare in quali predomina un particolare tema/termine piuttosto che un altro e in che misura.

I livelli di lettura

La combinazione di questi aspetti tramite le rappresentazioni precedentemente descritte dovrebbe consentire all'osservatore di avere, a prima vista, un'idea del panorama del giorno (o del periodo selezionato) per quanto riguarda avvenimenti e tematiche trattate dalla testata.

Per quanto riguarda questa prima visione generale, si è pensato alla regola delle 5 W del giornalismo. Elementi che rispondono a “who”, “what” e, in alcuni casi, “why” sono presentati nei nodi della rete delle relazioni, “when” dall'indicazione temporale e “where” dalla mappa geografica.

Il progetto, però, è stato strutturato in modo da consentire diversi livelli di lettura, con progressivo grado di approfondimento. Oltre al primo livello principale sono presenti essenzialmente altri due livelli di approfondimento. Il primo consente, a partire dalla rete delle relazioni, di selezionare un nodo e osservare in quali articoli è stato trattato quell'argomento e quanto quella stessa tematica è stata trattata nel corso del tempo.

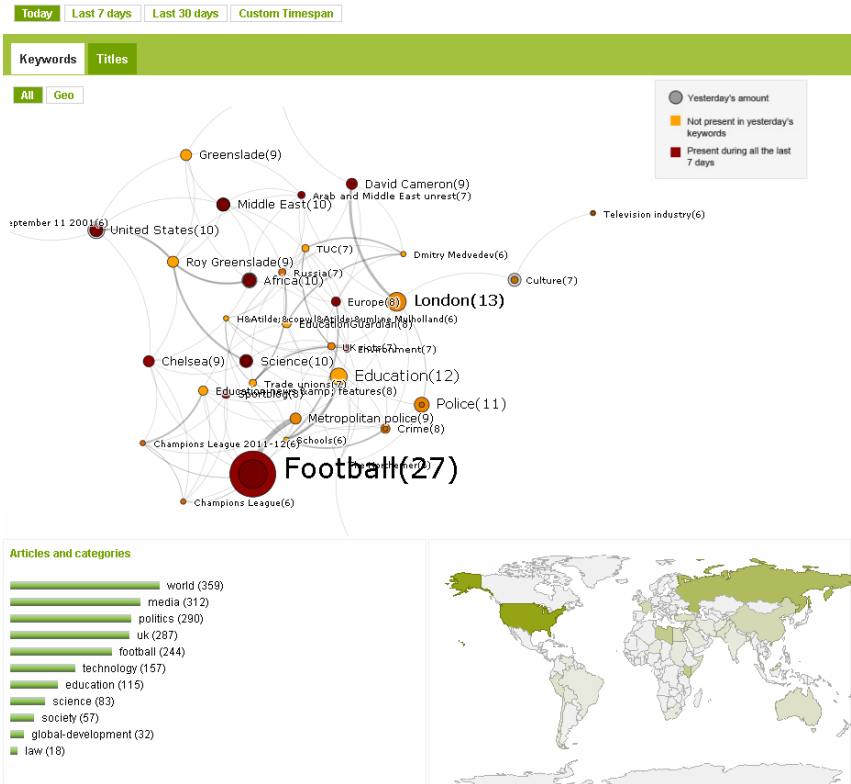


Fig. 62 - Schermata principale.

Da qui è possibile procedere in tre diversi modi: si può tornare alla rete delle relazioni, decidere di selezionare uno degli articoli oppure selezionare una data sulla timeline tematica.

Se si sceglie di approfondire ulteriormente, selezionando uno degli articoli giornalistici presentati, si potranno ottenere informazioni aggiuntive sull'articolo, derivanti da un'analisi semantica dello stesso. Una tree map mostrerà e gli argomenti trattati e le named entities citate all'interno del corpo dell'articolo, classificandoli in base al tipo, al numero di citazioni e alla loro posizione nel testo. A loro volta, gli elementi estrapolati da ogni singolo articolo potranno essere selezionati ed utilizzati come punto di partenza per una nuova ricerca e, quindi, per il proseguimento dell'esplorazione. Selezionando uno di questi termini, infatti, si tornerà al livello precedente in cui verranno mostrati, questa volta, gli articoli e la timeline relativi al nuovo elemento.

Si potranno avere così diversi tipi di percorso, che l'utente intraprenderà a seconda del proprio interesse e dei propri obiettivi, spaziando tra i tre livelli di approfondimento e tra le varie modalità. Il percorso sarà dunque personalizzato e caratterizzato da una combinazione dei fattori di relazione, tempo e luogo.

12/09/2011

[Back](#) [Details](#)

"Unions vote for 'conscience clause' to cover journalists"

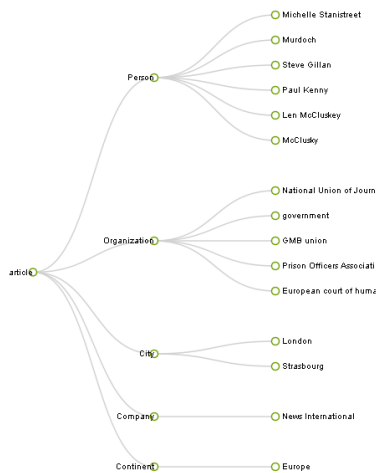


Fig. 64 - Schermata del secondo livello di dettaglio.

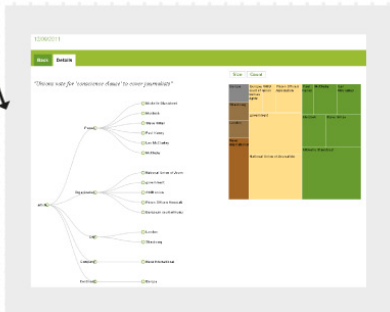


Livello principale

Primo livello di approfondimento



Articoli



Secondo livello di approfondimento

Fig. 65 - Struttura del primo livello di approfondimento dell'applicazione.

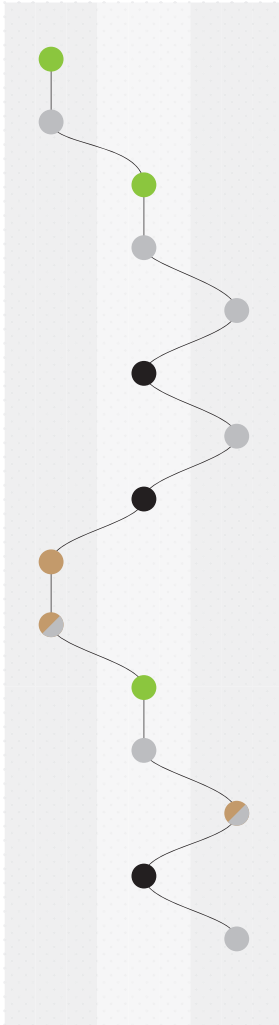
Fase di test con utenti

Il prototipo è stato fatto testare ad un campione di utenti tipo. Le operazioni svolte dagli utenti e i percorsi seguiti sono stati registrati ed annotati. Queste informazioni sono state utilizzate per realizzare alcuni schemi che permettessero di osservare e analizzare i pattern di utilizzo dell'applicazione, testando così la funzionalità e la completezza delle diverse parti e dei diversi livelli di approfondimento presenti. Il test con gli utenti è stato utile, oltre che per analizzare i pattern di esplorazione, anche per individuare e risolvere alcuni problemi e adattare alcuni aspetti per rendere più agevole l'interazione dell'utente con l'applicazione.

Vengono proposti qui alcuni di questi schemi.

- Utente 1: 25 anni, nessuna conoscenza specialistica in ambito giornalistico.
- Utente 2: 50 anni, nessuna conoscenza specialistica in ambito giornalistico.
- Utente 3: 30 anni, conoscenze specialistiche in ambito giornalistico.

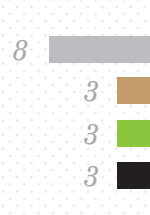
Livelli di dettaglio
1 2 3



relazioni ■
elemento geografico ■
elemento temporale ■
indietro ■

Tempo:
10 min

Elementi
selezionati:
15



Primo Utente

Fig. 66 - Schema di navigazione del primo utente durante la fase di test.

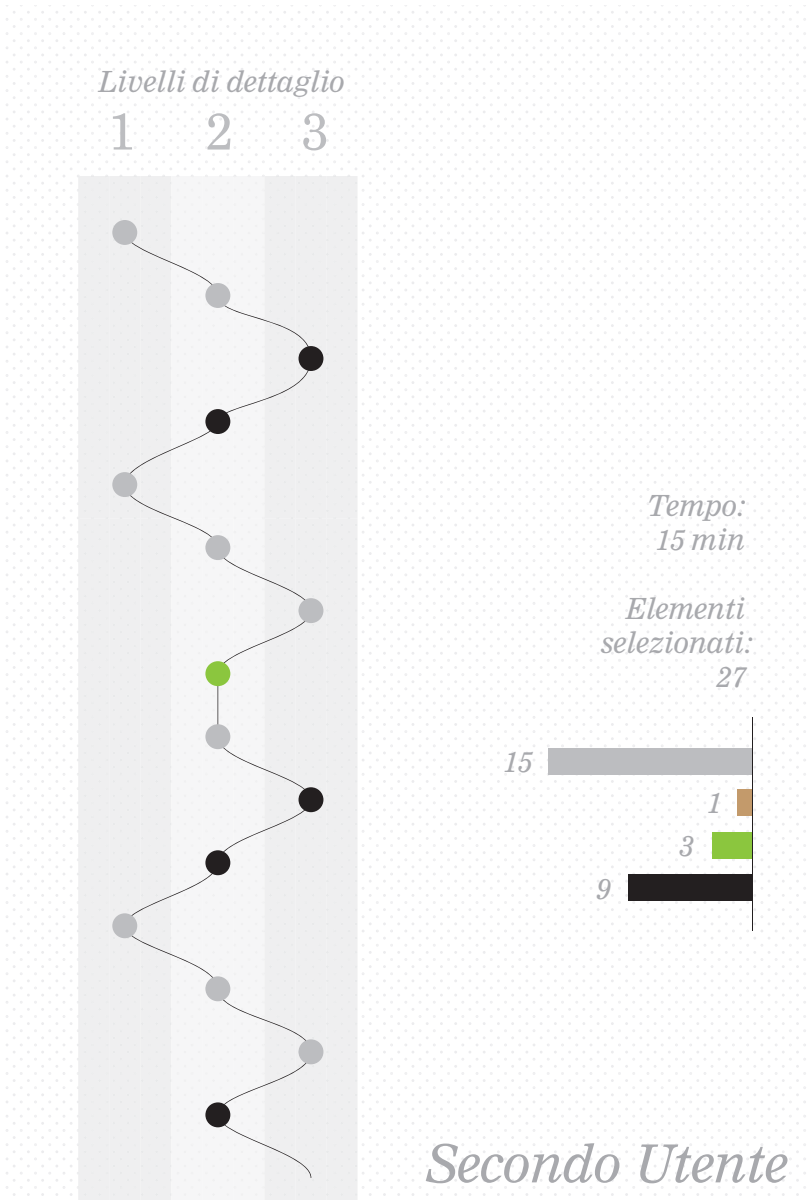
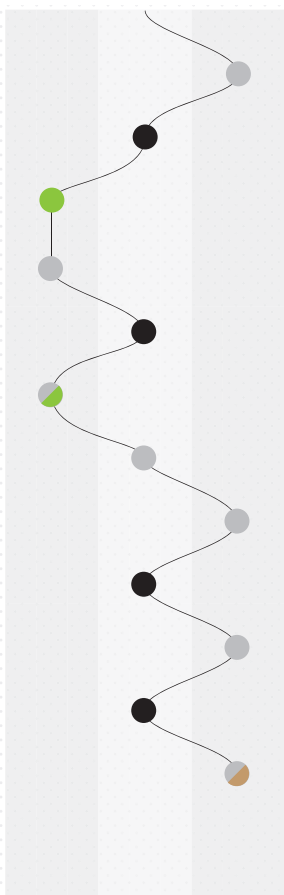


Fig. 67 - Schema di navigazione del secondo utente durante la fase di test. Prima parte.



Secondo Utente

Fig. 68 - Schema di navigazione del secondo utente durante la fase di test. Seconda parte.

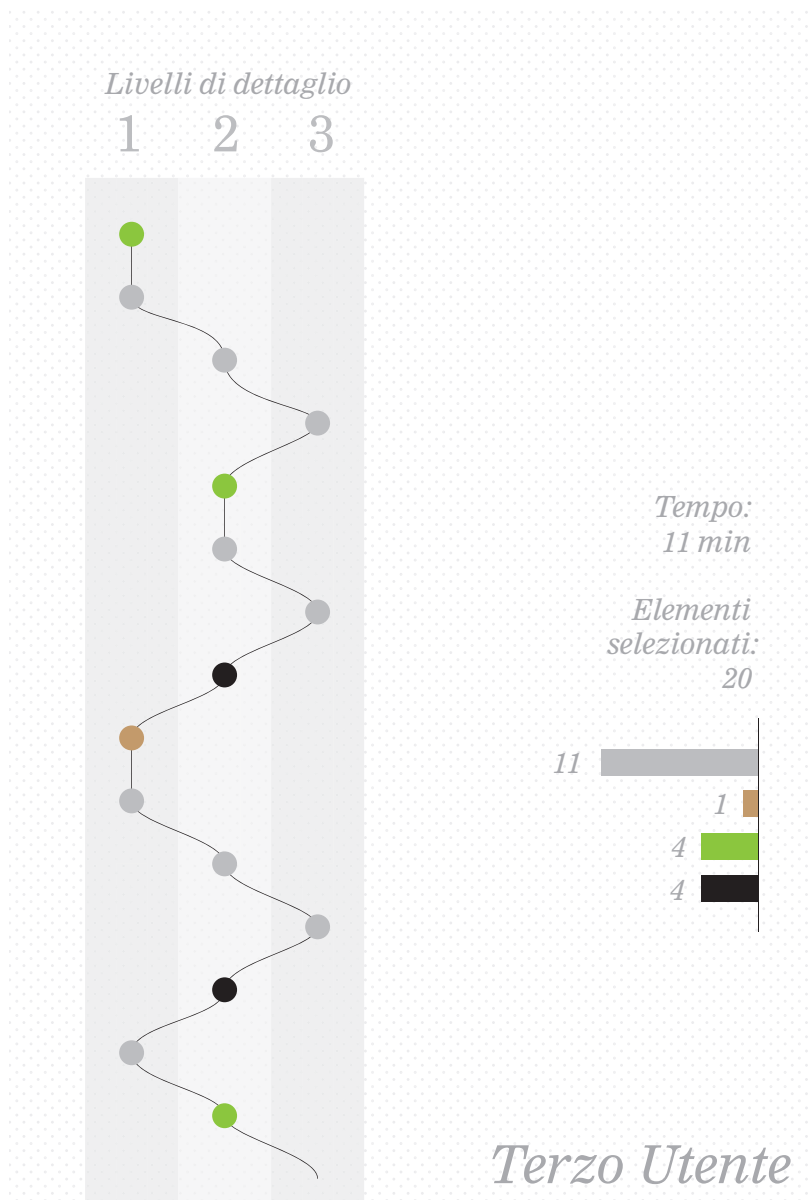
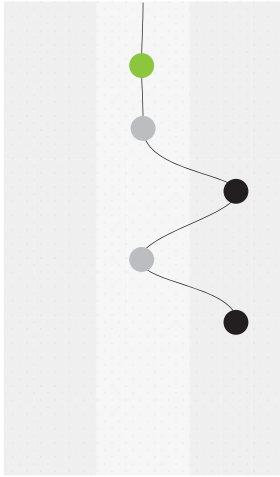


Fig. 69 - Schema di navigazione del terzo utente durante la fase di test. Prima parte.



Terzo Utente

Fig. 70 - Schema di navigazione del terzo utente durante la fase di test. Seconda parte.

Obiettivi raggiunti e possibilità di sviluppo

Nel complesso il progetto ha raggiunto una fase di sviluppo piuttosto avanzata. Le funzionalità fin'ora descritte sono state per la maggior parte realizzate, implementate e testate. Vi sono comunque alcuni punti aperti e alcune idee che non sono state portate all'effettiva realizzazione e che rimangono come possibilità di sviluppo future.

Come descritto precedentemente, si è scelto di utilizzare i dati e gli articoli diffusi da Guardian.co.uk come base per l'applicazione. Il progetto, tuttavia, è pensato per poter essere applicato anche ad altri siti e piattaforme di e-journalism (a patto che questi siano forniti di una API che permetta agli utenti o agli addetti del settore di estrapolare e raccogliere dati sugli articoli pubblicati).

Attualmente l'applicazione funziona su browser Firefox, a causa della limitata compatibilità di altri browser con alcune delle librerie usate. Sarebbe utile risolvere questo

problema, in modo da facilitare l'accesso all'applicazione da parte degli utenti.

Potrebbe inoltre essere utile estenderne la possibilità di utilizzo su diversi dispositivi, considerando che, ad esempio, per l'accesso ad internet e la consultazione di notizie online sono sempre più utilizzati mezzi come smartphone e tablet.

Un'ulteriore possibilità sarebbe quella di trattare una maggiore varietà di dati e informazioni, in formato non solo testuale, ma anche multimediale, in modo da arricchire l'esperienza della navigazione all'interno dei nodi della rete. Nel caso specifico del Guardian, poi, alcuni articoli proposti sul sito web (nella sezione DataStore) sono corredati di dataset inerenti l'argomento trattato. Sarebbe stato interessante riuscire a trovare una collocazione e un ruolo per questi dataset aggiuntivi all'interno del progetto.

Anche alcuni degli aspetti già trattati si prestano ad un ulteriore approfondimento o raffinamento. In particolare l'aspetto geografico, al momento trattato solo a livello di continenti e nazioni, senza un'analisi più dettagliata e precisa delle località a cui le notizie sono collegate.

Infine, un aspetto che non è stato portato a compimento per ostacoli di ordine tecnico, è quello un raffinamento della

modalità di visualizzazione dell'evoluzione temporale delle reti, in modo che il cambiamento nel tempo sia mostrato in modo fluido e uniforme. Un funzionamento di questo tipo renderebbe forse più facile ed immediata l'identificazione delle differenze tra diversi momenti e la comprensione di come queste siano avvenute.

Bibliografia

Paul Bradshaw, *How to be a data journalist*, Datablog
<[http://www.guardian.co.uk/news/datablog/2010/oct/01/
data-journalism-how-to-guide](http://www.guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide)>

Vilém Flusser, *La cultura dei media*, Mondadori, 2004
Margherita Pillan, Susanna Sancassani, *il bit e la
tartaruga*, Apogeo, 2004

Ben Fry, *Visualizing Data: Exploring and Explaining Data
with the Processing Environment*, O'Reilly, 2007

Henry Jenkins, *Cultura convergente*, Apogeo, 2007

Journalism.org, *New Media, Old Media: How Blogs
and Social Media Agendas Relate and Differ from the
Traditional Press*, 2010

Marsh K., *End of journalism as we know it*, The Guardian,
2010 < [http://www.guardian.co.uk/commentisfree/2010/
jan/04/citizenmedia-bbc](http://www.guardian.co.uk/commentisfree/2010/jan/04/citizenmedia-bbc) >

Margherita Pillan, Susanna Sancassani, *Costruire servizi digitali*, Apogeo, 2003

Patrizia Rodi, *Mediamondo: viaggio attraverso le comunicazioni di massa*, Lupetti, 2010

Simon Rogers, *Facts are Sacred: The power of data*, Amazon, 2011

Julie Steele, Noah Iliinsky, *Beautiful Visualization: Looking at Data through the Eyes of Experts*, O'Reilly, 2010

Sitografia

CrowdsourcingResults <<http://crowdsourcingresults.com/>>

Data.Gov <<http://data.gov>>

Datablog <<http://www.guardian.co.uk/news/datablog/>>

Fondazione <ahref <<http://http://www.ahref.eu>>

Google News <<http://news.google.com/>>

Infomous <<http://www.infomous.com>>

ManyEyes <<http://www-958.ibm.com/software/data/cognos/manyeyes/>>

Newsmap <<http://newsmap.jp/>>

Newsdots <http://slatest.slate.com/features/news_dots/default.htm>

Open Knowledge Foundation <<http://okfn.org/>>

ProPublica <<http://www.propublica.org/>>

SpaghettiOpenData <<http://www.spaghettiopendata.org/>>

Storify <<http://storify.com/>>

Storyful <<http://storyful.com/>>

TheGuardian <<http://www.guardian.co.uk>>

Visual.ly <<http://visual.ly>>

Ringraziamenti

Desidero innanzitutto ringraziare il Professor Ciuccarelli, per il tempo dedicato alla mia tesi e per gli insegnamenti nel corso dei precedenti anni di Laurea Magistrale.

Ringrazio poi tutti i professori che hanno tenuto i corsi che ho avuto modo di seguire durante i miei anni al Politecnico di Milano e Daniele Galiffa, che mi ha seguito durante il periodo di tirocinio svolto lo scorso anno presso VISup s.r.l.

Un grazie anche alla mia famiglia e agli amici che mi hanno supportato durante questo percorso.