

UNIVERSITÀ DEGLI STUDI DI CATANIA

DIPARTIMENTO DI SCIENZE UMANISTICHE

CORSO DI LAUREA IN SCIENZE PER LA COMUNICAZIONE INTERNAZIONALE

MARTINA CAVANNA

**UN GIORNALISMO DI PRECISIONE?
STORIA ED EVOLUZIONE DEL DATA
JOURNALISM INTERNAZIONALE**

PROVA FINALE

RELATORE

Chiar.mo Prof. Luciano Granozzi

ANNO ACCADEMICO 2012 - 2013

INDICE

ABSTRACT	1
RESUMEN	4
INTRODUZIONE	7
CAPITOLO PRIMO	
STORIA ED EVOLUZIONE DEL GIORNALISMO DI PRECISIONE	11
1.1. L'emergere di un concetto.....	11
1.2. Le inchieste più celebri di giornalismo di precisione.....	19
1.2.1. Philip Meyer "The people beyond 12 th Street: a survey of attitudes of Detroit negroes after the riot of 1967".....	20
1.2.2. Bill Dedman "The color of money".....	22
1.2.3. Stephen K. Doig "What Went Wrong".....	26
CAPITOLO SECONDO	
LA STRUTTURA DEL DATA JOURNALISM	31
2.1. Il data journalism come processo.....	31
2.2. Gli strumenti del mestiere.....	38
2.2.1. Raccolta e filtraggio.....	38
2.2.2. La visualizzazione.....	40
2.3. Cultura Open Data.....	46
2.3.1. Big Data.....	50

CAPITOLO TERZO

UNO SGUARDO GLOCALE.....	52
3.1. Il data journalism nel mondo.....	52
3.1.1. Stati Uniti.....	52
3.1.2. Francia.....	55
3.1.3. Spagna.....	56
3.1.4. Germania.....	58
3.2. Una parentesi italiana.....	59
3.2.1. Spaghetti Opendata.....	60
3.2.2. Fondazione ahref e la Data Journalism School.....	62
3.2.3. Datajournalismitaly.....	63
3.3. I data blog nel mondo e l'esperienza collettiva di DataJCrew...	64

CAPITOLO QUARTO

IL CASO INGLESE.....	68
4.1. Il Guardian Data.....	70
4.2. Le grandi inchieste del Guardian Data.....	74

CONCLUSIONI.....	80
-------------------------	-----------

GLOSSARIO.....	83
-----------------------	-----------

BIBLIOGRAFIA.....	86
--------------------------	-----------

SITI WEB.....	90
----------------------	-----------

ABSTRACT

It was 1968 when Philip Meyer helped the Detroit Free Press newspaper win a Pulitzer Prize for its reporting, by applying social science methods to investigate the underlying causes of the 1967 Detroit summer riots. The investigation aimed to disprove the theory according to which the rioters were the most frustrated and hopeless cases at the bottom rung of society, and that people who had attended college were just as likely to participate in the riot as those who had failed to finish high school. It was one of the first investigations supported by the data collected directly by a journalist through a survey research aimed at comparing the rioters with the non-rioters. One year later this kind of journalism obtained a name with the first edition of the book written by Meyer himself: *Precision Journalism*. The reporter turns into a scientist who, through the rigor of the data analysis method, relates to the story.

This work aims to analyze the start and evolution of precision journalism, or data-driven journalism, to provide a comprehensive framework that outlines all its features: from its emergence as a concept to the historical and structural evolution that characterized it as one of the most *avant-garde* methods in investigative reporting.

The first chapter highlights the salient passages that have led the emergence of precision journalism, in order to fully understand the context and dynamics. The history of data-driven journalism is marked by two major Pulitzer Prize-winning investigations that distinguished precision journalism as a stand-alone type of reporting. In 1988 Bill Dedman wrote “The color of money”, a series of articles on racial discrimination by mortgage lenders in middle-income black neighborhoods in Atlanta, Georgia. Steve Doig’s “What went wrong” report is an analysis of the damage patterns from Hurricane Andrew in 1992 that showed, through the support of charts and tables, how weakened building codes and poor construction practices contributed to the extent of the disaster.

The *modus operandi* of Meyer and Doig has now become the role model to be followed by all information journalists: data-driven journalism emerges from the union of investigative journalism and Open Data. This is the main concept developed in the second chapter: in an era of information redundancy the increasing availability of data requires reporters to have an increasingly wider range of skills involving the ability to collect, filter, and tell them to a wider audience. Online databases access have facilitated, simplified and accelerated the research work.

World governments are moving in this direction, by providing the citizens with databases that they can access freely. The precursor in this field is the United States with data.gov, while in other countries activists and communities are creating unofficial projects of great impact, waiting for their governments to follow the American model. We can mention independent newsrooms like Mediapart or Rue89 in France and communities such as Journalism++ in Germany and Spaghetti OpenData or DatajournalismItaly in Italy. Entering the realm of Open Data, we cannot take into account the context in which they develop. The United States owe its leadership in this field to the Freedom of Information Act, a 1966 law that authorizes journalists and ordinary citizens to access information and documents controlled by the United States government, ensuring the transparency of public administration towards citizens and the freedom of the press towards journalists. In our country law 241/90 is still very restrictive, but a petition for the adoption of an Italian FOIA has been made.

The second part of the chapter is dedicated to the pragmatic aspect of data-driven journalism: data collection is only the first of a series of steps that a data journalist must follow. Thanks to the Internet there are plenty of free tools that allow anyone, including beginners, to produce a data investigation. To this regard we offer a thorough look at the best tools suited to filtering, such as ScraperWiki and Open Refine and visualize, perhaps the more creative and fascinating part of the process. Among those tools the most common are Piktochart, Infogr.am, Many Eyes, Tableau Public and others.

The second part of the work provides a look into the global and local data-driven journalism experience. In the third chapter we list the most important official and independent newspapers that have understood the potential of this new way of doing journalism and have decided to open a space dedicated to it within their platforms. ProPublica in the United States, Le Monde datablog in France, El Mundo and LA NACION data in Spain, Zeit Online in Germany and lastly the most influential Italian communities, such as Spaghetti OpenData and Datajournalismitaly, are some cases of the five countries analyzed. Most of them are self-financed and self-produce their own reports with the work of volunteers and enthusiasts who discuss and share information through a mailing list system.

The fourth and final chapter focuses on a careful analysis of one of the first sections dedicated to data journalism, the Guardian Data, created from the valuable contribution of its first data editor Simon Rogers. The most widely read newspaper in England has always believed and invested in data: from the first edition of the Manchester Guardian in 1821, to the quote of the legendary editor Charles Prestwich Scott, written on the occasion of the first century of the journal in 1921: “Comment is free, but facts are sacred”. What data cannot do under any circumstances, unlike words, is tell untruths, transforming reality into a packaged reflection of priorities and specific interests.

This work is developed by two approaches: the theoretical one is realized with the support of literature searches and the study of the websites and newspapers analyzed, while the practical approach relates to the experience gained from the administration of the collective blog Data Journalism Crew. The goal of this work is to demonstrate that data-driven journalism is one of the most promising frontiers in journalistic investigation, an *avant-garde* technology and a revolution in terms of information ethics.

RESUMEN

Era el año 1968 cuando Philip Meyer ayudó al periódico Detroit Free Press para ganar un Premio Pulitzer por su reportaje, mediante la aplicación de los métodos de las ciencias sociales para investigar las causas reales de los disturbios acaecidos en el verano de 1967 en Detroit. La investigación tuvo como objetivo refutar la teoría de que los manifestantes eran los casos más frustrados y sin esperanza de la sociedad y que las personas que habían asistido a la universidad tenían la misma probabilidad de participar en los disturbios como los que habían podido terminar la escuela secundaria. Fue una de las primeras investigaciones realizada con el apoyo de los datos recogidos directamente por un periodista a través de una investigación para comparar los manifestantes con los que no eran alborotadores. Al año siguiente este tipo de periodismo obtiene un nombre con la primera edición del libro escrito por Meyer: *Precision Journalism*. El periodista se convierte en un científico que, por el rigor del método de análisis de datos, se refiere a la historia.

Este trabajo tiene como objetivo analizar el nacimiento y la evolución del periodismo de precisión, o el periodismo de datos, para proporcionar un marco amplio que lo describa en todos sus aspectos: desde su aparición como concepto hasta la evolución histórica y estructural que lo consagraron como uno de los métodos de vanguardia en el periodismo de investigación.

En el primer capítulo se destacan los pasajes más importantes que han llevado a la aparición del periodismo de precisión, con el fin de entender plenamente el contexto y la dinámica. La historia del periodismo de datos está marcada por dos grandes investigaciones ganadoras del Premio Pulitzer que consagraron el periodismo de precisión a un tipo independiente de los informes. En 1988 Bill Dedman escribió “The color of money”, una serie de artículos sobre la discriminación racial por parte de los prestamistas hipotecarios en los barrios negros de medianos ingresos en Atlanta, Georgia. “What Went

Wrong” es el informe firmado por Stephen Doig, con el objetivo de demostrar a través de la ayuda de gráficos y tablas que los daños causados por el huracán Andrew en 1992 se debieron en gran parte a la corrupción edilicia y no a la fuerza destructiva de la naturaleza.

Hoy en día el *modus operandi* de Meyer y Doig se ha convertido en el modelo a seguir por todos los periodistas de información: el periodismo de datos surge de la unión del periodismo de investigación y de los Open Data. Este es el principal concepto desarrollado en el segundo capítulo: en una era de redundancia de la información, la creciente disponibilidad de datos requiere a los periodistas una gama más amplia de habilidades, como recopilar, filtrar y contarlos a un público más amplio. Las bases de datos en línea facilitan, simplifican y aceleran el trabajo de investigación. Los gobiernos del mundo se están moviendo en esta dirección, ofreciendo a los ciudadanos bases de datos que se pueden acceder libremente. Precursor en este campo fueron los Estados Unidos con data.gov, mientras que en otros países los activistas y las comunidades han creado proyectos no oficiales de gran impacto, a la espera de que los gobiernos siguiesen el modelo americano. Hablamos de redacciones independientes como Mediapart o Rue89 en Francia y en comunidades como Journalism ++ en Alemania y DatajournalismItaly o Spaghetti OpenData en Italia. Entrando en el ámbito del Open Data, no podemos pasar por alto el contexto en que se desarrollan. Los Estados Unidos deben su liderazgo a la Freedom of Information Act, una ley de 1966 que permite a periodistas y ciudadanos acceder a las informaciones y a los documentos controlados por el gobierno de los Estados Unidos, que garantiza la transparencia de la administración pública y la libertad de prensa. En nuestro país la ley 241/90 sigue siendo muy restrictiva, pero una petición para la adopción de un FOIA italiano está despuntando. La segunda parte del capítulo está dedicada a los aspectos pragmáticos del periodismo de datos: la recopilación es solo el primero de una serie de pasos que un periodista de datos debe cumplir. Gracias a Internet hay muchas herramientas gratuitas que permiten a cualquier persona, a los principiantes también, producir una investigación de datos. En este sentido, ofrecemos un vistazo a fondo en las herramientas más adecuadas para filtrar, como ScraperWiki y Open Refine y visualizar, tal

vez la parte más creativa y fascinante del proceso. Entre esas herramientas las más comunes son Piktochart, Infogr.am, Many Eyes, Tableau Public y otros.

La segunda parte de la obra ofrece una mirada a la experiencia global y local del periodismo de datos. En el tercer capítulo se enumeran los periódicos oficiales e independientes más importantes, que han comprendido el potencial de esta nueva forma de hacer periodismo y han decidido abrir un espacio dedicado dentro de sus plataformas. ProPublica en los Estados Unidos, Le Monde Datablog en Francia, El Mundo y LA NACION data en España, Zeit Online en Alemania y las comunidades italianas más influyentes, como Datajournalismitaly y Spaghetti OpenData, son algunos casos de los cinco países analizados. La mayoría se autofinancia y autoproduce los reportajes con el trabajo de voluntarios y entusiastas que discuten y comparten informaciones a través de un sistema abierto de listas de correo.

El cuarto y último capítulo se centra en un análisis cuidadosa de una de las primeras secciones dedicadas al periodismo de datos, el Guardian Data, que nace de la valiosa contribución de su primer data editor Simon Rogers. El periódico más leído en Inglaterra siempre ha creído e invertido en los datos: desde la primera edición del Manchester Guardian en 1821, hasta la cita del legendario editor Charles Prestwich Scott, escrita en ocasión del primer siglo de la revista en 1921: “Comment is free, but facts are sacred”. Lo que los datos no pueden hacer en ninguna circunstancia, a diferencia de las palabras, es mentir, transformando la realidad en una reflexión empaquetada de prioridades e intereses específicos.

Este trabajo se desarrolla en función de dos enfoques : el teórico se realiza con el apoyo de las búsquedas bibliográficas y el estudio de los sitios web y periódicos analizados, mientras que el enfoque práctico se refiere a la experiencia adquirida en la administración del blog colectivo Data Journalism Crew. En este trabajo nos planteamos el objetivo de demostrar que el periodismo de datos es una de las fronteras más prometedoras del periodismo de investigación, una tecnología de vanguardia y una revolución en términos de ética de la información.

INTRODUZIONE

È il 1968 quando Philip Meyer vince il Pulitzer grazie a un'inchiesta sulla rivolta della 12th Street, mirata a dimostrare che gli scontri di Detroit non coinvolgevano solo le persone più frustrate e disperate della società afroamericana e che la gente di colore che aveva frequentato corsi universitari aveva tante possibilità di partecipare alle rivolte quanto quelle meno istruite. È uno dei primi reportage a basare le proprie tesi su dei dati raccolti direttamente da un giornalista attraverso dei sondaggi atti a mettere a confronto i rivoltosi con i non rivoltosi. Un anno dopo questo modo di fare giornalismo prende un nome con la prima edizione del libro scritto dallo stesso Meyer: *Precision Journalism*. Il giornalista diventa uno scienziato che, attraverso il rigore del metodo di analisi dei dati, si relaziona alla storia da raccontare.

Questo lavoro mira ad analizzare la nascita e l'evoluzione del fenomeno del giornalismo di precisione, o data-driven journalism, per fornire un quadro completo che lo delinea in tutti i suoi aspetti: dalla sua nascita come concetto all'evoluzione storica e strutturale che lo ha consacrato come uno dei modi più all'avanguardia di fare giornalismo d'inchiesta.

Il primo capitolo evidenzia i passaggi salienti che hanno portato all'emergere del giornalismo di precisione, per comprenderne appieno il contesto e le dinamiche. In particolare, la storia del data-driven journalism è segnata da due importanti inchieste vincitrici del premio Pulitzer che consacrano il giornalismo di precisione a tipologia autonoma d'inchiesta. Nel 1988 Bill Dedman scrive "The color of money", una serie di articoli sulla discriminazione razziale da parte di banche e istituti di credito ipotecario nella città di Atlanta, in Georgia. "What went wrong" è il reportage firmato da Stephen Doig, mirato a dimostrare attraverso il supporto di grafici e tabelle che i danni causati dall'uragano Andrew nel 1992 erano attribuibili in larga parte alla corruzione edilizia e non alla forza devastatrice della natura.

Il *modus operandi* di Meyer e Doig è diventato oggi il modello da seguire per i nuovi giornalisti dell'informazione: il data-driven journalism nasce dall'incontro del giornalismo investigativo con gli Open Data. Questo è il concetto fondamentale sviluppato nel secondo capitolo: in un'epoca di sovrabbondanza d'informazione la sempre maggiore disponibilità di dati richiede ai giornalisti una serie sempre più ampia di competenze che gli conferisca la capacità di raccogliarli, filtrarli e raccontarli a un pubblico sempre più vasto. Internet offre un terreno fertile per il giornalismo e la ricerca, grazie alla possibilità di raccogliere dati su individui, enti pubblici e ministeriali, persino sulle società. Grazie alle banche dati online i limiti di accesso alle informazioni sono abbattuti, facilitando, semplificando e velocizzando il lavoro di ricerca. I governi di tutto il mondo si stanno muovendo in questa direzione, mettendo a disposizione del cittadino delle banche dati a cui poter accedere liberamente. È indubbio che i modi e i tempi di questa evoluzione varino di paese in paese: antesignani in questo campo sono stati gli Stati Uniti con il sito data.gov, mentre negli altri paesi attivisti e comunità di appassionati hanno dato vita a progetti non ufficiali di grande impatto, in attesa che il proprio governo segua il modello americano. Si parla di siti indipendenti come Mediapart o Rue89 in Francia e di comunità come Journalism++ in Germania e Datajournalismitaly o Spaghetti Opendata in Italia.

Entrando nel territorio degli Open Data, non si può non tener conto del contesto in cui questi si sviluppano. Gli Stati Uniti devono la loro leadership al Freedom of Information Act, una legge sulla libertà di informazione risalente al 1966. Grazie al FOIA giornalisti, studiosi, ma anche semplici cittadini hanno ottenuto l'accesso agli archivi di Stato statunitensi, a documenti riservati e coperti da segreto di Stato, garantendo la trasparenza della pubblica amministrazione nei confronti del cittadino e il diritto di cronaca e la libertà di stampa dei giornalisti. San Francisco è stata la prima città a creare una legge sugli Open Data nel settembre 2010, seguita a ruota dalle altre città americane. Il nostro paese è meno preparato in materia di accesso alle informazioni: la legge n°241 del 1990 e i successivi aggiornamenti sono ancora molto restrittivi in termini di diritto di accesso agli atti e trasparenza, ma con la

contaminazione effettuata dal data-driven journalism anche in Italia è nata l'iniziativa per l'adozione di un FOIA. Non è sufficiente però che un dato sia accessibile: per essere Open, un dato deve essere tecnologicamente e legalmente abilitato a essere riusato, diventando così compatibile col concetto di dato aperto.

Nella seconda parte del capitolo si affronta quindi l'aspetto più pratico del data-driven journalism: la raccolta dati è solamente il primo di una serie di step che un giornalista dei dati deve compiere. Per fortuna la Rete mette a disposizione anche dei meno esperti degli strumenti gratuiti che permettono a chiunque di produrre un'inchiesta basata sui dati. A questo proposito si offre uno sguardo approfondito sugli strumenti più adatti al filtraggio, come ScraperWiki e Open Refine e alla visualizzazione, aspetto forse più creativo e affascinante del processo. Tra gli strumenti di visualizzazione i più diffusi a essere sottoposti ad analisi sono Piktochart, Infogr.am, Many Eyes, Tableau Public e altri ancora.

La seconda parte della prova finale fornisce uno sguardo globale e locale sul mondo del data-driven journalism. Nel terzo capitolo si enumerano le principali testate, ufficiali e indipendenti, che hanno compreso le potenzialità di questo modo nuovo di fare giornalismo e hanno deciso di dedicargli uno spazio all'interno delle loro piattaforme. ProPublica negli Stati Uniti, il datablog di Le Monde in Francia, El Mundo o LA NACION data per la Spagna, lo Zeit Online in Germania e infine le più influenti comunità italiane, come Datajournalismitaly e Spaghetti Opendata sono alcuni casi dei cinque paesi analizzati. La maggior parte di esse sono autofinanziate e autoproducono le proprie inchieste attraverso il lavoro di volontari e appassionati che discutono e condividono notizie attraverso delle mailing list aperte a chiunque.

Il quarto e ultimo capitolo si concentra infine su un'attenta analisi di una delle prime sezioni dedicate al data journalism, il Guardian Data, nato dal prezioso contributo del suo primo data editor Simon Rogers. Sui dati, il quotidiano più letto d'Inghilterra ha sempre creduto e investito: dalla prima edizione del Manchester Guardian del 1821, alla citazione del leggendario redattore Charles Prestwich Scott, scritta per il manifesto

in occasione del primo secolo di vita del giornale nel 1921: “Comment is free, but facts are sacred”. Ciò che i dati non possono fare in alcuna circostanza, a differenza delle parole, è mentire, fare della realtà il riflesso confezionato di priorità e interessi specifici.

La prova finale si sviluppa mediante due approcci: quello teorico realizzato con l’ausilio di ricerche bibliografiche e con l’analisi dei siti e dei quotidiani oggetto di verifica e uno pratico grazie all’esperienza maturata dalla gestione del blog collettivo Data Journalism Crew. L’obiettivo della ricerca è quindi quello di dimostrare che il data-driven journalism rappresenta una delle frontiere più promettenti dell’indagine giornalistica, un’avanguardia tecnica e una rivoluzione in termini di etica dell’informazione.

CAPITOLO PRIMO

STORIA ED EVOLUZIONE DEL GIORNALISMO DI PRECISIONE

1.1 L'emergere di un concetto

Il concetto di giornalismo di precisione è, come spesso accade nelle scienze sociali, molto più vecchio del termine stesso. Già pratica comune per diversi giornalisti, il termine “giornalismo di precisione” diventa popolare solamente nei primi anni Settanta, in seguito alla pubblicazione della prima edizione del libro di Philip Meyer¹, *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Inizialmente il titolo previsto per il libro era un altro ed è nella primavera del 1970, durante una visita alla Kansas State University, che il titolo viene semplificato da “The application of social and behavioral science research methods to the practice of journalism” a “Precision Journalism” da Everette E. Dennis². Dennis, autore e leader nel campo dell'istruzione superiore, era solito utilizzare il termine durante le sue conferenze universitarie per sottolineare il contrasto esistente tra il metodo scientifico proprio del giornalismo di precisione e l'approccio pseudoartistico del Nuovo Giornalismo³, che si serviva di tecniche proprie della letteratura per

¹ Grazie al libro, antesignano dell'uso dei computer e dell'analisi dei dati alla pratica della cronaca giornalistica, il giornalista, scrittore e docente viene considerato il padre dei giornalisti che utilizzano gli strumenti della ricerca sociologica e della statistica per realizzare inchieste. In suo onore l'*Institute Reporters and Editors* assegna ogni anno premi a lui intitolati per le migliori inchieste di giornalismo di precisione.

² Philip Meyer, *Paper Route: Finding My Way to Precision Journalism*, iUniverse, 2012, pag. 298.

³ Espressione coniata nel 1973 dal giornalista americano Tom Wolfe per riferirsi allo stile anticonvenzionale del giornalismo degli anni Sessanta e Settanta. Questo stile di scrittura giornalistica si presenta come una commistione di letteratura e giornalismo, con motivi tipici della narrativa, capaci

scrivere delle notizie. Con il suo manoscritto Meyer stabilisce invece un trend controcorrente: l'uso da parte di giornalisti di tecniche di ricerca delle scienze sociali per aumentare la profondità e l'accuratezza delle grandi storie, così che i giornalisti potessero confrontare grandi quantità di dati con relativa facilità, conferendo alla notizia il valore aggiunto dell'inconfutabile obiettività. Secondo Meyer, infatti, il giornalista non può avere a che fare con la dimensione della fiction e la disciplina dell'arte non è il genere che più si addice al giornalismo, che deve essere fatto di verità controllabili e non interpretate ideologicamente. La consapevolezza dell'irraggiungibilità dell'oggettività ideale dei fatti, aveva infatti portato alcuni dei media negli anni Sessanta a seguire la filosofia del Nuovo Giornalismo, che svincolava i giornalisti trasformandoli in narratori attraverso gli artefici della fiction. Al contrario, paragonato a uno storico e scienziato, il giornalista deve acquisire competenze tali da permettergli di creare notizie giornalistiche fatte di ricostruzioni documentate attraverso operazioni teoriche. I fatti e le notizie esistono già per loro natura, ma sta nell'acume del giornalista la capacità di farle parlare: egli è avvantaggiato dalla possibilità di poter prendere in prestito leggi da scienze come economia, sociologia, psicologia, statistica, per interrogare i fatti in maniera efficace. Le notizie non sono esenti dall'interpretazione personale, non esistono fatti imparziali se questi implicano una spiegazione giornalistica, ma questa non va confusa con la faziosità: la completezza dell'informazione non dipende dalla quantità di fatti raccolti in essa, ma dalla loro selezione obiettiva ed empirica. Il metodo scientifico, insiste Meyer, è l'unico strumento valido per far fronte alle degenerazioni provocate dalle

di catturare l'interesse del lettore. Una delle forme più consolidate è il romanzo-reportage, chiamato anche *nonfiction novel*.

prospettive dell'intrattenimento e della pubblicità⁴. Andando avanti di qualche anno però, il padre del giornalismo di precisione ha sentito il bisogno di fare alcune precisazioni circa la sua idea sul Nuovo Giornalismo, andando oltre la divisione assoluta tra giornalismo di precisione e giornalismo narrativo. Inizialmente questi due approcci, sembravano essere nemici naturali in conflitto tra loro, ma la sua tesi è che oggi, nel XXI secolo, si deve considerare la possibilità che potremmo avere bisogno di entrambi. Ciò che hanno in comune è il riconoscimento che i dati grezzi richiedono una struttura che li renda coerenti, anche se attraverso mezzi differenti: l'arte nel caso del giornalismo narrativo, la scienza nel caso del giornalismo di precisione. La difesa di quelli che fanno notizia troppo simile alla fiction è che portano "una verità più alta", ma il lettore merita di sapere quando la finzione sia il veicolo per la presunta verità. A tal proposito nel 1980, la giornalista Janet Cooke, scrisse un pezzo straziante su un bambino di soli 8 anni di nome Jimmy⁵, eroinomane sin dall'età di 5 anni, che gli portò il Pulitzer nel 1981. Successivamente si scoprì che il ragazzino non esisteva e che la storia era stata inventata: la Cooke rassegnò le dimissioni e restituì il premio. Entrambi i generi sono forme speciali che richiedono particolari competenze. La soluzione migliore, conclude Meyer, sarebbe la loro fusione in un nuovo, unico genere chiamato "*evidence-based narrative*"⁶: un processo che implica una buona narrazione basata su prove verificabili.

⁴ Philip Meyer, *Giornalismo e metodo scientifico. Ovvero il giornalismo di precisione*, Armando editore 2006, pag. 7-16.

⁵ L'articolo fu pubblicato il 19 aprile 1981 sul Washington Post <<http://www.uncp.edu/home/canada/work/markport/lit/litjour/spg2002/cooke.htm>>.

⁶ Testo adattato dell'Hedy Lamarr Lecture che Meyer ha condiviso all'Austrian Academy of Sciences <<http://www.nieman.harvard.edu/reports/article-online-exclusive/100044/Precision-Journalism-and->

Il giornalismo di precisione, fin dalla sua teorizzazione attraverso la pubblicazione del libro di Meyer, viene tuttavia continuato a chiamare negli Stati Uniti *data-driven journalism* (spesso abbreviato in ddj) o *computer-assisted reporting*⁷ e in qualche caso *database journalism*, anche se con quest'ultimo termine si tende soprattutto a porre l'attenzione sulla fase di ricerca delle notizie nei database. Nel mondo ispanico il concetto si è affermato con il nome di *periodismo de precisión*, grazie al contributo di José Luis Dader⁸, che ha curato la traduzione in spagnolo del libro di Philip Meyer nel 1993. In Italia il giornalismo di precisione inizia ad affermarsi concretamente solo dal 2010 quando, durante il Festival Internazionale del Giornalismo di Perugia, in occasione del primo panel italiano sull'argomento che ha riunito Philip Meyer, Stephen Doig, José Luis Dader, Ilvo Diamanti, Mario Tedeschini Lalli e Giorgio Meletti, si è scelto il nome di “giornalismo di precisione”.

Se già la scelta di un nome che definisca il genere ha causato dell'indecisione, fornire una spiegazione accurata di cosa sia esattamente il giornalismo di precisione non è un compito facile. Questo accade perché è un campo che, sviluppandosi negli anni, è arrivato ad abbracciare così tante discipline da diventare un argomento tanto variegato quanto complesso. Il giornalismo di precisione è un settore del giornalismo che riflette il ruolo sempre più importante che i dati

Narrative-Journalism-Toward-a-Unified-Field-Theory.aspx>. Una traduzione italiana è consultabile qui <<http://datajcrew.sudmediatika.it/?p=2103>>.

⁷ Corrispondente inglese del giornalismo computazionale. L'utilizzo del computer per raccogliere e analizzare i dati necessari per scrivere le notizie, già pratica comune negli anni Sessanta, ha portato ad affermare che il giornalismo di precisione affondi le proprie radici in questo approccio.

Barbanu, Sassu, *Manuale di giornalismo*, Laterza 2012, pag. 175, 176.

⁸ Docente ed editorialista spagnolo, all'Università di Salamanca istituisce il primo corso europeo di strumenti e tecniche di giornalismo di precisione. È attualmente professore di opinione pubblica e giornalismo di precisione presso l'Università Complutense di Madrid.

numerici hanno acquisito nell'era digitale, utilizzati per la produzione e la distribuzione delle informazioni. Ma definirlo come semplice giornalismo in base al quale i contenuti delle notizie sono organizzati intorno a pezzi di dati strutturati è un'affermazione troppo grossolana. Si tratta infatti di un processo che vede una maggiore interazione tra i giornalisti, produttori di storie attraverso dati numerici o database utilizzati come materia prima, e diversi altri campi di studio come la programmazione, la statistica e l'informatica. Secondo il giornalista e sviluppatore web Adrian Holovaty⁹, una differenza fondamentale tra il database journalism e il giornalismo tradizionale è che quest'ultimo produce articoli come prodotto finale, mentre il primo crea database di fatti che vengono continuamente aggiornati e migliorati. La maggior parte del materiale raccolto dai giornalisti è "informazione strutturata: il tipo di informazioni che possono essere tagliate e fatte a dadini, in modo automatico, dai computer"¹⁰.

In conclusione, a causa della complessità legata alla varietà dei campi che tocca, non esiste una definizione assoluta di ciò che rappresenta il data journalism. Vi sono tuttavia diversi modi in cui questo può essere visto¹¹. Uno dei modi più semplici di rispondere alla domanda "Che cosa è il data journalism?" è quello di mostrare come riferimento una mappa o un grafico interattivo appositamente creati per illustrare una storia di data journalism. In questo senso, il giornalismo dei dati può essere definito come un particolare tipo di produzione. Chi lavora con i dati,

⁹ Originario dell'Illinois, si è laureato alla Missouri School of Journalism nel 2001, iniziando a lavorare come sviluppatore web e giornalista presso il Lawrence Journal-World e il Washington Post. Celebre per aver creato EveryBlock, una startup web in grado di fornire notizie microlocal, grazie al suo saggio ha ispirato il sito, vincitore di un Pulitzer, PolitiFact.

¹⁰ La lettera integrale, pubblicata nel 2006 e considerata il manifesto del database journalism, è reperibile qui <<http://www.holovaty.com/writing/fundamental-change/>>.

¹¹ Tre modi di vedere il data journalism, di Tony Hirst <<http://schoolofdata.org/2013/06/11/several-takes-on-defining-data-journalism/>>.

però, sa bene che la produzione di un grafico è spesso la parte più semplice del processo e che la maggior parte del tempo viene spesa per trovare e analizzare i dati, oppure ponendosi delle domande sui dati al fine di trovare la storia celata al loro interno. Di conseguenza, il giornalismo dei dati non è tanto ciò che viene prodotto dai giornalisti dei dati, quanto l'insieme delle competenze di cui i giornalisti hanno bisogno per produrre le mappe o i grafici che vengono indicati come esempi di data journalism. Per delineare queste competenze basta dare un'occhiata agli annunci di lavoro¹² pubblicati da diverse testate giornalistiche in cerca di specifiche qualità adatte a ricoprire il ruolo del giornalista dei dati nella propria redazione. Secondo il Trinity Mirror, il più grande gruppo editoriale britannico, il candidato prescelto deve essere in grado di coniugare un elevato grado di competenza tecnica – in termini di ricerca, interrogazione e visualizzazione dei dati – con le più tradizionali competenze giornalistiche, come riconoscere le storie celate all'interno dei dati o la produzione di contenuti che siano veramente utili per i consumatori. Il candidato è chiamato a contribuire a creare e sviluppare pacchetti basati sui dati, trovare e raschiare le principali fonti di dati. Per il Financial Times, principale giornale economico-finanziario del Regno Unito, un giornalista esperto di dati¹³ deve avere una significativa esperienza nell'acquisizione, nel trattamento e nella presentazione dei dati nel contesto della notizia. Deve possedere una conoscenza enciclopedica delle attuali migliori pratiche nel campo del data journalism e della visualizzazione interattiva dei dati. Discutere i dati è parte quotidiana di questo lavoro, quindi è richiesta una conoscenza di

¹² Uno sguardo agli annunci di lavoro, di Tony Hirst <<http://blog.ouseful.info/2013/05/31/data-journalism-jobs-in-the-air/>>.

¹³ L'annuncio del Financial Times <<https://www.careersatft.com/interactive-data-journalist-1647/JobDetails.aspx>>.

programmi come Excel, SQL, Open Refine o Stata. Bisogna avere dimestichezza con i linguaggi HTML e CSS. Ben più importante delle proprie competenze, tuttavia, è la comprovata capacità di risolvere i problemi in modo indipendente e di tenere costantemente aggiornate le proprie abilità in un settore in rapida evoluzione.

Guardando a questi due esempi, gli annunci attuali sembrano richiedere competenze associate con lo sviluppo di applicazioni basate sui dati interattivi, ponendo l'accento su una serie di competenze di web design che associano la pratica del giornalismo basato sui dati alla produzione di grafici, mappe e altri strumenti utilizzati per illustrare una storia.

Il data journalism va inteso anche come un processo che, in qualche modo, coinvolga dei dati, anche se non necessariamente in maniera esclusiva. Mentre ci possono essere notizie interamente incentrate su essi, a volte i dati potrebbero condurci a una storia che però viene meglio raccontata in altri modi. I dati, in altre parole, possono semplicemente svolgere il ruolo di fonte e, in questo senso, il data journalism come il giornalismo investigativo è più orientato al processo, anche se potenzialmente – ma non sempre – con tempi più brevi. Sotto questo punto di vista, le competenze richieste a un giornalista dei dati possono assumere la forma semplice di competenze informatiche avanzate, come la possibilità di eseguire ricerche utilizzando potenti motori di ricerca web, o servirsi di strumenti di filtraggio dati o numeri come Open Refine. Il processo potrebbe ugualmente comportare l'uso di strumenti di visualizzazione per dare un senso al set di dati, o per generare ulteriori domande che potrebbero essere poste dai lettori stessi. I processi di analisi statistica possono essere necessari per capire se un set di dati ha

effettivamente qualcosa da dire, o se in realtà un dato è o non è un numero che celi qualcosa di importante a livello giornalistico.

Viviamo in un mondo digitale in cui quasi tutto – foto, video, audio, voti, omicidi, persino la nostra cerchia di amici – può essere e viene descritto con due numeri: zero e uno. Ciò che rende il giornalismo dei dati diverso sono le nuove possibilità che si creano dalla combinazione del tradizionale fiuto per la notizia e la capacità di raccontare una storia avvincente, con la portata e la varietà di informazioni digitali oggi disponibili. Queste possibilità possono avvenire in qualsiasi fase del processo: attraverso la programmazione per automatizzare il processo di raccolta e fusione delle informazioni, attraverso un’infografica coinvolgente che racconti una storia complessa in maniera avvincente, oppure ancora rendendo trasparente il processo di raccolta delle informazioni rendendo i dati aperti e quindi usufruibili da chiunque.

Anche il giornalista Paul Bradshaw pone l’accento sull’ambiguità dei termini “dati” e “giornalismo”. Perché i dati potrebbero essere sia la fonte del data journalism, che lo strumento con cui la storia è raccontata, oppure ancora potrebbero essere considerati entrambe le cose. Come ogni fonte, dovrebbe essere trattata con scetticismo e, come ogni strumento, si dovrebbe essere consapevoli di come si potrebbero modellare e limitare le storie che vengono create con esso¹⁴.

Il data journalism rappresenta l’arte di scovare delle storie da un particolare tipo di fonte – i dati –, l’abilità di interpretare il tutto e di raccontarlo sotto forma di inchieste realizzate con gli strumenti della matematica, della statistica e delle scienze sociali e comportamentali,

¹⁴ Jonathan Gray, Lucy Chambers, Liliana Bounegru, *The Data Journalism Handbook*, op. cit., pag. 8.

applicate alla pratica del giornalismo¹⁵. Una pratica che si avvicina alla ricerca scientifica, nella quale il cronista adotta un metodo scientifico e applica il rigore del metodo usato per raccontare una storia con la forza della precisione e dell'obiettività¹⁶.

1.2 Le inchieste più celebri di giornalismo di precisione

Di tutte le aree tematiche, quelle che suscitano maggior interesse da parte dei lettori sono sicuramente quelle di interesse pubblico, che trattano di politica, spesa pubblica o salute, campi in cui il giornalismo di precisione ha avuto un ruolo decisivo. I primi esempi di inchieste di giornalismo di precisione risalgono alla fine degli anni Sessanta, ma con l'affermarsi del genere vi è un crescente elenco delle tematiche che il giornalismo di precisione può ricoprire in maniera efficace. Il genere raggiunge la consacrazione a tipologia autonoma di inchiesta dopo l'assegnazione del premio Pulitzer a importanti inchieste condotte da Philip Meyer, Bill Dedman e Steve Doig.

¹⁵ Il giornalismo di precisione secondo Steve Doig sul dizionario online della Comunicazione <http://www.blackwellreference.com/public/tocnode?id=g9781405131995_yr2010_chunk_g978140513199521_ss96-1>.

¹⁶ Philip Meyer, *Giornalismo e metodo scientifico. Ovvero il giornalismo di precisione, op. cit.*, pag. 18-20.

1.2.1 Philip Meyer “The people beyond 12th Street: a survey of attitudes of Detroit negroes after the riot of 1967”

Nel 1968 uno staff di giornalisti del Detroit Free Press¹⁷, tra cui figurava Philip Meyer che era stato mandato dal Knight Newspapers¹⁸ a dare una mano a causa del sovraccarico di lavoro da coprire, si aggiudica il premio Pulitzer in Giornalismo Locale e di Ultim’ora realizzando un’inchiesta sulla rivolta della 12th Street avvenuta a Detroit, nel Michigan, nel 1967. Iniziati durante le prime ore del mattino del 23 luglio 1967 con un raid della polizia in un bar aperto oltre l’orario di chiusura e privo dell’apposita licenza di vendita per le bevande alcoliche, all’angolo tra la 12th Street e Clairmount Street, gli scontri di Detroit hanno causato un tale disordine pubblico da trasformarsi in una delle rivolte civili più letali e distruttive della storia degli Stati Uniti, con una durata di cinque giorni e un bilancio di 43 morti, 1.189 feriti, più di 7.200 arresti, e più di 2.000 edifici distrutti con danni calcolati intorno ai 500 milioni di dollari.

Una buona opportunità per formulare delle teorie si ha durante i periodi di tensione sociale e quella rivolta rappresentava il terreno fertile per trovare ipotesi da verificare. Le scienze sociali allora erano sull’orlo di una rivoluzione, i metodi quantitativi erano più che mai produttivi grazie al rapido sviluppo dei linguaggi statistici di alto livello destinati all’utente e Meyer, immerso nel metodo scientifico, aveva appena imparato ad utilizzare il *data-text*¹⁹ durante il suo anno di studi ad

¹⁷ Fondato nel 1831, il Detroit Free Press è il più grande quotidiano di Detroit, Michigan, USA. Ha ricevuto nove premi Pulitzer e quattro Emmy Awards.

¹⁸ Fondata da John S. Knight nel 1933, è una delle aziende antenate del Knight Ridder, una società di media americana specializzata nella pubblicazione di quotidiani digitali e cartacei.

¹⁹ Linguaggio di programmazione per la ricerca in scienze sociali scritto da membri della facoltà del dipartimento di relazioni sociali per la IBM 7090 della Harvard. A disposizione di chiunque al costo

Harvard. Molte teorie risiedevano nella saggezza popolare, che attribuiva la causa dello scoppio della rivolta razziale alle persone più frustrate e senza speranze provenienti dal fondo della scala economica, prive di altri mezzi di avanzamento o di espressione. Un'altra teoria sosteneva che una parte considerevole della popolazione afroamericana di Detroit consisteva di migranti dal sud rurale, ragion per cui alle difficoltà di socializzazione negli ambienti urbani del nord seguirono i disordini. Contro queste e altre teorie²⁰, il giornale decise di confrontare i rivoltosi con i non rivoltosi attraverso una ricerca analitica fatta con dei sondaggi a un campione di famiglie nella zona della rivolta. Dalle ricerche emerse che gli individui che avevano frequentato l'università, il 18 per cento, avevano le stesse probabilità di partecipare ai disordini rispetto a chi non aveva completato nemmeno la scuola secondaria e che chi era stato cresciuto nel nord della città, il 25 per cento dei rivoltosi, aveva addirittura maggiori probabilità di sommossa rispetto ai vicini del sud, che rappresentavano solamente l'8 per cento dei rivoltosi²¹. L'istruzione e lo status economico non erano predittivi del comportamento rivoltoso. Se le teorie non fossero state dimostrate e se i giornalisti non avessero operato con il supporto e il rigore della scienza, la storia dei disordini di Detroit non sarebbe stata altro che una raccolta di fatti solo apparentemente correlati.

irrisorio di circa \$10, non ebbe futuro perché, quando si affermò la terza generazione di computer, non poté essere adattato in tempi brevi ai nuovi dispositivi e fu superato da altri linguaggi di alto livello come lo *Statistical Analysis System* e lo *Statistical Package for the Social Sciences*.

Philip Meyer, *Giornalismo e metodo scientifico. Ovvero il giornalismo di precisione*, op.cit., pag. 103.

²⁰ Meyer accenna alla *riff-raff theory*, alla *non-assimilation theory* e alla *rising aspirations theory* nel testo adattato dell'Hedy Lamarr Lecture <<http://www.nieman.harvard.edu/reports/article-online-exclusive/100044/Precision-Journalism-and-Narrative-Journalism-Toward-a-Unified-Field-Theory.aspx>> Una traduzione italiana è consultabile qui <<http://datajcrew.sudmediatika.it/?p=2103>>.

²¹ Philip Meyer, *Giornalismo e metodo scientifico. Ovvero il giornalismo di precisione*, op. cit., pag. 30-31.

1.2.2 Bill Dedman “The color of money”

Giornalista investigativo e autore americano, Bill Dedman²² vince il Pulitzer in Giornalismo Investigativo nel 1989 con “The color of money”, una serie di articoli pubblicati sull’Atlanta Journal-Constitution²³, giornale allora redatto da Bill Kovach, sulla discriminazione razziale da parte di banche e istituti di credito ipotecario nella città di Atlanta, in Georgia.

La prima serie, pubblicata in quattro numeri consecutivi²⁴ dal primo al quattro maggio 1988, ha rivelato che le banche, gli istituti di credito e le casse di risparmio, pur concedendo prestiti nei quartieri bianchi più poveri della città, non faceva altrettanto nei quartieri neri della classe media o più abbiente. A parità di reddito, gli individui bianchi ottenevano un mutuo cinque volte più spesso rispetto agli individui neri. Secondo lo studio, effettuato su sei anni di rapporti di mutuo, per un totale di 6.2 miliardi di dollari prestati, la razza – non il valore della casa o il reddito familiare – guidava le tendenze al prestito delle più grandi istituzioni finanziarie della metropoli. La ricerca, durata cinque mesi, ha analizzato 109.000 mutui concessi tra il 1981 e il 1986 in 64 quartieri (39 bianchi, 14 neri e 11 integrati) a reddito medio, ovvero con un guadagno compreso tra i \$12.849 e i \$22.393 nel 1979, anno di riferimento per il censimento del 1980. Per garantire l’affidabilità dei risultati i quartieri erano comparabili in termini di reddito e del valore

²² Nato e cresciuto nel Tennessee, Dedman ha iniziato a fare giornalismo a soli 16 anni come fattorino presso il Chattanooga Times. Ha lavorato come giornalista per il Washington Post e il Boston Globe. Fu il primo direttore della segnalazione assistita da computer per The Associated Press e ha scritto per il New York Times. Ha anche creato Power Reporting, un sito web di banche dati per aiutare i giornalisti nel campo della ricerca.

²³ L’Atlanta Journal-Constitution (AJC) è l’unico grande quotidiano nell’area metropolitana di Atlanta, in Georgia, negli Stati Uniti. È il risultato della fusione tra The Atlanta Journal e The Atlanta Constitution, avvenuta nel 1982.

²⁴ Disponibili in formato digitale sul sito Power Reporting <<http://powerreporting.com/color/>>.

dell'alloggio: per esempio, per tenere conto della rapida crescita registrata nella periferia di Atlanta, lo studio ha escluso ogni quartiere che è cresciuto di oltre il 10 per cento nel numero di case unifamiliari dal 1980 al 1987. Anche con questi accorgimenti, le differenze erano ben visibili. Nel 1984 i prestiti concessi in 1.000 strutture unifamiliari in quartieri bianchi erano 4,0 di più rispetto ai quartieri neri, e la cifra aumentava nei due anni successivi da 4,7 fino a 5,4 nel 1986. Contro chi affermava che queste differenze fossero causate da una concentrazione maggiore di vendite nei quartieri bianchi, la ricerca era stata accompagnata da uno studio parallelo sui beni immobili venduti nel 1986 in 16 dei 64 quartieri, che aveva portato a conclusioni simili: le banche avevano finanziato gli acquisti immobiliari quattro volte più spesso nei quartieri bianchi rispetto a quelli neri. Nelle aree bianche a medio reddito, le banche avevano concesso il 35 per cento dei mutui, mentre nelle aree nere a pari reddito, avevano concesso solo il 9 per cento dei mutui. Persino i quartieri bianchi a basso reddito avevano ricevuto più prestiti rispetto ai quartieri neri a reddito medio-alto, con il 31 per cento dei mutui in più rispetto al 17 per cento dei quartieri neri.

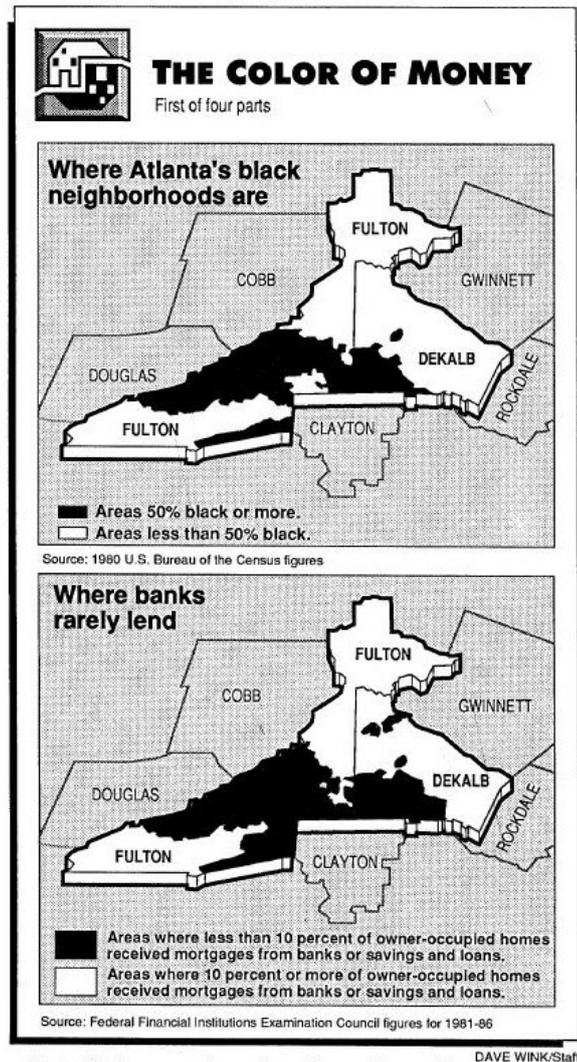


Figura 1 – Due mappe dell'Atlanta Journal-Constitution, pag.3

Come si evince dalle due mappe messe a confronto, nelle aree colorate di nero, quelle in cui si registrava una concentrazione superiore al 50 per cento di abitanti neri, a meno del 10 per cento dei proprietari era stato concesso un mutuo. Al contrario, nelle aree colorate di bianco in cui si registrava una concentrazione inferiore al 50% di individui neri, più del 10 per cento dei proprietari di abitazione aveva ottenuto un mutuo. Alla fine dell'analisi, solo due degli ottantotto istituti analizzati hanno accettato di divulgare le proprie cifre, che suggerivano che i neri facevano richiesta di prestito in proporzione inferiore rispetto ai bianchi,

ma le loro richieste venivano rigettate quattro volte più spesso rispetto a quelle dei bianchi. La Georgia Federal Bank nel 1987 ha respinto 241 su 4.990 candidati bianchi, o il 5 per cento, e 51 di 238 candidati neri, o il 21 per cento. La Fulton Federal Savings and Loan, dal 1985 al 1987, ha respinto 1.301 su 12.543 candidati bianchi, o il 10 per cento, e 363 di 1.022 candidati neri, ovvero il 36 per cento²⁵.

Grazie all'ampio uso di mappe, grafici e tabelle che dimostravano in maniera chiara le percentuali e le differenze dei prestiti concessi a bianchi e neri, l'inchiesta suscitò una tale eco da spostare l'attenzione sugli istituti di credito di tutta la nazione²⁶. I lettori furono entusiasti del metodo utilizzato che conferiva alla notizia un'oggettività indiscutibile: per quel periodo, durante il quale ancora si guardava ai computer con diffidenza, Bill Dedman rappresenta un vero antesignano. L'inchiesta non era stata creata solamente con interviste fatte a banchieri, proprietari di case, periti, legali o agenti immobiliari. Il giornale aveva ottenuto i rapporti compilati dal Federal Financial Institutions Examination Council²⁷ grazie al Freedom of Information Act²⁸ e li aveva poi combinati con i dati demografici del censimento del 1980, aggiornati con le informazioni del 1987 dell'Atlanta Regional Commission²⁹. Tutte le analisi statistiche erano state dirette da ricercatori universitari specializzati nel settore economico ed eseguite al computer adattando alcuni metodi e sistemi di classificazione utilizzati dai ricercatori della

²⁵ The Atlanta Journal and The Atlanta Constitution 1 maggio 1988, pag. 1-3.

²⁶ A tal proposito il 22 gennaio 1989 venne pubblicato un articolo dal titolo "Blacks turned down for home loans from S&Ls twice as often as whites", disponibile in formato digitale su <<http://powerreporting.com/color/53.html>>.

²⁷ Organismo di coordinamento tra le varie agenzie di vigilanza bancaria e l'applicazione uniforme dei principi di sorveglianza sul sistema bancario e finanziario negli Stati Uniti.

²⁸ Legge emanata negli Stati Uniti il 4 luglio 1966 dal presidente Lyndon B. Johnson, che fornisce il diritto di accesso totale o parziale a documenti classificati delle amministrazioni pubbliche a chiunque voglia sapere come opera il Governo federale.

²⁹ The Atlanta Journal and The Atlanta Constitution 1 maggio 1988, pag. 36.

Johns Hopkins University e della Temple University. Sebbene lo studio non sia stato in grado di includere tutti i mutui concessi o negati, a causa di limitazioni alle leggi di quel periodo e di alcuni istituti che si sono rifiutati di rilasciare per intero i propri dati, la ricerca è riuscita a includere e ad analizzare 82.610 mutui per l'acquisto di una casa e 26.721 prestiti per ristrutturare la propria abitazione, per un totale di 6.241.374 milioni di dollari in prestiti da parte di banche e istituti di credito ipotecario in sei anni. Una cifra che, senza la rapidità e la precisione dei computer, non sarebbe mai stata completata in soli cinque mesi.

1.2.3 Stephen K. Doig “What Went Wrong”

Chiamato semplicemente Steve Doig³⁰, il giornalista e attualmente professore presso la scuola di giornalismo Walter Cronkite dell'Università dell'Arizona (ASU), il 24 agosto del 1992 si ritrova direttamente coinvolto nella tragedia provocata dall'uragano Andrew, classificato come il secondo uragano più distruttivo nella storia degli Stati Uniti e l'ultimo di tre uragani di Categoria 5 che hanno colpito gli Stati Uniti durante il XX secolo. Provocò la morte di 65 persone e danni per 26,5 miliardi di dollari di allora, la maggior parte dei quali registrati nella Florida meridionale e in particolare nella città di Miami dove, tra le altre, fu devastata anche la casa di Doig. Come accade con le impronte digitali ritrovate sulla scena del crimine, dalle fotografie aeree del

³⁰ Nato in California, dopo gli studi ha trascorso un anno come corrispondente di guerra per l'esercito degli Stati Uniti durante la guerra del Vietnam, per la quale ha ricevuto una stella di bronzo al valor per il servizio svolto. Nel 1974 in Florida comincia la carriera giornalistica al Daytona Beach News-Journal. Dal 1977 al 1996 è cronista per il Miami Herald, dove scrive di educazione, scienza, servizi meteorologici e aviazione, fino a quando non si costituisce all'interno del giornale uno staff speciale per le inchieste, del quale viene chiamato a far parte.

disastro emergeva un quadro chiaramente disomogeneo: le costruzioni avevano subito danni più o meno gravi che variavano non a seconda della direzione e della forza del vento, ma a seconda della loro suddivisione territoriale. La progettazione e la qualità delle costruzioni avevano largamente influito sul livello di devastazione prodotto dall'uragano. A seguito dell'accaduto Doig, con una squadra di giornalisti del Miami Herald³¹, decide di realizzare l'inchiesta intitolata "What Went Wrong"³², atta a far luce sugli abusi e sulle frodi edilizie di Miami, che avevano portato a realizzare nel tempo case sempre meno sicure per ottenere ingenti risparmi sui materiali di costruzione. Il rapporto, raccolto in uno speciale di sedici pagine pubblicato il 20 dicembre 1992, dimostra che più della forza dell'uragano, a distruggere i tetti della città era stata la corruzione nell'edilizia, che il team di giornalisti aveva sottoposto a rigorose analisi quantificate mediante verifiche agli edifici danneggiati, alle registrazioni meteorologiche e ai certificati di agibilità edilizia. Nei tre mesi successivi al disastro, dei computer avevano analizzato 60.000 verbali di accertamento danni effettuati dalla contea, creando una mappa colorata che mostrava come i 420 quartieri avevano reagito all'uragano, alla quale fu sovrapposta una seconda mappa delle zone di vento stimate.

³¹ Il più grande quotidiano nel sud della Florida, circola anche in tutta l'America Latina e Caraibi: il giornale è stato vincitore di 20 premi Pulitzer da quando fu fondato nel 1903.

³² Qui tutti gli articoli <<http://www.flickr.com/photos/juggernautco/sets/72157607210036175/>>.

Hurricane Andrew's damage was intense in the areas where its most powerful winds came ashore. Still there were pockets of large-scale destruction in areas far from the strongest winds.

Less wind, lots of damage

HIGH WINDS, LOW PRESSURE

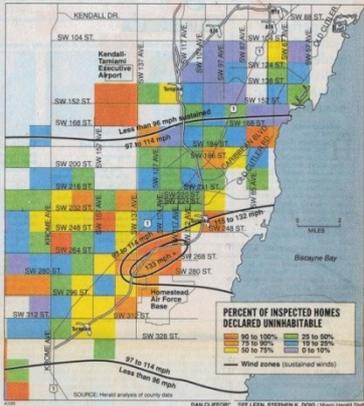
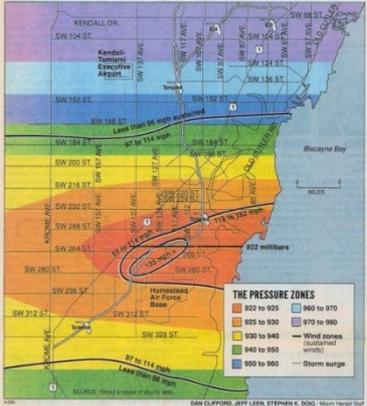
The map charts the intensity of Hurricane Andrew by displaying the wind and barometric pressure conditions that existed as the storm came ashore. The barometric pressure gradients — called isobars — are based on unofficial readings from amateur meteorologists that have been analyzed by meteorologist Edward Rappoport, of the National Hurricane Center in Coral Gables. Pressure is considered the most accurate measure of hurricane strength, with the lowest

pressures indicating the strongest storms. The 922 millibar reading, centered near Naranja Lakes, marks Andrew's position as the first strong hurricane to hit the continental United States this century. The dotted line that meanders near the coast is the extent of the tidal surge of seawater. The wind zones areas are based on preliminary estimates of sustained wind strength by NOAA scientist Peter Black.

PATCHWORK OF DESTRUCTION

This map shows the percentage of all the homes inspected in each square mile of South Dade that have been declared uninhabitable, whether repairable or not. A square mile is shaded only if at least 10 percent of the housing units there had been inspected and there are at least 25 homes. Presumably, high-damage squares are clustered around the zone of peak wind

intensity and lesser-damage squares tend to predominate north and inland where Andrew's sustained winds weren't as ferocious. But also note the pockets of high-percentage uninhabitable homes in the northern portion of the hurricane zone, as well as the spots near the maximum intensity zone where a surprisingly small percentage of homes were rendered uninhabitable.



UNINHABITABLE HOMES

This dot-density map indicates the number and approximate location of nearly 15,000 homes south of Kendall Drive declared 'uninhabitable' by county inspectors in the weeks following Andrew. The red dots represent homes damaged so badly that they need to be demolished. The blue dots represent uninhabitable homes that were deemed 'repairable'. Notice the dot patterns in relationship to the wind

zones on the map. For instance, in the area around the wind zone, the oval just north of Homestead Air Force Base, there is a hazy pattern of destruction, as would be expected. But there also is heavy damage in the zone north of SW 186th St., an area where hurricane scientists believe the sustained winds stayed below 50 mph, well below the maximum winds specified in the South Florida Building Code.

SOME BETTER, SOME WORSE

Homes built since 1980 were constructed like those built earlier, the average level of damage for both should be approximately the same. Wind this map shows is where homes built since 1980 were damaged to a significantly greater or lesser extent than those nearby. A square is colored red if the average damage to homes built since 1980 was at least 50 percent higher than average damage to all homes. A square is colored green if the average damage was at least

50 percent less than the average. Or, for instance, if there are a thousand homes in a specific square mile — including 500 built since 1980 — and half of all homes were uninhabitable, then 250 built since 1980 should be uninhabitable. This map shows only those square miles where the damage difference is at least 50 percent greater (or 50 percent less). The square in this example would be red if 500 or more homes were uninhabitable, green if there were 100 or fewer.

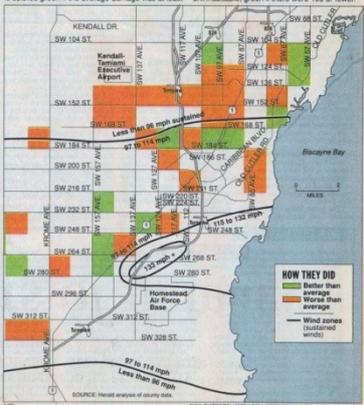
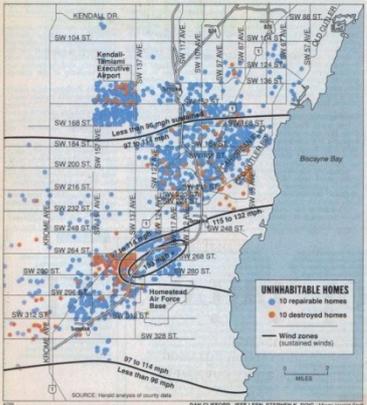


Figura 2 – Le mappe del Miami Herald, pag. 16

A quel punto il quadro fu chiaro e inconfutabile: molti dei quartieri più seriamente colpiti si trovavano lontano dalle zone con i venti più forti e le case più nuove avevano subito danni più gravi rispetto a quelle di più antica costruzione. Le case costruite dal 1980 in poi erano il 68 per cento

La mappa, incorniciata da un grafico, elencava tutte le 420 suddivisioni mettendo a confronto le unità abitative in suddivisione, la percentuale di case ispezionate, la percentuale di case ritenute inagibili, l'anno medio di costruzione e il valore medio valutato. Un diagramma riassuntivo con tonalità che andavano dal bianco al grigio antracite, mostrava la velocità del vento, dimostrando che nelle aree con venti più lievi, da 85 a 127 miglia all'ora, le case più nuove costruite dopo il 1979 avevano una probabilità tre volte maggiore di diventare inagibili rispetto a quelle costruite in precedenza. Doig dimostrò così che la corruzione dell'edilizia urbana aveva modificato a proprio favore i regolamenti edilizi, costruendo negli anni case sempre meno sicure, a rischio di crollo con l'arrivo di un qualunque disastro naturale. Dopo l'evento dell'uragano, Doig non fece semplici constatazioni sulla base di ciò che Andrew aveva provocato o di quello che gli intervistati avevano affermato. Egli andò oltre, procurandosi tutti i database di cui aveva bisogno per confrontare i dati al fine di dimostrare la sua tesi. Grazie a questa inchiesta, definita da Philip Meyer come l'esempio più importante di giornalismo di precisione³⁴, il suo modo di procedere diventa l'esempio del tipo di servizio pubblico che dovrebbe offrire il giornalismo.

³⁴ *Ibid.*, pag. 31.

CAPITOLO SECONDO

LA STRUTTURA DEL DATA JOURNALISM

2.1 Il data journalism come processo

Con l'avvento dei personal computer negli anni Ottanta si intensificano i rapporti tra i giornalisti e la tecnologia basata sui computer. Internet offre un terreno fertile per il giornalismo e la ricerca, grazie alla possibilità di raccogliere dati sugli individui, sugli enti pubblici e ministeriali, persino sulle società: il mondo appare come una raccolta infinita di dati. Grazie alle banche dati online i limiti di accesso alle informazioni sono abbattuti, facilitando, semplificando e velocizzando il lavoro di ricerca³⁵. L'era digitale comporta quindi lo sviluppo di un concetto più ampio di *precision journalism* ed è per questo che si evolve prendendo il nome di *data-driven journalism*. I fattori coinvolti nel processo di formazione della notizia aumentano insieme ai ruoli da ricoprire, rendendo il giornalismo dei dati sempre più un lavoro di squadra. Secondo Rich Gordon³⁶, capo redattore web del Miami Herald, le testate che vogliono intraprendere la strada del data-driven journalism devono investire in particolare su tre figure professionali, specializzate in tre diversi settori³⁷:

- Il giornalista computazionale: un giornalista specializzato nell'uso dei dati grazie al supporto dei computer. Il giornalismo computazionale, infatti, aumenta l'efficienza dei giornalisti che si occupano di reportage e approfondimenti investigativi, permettendogli di analizzare i dati nascosti nella Rete per scoprire i risultati più interessanti e giornalmisticamente rilevanti. In genere sa utilizzare i fogli di calcolo e i software di mappatura.

³⁵ Marco Pratellesi, *New Journalism, Teorie e tecniche del Giornalismo Multimediale*, Mondadori, Milano, 2008, pag. 128-129.

³⁶ Appassionato di data journalism da molti anni, è professore e direttore del Digital Innovation presso la Medill School of Journalism.

³⁷ Gordon ha presentato quest'idea durante la conferenza dell'American Society of News Editors a Washington il giugno scorso <<http://knightlab.northwestern.edu/2013/06/28/want-to-build-a-data-journalism-team-youll-need-these-three-people/>>.

- Lo sviluppatore di app notizie: un programmatore in grado di comprendere quali siano il modo e la forma migliore per presentare i dati all'utente finale. Possiede le competenze pratiche per la creazione di programmi con il quale l'utente interagisce indirettamente, oltre che una profonda conoscenza dei linguaggi di formattazione più comuni come HTML, CSS e JavaScript.
- Lo specialista della visualizzazione dati: evoluzione del grafico, deve essere uno specialista nel presentare i dati in maniera interessante e accessibile. La grafica web può essere cliccabile e interattiva, ragion per cui anche lui deve possedere le nozioni dei principali linguaggi di programmazione.

Queste figure sono tutte evoluzioni di ruoli che hanno radici storiche: la scrivania del giornalista computazionale una volta era la città, l'esperto di visualizzazione dati era un graphic designer; tutti e tre i ruoli hanno in comune un'evoluzione che richiede specifiche competenze di programmazione. Per le redazioni è sicuramente un bel vantaggio riuscire a trovare tutte queste qualità racchiuse in un singolo individuo, ma nella maggior parte dei casi si troveranno suddivise in tre diverse persone. È per questo che le redazioni devono lavorare con ciò che hanno e investire nella formazione del personale sviluppando le qualità per cui sembrano naturalmente portati.

Raccontare storie basate sui dati è il *leitmotiv* del data-driven journalism. Questa nuova connessione tra dati e notizia permette ai giornalisti di raccontare storie che altrimenti non sarebbe facile raccontare, trovare nuovi punti di vista tramite la ricerca, l'elaborazione e la presentazione dei dati. È il giornalista Mirko Lorenz a definire il processo del data-driven journalism come un *workflow*, un flusso di lavoro in cui i dati rappresentano la base per la ricerca, la visualizzazione e la narrazione della storia³⁸. Attraverso queste fasi, la storia non rappresenta più il prodotto del giornalismo, ma il processo stesso.

³⁸ Il concetto è stato esposto durante il Data-driven journalism round table di Amsterdam il 24 agosto 2010. L'evento è stato organizzato dall'European Journalism Centre in collaborazione con Liliana Bounegru e altre personalità influenti del mondo del data-driven journalism. Mirko Lorenz, *Data-driven journalism: What is there to learn?*, Amsterdam 2010 pag. 12, 13.

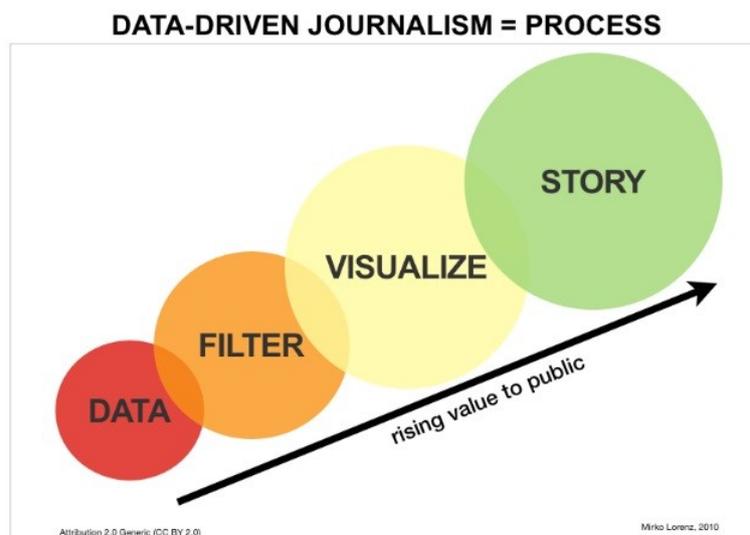


Figura 4 – Il processo del data-driven journalism come un flusso

In questo processo di rifinitura i dati vengono raccolti, filtrati e visualizzati trasformandosi nella storia finale. Man mano che il processo va avanti aumenta il suo valore agli occhi del pubblico, specialmente quando una grande quantità di dati viene elaborata e tradotta in una storia chiara e semplice da comprendere.

Nel 2011 Paul Bradshaw, da sempre attento osservatore del mondo del giornalismo, è andato oltre il modello proposto da Lorenz, analizzando i principali aspetti del data-driven journalism attraverso due modelli³⁹. Il primo illustra il processo che porta alla realizzazione di un articolo attraverso un diagramma a piramide rovesciata; il secondo analizza, attraverso un diagramma a piramide regolare, le sei modalità di comunicazione proprie del data-driven journalism.

³⁹ I modelli sono analizzati in due articoli pubblicati sull'onlinejournalismblog e tradotti da LSDI <<http://onlinejournalismblog.com/2011/07/13/the-inverted-pyramid-of-data-journalism-part-2-6-ways-of-communicating-data-journalism/>> e <<http://paulbradshaw.wordpress.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>>.

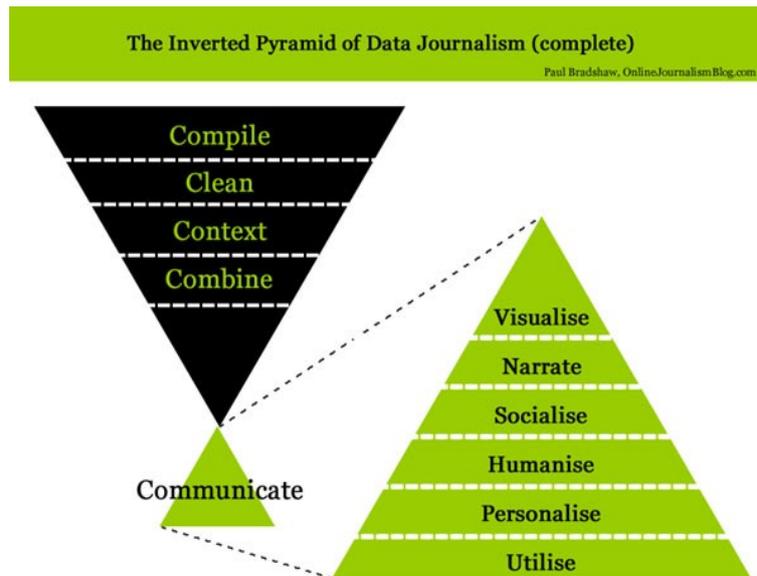


Figura 5 – Il processo del data-driven journalism secondo Paul Bradshaw

I. Raccogliere

Le fasi che precedono un lavoro di data-driven journalism partono sempre da un interrogativo da fugare attraverso i dati o da una serie di dati da avvalorare attraverso interrogativi e analisi. In entrambi i casi, tutto comincia con la raccolta dei dati, siano essi estrapolati da documenti e convertiti in informazioni analizzabili, estratti da API⁴⁰, ottenuti da fonti ufficiali o tramite sondaggi, oppure ancora attraverso strumenti specializzati nello *scraping*⁴¹ quali Scaperwiki, che consentono di penetrare i data base nascosti nel web.

II. Pulire

Una volta ottenuti, i dati contengono spesso errori umani. Possono essere incompleti, duplicati, contenere errori di battitura o errori legati ad una

⁴⁰ L'*Application Programming Interface* (in italiano Interfaccia di Programmazione di un'Applicazione), secondo la definizione fornita da Paul Bradshaw, è una tecnologia che permette di richiedere dati (ogni volta che si utilizzano Twitter, Facebook o Google Maps si utilizzano le loro API per fare domande sui propri dati) e che permette ad alcuni servizi di interagire tra loro". Paul Bradshaw, *Scraping for journalists, How to grab information from hundreds of sources, put it in data you can interrogate - and still hit deadlines*, Leanpub 2013, pag. 478.

⁴¹ Il raschiamento è una delle tecniche più potenti usata dai giornalisti dei dati che permette la raccolta di dati dispersi fra vari materiali documentari. Lo *scraping* permette anche di catturare i dati che le organizzazioni avrebbero preferito non condividere.

errata formattazione. È quindi necessario pulirli per renderli affidabili ed accrescerne la qualità. Vi sono diversi strumenti che si prestano bene a questo lavoro, come Excel o Google Docs, in cui basta ricorrere ai filtri o alla funzione “trova e sostituisci”. Open Refine consente invece operazioni più articolate.

III. Contestualizzare

Come ogni fonte, i dati non sono sempre affidabili. Bisogna pertanto chiedersi sempre chi li ha raccolti, quando, a quale scopo e con quale metodologia. Ad esempio, conoscere il numero di decessi avvenuti nell’ospedale di una determinata città è interessante, ma diventa significativo solo contestualizzando il dato rispetto al numero di abitanti, al numero di decessi precedentemente registrati e così via. La storia assume un maggiore impatto giornalistico e una maggiore rilevanza civica solo se contestualizzato.

IV. Incrociare

È possibile trovare una buona storia in un singolo dataset, ma spesso è necessario incrociarne diversi per ottenere una storia soddisfacente. La combinazione più classica dei diversi dataset è il *mash up*⁴² con mappe che consentono di visualizzare nello spazio i contenuti, personalizzabili dall’utente. Selezionando determinate opzioni, i mash up permettono infatti alle persone di raffinare i dati sulla base dei propri interessi. Per esempio, la mappa #doveticuri creata da Wired Italia permette all’utente di selezionare la propria città per ottenere informazioni sulla condizione del proprio ospedale e confrontarla con la media nazionale.

Il secondo diagramma, questa volta a piramide regolare, è interamente dedicato al processo di comunicazione proprio del data-driven journalism. Bradshaw identifica sei modalità con cui il giornalismo dei dati può comunicare ai lettori.

⁴² Letteralmente “poltiglia”, in termini informatici il *mash-up* indica un’applicazione che usa contenuti da più sorgenti per creare un servizio completamente nuovo. Uno degli elementi del Web 2.0, richiede minime conoscenze tecniche e per questo raccoglie spesso contributori inusuali.

I. Visualizzazione

La visualizzazione è il modo più rapido ed efficace per comunicare i risultati della propria inchiesta al lettore. La semplicità e l'istantaneità dell'infografica implica tuttavia che i lettori non si soffermino ad analizzarla a lungo. Occorre quindi essere chiari, evitare il 3D e assicurarsi che la grafica sia auto esplicativa. È una ricerca in divenire in cui statistica, taglio giornalistico e sensibilità artistica si intrecciano. Prima ancora di pensare a creare una visualizzazione esteticamente accattivante, l'interrogativo che bisogna sempre porsi è: "Racconta davvero una storia?". In caso la risposta fosse negativa, si starà solo facendo della buona grafica creativa.

II. Narrazione

Così come per la visualizzazione, anche nel caso della narrazione "meno è meglio". È necessario ponderare l'obiettivo che si vuole raggiungere nel comunicare i dati raccolti e di conseguenza impostare in maniera efficace la narrazione. Il data-driven journalism mira a comunicare qualcosa a qualcuno e quindi è bene tenere sempre a mente il destinatario del proprio lavoro, per non rischiare di trasformare la storia in una bella visualizzazione priva di rilevanza giornalistica.

III. Comunicazione sociale

La comunicazione è un atto sociale e lo dimostra sia il successo dell'infografica, che le molteplici iniziative di *crowdsourcing*⁴³ rivolte alla raccolta dei dati. Uno dei migliori esempi è il Guardian, che nel 2008 ha coinvolto i lettori nell'inchiesta sullo scandalo delle spese parlamentari in Regno Unito, che portò alle dimissioni del presidente dell'assemblea. Il successo dell'esperimento è continuato con numerose altre inchieste, anche più complesse, ospitate dalla piattaforma dedicata al data-driven journalism chiamata Guardian Datastore.

⁴³ Dall'inglese *crowd* (folla) e *outsourcing* (esternalizzazione di una parte delle attività), consiste nel coinvolgimento della società civile agevolando nuove forme collaborative tra giornalisti e lettori. Il termine è un neologismo apparso per la prima volta in un articolo di Jeff Howe per Wired nel 2006. Barbanu, Sassu, *Manuale di giornalismo*, op. cit., pag. 178.

IV. Umanizzazione

Radio e televisione ricorrono spesso ai casi di studio per presentare vicende basate su grandi numeri. Se i tempi di attesa per i servizi legati alla sanità si sono allungati, allora si intervista una persona che ha vissuto un'esperienza negativa in quel senso. In altre parole, si umanizzano i dati. Bradshaw spiega che quando si lavora su numeri di scala difficilmente comprensibili a livello umano, allora è altrettanto complicato riuscire a coinvolgere le persone, a prescindere dall'impatto della grafica. Pertanto è necessario tenere a mente che una semplice intervista a una persona che è stata influenzata da quel tipo di dati può fare la differenza relativamente alla forza d'urto dell'articolo.

V. Personalizzazione

Grazie al giornalismo dei dati un utente può potenzialmente controllare l'informazione che gli viene presentata. Molto comune è la personalizzazione geografica: l'utente è invitato a inserire il proprio codice postale o altre informazioni di natura geografica, al fine di stabilire l'andamento di determinate questioni nella propria città. La tendenza alla personalizzazione, tuttavia, fa suonare anche qualche campanello di allarme, primo fra tutti quello relativo all'oscurantismo prodotto da motori di ricerca, social network e siti di informazione che decidono a priori, sulla base dei dati raccolti sulla nostra persona, le informazioni da visualizzare.

VI. Utilizzo

L'informazione accresce il proprio valore quanto più vicina è all'azione dell'utente. Creare utilità partendo dai dati è un processo relativamente costoso, ma che sta prendendo sempre più piede. L'aumento delle testate o dei siti governativi che adottano degli standard nell'archiviazione dei dati, facilita la creazione di applicazioni che estraggono automaticamente le informazioni dai dataset. Nel nostro paese, per esempio, l'Istituto Nazionale di Statistica mette a disposizione i propri dati su I.Stat⁴⁴, il

⁴⁴ Le statistiche sono organizzate per temi strutturati in un albero gerarchico a due livelli. I dati sono presentati in forma aggregata in tavole multidimensionali, con la possibilità di comporre tabelle e grafici personalizzati. Il sistema è interrogabile secondo le seguenti modalità: per parola chiave, per

*data warehouse*⁴⁵ delle statistiche, accessibile da chiunque e costantemente aggiornato.

2.2 Gli strumenti del mestiere

Dai diagrammi illustrati nel precedente paragrafo si evince che comunicare è il fine ultimo del processo di produzione della storia che, inoltre, appare complesso e ricco di fasi imprescindibili. Quando il giornalismo dei dati ha fatto capolino nel panorama mondiale dell'informazione, i giornalisti hanno dovuto affrontare una fase iniziale di autoformazione basata sulla sperimentazione. Via via che il fenomeno è andato affermandosi e i giornalisti specializzandosi, anche la Rete ha messo a disposizione *tools* gratuiti e intuitivi che permettono a chiunque, anche a un principiante, di costruire una storia basata sui dati. Questo paragrafo è dedicato all'analisi di alcuni degli strumenti più diffusi e utilizzati nelle diverse fasi del processo.

2.2.1 Raccolta e filtraggio

Il data-driven journalism non è esente dalla necessità di consultare fonti, proprio come accade nel giornalismo classico. Una volta individuata la fonte più adatta che potrebbe possedere i dati oggetto del nostro interesse, la raccolta avviene sostanzialmente in tre modi: la fonte pubblica automaticamente i dati sul proprio data warehouse permettendo all'utente di scaricarli; la fonte non rilascia i dati, ma questi possono essere ottenuti attraverso una richiesta scritta; attraverso strumenti adatti allo scraping.

Nel primo caso si parla della maggior parte delle agenzie di statistica nazionale, come ISTAT in Italia, ma ci sono anche enti internazionali, come Eurostat, The World Bank o l'Organizzazione Mondiale della

tema e per territorio. Un ampio apparato di metadati agevola il reperimento e la comprensione delle statistiche da parte degli utenti <<http://dati.istat.it/Index.aspx>>.

⁴⁵ Un *data warehouse* è un magazzino centrale di dati di un'organizzazione, progettato per consentire all'utente di ottenere e lavorare grosse quantità di dati.

Sanità. Il Governo e i dipartimenti locali, regionali e nazionali sono un'altra fonte importante: i più conosciuti sono data.gov, data.gov.uk, dati.senato.it e molti altri ancora. Enti di beneficenza e organizzazioni non-profit, istituzioni accademiche, giornali, sindacati, partiti politici, tutti raccolgono regolarmente dati. Qualora alcuni di questi enti non rilasciassero automaticamente i propri dati, è possibile ottenerli formulando un'apposita richiesta: il diritto di accesso all'informazione è un concetto fondamentale nel data-driven journalism, che vede nei dati aperti la sua ragione di esistere. Essendo un argomento molto complesso, torneremo successivamente sul concetto in maniera più approfondita.

Lo scraping, l'uso di un computer per catturare informazioni da fonti online, è il metodo più utilizzato, grazie alla sua immediatezza e accuratezza. Il raschiamento permette di raccogliere e analizzare qualunque informazione sia stata digitalizzata, in qualunque formato: tabelle sepolte in PDF e pagine web, informazioni nascoste o sparse per centinaia di fogli di calcolo con nessun link di rimando⁴⁶. In un'epoca di sovrabbondanza d'informazione la sempre maggiore disponibilità di dati richiede al giornalista la capacità di riuscire a estrarre l'informazione-diamante tra le tonnellate di dati presenti nel database. Inoltre spesso i giornalisti devono correre contro il tempo scandito da *deadline* molto ravvicinate e per questo possono velocizzare il loro lavoro attraverso strumenti online ideati appositamente per automatizzare il filtraggio e presentare i dati raccolti in una struttura coerente.

- Data Wrangler: strumento interattivo per la pulizia e la trasformazione dei dati che consente di spendere meno tempo nella formattazione e più tempo nell'analisi dei dati che possono essere esportati in vari formati.
- Open Refine: precedentemente conosciuto come Google Refine, permette di accorciare e facilitare di molto il lavoro di pulizia dei dati, trasformandoli in diversi formati e collegandoli a database come Freebase⁴⁷. Refine è particolarmente efficace nel raggruppamento di

⁴⁶ Paul Bradshaw, *Scraping for journalists, How to grab information from hundreds of sources, put it in data you can interrogate - and still hit deadlines*, op. cit., pag. 1, 2.

⁴⁷ Freebase è una base di conoscenza collaborativa costituita da metadati composti principalmente dai membri della sua comunità. Il suo scopo è creare una risorsa globale che consenta l'accesso

tutte le varianti – e di tutti gli errori ortografici – contenuti in un nome, in un'unica identità. Così, per esempio “Mario R. Rossi”, “Mario Rossi R.” e “Rossi, Mario R.” diventeranno tutti l'equivalente di “Rossi Mario R.”.

- Python: software Open Source che permette l'elaborazione, l'analisi statistica e la creazione di applicazioni ad hoc per la visualizzazione dei dati. Python è un linguaggio di programmazione ad alto livello, orientato soprattutto agli sviluppatori più esperti.
- Scraperwiki: piattaforma per la costruzione collaborativa di programmi per estrarre e analizzare i dati pubblici. “Scraper” si riferisce a *screen scraper*, programmi che eseguono *data mining*⁴⁸ nella Rete. Dal 2013 il sito è stato aggiornato, rendendo gli scrapers privati per default. Questi possono essere creati utilizzando una varietà di linguaggi di programmazione, tra cui Python e JavaScript.

2.2.2 La visualizzazione

Nella loro forma più semplice, i dati possono essere visualizzati attraverso i grafici. Secondo Tufte⁴⁹, i grafici a linee e i grafici a torta sono generalmente impiegati per descrivere, mentre quelli a dispersione, anche conosciuti come *scatter plots*, sono più adatti per spingere il lettore ad analizzare i dati. I diagrammi a barre generalmente presentano poche informazioni di base e risultano efficaci per i confronti visivi, sfruttando l'innata capacità umana di confrontare lunghezze da prospettive *side-by-side*. Per creare un grafico a barre sono necessarie almeno due colonne o righe contenenti le serie di valori numerici da mettere a confronto. I grafici a linee sono invece più adatti a visualizzare un mutamento nel tempo e per questo vengono usati con dati di serie temporali. I grafici a torta sono utili quando si vuole mostrare come i valori all'interno di una serie si rapportano al totale. I grafici analitici

all'informazione in maniera più efficace. I dati sono disponibili liberamente e gratuitamente sotto forma di API aperte, RDF endpoint e altro.

⁴⁸ Processo di estrazione di informazione da dati strutturati tramite l'applicazione di algoritmi che individuano le associazioni nascoste tra le informazioni, rendendole visibili.

⁴⁹ Statistico e scultore statunitense, famoso per i suoi scritti nell'ambito dell'infografica, è stato soprannominato il Leonardo da Vinci dei dati dal New York Times e il Galileo della grafica dal BusinessWeek.

presentano un maggior numero di dati e obbligano l'utente all'esplorazione⁵⁰. La scelta del grafico influisce quindi sul tipo, sulla quantità e sul grado di sofisticazione dell'informazione rappresentata: a tale proposito è bene tenere a mente in che modo avviene nel lettore il processo di elaborazione mentale del grafico.

I grafici possono essere integrati nelle infografiche, nate dall'incrocio delle arti grafiche con il giornalismo e l'informatica. Esse possono essere bidimensionali o tridimensionali, animate o fisse.

Secondo l'esperto Alberto Cairo⁵¹, il giornalista che vuole visualizzare la propria storia deve ragionare prima di tutto come un artigiano, un ingegnere: non può ragionare come un'artista rischiando di prediligere l'aspetto estetico dell'infografica⁵². Un giornalista può trarre ispirazione dalla grande narrativa, ma non lascerà mai che i suoi articoli diventino letteratura. Lo stesso concetto è valido per la visualizzazione che è, innanzitutto, un'arte funzionale. Di conseguenza gli obiettivi più importanti di una buona visualizzazione dati saranno due: presentare le informazioni e permettere agli utenti di esplorarle. In altre parole la visualizzazione è uno strumento nelle mani del giornalista per comunicare con gli utenti e uno strumento nelle mani dei lettori per analizzare ciò che gli viene presentato. Il raggiungimento dell'equilibrio tra queste due dimensioni determina la buona riuscita del lavoro.

Fino a poco tempo fa l'elaborazione visiva dei dati raccolti era pane per i denti di designer esperti o scienziati informatici, ma lo sviluppo di strumenti gratuiti nella Rete ha reso molto più facile l'arduo compito di produrre visualizzazioni complesse.

- Easel.ly: ancora nella sua fase beta, l'applicazione web permette di creare infografiche personalizzate partendo da zero oppure scegliendo uno tra i temi a disposizione. Una volta registrati si potrà accedere a

⁵⁰ Tufte Edward Rolf, *The visual display of quantitative information*, Graphics Press, 1983 pag. 13.

⁵¹ Alberto Cairo è docente presso la School of Communication dell'Università di Miami. Tiene corsi di informazione e visualizzazione grafica ed è affascinato dalla convergenza tra comunicazione visiva, giornalismo e scienze cognitive. Nel 2012 e nel 2013, Cairo ha organizzato il primo Massive Open Online Course (MOOC) mai prodotto al mondo, in collaborazione con il Knight Center dell'Università del Texas.

⁵² Alberto Cairo, *L'arte funzionale. Infografica e visualizzazione delle informazioni*, Pearson 2013, pag. 73-75.

tutti gli strumenti dalla barra del menu in alto. Se si sceglie di partire da uno dei modelli proposti da Easel.ly è anche possibile rimuovere tutto il contenuto semplicemente cliccando su “clear”. Si possono inserire, trascinandoli sul foglio di lavoro, oggetti (divisi per categorie), forme, sfondi e si possono anche caricare proprie immagini. Una volta completata, è possibile scaricare la propria infografica in formato .jpg dalla homepage, oppure condividerla con la comunità di Easel.ly.

- Infogr.am: una volta loggati basterà cliccare su “new infographic”, scegliere un modello cliccando su “use it” e il processo creativo avrà inizio. Rispetto alle altre applicazioni Infogr.am fornisce uno spazio piuttosto limitato. Non si può inoltre partire da zero, ma si può solo scegliere uno dei temi proposti e procedere con la visualizzazione. Seguendo la barra a destra è possibile inserire grafici, mappe, testi, immagini e video, altamente personalizzabili. Grazie alle opzioni sulla sinistra della pagina si potrà scegliere la risoluzione dell’infografica, che va da un minimo di 500px ad un massimo di 700. Le infografiche vengono salvate automaticamente dal sito, una volta tornati alla home cliccando su “library” è possibile accedere al proprio profilo dove troverete tutte le vostre opere: potrete eliminarle, duplicarle, modificarle o pubblicarle. Non è possibile scaricare l’infografica sul proprio computer.
- Piktochart: tutti i temi free disponibili sono esposti sotto forma di anteprima, sulla quale basterà cliccare per caricarla e iniziare il proprio lavoro. Dal pannello di lavoro posizionato sulla sinistra è possibile aggiungere immagini, testi e forme, mentre il menu in alto permette di duplicare, copiare, bloccare i riquadri e così via. Una volta completato il lavoro si può salvare cliccando sull’apposito bottone in alto, oppure scaricarlo come immagine in formato .png.
- Prezi: in realtà è uno strumento molto creativo ideato per le presentazioni, ma uno dei suoi punti di forza risiede nella possibilità di creare infografiche animate con effetto zoom. Una volta scelto il template è possibile aggiungere testo, immagini, forme e video, ad un altissimo livello di personalizzazione. Una volta salvata, cliccando sulle varie componenti dell’infografica è possibile interagire con essa

seguendo il percorso animato che abbiamo creato. Con il pulsante in alto a destra è possibile scaricarla come file pdf.

- Venngage: su questa piattaforma è possibile caricare i propri dati, supportati in formato xlsx e xls di excel. Da sinistra si accede al menu di personalizzazione con la scelta di forme, testi, sfondi, immagini e grafici. Cliccando su “page” si possono definire le misure dell’infografica e scaricarla in tre formati di qualità. È anche possibile pubblicarla personalizzandone l’url.
- Visual.ly: “seeing is believing” è il motto di questo social network interamente dedicato alle infografiche fondato da Stew Langille. Dopo la registrazione cliccando su “create” è possibile scegliere il modello che più si adatta alla nostra infografica: scoprirete la vita di un *hashtag* o quanto siete popolari su Twitter rispetto ad un vostro amico o ancora, per esempio, la vostra vita sociale su Facebook. La particolarità di Visual.ly sta infatti nella possibilità di attingere a dei database di dati pubblici creando delle infografiche pronte. È possibile scaricare il proprio lavoro sul computer, in formato pdf, oppure condividerlo sul web. Un altro aspetto molto importante di questo software è la possibilità di mettere in contatto l’utente con editori di contenuti professionisti, semplicemente cliccando su Marketplace e compilando il form.
- Datawrapper.de: finanziato da ABZV, fondazione tedesca per la formazione dei giornalisti, è un progetto Open Source per la creazione di grafici e altre visualizzazioni di dati, recentemente aggiornato in modo da permettere ai siti di notizie di creare automaticamente grafici con i propri dati feed, capacità particolarmente utile per siti che contengono news finanziarie. Il processo è semplice e meccanico: seguendo quattro soli step basterà importare i dati, verificarli assicurandosi che siano stati importati correttamente, visualizzarli e infine pubblicarli ed esportarli.

Il modo migliore per rappresentare dati geografici è attraverso le mappe, evoluzione digitale della cartografia. Attraverso il principio della visualizzazione, la mappa diventa strumento di ricerca capace di catalizzare l’attenzione dello spettatore sul contesto spaziale in cui si sviluppa il problema oggetto dell’investigazione giornalistica. Negli

ultimi anni la diffusione di dati geografici nel dominio pubblico è aumentata significativamente, così come il numero di strumenti con cui questi si possono rappresentare. Alcuni richiedono una conoscenza minima dei linguaggi di programmazione per produrre qualcosa di efficace, mentre altri sono a prova di principiante, in grado di presentare il risultato seguendo una serie di passaggi meccanici⁵³. Aspetto molto importante della mappatura è la scelta del colore: una vera e propria disciplina che, nella sua complessità, può determinare la buona o la cattiva riuscita della visualizzazione⁵⁴. Gli uomini, per esempio, tendono a preferire l'arancione al giallo e il blu al rosso, viceversa per le donne che preferiscono il rosso e il giallo agli altri due colori. I bambini sono catturati dai colori ad alta saturazione, mentre gli adulti più maturi, col tempo, tendono a preferire più i colori pastello. La gente risponde emotivamente a certi colori e alcune di queste risposte sono talmente comuni e prevedibili da diventare strumenti di mappatura. Le preferenze di colore variano a seconda della cultura e delle caratteristiche demografiche. La maggior parte dei colori si riferiscono a diversi concetti, positivi e negativi. Il rosso, per esempio, è associato con il fuoco, il sangue, l'amore, la rabbia o il comunismo e ciò probabilmente dipende molto dal contesto: in molti dipingerebbero l'Unione Sovietica, Cuba o la Cina di rosso in un mappamondo immaginario. Allo stesso modo, il nero potrebbe connotare l'idea di morte, mentre il blu la depressione o l'aristocrazia. L'efficacia dell'uso del colore come simbolo della mappa serve a rafforzare o meno il suo ruolo di rappresentazione grafica. Ancora una volta è quindi opportuno tenere a mente il destinatario finale a cui è indirizzata la storia.

- Google Fusion Tables: è lo strumento usato dalle più influenti testate online, tra cui il Guardian Data. Il suo funzionamento si divide essenzialmente in tre passaggi: importare i dati, mapparli e personalizzarne l'aspetto. È possibile scegliere il dataset da caricare tra gli *spreadsheets*⁵⁵ messi a disposizione da Google oppure

⁵³ Nathan Yau, *Visualize this: the flowing data guide to design, visualization and statistics*, 2011 pag. 81-83.

⁵⁴ Mark Monmonier, *How to lie with maps*, 1996 pag. 147-156.

⁵⁵ Corrispondente inglese del foglio elettronico, o foglio di calcolo, ideato da Dan Bricklin. È un software che permette di effettuare calcoli, elaborare dati e tracciare efficaci rappresentazioni grafiche.

caricandolo dal proprio computer; verificata la corretta formattazione, i dati caricati appariranno in un foglio di Fusion pronto per essere mappato. Il vantaggio di questo strumento è che rileva automaticamente i dati di localizzazione in una scheda chiamata “Map of latitude”. Cliccandoci i dati saranno rappresentati geograficamente. Infine è possibile personalizzare il tutto cambiando i nomi alle etichette, la forma dei puntatori, i colori utilizzati nella mappa e molto altro ancora.

- **Many Eyes:** anche in questo caso il programma è costruito sulla base dei tre passaggi fondamentali – caricare, mappare, personalizzare. Messo a disposizione da IBM, il punto di forza di Many Eyes è l’analisi del testo (per esempio è possibile creare tag cloud di brani o mappe mentali che evidenziano la centralità di alcune parole) e di visualizzazione geografica su mappa dei dati. Con pochi click, avendo una base di dati già pronta, si possono realizzare mappe geografiche che evidenziano l’emergenza di un fenomeno. Il tutto, una volta terminato, può essere inserito all’interno di qualsiasi pagina Web. Cosa ben più importante è che Many Eyes può essere utilizzato anche come fonte per ottenere dataset: mettendo a disposizione tutti i dati caricati dagli utenti, a oggi ne raccoglie più di 425mila.
- **Tableau Public:** software creato appositamente per computer, permette di trasformare i dati in visualizzazioni interattive e intuitive. Per creare una visualizzazione basterà caricare i dati e cliccarci due volte per riportarli sulla mappa. Ma c’è di più: una volta salvato il lavoro, Tableau permette di intrecciare tra loro più mappe semplicemente selezionando l’opzione “Usa come filtro”. In questo modo, per esempio, si potrà facilmente risalire ai dati di disoccupazione regionale filtrandoli con i dati di disoccupazione nel corso degli anni, osservando l’evoluzione del fenomeno nel tempo.

Il principio su cui si basa il foglio di calcolo è semplice: fornire una tabella, detta anche foglio di lavoro, formata da celle in cui si possono inserire dati, numeri o formule.

2.3 Cultura Open Data

Il data-driven journalism nasce dall'incontro del giornalismo investigativo con gli Open Data. Secondo la definizione fornita dalla Open Definition⁵⁶, i dati aperti sono “dati che possono essere liberamente utilizzati, riutilizzati e ridistribuiti da chiunque, soggetti eventualmente alla necessità di citarne la fonte e di condividerli con lo stesso tipo di licenza con cui sono stati originariamente rilasciati”. Sulla base di questo principio, i dati devono essere disponibili nella loro interezza, preferibilmente scaricabili da Internet in un formato utile e modificabile. Ciò non è però ancora sufficiente per definirli “aperti”: i dati devono essere riutilizzabili al cento per cento, quindi privi di qualunque clausola restrittiva – per esempio il divieto di uso a fini commerciali o la restrizione d'uso solo per determinati scopi – e soprattutto devono poter essere ridistribuiti da chiunque, permettendone la combinazione con altre basi di dati. Definire con esattezza il concetto di Open Data è fondamentale per comprendere lo straordinario valore che essi rappresentano nella nostra società: essi alimentano la trasparenza e il controllo democratico dell'informazione, permettendo ai cittadini di sapere cosa sta facendo il proprio governo e ostacolare la corruzione che, nel caso dell'Italia, costa circa 60 miliardi di euro l'anno. Consentono il miglioramento dell'efficacia e dell'efficienza dei servizi pubblici: con l'apertura dei dati, il governo può aiutare a guidare la creazione di servizi innovativi che offrono un valore sociale e commerciale per il cittadino. Gli Open Government Data, ovvero i dati rilasciati dalla pubblica amministrazione, stimolano i cittadini ad assumere un atteggiamento attivo nell'ambito della società civile: non si tratta più di sapere cosa accade durante il processo di *governance*, ma essere in grado di contribuirvi.

È indubbio che i modi e i tempi di questa evoluzione varino di paese in paese. Antesignani in questo campo sono stati gli Stati Uniti con il sito data.gov, mentre negli altri paesi attivisti e comunità di appassionati hanno dato vita a progetti non ufficiali di grande impatto, in attesa che il proprio governo segua il modello americano. Gli Stati Uniti devono la

⁵⁶ La prima versione della Open Knowledge Definition è stata sviluppata dalla Open Knowledge Foundation. Tuttavia il progetto è basato sulla comunità e chiunque può partecipare.

loro leadership al Freedom of Information Act⁵⁷, la più importante legge americana in materia di accesso pubblico all'informazione statale, emanata il 4 Luglio 1966 dal Presidente Lyndon Johnson. Questa legge impone alle amministrazioni pubbliche una serie di regole per permettere a chiunque di sapere come opera il Governo Federale, comprendendo l'accesso totale o parziale a documenti classificati e non. Tutte le agenzie di Stato hanno il dovere di divulgare i documenti che vengono loro richiesti tramite un'istanza scritta effettuata da un cittadino degli Stati Uniti o appartenente ad una nazione straniera, da organizzazioni, società, corporazioni, da un governo locale, un'associazione o un'università. Grazie al FOIA giornalisti, studiosi, ma anche semplici cittadini hanno ottenuto l'accesso agli archivi di Stato americani, a documenti riservati e coperti da segreto di Stato, garantendo la trasparenza della pubblica amministrazione nei confronti del cittadino e il diritto di cronaca e la libertà di stampa dei giornalisti. Nel corso degli anni moltissime vicende sono venute a galla grazie all'uso di questa legge: dallo sfruttamento del lavoro minorile, alle scarse cure somministrate ai pazienti degli ospedali militari, alla scoperta dei difetti strutturali del serbatoio della Ford Pinto che negli anni Settanta causarono numerosi incidenti mortali, prima che la vettura venisse ritirata dal mercato.

Attualmente il governo risponde a circa 22 milioni di richieste all'anno, la maggior parte inoltrate da veterani di guerra e cittadini anziani che cercano informazioni riguardo il loro stato di servizio e i loro assegni di indennità. Gli Stati Uniti sono stati la terza nazione al mondo, dopo la Svezia con l'Atto sulla Libertà di Stampa del 1766 e la Finlandia nel 1951, a dotarsi di questo tipo di legge, ma la vera espansione si è registrata dopo il 1989 la società civile dell'Europa centrale e orientale ha reclamato questo diritto come parte del cambiamento durante le transizioni post-comuniste.

Il nostro paese è meno preparato in questo campo: non ci sono, purtroppo, *repository* di dati pubblici altrettanto ricche come il data.gov statunitense o il data.gov.uk britannico e le norme che garantiscono

⁵⁷ Il Foia si è adeguato alle evoluzioni della società: nel 1996 è stato emanato l'Electronic Freedom Of Information Act (E-Foia). Con questo atto gli enti governativi sono stati obbligati ad adeguarsi alle nuove tecnologie informatiche e a rendere disponibili elettronicamente i propri documenti.

l'accesso ai dati pubblici, come la legge n°241 del 7 agosto 1990⁵⁸ e successivi aggiornamenti, sono ancora molto restrittive in termini di diritto di accesso agli atti e trasparenza. A tal proposito Diritto Di Sapere⁵⁹ ha pubblicato un progetto sostenuto dalla Open Society Foundations⁶⁰ che mira a estendere il diritto di accesso all'informazione. LegalLeaks⁶¹ è una guida ricca di consigli pratici per cittadini e giornalisti che vogliono ottenere l'accesso all'informazione, un manuale di facile lettura che mira a fornire risposte alle domande più immediate dei non addetti ai lavori, ma comprende anche riferimenti alla legislazione e strumenti di approfondimento. Si affrontano questioni pratiche come quelle riguardanti i costi e i tempi della richiesta e si forniscono suggerimenti utili per integrare l'uso delle leggi per l'accesso nella propria raccolta di informazioni.

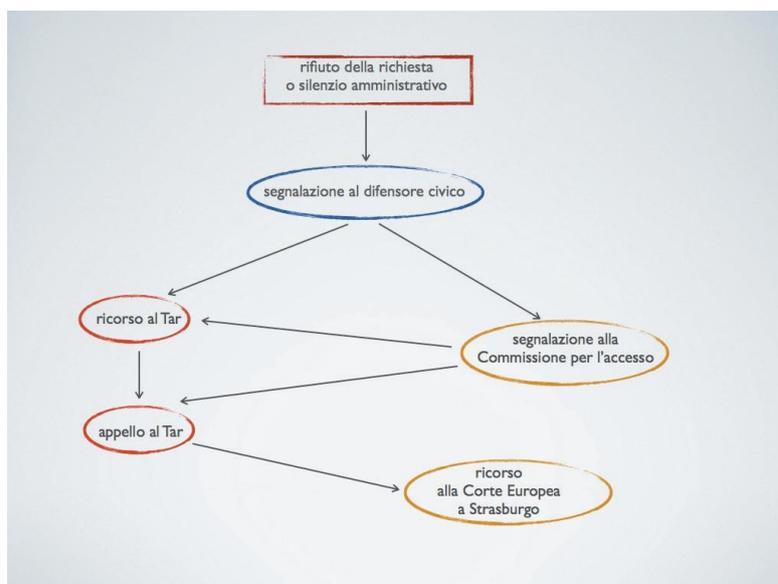


Figura 6 – I passi del ricorso in Italia

⁵⁸ Il diritto di accesso, nel nostro paese, è legato al possesso di una situazione legittimante. L'amministrazione, prima di rispondere positivamente alla richiesta, deve verificare ed informare eventuali controinteressati. La natura del diritto di accesso, legato a situazioni individuali, ma funzionale anche alla cura di interessi pubblici, si è attenuato in seguito alla riforma del 2005, che ha permesso ai cittadini di attuare un controllo democratico sull'attività dell'amministrazione.

⁵⁹ L'associazione, fondata nel 2012 senza scopo di lucro né discriminazione di carattere politico, di religione o di razza, aperta a uomini e donne, enti ed istituzioni private e pubbliche, ha lo scopo di promuovere l'accesso all'informazione, la trasparenza e l'integrità della pubblica amministrazione.

⁶⁰ Fondata nel 1979 dal filantropo George Soros, si impegna da allora per costruire una società vibrante e tollerante i cui governi siano responsabili e aperti alla partecipazione di tutti i cittadini.

⁶¹ Il manuale è stato pubblicato per la prima volta nel gennaio 2013, ma viene aggiornato ogni qual volta si registri una modifica alle legislazioni in materia di accesso all'informazione <<http://www.dirittodisapere.it/wp-content/uploads/2013/02/LEGALLEAKSaprile13.pdf>>.

L'ultimo aggiornamento⁶² è avvenuto nell'aprile 2013 con un'analisi del decreto legislativo 33/2013, anche conosciuto come Decreto trasparenza. Esso, infatti, fornisce una nuova definizione di trasparenza, intesa come “accessibilità totale delle informazioni concernenti l'organizzazione e l'attività delle pubbliche amministrazioni, allo scopo di favorire forme diffuse di controllo sul perseguimento delle funzioni istituzionali e sull'utilizzo delle risorse pubbliche”. Il sistema diventerebbe quindi totalmente accessibile per alcune informazioni, se non fosse per il paradosso che mantiene in vigore l'articolo 24 comma 3 della legge 241/90, che rifiuta il controllo diffuso per tutti gli altri atti. Sembra FOIA ma non è, poiché non introduce alcuna norma che possa definirlo tale.

Sempre nell'aprile 2013, Diritto Di Sapere in collaborazione con la ONG e Access-Info Europe, ha pubblicato *The Silent State*⁶³, il primo rapporto sullo stato dell'accesso all'informazione della Pubblica Amministrazione in Italia. Lo studio, iniziato nel gennaio 2013, raccoglie le risposte di circa 300 richieste di dati effettuate da 33 volontari fra cittadini e media. Solo il 13% delle richieste ha avuto un riscontro pienamente soddisfacente, mentre nel 73% dei casi le richieste sono state considerate non conformi agli standard internazionali del diritto all'accesso. In particolare, il 65% di queste sono state classificate come silenzio amministrativo, per non aver ricevuto una risposta dalla Pubblica Amministrazione entro i 30 giorni previsti dalla legge.

Infine, è necessario segnalare che, grazie alla contaminazione effettuata dal data-driven journalism, anche in Italia è partito l'appello per l'adozione di un FOIA italiano⁶⁴. L'iniziativa, lanciata sotto forma di petizione online nel 2012, chiede il riconoscimento del diritto di tutti di chiedere conto delle scelte e dei risultati del lavoro amministrativo, con una legge sul modello del Freedom of Information Act presente in moltissimi paesi democratici. Nel luglio 2013 la presidente della Camera Laura Boldrini ha incontrato i promotori dell'iniziativa e ha assicurato

⁶² Diritto Di Sapere, *LegalLeaks. Media e democrazia come usare il diritto di accesso all'informazione*, aprile 2013, pag. 42, 43.

⁶³ Originariamente realizzato in inglese, il rapporto “Silenzio di Stato” è stato tradotto in italiano nel giugno 2013 <<http://www.dirittodisapere.it/rapporto/>>.

⁶⁴ Tra i sessantadue promotori figurano anche Mario Tedeschini Lalli, Guido Romeo, Stefano Epifani, Andrea Fama e molti altri ancora <<http://www.foia.it/>>.

che seguirà con attenzione l'iter delle proposte che sul tema dovessero essere presentate alla Camera nei prossimi mesi. A oggi, purtroppo, ancora nessun miglioramento è stato registrato.

2.3.1 Big Data

La tecnologia si è sviluppata al punto da permetterci di ottenere informazioni basate su una marea di dati, con cui far crescere il proprio business, abbattere i tempi della burocrazia e migliorare i servizi, oppure ancora controllare consumi, servizi e molto altro ancora. Una volta entrati nel mondo dei dati aperti, quella dei Big Data è una tappa obbligata. Non esiste una definizione rigorosa dei Big Data, ma la maggior parte di esse concordano nel definirli come enormi volumi di dati provenienti da varie fonti che, se analizzati, forniscono informazioni dettagliate in tempo reale. Secondo i calcoli, si stima che ogni giorno vengano immagazzinati più di 2,5 trilioni di dati. I creatori di motori di ricerca sono stati i primi a scontrarsi con questa pesante realtà, il cui principale problema è rappresentato dall'impossibilità di raccogliere una tale quantità di dati con i metodi convenzionali. È così che il cambiamento quantitativo ne ha prodotto uno qualitativo: sono nate nuove tecnologie, come MapReduce di Google e il suo equivalente Open Source Apache Hadoop, prodotto da Yahoo. Questi strumenti consentono la gestione di una maggiore quantità di dati rispetto al passato e, soprattutto, grazie a essi i dati non devono essere messi in file o tabelle ordinate per essere lavorati⁶⁵. L'innovazione chiave di MapReduce è che permette una elaborazione distribuita delle operazioni di mappatura e riduzione. Per comprenderne il funzionamento basta scomporre il suo nome: nella fase "map" i dati di input vengono elaborati, voce per voce, e trasformate in un set di dati intermedio; nella fase "reduce", questi dati intermedi sono ridotti a un insieme di dati riassunti, che è il risultato finale desiderato⁶⁶. Essendo ogni operazione di mappaggio e di riduzione indipendente dalle altre, tutte le operazioni

⁶⁵ Viktor Mayer-Schönberger, Kenneth Cuker, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, Boston, New York 2013, pag. 5-8.

⁶⁶ O'Reilly Radar Team, *Big Data Now: Current Perspectives from O'Reilly Radar*, O'Reilly team, 2011, pag. 14, 15.

possono essere eseguite in parallelo risolvendo il problema della gestione di una quantità troppo grande di dati per un singolo computer. Il processo di esecuzione di un lavoro MapReduce con Hadoop comporta invece diverse operazioni che possono risultare più complesse e quindi rendere necessario il coinvolgimento di un esperto programmatore.

Come suggerisce Edd Dumbill, principale analista a O'Reilly Radar, un buon esempio di sviluppo del rapporto tra i Big Data e il web semantico⁶⁷ è Google. Nella sua fase matura, la missione di Google è quella di rendere i propri risultati sempre più utili per gli utenti. Per raggiungere questo obiettivo, il loro software usando dati ricavati da siti come schema.org, tentando di comprendere meglio il mondo e personalizzare ulteriormente l'esperienza di ricerca da parte dell'utente⁶⁸.

Il tema dei Big Data è ambiguo, poiché implica anche considerazioni di privacy per gli utenti. I Big Data, infatti, ci seguono ogni volta che compiamo una ricerca su Google, ogni volta che i *cookies*⁶⁹ immagazzinano informazioni sui siti che esploriamo su internet, ogni volta che compriamo un oggetto o un biglietto online, in ogni acquisto che facciamo con la carta di credito, in ogni momento in cui lo smartphone ci rende localizzabili e traccia i nostri spostamenti.

⁶⁷ Termine coniato da Tim Berners-Lee per indicare la trasformazione del World Wide Web in un ambiente dove i documenti pubblicati sono associati a informazioni e dati che ne specificano il contesto semantico in un formato adatto all'interrogazione e all'interpretazione (per esempio tramite motori di ricerca quali Google) e, più in generale, all'elaborazione automatica.

⁶⁸ O'Reilly Radar Team, *Big Data Now: Current Perspectives from O'Reilly Radar*, op. cit., pag. 57.

⁶⁹ Righe di testo usate per tracciare sessioni e memorizzare informazioni specifiche riguardanti gli utenti che accedono al server, come ad esempio siti web preferiti o, in caso di acquisti via internet, il contenuto dei loro "carrelli della spesa".

CAPITOLO TERZO

UNO SGUARDO GLOCALE

3.1 Il data journalism nel mondo

Così come accade per ogni cambiamento innovatore, il giornalismo dei dati è andato sviluppandosi con modi e ritmi diversi in luoghi differenti. Nella maggior parte dei casi si può parlare soprattutto del contributo fornito da attivisti e da singoli cittadini, ma sempre più governi stanno apportando le dovute modifiche per rendere i propri database un patrimonio pubblico in continuo accrescimento. Di seguito sono illustrati i casi più famosi in Stati Uniti, Francia, Spagna e Germania, lasciando per ultimo il caso dell'Inghilterra, a cui verrà dedicato per intero il quarto capitolo.

3.1.1 Stati Uniti

Negli Stati Uniti, così come in Gran Bretagna, la pubblica amministrazione ha aperto i propri database per rendere accessibili le proprie informazioni a tutti i cittadini. Data.gov è stata la prima iniziativa istituzionale al mondo in materia di Open Data, lanciata nel maggio 2009 con lo scopo di accrescere l'accesso pubblico alle banche dati aperte e ad alto valore aggiunto prodotte dal Ramo Esecutivo del Governo Federale USA. Il sito è in grado di fornire 379.831 banche dati, 922 applicazioni del governo e 236 applicazioni sviluppate dagli utenti e raccoglie i dati di 172 agenzie⁷⁰, affidando le sue prospettive di crescita alla partecipazione del cittadino, invitato a collaborare attraverso feedback o richieste.

⁷⁰ Andrea Fama, *Open data - data journalism. Trasparenza e informazione al servizio delle società nell'era digitale*, Simplicissimus Book Farm srl, 2011, pag. 57.

Alla mezzanotte del primo ottobre del 2013 la nazione della democrazia ha subito una dura battuta d'arresto: per paralizzare la riforma sanitaria *Obamacare*, approvata più di tre anni fa, i Repubblicani che detengono la maggioranza della Camera hanno dichiarato che non avrebbero votato l'innalzamento del tetto al bilancio se Obama non avesse ritardato di un anno l'entrata in vigore della riforma della discordia. E così è stato: lo *shutdown*, il blocco dell'intera pubblica amministrazione⁷¹, non solo ha messo a rischio centinaia di migliaia di posti di lavoro, ma ha anche colpito molti servizi offerti dal governo: uno fra tutti la chiusura del sito *data.gov* e di tutti i siti ospitati da *census.gov*, risorsa di inestimabile valore per chiunque volesse fare una richiesta di dati utile per la propria inchiesta.

Tra le testate giornalistiche ufficiali, il blog più aggiornato è sicuramente quello del *The New York Times*⁷², curato dagli sviluppatori della redazione online della testata. Ampi spazi sono dedicati a utili tutorial, ma vengono presentati anche progetti e condivise recensioni su libri ed eventi dedicati al data journalism.

ProPublica è una testata web non-profit indipendente e priva di un giornale cartaceo di riferimento, fondata nel 2007 da Paul Steiger⁷³ allo scopo di produrre dell'informazione che denunci lo sfruttamento dei più deboli e gli insuccessi dei potenti. La sua *mission* è seguire la tradizione del giornalismo come servizio pubblico per stimolare cambiamenti positivi. La piattaforma è finanziata esclusivamente dalle generose donazioni effettuate da alcuni filantropi e fondazioni. Tra questi, i miliardari californiani Herbert e Marion Sandler che, preoccupati per il futuro del giornalismo d'investigazione, hanno donato 30 milioni di dollari in tre anni per il progetto. I poco più di trenta collaboratori mirano

⁷¹ La notizia del mancato accordo sul bilancio è stata commentata da numerose testate italiane. Lo *shutdown* non si verificava da 17 anni. <http://www.corriere.it/esteri/13_ottobre_01/usa-shutdown-chiudono-i-monumenti_e56625c0-2aaf-11e3-b898-f13adc0c04f6.shtml>.

⁷² Il primo post del datablog Open risale al luglio 2007 <<http://open.blogs.nytimes.com/>>.

⁷³ Classe '42, è stato caporedattore del *The Wall Street Journal* dal 1991 fino al 15 maggio 2007. Steiger è laureato presso la *Hun School of Princeton* ed è stato membro del *Trumbull College* alla *Yale University*, dove è stato redattore della *Yale News and Review* e membro della *Manuscript Society*. Attualmente è Presidente Esecutivo di ProPublica.

a sviluppare storie di forte impatto civico per o con grandi marchi d'informazione come CNN, Usa Today, Huffington Post e New York Times, cedendo loro gratuitamente i contenuti prodotti: un modo di fare giornalismo che in Italia ancora non esiste⁷⁴. Secondo Sergio Maistrello, giornalista ed esperto in giornalismo online, i miliardari italiani non hanno ancora sviluppato una concezione dei media tale da portarli a donare dei soldi nell'interesse della società.

Nei suoi pochi anni di vita la piattaforma ha già raccolto quasi trenta premi Pulitzer. Il contributo più importante è stato dato grazie all'inchiesta "Dollars for Docs"⁷⁵, uno dei migliori esempi in circolazione di data journalism. Per molto tempo le compagnie farmaceutiche americane hanno mantenuto segreti i dettagli dei pagamenti erogati ai medici e agli altri operatori sanitari per promuovere i propri farmaci. Alcune case farmaceutiche avevano già iniziato a pubblicare i propri dati grazie a una sentenza dei tribunali americani, ma le informazioni erano male organizzate e quindi praticamente impossibili da consultare per i cittadini e per gli addetti ai lavori. I giornalisti Jeremy B. Merrill, Charles Ornstein, Tracy Weber, Sisi Wei e Dan Nguyen hanno deciso di analizzare i pagamenti di otto di quelle case farmaceutiche, creando così un database senza precedenti per un totale di 320 milioni di dollari destinati a 17mila medici americani per i farmaci prescritti nella loro area di competenza. Un archivio online continuamente aggiornato dove ogni cittadino può verificare se esista la possibilità che il proprio medico abbia o meno un conflitto di interessi.

⁷⁴ Sergio Maistrello, *Giornalismo e nuovi media. L'informazione al tempo del citizen journalism*, Apogeo, Milano 2010, pag. 180-182.

⁷⁵ Questa e tutte le altre inchieste realizzate dalla fondazione non-profit sono raccolte sul sito <<http://projects.propublica.org/docdollars/>>.

3.1.2 Francia

Il *journalisme de données* ha iniziato a svilupparsi in Francia soprattutto attraverso siti di notizie indipendenti come Mediapart⁷⁶, che ha creato nel settembre 2008 una mappa sulla crisi sociale. Google Maps permette agli utenti di verificare i tagli ai posti di lavoro verificatisi dopo il settembre 2008 e di collaborare attivamente inviando via e-mail le proprie informazioni che sono sistematicamente aggiunte alla mappa. In meno di un anno l'indagine ha ottenuto più di un milione e mezzo di visualizzazioni. Un altro importante contributo è stato dato da Rue 89⁷⁷, che nel 2010 ha lanciato una mappa che raccoglie le opinioni del dibattito sull'identità nazionale. Nell'aprile 2009 nasce OWNI⁷⁸, portale web indipendente specializzato in data journalism, visualizzazioni, indagini e inchieste attraverso i dati. Può essere considerato il pioniere del data journalism in Francia, grazie all'applicazione Spy Files⁷⁹, inventata per orientarsi fra le centinaia di migliaia di cablogrammi divulgati da Wikileaks. Nel febbraio del 2012 i>TELE, in collaborazione con il team di OWNI, ha dato vita a Le Véritomètre⁸⁰, un'applicazione di *fact checking*⁸¹ utile a verificare una ad una le affermazioni dei candidati alla presidenza del paese nel corso delle loro apparizioni televisive. L'impegno di OWNI lo ha reso un gioiello nel

⁷⁶ Mediapart è un giornale online d'inchiesta e di opinione francese creato nel 2008 dall'ex redattore capo di Le Monde Edwy Plenel. Il sito è diviso due sezioni principali: Le Journal, gestito da giornalisti professionisti e Le Club, un forum collaborativo curato dalla comunità. Nel 2011, ha lanciato Mediapart FrenchLeaks, ispirato da Wikileaks. Mediapart ha svolto un ruolo centrale nella rivelazione di diversi grandi scandali politici francesi, tra i quali la vicenda Bettencourt e il caso Cahuzac: nel 2012, Mediapart ha reso pubbliche delle registrazioni audio che hanno compromesso l'allora ministro del bilancio Jérôme Cahuzac.

⁷⁷ Rue89 è un sito francese d'informazione fondato nel 2007 da ex giornalisti del quotidiano Libération. Il nome Rue89 è stato scelto come riferimento alla libertà, attraverso la Rivoluzione francese (1789) e la caduta del muro di Berlino (1989), mentre il termine francese Rue, strada, rappresenta il simbolismo del luogo di incontro e discussione.

⁷⁸ Il nome del sito web di informazione francese è l'acronimo di *Objet Web Non Identifié*, il corrispondente francese di Ufo <<http://owni.fr/wp-signup.php?new=data>>.

⁷⁹ All'interno del sito è possibile scegliere se cercare tra i war logs di Iraq e Afghanistan e i documenti relativi al cablegate 2010 <<http://wikileaks.owni.fr/>>.

⁸⁰ Il successo dell'applicazione web ha spinto anche altri paesi a dare vita allo stesso progetto <<http://owni.fr/2012/02/16/veritometre-factchecking-presidentielle/>>.

⁸¹ Espressione inglese che significa verifica dei fatti. Il fact checking deriva dalla tradizione del giornalismo professionale: è un metodo empirico per dimostrare se dichiarazioni e fatti sono attendibili, attraverso la raccolta e il confronto di fonti.

panorama dell'informazione online, che purtroppo ha dovuto arrestare la sua corsa: un piano di ristrutturazione brutale ha comportato il licenziamento di tredici dei diciassette dipendenti, quasi tutti i giornalisti, e la trasformazione del sito in agenzia di servizi web più redditizi. L'amministratore delegato Eric Series ha così deciso di chiudere il sito nel dicembre 2012, dopo una situazione di tensione che durava ormai da diverso tempo. La causa principale, stando all'intervista rilasciata dalla storica firma di Owni Sylvain Lapoix a CaféBabel⁸², sarebbe stata la mancanza di un business model effettivo. Dal punto di vista economico è molto costoso fare del data-driven journalism, perché richiede numerose e rare.

Per quel che riguarda le testate giornalistiche ufficiali, Le Figaro dedica un'intera sezione della sua redazione online alle infografiche⁸³, mentre Le Monde ha aperto il blog "J'ai du bon data" che viene frequentemente aggiornato con post e recensioni riguardanti articoli, eventi o libri sul giornalismo dei dati.

3.1.3 Spagna

In Spagna il *periodismo de datos* si diffonde soprattutto grazie all'importante contributo offerto dal docente della Universidad Complutense de Madrid José Luis Dader, che traduce in spagnolo l'opera di Meyer, Precision Journalism, e promuove diversi corsi universitari sul giornalismo dei dati. Secondo Dader il giornalismo di precisione rappresenta la naturale evoluzione di quelli che erano i limiti del giornalismo d'inchiesta che, nella maggior parte dei casi, si limitava a seguire le proprie fonti fidate senza preoccuparsi di verificare i dati. Secondo il docente è quindi necessario investire non solo in tecnologia, ma anche nella formazione del personale in modo che l'accesso alle banche dati elettroniche e ai

⁸² Il sito internet cafébabel e i relativi progetti editoriali sono co-finanziati grazie al supporto della Commissione Europea e di altri partner <<http://www.cafebabel.fr/style-de-vie/article/owni-soit-il-faire-du-datajournalisme-nest-pas-indispensable.html>>.

⁸³ La sezione si dedica ad un'ampia varietà di temi, dalla politica allo sport alla cultura <<http://plus.lefigaro.fr/tag/infographie>>.

servizi di informazione on-line diventi una abitudine per confrontare e analizzare le informazioni⁸⁴.

In occasione nella prima edizione dei Data Journalism Awards, tenutasi nel 2012, il quotidiano El Mundo si classifica tra i finalisti con “operación SAGA” un’inchiesta iniziata nel 2011 dalla Guardia Civile che ha arrestato Eduardo Bautista e José Luis Rodríguez Neri, rispettivamente presidente e direttore della SGAE⁸⁵, per la presunta sottrazione di fondi alle imprese private appartenenti ai membri del Consiglio. Il grafico creato dai giornalisti de El Mundo descrive tutta la rete di relazioni e il flusso di denaro e fatture che sono state trasferite tra le diverse società per evitare le tasse.

Ispirato al The Guardian Datablog, al NYTimes Open Blog e a ProPublica, LA NACION data è il blog dell’omonima testata spagnola, curato da sei donne e un uomo, che si propone di raccogliere e condividere contenuti, visualizzazioni e inchieste supportate dai dati. Il giornalismo dei dati, scrive la redazione spagnola⁸⁶, è praticato da anni ormai, ma il nuovo impulso è dato dalla tecnologia digitale in grado di analizzare e interagire con i dati in modo molto più accessibile e collaborativo. È proprio sulla base di questo principio che i lettori sono invitati a partecipare, collaborando attraverso i commenti o inviando via e-mail o attraverso i canali social i dati mancanti o le proprie richieste.

⁸⁴ José Luis Dader, *Periodismo de precisión: la vía socioinformática de descubrir noticias*, Ed. Síntesis, Madrid 2002, pag. 56.

⁸⁵ La Sociedad General de Autores y Editores in italiano Società generale degli autori e degli editori, è il corrispettivo spagnolo della SIAE: un ente pubblico economico preposto alla protezione e all’esercizio dei diritti d’autore <<http://www.elmundo.es/elmundo/2011/graficos/sep/s3/sgae.html>>.

⁸⁶ “I dati servono per il servizio, il controllo e l’innovazione. Anche per la trasparenza, l’apprendimento e la conoscenza. Gli diamo vita se li convertiamo in informazioni.” <<http://blogs.lanacion.com.ar/data/about/>>.

3.1.4 Germania

La figura di maggiore spicco nel settore del *datenjournalismus* in Germania è sicuramente Mirko Lorenz, giornalista, architetto dell'informazione e formatore di giornalisti specializzati attraverso lezioni sul giornalismo dei dati tenute in tutto il mondo. Ha contribuito al lancio di Datawrapper nel febbraio 2012, uno strumento online per la creazione di grafici e altre visualizzazioni dati. Attraverso la sua opera di formazione sta tentando di riunire squadre di giornalisti specializzati in tutta Europa, radunati sotto il nome di Journalism++⁸⁷, per scatenare il potere della comprensione dei dati. È stato proprio lui, durante una presentazione⁸⁸ tenutasi nell'agosto del 2010 in occasione del Data-driven Journalism Roundtable organizzato dall'European Journalism Centre, a definire il processo del data journalism come un flusso di lavoro in cui i dati rappresentano la base per l'analisi, la visualizzazione e la narrazione della storia.

Lo Zeit Online, piattaforma web del settimanale tedesco Die Zeit⁸⁹, crede fortemente nei progetti di data journalism, potenti strumenti per portare traffico al sito coinvolgendo il pubblico in un modo innovativo. Un progetto è diventato virale in occasione del disastro della centrale nucleare di Fukushima quando lo Tsunami ha colpito il Giappone. Il team dello Zeit Online ha creato una mappa⁹⁰ che mostrava quante persone, residenti nei pressi di una centrale nucleare in Germania, sarebbero state coinvolte in un'evacuazione in caso di un'emergenza simile. Wolfgang Blau, capo redattore dello Zeit Online per due anni, ha sempre ritenuto

⁸⁷ Il manifesto di Journalism++ riporta i dieci punti che rappresentano le persone che ne fanno parte: giornalisti, designers, project managers e sviluppatori che credono fermamente che l'“Open” sia meglio del “Closed” e che i dati siano fondamentali per comprendere la società moderna <<http://jplusplus.org/>>.

⁸⁸ La presentazione non pretende di trovare tutte le risposte al data-driven journalism, ma mira anche a creare domande e dibattiti sull'argomento <<http://www.slideshare.net/mirkolorenz/data-driven-adam>>.

⁸⁹ Fondato ad Amburgo nel 1946 dal 2004 è diretto dall'italo-tedesco Giovanni di Lorenzo.

⁹⁰ Progetti come questo sono facilmente traducibili in altre lingue. Una versione inglese è stata adattata alla prossimità delle persone alle centrali nucleari in America <<http://opendata.zeit.de/atomreaktoren/#/en/>>.

la trasparenza, il coinvolgimento degli utenti e la conseguente la credibilità della testata aspetti fondamentali della loro filosofia.

Il progetto PISA based Wealth Comparison⁹¹ è una visualizzazione interattiva che consente di confrontare il tenore di vita tedesco con diversi paesi membri dell'OCSE⁹² per fornire un modello di confronto unificato. Gli indicatori trasformati in simboli intuitivi includono la ricchezza (numero di televisori posseduti, auto e stanze da bagno), la situazione familiare (la percentuale di famiglie con un solo figlio, la disoccupazione dei genitori e l'impiego della madre); l'accesso alle fonti della conoscenza (internet a casa, frequenza di utilizzo di e-mail e la quantità di libri posseduti); tre indicatori supplementari sul livello di sviluppo di ciascun paese. Gregor Aisch ha infine dato vita a una visualizzazione interattiva dal design accattivante. Il risultato della collaborazione è stato un grafico interattivo intuitivo e utile che ha suscitato l'interesse di una grande quantità di lettori⁹³.

3.2 Una parentesi italiana

Come spesso accade in più occasioni e ambiti, anche in questo caso il nostro paese ha cominciato a interessarsi di data journalism e Open Data in ritardo rispetto agli altri paesi del mondo. Il concetto ha cominciato a fare presa nella comunità italiana intorno al 2010, quando, durante il Festival del Giornalismo di Perugia, si è tenuto il primo panel sul giornalismo di precisione. Rispetto al resto del mondo il giornalismo dei dati è quindi ancora alla sua fase più

⁹¹ I dati sono tratti dal PISA 2009, il rapporto dell'OECD sulla classifica mondiale dell'istruzione, pubblicato nel dicembre 2010. Il rapporto si basa su un questionario che pone a studenti di quindici anni domande sulla loro condizione di vita domestica. <<http://opendata.zeit.de/pisa-wohlstands-vergleich/visualisierung.php#/en/DEU-OECD>>.

⁹² L'Organizzazione per la cooperazione e lo sviluppo economico è un'organizzazione internazionale composta da 34 paesi membri accomunati da un governo di tipo democratico e un'economia di mercato. L'organizzazione crea occasioni di confronto per stimolare il progresso economico e il commercio mondiale dei paesi membri.

⁹³ Jonathan Gray, Lucy Chambers, Liliana Bounegru, *The Data Journalism Handbook*, op. cit., pag. 45, 46.

giovane⁹⁴. Le inchieste più rilevanti sono spesso state eseguite da piccole organizzazioni indipendenti, da attivisti o da una esigua manciata di giornalisti. È però importante sottolineare che le abilità legate a questo tipo di giornalismo non sono ancora parte del curriculum delle scuole italiane di giornalismo, né sono richieste tra i requisiti di ammissione. Inoltre, in Italia non abbiamo ancora un Freedom of Information Act come quello in vigore negli Stati Uniti dal 1966 e, per questo, i giornalisti sono spesso costretti a dover fronteggiare una cultura della segretezza e la scarsa collaborazione quando si avvicinano alle istituzioni pubbliche per richiedere i dati necessari.

Nonostante ciò il nostro paese ha iniziato a offrire alcuni workshops, scuole e corsi, di breve e lunga durata, dedicati al data journalism, mentre diverse testate giornalistiche hanno deciso di concedere una possibilità a questo nuovo tipo di giornalismo, dedicando uno spazio non ufficiale delle loro redazioni online alle notizie riguardanti il giornalismo dei dati.

3.2.1 Spaghetti Opendata

Spaghetti Opendata⁹⁵, nome che racchiude in sé l'emblema della nostra nazione, è una comunità di volontari completamente autofinanziata, nata nel settembre 2010 da una proposta⁹⁶ di Alberto Cottica⁹⁷ che ha suscitato un certo interesse in Rete. La mailing list è il punto di ritrovo di questi cittadini interessati al rilascio di dati pubblici in formato aperto, in modo da renderne

⁹⁴ Articolo di Guido Romeo sul perché l'Italia è il paese delle opportunità per il data journalism <http://datadrivenjournalism.net/news_and_analysis/why_italy_is_the_land_of_opportunity_for_data_journalism>.

⁹⁵ Il sito della comunità è <<http://www.spaghetiopendata.org/it>>, mentre è possibile iscriversi alla mailing list attraverso il seguente link <<https://groups.google.com/d/forum/spaghetiopendata>>.

⁹⁶ A seguito di un insolito aperitivo, Cottica ha deciso di lanciare pubblicamente l'idea di raccogliere e condividere dati pubblici e aperti per permettere, a chi ne fosse in grado, di lavorarli <<http://www.cottica.net/2010/09/16/spagheti-open-data-reloaded-italiano/>>.

⁹⁷ Economista, esperto di politiche pubbliche collaborative e partecipazione online. Lavora al Consiglio d'Europa e all'Università di Alicante. Si impegna per rendere l'azione di governo più aperta e intelligente, utilizzando Internet per attingere all'intelligenza collettiva dei cittadini. Appassionato di matematica delle reti per imparare a progettare dinamiche sociali emergenti.

facile l'accesso e il riuso. Tra loro ci sono funzionari pubblici, sviluppatori, giornalisti dei dati di professione, ma anche semplici curiosi che si sono appassionati al tema. Lo scopo di SOD è avere un punto di accesso agli Open Data italiani, ai dati pubblici da trasformare in Open Data e proporre una modalità di approccio ai futuri siti di Open Data, riutilizzabili a costo zero da chiunque. Non intendono sostituirsi alle autorità pubbliche, ma in attesa del lancio di data.gov.it, Spaghetti Opendata vuole essere una soluzione ponte, utile nel tempo necessario allo Stato italiano per lanciare una propria iniziativa.

L'approccio proposto dalla comunità è quello di massima apertura a tutti i livelli. I dati raccolti sono inseriti o in fase di inserimento nel CKAN italiano⁹⁸; l'interfaccia e il template HTML sono in formato aperto e facilmente esportabili in vari formati. Uno degli scopi fondamentali di SOD è quello di creare confronto e consapevolezza sul tema degli Open Data e sulla loro riusabilità. Per essere aperto, un dato grezzo deve essere tecnologicamente e legalmente abilitato a essere riusato, diventando così compatibile col concetto di Open Data⁹⁹. Ciò significa che deve essere condiviso con una licenza che ne abiliti tale riuso e che sia in un formato aperto non proprietario per garantire la massima libertà. Per mostrare il passaggio tecnico verso il *Web of Data*, è stata inserita la legenda *5 star deployment scheme for Open Data*¹⁰⁰ ideato da Tim Berners-Lee. Per quelli di Spaghetti Opendata è importante incentivare il riuso non solo dei dati, ma anche delle tecnologie. Dietro la struttura c'è il linguaggio HTML e Simile Exhibit, un framework molto potente nato nel dicembre 2006 che

⁹⁸ CKAN Italia è un catalogo gestito dalla community, che contiene dataset utili reperibili in Internet. A seconda del tipo di dati (e delle loro condizioni d'uso), CKAN Italia può anche salvarne una copia o archivarli in un database e mostrare semplici visualizzazioni. Il sito è basato su un potente software open-source di catalogazione dei dati, chiamato CKAN, sviluppato dalla Open Knowledge Foundation. Maggior informazioni qui <<http://it.ckan.net/about>>.

⁹⁹ Andrea Fama, *Open data - data journalism. Trasparenza e informazione al servizio delle società nell'era digitale*, op. cit., pag. 72.

¹⁰⁰ Il processo permette di suddividere i dati in cinque categorie, ognuna corrispondente al numero di stelle assegnate. Una stella equivale a mettere la risorsa in Rete in qualunque formato scaricabile con una licenza di riuso. Tre stelle è il formato degli Open Data, permette di usare formati non proprietari (per esempio CSV al posto di Excel). Cinque stelle equivale agli Open Linked Data, che permette di collegare i propri dati ad altri dati per fornire un contesto più ampio. Tutto il processo dei dati collegati è accuratamente spiegato su <<http://5stardata.info/>>.

permette di caricare i dati da fonti diverse sparse per la Rete e che utilizza delle modalità interattive semplici per filtrare e navigare i dati. Non richiede infatti particolari strumenti o competenze: tutto avviene nel proprio browser e si carica al primo aprirsi della pagina. I dati di Spaghetti Opendata si caricano da alcuni fogli elettronici editati a più mani inseriti in Google Docs.

3.2.2 Fondazione <ahref e la Data Journalism School

La Fondazione <ahref sostiene dal 2010 iData¹⁰¹, un progetto che mira a sviluppare la prima piattaforma italiana Open Source e non-profit su e per i civic media con una particolare attenzione al mondo dei dati e al data journalism¹⁰². La piattaforma sarà collegata a un ventaglio di community che potranno collaborare per la raccolta, la produzione e l'elaborazione dei dati, con particolare attenzione ai dati legati all'ambiente, alla salute e allo sviluppo economico. I progressi sono poi riportati nel blog Open Data della Fondazione <ahref.

Sulla scia dell'applicazione francese Le Véritomètre e ispirato dagli esempi di successo di Newstrust e Politifact sviluppati negli Stati Uniti, è nato Factchecking.it¹⁰³, una piattaforma media civico che permette a tutti i cittadini di contribuire alla verifica diretta delle dichiarazioni di personaggi pubblici, delle notizie pubblicate sulla carta stampata e trasmesse via radio e tv e di qualsiasi informazione o documento pubblico. La Fondazione è convinta che i quattro principi – accuratezza, imparzialità, indipendenza, legalità – proposti da <ahref possano essere una valida linea guida nel processo di verifica dell'informazione. Dal sito è infatti possibile creare il proprio *fact checking* riportando la frase che si crede errata o dubbia, citandone fonti e autore. Una volta

¹⁰¹ La descrizione dettagliata del progetto è presente sul sito della fondazione <ahref <<http://www.ahref.eu/it/ricerca/progetti/idata>>.

¹⁰² Andrea Fama, *Open data - data journalism. Trasparenza e informazione al servizio delle società nell'era digitale*, op. cit., pag. 65.

¹⁰³ Progetto presentato in anteprima il 28 aprile 2012 durante il Festival internazionale del giornalismo di Perugia <<https://factchecking.civiclincs.it/it/>>.

pubblicato starà alla comunità iniziare il lavoro di verifica collaborativa cercando altre fonti che smentiscano o confermino la notizia.

Inoltre, grazie alla collaborazione con Istat, la Fondazione <ahref è arrivata a lanciare la terza edizione della Data Journalism School¹⁰⁴, un corso di tre giorni fatto di workshops e lezioni teoriche tenute da docenti del calibro di Elisabetta Tola e Guido Romeo. La Data Journalism School è aperta a studenti, giornalisti e operatori del mondo dell'informazione che vogliono muovere i primi passi o perfezionarsi nel giornalismo dei dati.

3.2.3 Datajournalismitaly

Datajournalismitaly¹⁰⁵ è la mailing list italiana che, proprio come Spaghetti Opendata, si interessa di data journalism. La piccola, ma crescente, comunità di interessati al giornalismo dei dati conta ormai più di trecento membri. Nata nel 2011 dall'incoraggiamento di Maurizio Napolitano¹⁰⁶, ambasciatore italiano della Open Knowledge Foundation, la lista vuole essere un punto di riferimento per cittadini e giornalisti (intesi come chiunque si preoccupi di trovare e verificare una notizia) fornendo esempi di data journalism in Italia e all'estero, parlando di strumenti di analisi e visualizzazione, fonti di dati e modi per ottenerli quando non sono immediatamente disponibili e dare appuntamenti sul tema e occasioni di formazione. Da marzo 2013 la comunità ha anche aperto il proprio sito¹⁰⁷ sotto la guida di Elisabetta Tola, direttrice responsabile, Marco Boscolo, giornalista freelance e Marco Montanari, sviluppatore. Insieme a un team di

¹⁰⁴ Il programma completo e l'elenco dei docenti <<http://www.ahref.eu/it/ricerca/educazione>>.

¹⁰⁵ Si può accedere alla mailing list dal link <<http://groups.google.com/group/datajournalismitaly>>.

¹⁰⁶ Tecnologo della FBK – Fondazione Bruno Kessler, si interessa di software libero/open source, social media, sistemi informativi geografici e Open Data. Da maggio 2011 è uno degli ambasciatori di Open Knowledge Foundation.

¹⁰⁷ Il sito nasce grazie a un consiglio dato dai pionieri del data-driven journalism come Simon Rogers, Steve Doig e Sara Cohen: “Start simple, start small”: non bisogna avere enormi risorse, strumenti complicati e *datasheet* infiniti. L'essenziale è trovare una storia importante e raccontarla <<http://www.datajournalism.it/>>.

collaboratori, lo scopo è focalizzarsi essenzialmente su inchieste data-driven nate da un database, da piccole redazioni o da blog autoprodotti, ma anche direttamente da loro stessi. Il sito raccoglie anche i migliori esempi di data-driven journalism in Italia e nel mondo e una sezione dedicata agli strumenti, ai tutorial, ai software e alle risorse già in circolazione per facilitare il lavoro del data journalist.

3.3 I data blog nel mondo e l'esperienza collettiva di DataJCrew

Negli ultimi tempi sempre più testate giornalistiche, ma anche professionisti indipendenti e semplici cittadini, hanno deciso di aprire un blog in cui poter parlare di data journalism e Open Data.

Osservando il panorama mondiale, un contributo fondamentale è offerto da DataDrivenJournalism¹⁰⁸, un centro di notizie e risorse offerte da una comunità di giornalisti, editori, designer e sviluppatori che utilizzano i dati al servizio del giornalismo. Attraverso corsi di formazione, conferenze e risorse online e l'incoraggiamento della collaborazione con la comunità di professionisti, l'obiettivo è sensibilizzare quanti più giornalisti, specialisti e sviluppatori a fare un uso migliore dei dati integrandoli ulteriormente nel loro lavoro. Sul sito sono raccolte notizie sui più interessanti progetti di data-driven journalism da tutto il mondo, ma anche informazioni sugli ultimi strumenti utili a lavorare con i dati in modo efficace e segnalazioni sugli eventi e le opportunità di lavoro. Il team annovera tra i suoi componenti anche Liliana Bounegru, uno degli autori del Data Journalism Handbook, la bibbia del giornalismo dei dati nata durante un workshop di due giorni al MozFest 2011 di Londra. Il manuale raccoglie il contributo di tantissime testate giornalistiche ed è stato prodotto in collaborazione con la Open Knowledge Foundation e pubblicato da O'Reilly Media.

¹⁰⁸ Il programma, nato nel 2010, fa parte della Data Driven Journalism Initiative dell'European Journalism Centre, uno dei principali enti promotori del data-driven journalism nel mondo <<http://ejc.net/projects/ongoing/data-driven-journalism>>.

La Open Knowledge Foundation¹⁰⁹ è una organizzazione non-profit fondata nel 2004 e dedicata alla promozione di dati e contenuti aperti. Il lavoro dell'OKF comprende alcuni tra i progetti più innovativi e di successo, come il progetto CKAN usato, tra i tanti, anche da data.gov e da data.gov.uk. Nel 2011 la Fondazione ha dato vita all'Open Data Challenge, la più grande competizione d'Europa sugli Open Data che ha registrato 430 iscrizioni da 24 Stati. La loro visione è basata sull'importanza della conoscenza aperta, intendendo con ciò qualsiasi contenuto, informazione o dato libero di essere usato, riutilizzato e ridistribuito senza alcuna restrizione legale, tecnologica o sociale. I principi fondamentali sono quindi l'accesso gratuito e aperto al materiale e la libertà di ridistribuirlo e riutilizzarlo. È con la conoscenza aperta che i dati diventano utili, usabili e potenti, ovvero in grado di aiutare qualcuno a risolvere un problema reale.

Il giornalista Paul Bradshaw¹¹⁰ gestisce l'Online Journalism Blog, uno spazio personale in cui sono raccolti commenti e analisi su diversi argomenti, tra cui figura anche il data-driven journalism. I lettori possono contribuire attivamente attraverso i commenti e gli spunti di riflessione lanciati da Bradshaw sono spesso estrapolati da ebook scritti da lui stesso, come *Scraping for Journalists* e *Data Journalism Heist*.

DataKind¹¹¹ è una comunità indipendente di pionieri e scienziati dei dati, fondata da Jake Porway¹¹² nella speranza di creare un mondo in cui ogni organizzazione sociale abbia accesso ai dati per meglio servire l'umanità. Tra i servizi offerti vi sono i DataDives, appuntamenti fine settimanali mirati ad aiutare organizzazioni non-profit a risolvere i loro problemi legati ai Big Data, come nel caso

¹⁰⁹ La sede italiana dell'OKF è gestita da Stefano Costa, Maurizio Napolitano e Francesca De Chiara <<http://okfn.org/>>.

¹¹⁰ Bradshaw lavora anche alla Birmingham City University e alla Scuola di Giornalismo della City University di Londra. È anche il cofondatore di Help Me Investigate, un sito di giornalismo investigativo finanziato da Channel 4 e Screen WM. Ha scritto per il journalism.co.uk, il Guardian Datablog e il Nieman Reports. È inoltre l'autore dell'Online Journalism Handbook <<http://onlinejournalismblog.com/>>.

¹¹¹ Si può sostenere il progetto attraverso donazioni o collaborando, compilando l'apposito form per mettersi direttamente in contatto con il team <<http://www.datakind.org/>>.

¹¹² Recentemente ha lavorato al New York Times Company Research & Development Lab come specialista di dati. Ha conseguito una laurea in informatica presso la Columbia University per poi specializzarsi in statistica presso la UCLA.

della lotta alla povertà e alla frode in Kenya affrontata dalla World Bank. Altri interessanti progetti sono DataCorps, un team specializzato nell'uso dei dati al servizio dell'umanità tramite collaborazioni con associazioni governative e fondazioni, e DataChapters, le sezioni di DataKind in giro per il globo.

Alcuni importanti esempi italiani sono il Datablog¹¹³ de La Stampa o l'OpenDataBlog¹¹⁴ de Il Sole24Ore, incentrato per lo più su visualizzazioni e infografiche di carattere economico che si rifanno agli articoli del quotidiano. Il blog di Dataninja¹¹⁵ è invece curato dai giornalisti Alessio Cimarelli e Andrea Nelson Mauro che si dedicano a tutte le diverse fasi del data journalism: dallo scraping e dall'analisi dei dati alla visualizzazione della storia.

A Wired Italia sono stati realizzati due progetti di data journalism molto importanti a livello civico: #scuolesicure¹¹⁶ è una mappa interattiva sulla valutazione della sicurezza sismica degli edifici scolastici, mentre l'indagine #doveticuri¹¹⁷ ha portato alla creazione della prima mappa interattiva sui migliori – e peggiori – ospedali d'Italia. Un resoconto dettagliato del pericolo di morte in cui incorrono i pazienti, ottenuto con i dati raccolti dall'Agenzia nazionale per i servizi sanitari regionali (Agenas) che monitora ogni anno tutte le attività degli ospedali italiani. Dati di grande valore anche dal punto di vista mondiale, poiché vengono utilizzati per studiare apposite politiche sanitarie e ottimizzazioni delle spese della sanità, sempre più considerevoli per le casse delle nostre regioni.

¹¹³ Online dal 2001, il blog è curato da diversi giornalisti <<http://www.lastampa.it/Blogs/datablog>>.

¹¹⁴ Attivo dal 2011 è stato uno dei primi blog gestiti da una testata giornalistica italiana <<http://opendatablog.ilsole24ore.com/#axzz2gUKgwtbI>>.

¹¹⁵ Il blog collabora anche con testate editoriali e con la pubblica amministrazione <<http://www.dataninja.it/blog/>>.

¹¹⁶ Per Wired non è stato affatto semplice ottenere questi dati, che risultano essere ancora scarsi. La speranza è che, con un progetto aperto alla collaborazione di chiunque, la mappatura delle scuole si concluda entro i termini previsti, a differenza di quanto accadrebbe se si dovessero attendere i dati forniti dalle amministrazioni <<http://daily.wired.it/news/politica/2012/09/17/mappa-sicurezza-simica-scuole-173456.html>>.

¹¹⁷ Sono stati presi in considerazione 19 indicatori, tra cui i tipi più comuni di infarti, ictus e tumori, e si fa un confronto diretto tra la percentuale dei decessi della struttura ospedaliera, la media nazionale e il numero dei ricoverati per la patologia <<http://daily.wired.it/news/scienza/2013/04/10/mappa-ospedali-italiani-7237824.html>>.

Nel maggio del 2012 un gruppo eterogeneo composto da studenti, neolaureati e lavoratori dislocati fra la Sicilia e il resto del mondo, ma con in comune l'interesse per l'informazione, il giornalismo innovativo e la comunicazione digitale, ha dato vita all'esperienza collettiva del blog DataJCrew. Nato da un'idea della giornalista Rosa Maria Di Natale, l'obiettivo principale dell'esperimento è quello di fungere da ritrovo per tutti gli utenti interessati alle inchieste basate sul giornalismo dei dati. Privo di alcuna sede fisica o supporto cartaceo, Data Journalism Crew esiste grazie a un gruppo ospitato da Yahoo! Gruppi e a regolari *hangouts* lanciati da Google Plus, strumenti gratuiti offerti dalla Rete, con i quali è semplice e istantaneo comunicare. Il team pubblica quotidianamente articoli estrapolati dai numerosi siti web e blog monitorati: tra i tanti figurano il Guardian Datastore, LSDI, l'OJB di Paul Bradshaw, ProPublica, datadrivenjournalism.net e la lista prosegue fino a comprendere i principali canali che si occupano di data-driven journalism in Italia, Francia, Inghilterra, Stati Uniti, America Latina e Spagna. Il blog, totalmente indipendente e autofinanziato, ha anche autoprodotta il suo primo esperimento data grazie al prezioso lavoro di Daniele Palumbo, che ha raccolto tutti i corsi di data-driven journalism nel mondo in una mappa¹¹⁸, regolarmente aggiornata e tradotta in inglese, francese e spagnolo. L'esperimento è stato realizzato raccogliendo i dati su Excel, che sono poi stati uniti ai dati geografici tramite Google Fusion e infine visualizzati con Google Maps.

A meno di un anno dalla pubblicazione, il blog si è distinto per essere riuscito a offrire, usando la sperimentazione sui dati e una selezione accurata delle informazioni, uno spaccato del giornalismo mondiale suggerendo all'utente ciò che c'è di meglio sui diversi argomenti di attualità e cultura grazie a un attento lavoro di promozione social.

¹¹⁸ Seguendo uno dei principi fondamentali del data-driven journalism, la mappa è un esperimento *in progress*, ovvero continuamente aggiornata grazie alle segnalazioni inviate dagli utenti <<http://datajcrew.sudmediatika.it/?p=1209>>.

CAPITOLO QUARTO

IL CASO INGLESE

Nel 1921 in occasione del centenario del The Guardian, il quotidiano più letto del Regno Unito, allora conosciuto come Manchester Guardian, il noto editore Charles Prestwich Scott¹¹⁹ pubblicò un importante saggio in cui si leggeva: “*Comment is free, but facts are sacred*”¹²⁰. Riflettendo sul ruolo del giornale, egli riteneva infatti che il suo scopo primario fosse quello di segnalare delle notizie precise. Novant’anni dopo, pubblicare quei fatti sacri è diventato un nuovo modo di fare giornalismo che ha preso il nome di data-driven journalism. La redazione del quotidiano inglese ha per prima riconosciuto il grande potenziale che si cela dietro questo innovativo modo di riportare i fatti.

È il 5 maggio 1821 quando il giornale tenta il suo primo esperimento¹²¹ di data-driven journalism, pubblicando sul quotidiano un articolo a corredo di una tabella che elencava tutte le scuole di Manchester e Salford, le loro rette e il numero degli alunni.

¹¹⁹ C. P. Scott fu un giornalista britannico, editore e politico. Ha diretto il Manchester Guardian, oggi conosciuto come Guardian, dal 1872 fino al 1929 e ne è stato proprietario dal 1907 fino alla sua morte, sopraggiunta nel 1932. È ricordato inoltre per aver perseguito un programma liberale progressista nelle pagine del giornale.

¹²⁰ “Il commento è libero, ma i fatti sono sacri”. Il saggio “A Hundred Years” fu pubblicato il 5 maggio 1921 <<http://www.theguardian.com/commentisfree/2002/nov/29/1>>.

¹²¹ L’inchiesta nacque perché all’epoca le statistiche ufficiali della città venivano raccolte da appena quattro sacerdoti, un compito impossibile che portava spesso dati inesatti e difettosi. Simon Rogers, *Facts are sacred, the power of data*, 2013, pag. 41-44.

numero reale degli alunni che usufruivano dell'istruzione gratuita supera di molto gli 8.000 bambini riportati nelle stime ufficiali: in realtà si trattava di circa 25.000 soggetti. Senza conoscere lo stato effettivo delle cose, il sistema di istruzione inglese non avrebbe mai potuto migliorare.

Dove si trova la differenza tra allora e il presente? Prima i dati erano pubblicati in supporti cartacei la cui stampa era molto costosa, in cui si faceva riferimento ai grafici come "figure"; adesso parliamo di spreadsheet e di formati file elaborati sui computer. Gli strumenti a nostra disposizione per analizzare i dati possono essere cambiati, ma la motivazione per cui lo facciamo è rimasta esattamente la stessa: trasformare lo stato delle cose per vivere in una società migliore.

4.1 Il Guardian Data

La redazione online del Guardian è separata da quella del cartaceo, ma insieme condividono la stessa location: il trasferimento nella nuova sede è avvenuto alla fine del 2008, rafforzando ulteriormente l'integrazione tra la redazione del cartaceo, quella del sito web e quella dell'inserito domenicale The Observer. Pur avendo una propria piattaforma di riferimento, questi tre gruppi condividono spazi, sistema editoriale e suddivisione tematica¹²². In effetti il Guardian è una delle testate tradizionali che meglio ha saputo sfruttare le opportunità offerte dal web: già nel 2006 Alan Rusbridger, direttore del Guardian dal 1953, ha lanciato il motto "*Web first*"¹²³ per sottolineare come i giornalisti debbano scrivere prima per il web, eccezionale mezzo di pubblicazione e distribuzione per raggiungere i propri utenti.

¹²² Sergio Maistrello, *Giornalismo e nuovi media. L'informazione al tempo del citizen journalism*, op. cit., pag. 113, 114.

¹²³ "Il Web prima di tutto" la decisione è stata condivisa con un articolo del 7 giugno 2006 <<http://www.theguardian.com/media/2006/jun/07/theguardian.pressandpublishing>>.

L'anno successivo al trasferimento è stato inaugurato il Guardian Data¹²⁴, quando nel marzo del 2009 il suo primo data editor Simon Rogers ha spiegato con un post all'interno del suo blog le motivazioni che hanno spinto il Guardian ad aprire una sezione dedicata al data-driven journalism. Quante volte le testate hanno lavorato a una serie di dati che, una volta superata la fase di pubblicazione, sono stati depositati all'interno di un hard disk mai più riconsultato? Questa è la motivazione principale per cui il blog è stato aperto: mantenere i dati aggiornati e aperti. Fondamentale è la condivisione e il coinvolgimento degli utenti: il Datablog viene definito come un luogo in cui discutere apertamente e decidere cosa fare dei dati messi a disposizione sul sito. A tale scopo i dati usati per gli articoli, raccolti in fogli di calcolo, sono caricati su Google Docs: ciò facilita di molto la possibilità di aggiornare rapidamente i dataset, offrendo ai lettori la possibilità di scaricarli con le statistiche più recenti.

L'obiettivo del Datastore è proprio quello di rendere i dati più accessibili alle persone. Trovarli e scaricarli è un processo veramente semplice e veloce: all'interno di ogni articolo si trova un link al foglio di calcolo in cui è possibile visualizzare tutti i dati. La scelta di Google Spreadsheets non è dettata solo dalla funzionalità della condivisione di tipo *user-friendly*, ma anche dall'accesso diretto ai dati che il servizio offre. Esportare i dati è un gioco da ragazzi. Tutti gli utenti hanno il permesso di scaricarli, basta cliccare sull'opzione "scarica come" alla voce del menu "File". Diversi formati di output sono disponibili, come Excel, HTML, Acrobat PDF, testo e CSV. All'interno del Datastore, alla voce "Data A-Z" è possibile trovare l'elenco alfabetico completo di tutti i temi affrontati dal Guardian: se, per esempio, l'utente vuole trovare tutti gli articoli e gli spreadsheet relativi all'Afghanistan, basterà cliccarci sopra e si aprirà una nuova pagina contenente tutti gli articoli inerenti all'argomento.

Nel marzo 2012 è stata inaugurata una nuova sezione del sito: lo Show and Tell, sponsorizzato da Google, è una vetrina che raccoglie e ricondivide il meglio del mondo delle infografiche e delle visualizzazioni

¹²⁴ Con il termine ci si riferisce sia al Datablog che al Datastore, *directory* in cui vengono raccolti tutti i dati usati e pubblicati dal sito <<http://www.guardian.co.uk/data>>.

dati in giro per la Rete. Ancora una volta il lettore è stimolato a partecipare: chi vuole è invitato a condividere i propri lavori di visualizzazione nell'apposito gruppo Flickr del Guardian. La sezione Editor's picks, invece, raccoglie gli articoli più interessanti secondo il capo redattore: dall'economia, alla salute, allo sport all'educazione, i temi trattati variano di volta in volta.

Dietro tutte le storie pubblicate dal Guardian c'è un processo¹²⁵, suddiviso sostanzialmente in quattro fasi. I dati vengono raccolti da siti governativi, tramite ricerche effettuate dai giornalisti, o da diverse altre fonti, come l'UNESCO, l'ONS, la World Bank e molte altre.

Spesso il lavoro di raccolta e pulizia dati è sottovalutato, ma richiede in realtà un grande impegno. Il Datastore svolge la maggior parte del lavoro sporco al posto dell'utente, identificando e ripulendo le migliaia di dataset da colonne non necessarie, formati errati, differenti unità di misura e quant'altro. Il team comincia allora a studiare la potenzialità di questi dati, osservandoli da un punto di vista giornalistico e pensando con quali altri dataset potrebbero essere uniti per ottenere una storia interessante.

Una volta compiuti i calcoli necessari – spesso si tratta per lo più di variazioni percentuali – si arriva al processo di output: è la fase in cui si sceglie se i dati saranno mostrati sotto forma di articolo o di visualizzazione e, eventualmente, quali strumenti utilizzare per crearla.

¹²⁵ Simon Rogers, *Facts are sacred, the power of data*, op. cit., pag. 235-240.

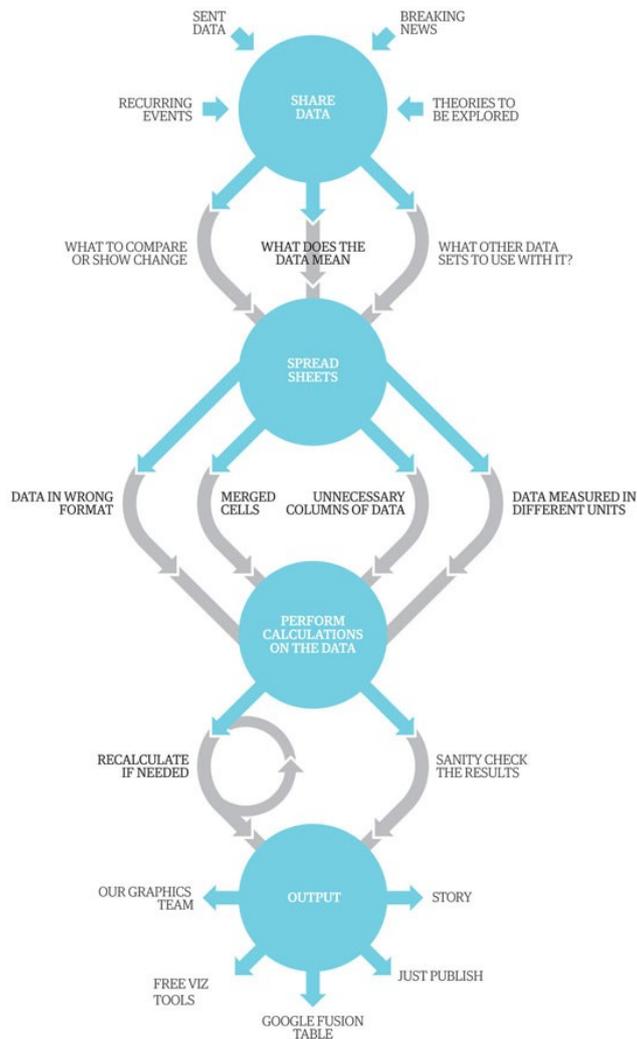


Figura 8 – Il processo di produzione del Guardian Data

Google Spreadsheets, Open Refine, Data Wrangler, Google Fusion Tables e Tableau Public sono solo alcuni degli strumenti preferiti utilizzati dal team data del Guardian: lo stesso Rogers ha più volte affermato che per fare del data-driven journalism non sono necessari costosi strumenti e che basta sfruttare quelli che la Rete mette a disposizione di tutti gli utenti in forma gratuita.

Il Guardian Data è il volto moderno dell'approccio ai dati. In esso sono contenuti tutti i principi e le caratteristiche fondamentali del data-driven journalism: dalla grafica accattivante usata per rendere visivamente comprensibili temi complessi, alle informazioni su eventi dedicati, all'analisi approfondita dei grandi fatti di cronaca. La parte più

importante del processo sono la trasparenza e la condivisione dei dati, che permettono ai lettori stessi di analizzare e migliorare i dati offrendo il loro prezioso contributo. Aprire i dati, questo è il *main focus* del Datablog, sin dal suo primo post pubblicato nel 2009.

4.2 Le grandi inchieste del Guardian Data

Dietro al successo del Guardian Data si cela il grande impegno di Simon Rogers, già curatore delle notizie per il Guardian dal 2001. Il suo team è composto da giovani e brillanti ricercatori e giornalisti del calibro di John Burn-Murdoch, Ami Sedghi, James Ball e Mona Chalabi, che pubblicano ogni giorno diversi post contenenti analisi e visualizzazioni su importanti temi di attualità: istruzione, politica e spesa pubblica, società, sport, cultura, ambiente e molti altri ancora. Gli studi che hanno riscosso un notevole successo sono molti: l'evasione fiscale nel mondo, i viaggi dei migranti e le tappe più pericolose, le vittime degli scontri in Egitto e Siria, le vittime della guerra in Afghanistan, il rapporto sulla violenza contro le donne, la classifica delle migliori università del mondo, sono solo alcune delle più importanti.

All'interno del sito si trovano anche le “series”, le grandi inchieste pubblicate in sezioni monotematiche all'interno delle quali sono raccolti tutti gli articoli, le ricerche, le infografiche e le visualizzazioni inerenti al tema trattato. Un esempio emblematico è la “US election data”¹²⁶, lo speciale sulle elezioni americane del 2012 definite dal Guardian non soltanto come uno scontro tra i candidati alla presidenza, ma come una vera e propria battaglia di visualizzazioni data-driven. La notte delle elezioni ha infatti rappresentato una delle occasioni in cui i giornalisti dei dati hanno dovuto competere fra loro per raccontare la stessa notizia. Il Guardian ha coperto l'evento pubblicando diversi dataset contenenti le somme spese dai candidati per le loro campagne, il numero di elettori iscritti per Stato, i dettagli completi delle schede elettorali e diverse

¹²⁶ Ancora una volta il Guardian ha chiesto l'aiuto dei suoi lettori: all'interno della US election data è possibile esplorare e analizzare i dati per trovare le statistiche più interessanti e segnalarle al giornale <<http://www.theguardian.com/news/series/us-election-data> >.

mappe interessanti: una, per esempio, paragona gli Stati a favore di Obama a quelli con alti tassi di povertà, disoccupazione e scarsa copertura sanitaria; un'altra mostra i risultati delle presidenziali offrendo la possibilità di filtrare i dati per conoscere il risultato di ogni singolo Stato.

Una grande serie dedicata allo sport è la “London 2012 Olympics data”¹²⁷, le prime Olimpiadi Open Data perché ricche di articoli pieni di dati statistici: dal confronto di ogni Stato vincitore di una medaglia con le Olimpiadi del 2008, all'analisi di ogni atleta in base a età, peso, sesso e altezza. Durante l'evento è stata creata anche un'alternativa tabella interattiva delle medaglie: alternativa perché il Guardian Data si è staccato dalla classica misurazione della performance di una squadra olimpionica – basata sul numero di medaglie vinte – calcolando il risultato in base alla dimensione della popolazione di un paese, al suo prodotto interno lordo o comparandolo con la dimensione del proprio team. I risultati si sono rivelati ben diversi da quelli reali¹²⁸, e hanno prodotto un vasto coinvolgimento del pubblico.

Il primo grande esperimento di crowdsourcing è avvenuto il 23 giugno del 2009, quando il team del Guardian Data ha pubblicato un'inchiesta che analizzava le spese effettuate in quattro anni dai 646 membri del Parlamento della House of Commons. L'esperimento ha portato alla pubblicazione di 700.000 ricevute fiscali che, anziché coprire le spese di viaggio verso Westminster, erano state usate per scopi ben diversi. Per mesi solamente il Daily Telegraph aveva ottenuto l'accesso limitato alla documentazione dettagliata contenente tutte le spese di ogni singolo membro del Parlamento grazie a una fonte anonima. La partecipazione dei lettori è sopraggiunta a causa della difficoltà nell'analizzare questi documenti. L'idea è arrivata da Simon Willison, che ha suggerito di

¹²⁷ Tutte le visualizzazioni, i dati e le statistiche sulle Olimpiadi sono state raccolte nella sezione <<http://www.theguardian.com/sport/series/london-2012-olympics-data>>.

¹²⁸ Nella realtà gli USA si sono classificati primi con 104 medaglie, ma in base alla popolazione e al PIL sarebbe dovuta arrivare prima Grenada, mentre in base alla grandezza del team avrebbe dovuto vincere più medaglie la Cina <<http://www.theguardian.com/sport/datablog/interactive/2012/aug/07/olympics-2012-alternative-medal-table-visualised>>.

pubblicare queste ricevute cartacee in un microsito¹²⁹ dove i cittadini potevano commentare e interagire con i dati, aiutando a migliorarli evidenziando, per esempio, le spese maggiori o quelle più sospette. Un solo utente è stato capace di revisionare più di 29.000 pagine: affidando i dati all'arena pubblica, i cittadini sono arrivati a conoscere le spese pubbliche dei membri del Parlamento persino meglio dell'intera redazione del giornale. L'inchiesta continua tutt'oggi: la sezione è quotidianamente aggiornata con le spese di tutti i deputati dei diversi schieramenti politici e con articoli e interviste a esperti.

Il punto di svolta per il Datablog arriva nella primavera del 2010, quando Wikileaks rilascia centinaia di migliaia di documenti sulle guerre in Afghanistan e Iraq e sui rapporti diplomatici dell'ambasciata americana. I primi ad essere sottoposti a verifica sono i 92.201 war logs consegnati in un formato Excel pessimamente formattato, ottenuti da David Leigh e Nick Davies a seguito di un negoziato con Julian Assange. Il termine ufficiale per definire quei documenti è SIGACTS: il database contenente tutti le operazioni militari americane più significative. Dai risultati, analizzati da David Leigh, Nick Davies, Declan Walsh, Simon Tisdall e Richard Norton-Taylor, è emerso che il sud, dove si trovavano le truppe britanniche e canadesi, è stata la zona più colpita. Il database permette inoltre di effettuare ricerche tramite parole chiave o storie chiave come, per esempio, gli attacchi IED¹³⁰. L'anno successivo la storia si è ripetuta per i 391.000 war logs dell'Iraq, riguardanti le operazioni militari avvenute tra il gennaio 2004 e il dicembre 2009. L'obiettivo è sempre lo stesso: analizzare i dati, permettere al team di giornalisti di costruire notizie di importante impatto civico e mettere i dati a disposizione degli utenti. Questa volta ciò che si evidenzia nell'inchiesta è l'enorme volume di morti, la maggior parte dei quali civili¹³¹. Infine, il dicembre del 2010 è stato segnato dal cablegate, il rilascio di 251.284 cavi inviati dalle

¹²⁹ La sezione è stata chiamata Investigate your MP's expenses ed è tutt'oggi aggiornata <<http://www.theguardian.com/politics/mps-expenses>>.

¹³⁰ Gli ordigni esplosivi improvvisati, più comunemente conosciuti come IED (Improvised Explosive Device) sono ordigni realizzati in maniera improvvisata principalmente usati nelle tecniche di guerriglia. Tutti i dati sugli incidenti possono essere scaricati su un foglio di calcolo <<http://www.theguardian.com/world/datablog/2010/jul/25/wikileaks-afghanistan-data>>.

¹³¹ Su 109.032 morti, 66.081 vittime erano civili <<http://www.theguardian.com/news/datablog/2010/oct/23/wikileaks-iraq-data-journalism>>.

ambasciate di tutto il mondo su SIPRnet¹³², un enorme dataset di documenti ufficiali. Di questi, 97.070 sono stati classificati come “riservati”, 28.760 sono contrassegnati come PTER, acronimo che indica la prevenzione al terrorismo e il più antico dei cabli risale al 1966, mentre la maggior parte dei documenti, 56.813 per l’esattezza, si riferiscono solo al 2009, il primo anno dell’amministrazione Obama. Analizzarli ed estrapolarne le storie è un compito enorme, tutt’ora in fase di svolgimento. Per questo, ancora una volta, il Guardian Data ha chiesto l’aiuto dei suoi lettori affinché condividano le loro idee su ciò che desiderino diventi oggetto di indagine nella loro inchiesta¹³³.

Per cinque giorni, nel mese di agosto del 2011, in Inghilterra si sono verificati una serie di disordini, con numerosi episodi di sciacallaggio e rivolte. Cominciati nel quartiere di Tottenham, a seguito dell’uccisione di un giovane sospettato durante una sparatoria con la polizia, si sono poi estesi a Londra, Liverpool, Birmingham, Nottingham, Manchester e Salford.

Ispirandosi allo studio effettuato da Philip Meyer e dal Detroit Free Press sui disordini di Detroit nel 1967, Reading the Riots¹³⁴ è l’inchiesta realizzata dal Guardian Data sulle cause e le conseguenze delle rivolte estive. Lo studio include interviste confidenziali di un gran numero di persone direttamente coinvolte nei disordini, mentre parallelamente è stato effettuato uno studio da ricercatori della Manchester University, su un secondo database di oltre due milioni e mezzo di tweet relativi ai disordini. Durante le interviste sono stati raccolti dati demografici sugli intervistati, per un totale di più di un milione di parole contenute nei racconti e nelle testimonianze. I dati sono stati poi visualizzati in una mappa che ha permesso al team di vedere il quadro complessivo della rivolta.

¹³² Il Secret Internet Protocol Router Network è un sistema di reti di computer interconnessi utilizzato dal Dipartimento della Difesa e dal Dipartimento di Stato degli Stati Uniti per la trasmissione di informazioni riservate, in un ambiente “completamente sicuro”.

¹³³ Gli utenti possono navigare all’interno del database attraverso parole chiave o cliccando su di uno Stato nella mappa < <http://www.theguardian.com/world/interactive/2010/nov/28/us-embassy-cables-wikileaks> >.

¹³⁴ L’inchiesta è stata fatta in associazione con la London School of Economics e supportata dalla Joseph Rowntree Foundation e dalla Open Society Foundations <<http://www.theguardian.com/uk/series/reading-the-riots>>.

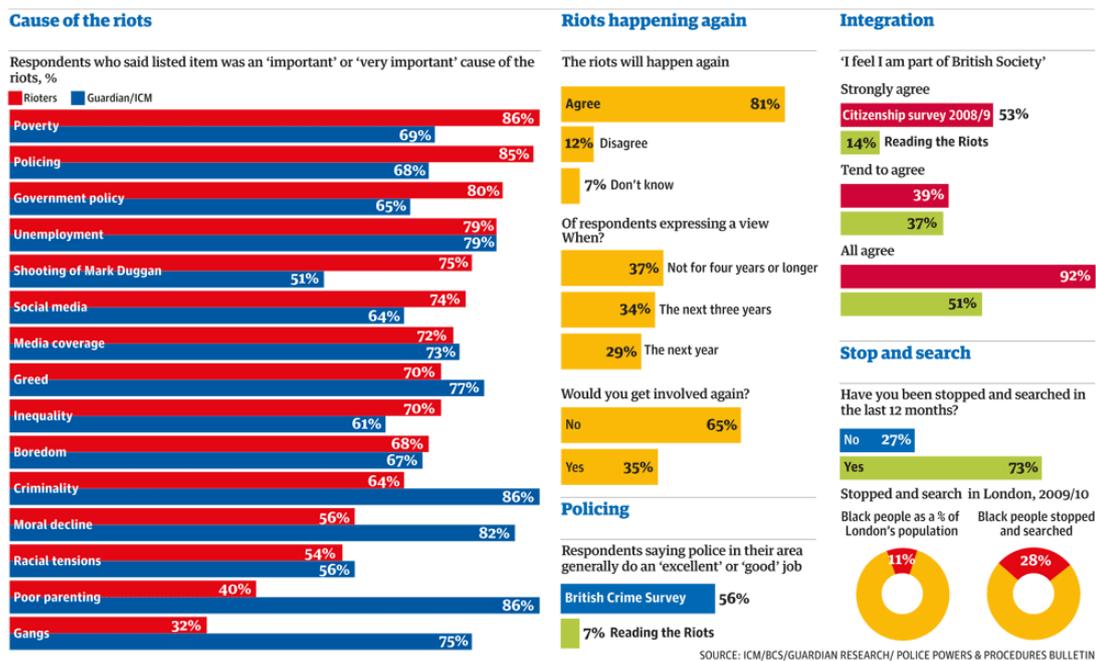


Figura 9 – I numeri dei disordini in Inghilterra dell'agosto 2011

Dai primi risultati ottenuti dai tribunali, basati su un campione di 123 casi, è emerso che la maggior parte dei rivoltosi erano uomini tra i 18 e il 24 anni. Le rivolte sono state alimentate da una serie di fattori, tra cui la mancanza di opportunità per i giovani, un'educazione insufficiente da parte dei genitori, l'incapacità del sistema giudiziario di riabilitare i condannati e l'atteggiamento sospettoso e materialista della polizia. Navigando all'interno della sezione del sito dedicata all'inchiesta, salta subito all'occhio quanto la copertura della notizia sia stata profonda: non solo le fasi dei disordini sono dettagliatamente descritte, ma vi sono anche una serie di articoli contenenti video, commenti, breaking news e analisi di tutti i dati raccolti durante le interviste.

Lo stesso anno Rogers è stato nominato Best UK Internet Journalist dall'Oxford Internet Institute, mentre il Datastore ha ricevuto importanti riconoscimenti, tra cui il Knight-Batten Award per l'innovazione nel giornalismo.

Il 18 aprile 2013 Simon Rogers ha passato il testimone al giovane James Ball¹³⁵. Dietro questa scelta si cela la proposta di entrare a far parte del team di Twitter come primo data editor nella storia del social network. A soli 27 anni James Ball, già parte del brillante team di Rogers, è stato nominato nuovo data editor della sezione dedicata al data journalism. Laureatosi nel 2008, ha lavorato due anni per Wikileaks, The Grocer e al Bureau of Investigative Journalism, fino ad approdare nel febbraio 2011 al Guardian Datablog, per il quale ha collaborato a numerose e multi premiate inchieste di alto profilo, come Offshore secrets, Guantanamo files e Reading the Riots. Il suo compito sarà quello di continuare a curare il Datastore, ma è anche stato incaricato di incorporare il data-driven journalism alla redazione del Guardian in maniera sempre più efficace. In occasione del Festival Internazionale del Giornalismo tenutosi a Perugia, ha così commentato la newsroom integrata del giornale inglese: “L’obiettivo del data editor è portare ogni giornalista a dare più valore ai dati e, una volta raggiunto lo scopo, estinguersi”.

Per Ball la risposta alla supportabilità dei giornali è strettamente connessa alla precisione del lavoro, ampiamente presente nel data journalism. La chiave per il futuro è la valorizzazione della comunità: collaborazioni, eventi e campagne mirate, a fianco di associazioni per guadagnare e colpire bersagli precisi.

¹³⁵ L’annuncio è avvenuto attraverso un post intitolato “*Farewell Guardian, hello Twitter*” pubblicato dallo stesso Rogers sul suo blog <<http://simonrogers.net/2013/04/18/farewell-guardian-hello-twitter/>>

CONCLUSIONI

In un'era dove l'informazione viaggia veloce nella Rete, i giornalisti hanno messo in atto un processo di formazione che li spinge a diventare più consapevoli e rapidi nel riconoscere il potenziale di una notizia. I dati sono ovunque: i governi, le società, persino i singoli individui producono giornalmente miliardi di informazioni. Ne siamo letteralmente sommersi. In questo contesto la precisione del lavoro è tutto: il data-driven journalism ha fatto proprio questo principio, offrendo un valore aggiunto alle proprie inchieste. La forza del metodo di analisi conferisce precisione e, quindi, obiettività alla notizia. I giornalisti diventano il ponte e la guida tra coloro che possiedono i dati e il pubblico che vorrebbe accedervi e comprenderli. Il giornalista dei dati diventa una fonte fidata, un rifugio sicuro per l'informazione di qualità. Esso implementa le capacità tecniche e tecnologiche del giornalista che, grazie al digitale, è chiamato a rispondere a uno standard di qualità professionale sempre più alto e specializzato. Laddove il data journalism è una realtà già affermata, i giovani sono sempre più consapevoli che intraprendere una carriera del genere comporta una formazione professionale dove il giornalista non è più un "narratore romantico", ma un esperto in strumenti specifici, software, analisi ed elaborazione dei dati. L'Italia è ancora molto indietro in questo senso, ma i pionieri del data journalism sono molto attivi e si impegnano duramente affinché qualcosa cambi anche nel nostro paese. Ne sono un esempio la [Fondazione](#) o le comunità di Spaghetti Opendata e Datajournalismitaly, senza dimenticare la sezione italiana della Open Knowledge Foundation che condividono il loro sapere per diffondere la cultura dell'Open Data e del data journalism.

La scelta di concentrarci nell'analisi del caso inglese del Guardian Data non è casuale: si dice che i dati siano un fenomeno nuovo, proprio dell'era dell'informazione, ma in realtà essi sono sempre esistiti e lo dimostra il Manchester Guardian che ha pubblicato la sua prima inchiesta basata sui dati nel 1821. La differenza risiede solo nell'evoluzione da supporto cartaceo a formato digitale. La testata

inglese ha dimostrato di essere un'antesignana in questo campo, facendo della trasparenza e del crowdsourcing il proprio cavallo di battaglia.

Se prima l'obiettivo principale era capire cosa fosse il data-driven journalism, adesso sempre più spesso ci si chiede come fare del buon giornalismo dei dati. Inizialmente la gente si chiedeva cosa fosse questo tipo di giornalismo e a cosa avrebbe portato, ma il tempo ha dato le sue conferme e adesso il giornalismo dei dati, con i suoi strumenti e il suo modo innovativo di trovare e raccontare la notizia, sta diventando uno standard nel mondo dell'informazione. Il giornalismo dei dati non è solo grafici e visualizzazioni: si tratta di raccontare una storia nel miglior modo possibile. "I fatti sono sacri" e ciò vuol dire che la notizia non può più essere manipolata attraverso la retorica delle parole che potrebbero influenzare il lettore: aprendo i dati che stanno dietro ogni inchiesta, il lettore diventa capace di trarre le proprie conclusioni senza il rischio di essere fuorviato. Grazie ai dati il giornalismo diventa un modo più efficace di raccontare una storia utilizzando la potenza dei dati, capace di rivoluzionare intere società. I dati, infatti, non stanno solo cambiando il modo di fare giornalismo, ma anche il nostro mondo.

Attraverso questo lavoro di ricerca e analisi dell'evoluzione del data-driven journalism nel mondo, abbiamo voluto dimostrare che la *data revolution* non solo ha cambiato il modo di fare giornalismo, ma ha cambiato in profondità anche la mentalità del cittadino e della società. È l'idea stessa di cittadinanza che si sta evolvendo verso una dimensione più attiva e consapevole e ciò che sta accadendo all'informazione ne è soltanto lo specchio: non è più sufficiente che un articolo sia pubblicato da un giornale per renderlo affidabile; ci vogliono i dati, le informazioni che si celano dietro la creazione di quella notizia. Non è più accettabile che il governo non condivida con i suoi cittadini i dati del proprio operato. È la realizzazione pratica del passaggio da cittadino informato a cittadino informante.

Centrale al tema in questione è la trasparenza delle istituzioni, che, rendendo aperti i propri dati, ha reso l'informazione più democratica e ha abbattuto quei limiti di accesso che sembravano essere insormontabili. Un nuovo modo di informare, a cui si è arrivati non senza fatica e che

comporta ancora notevoli sforzi. Sono molti, infatti, i paesi che ancora lottano per uniformarsi agli standard di una vera società democratica, emulando il modello americano e, perché no, tentando anche di superarlo. La trasparenza non è una concessione, ma un diritto che, in quanto tale, deve essere esercitato e difeso.

GLOSSARIO

Big Data: termine usato per indicare grandi aggregazioni di dati, la cui mole richiede strumenti differenti da quelli usati per il data journalism tradizionale. Il progressivo aumento della dimensione dei data set è legato alla necessità di analisi su un unico insieme di dati. Le tre caratteristiche principali dei Big Data sono il volume, la varietà e la velocità, essenziale effettuare analisi e ottenere dei dati in tempo reale.

CSV: Il comma-separated values è un formato di file basato su file di testo utilizzato per l'importazione e l'esportazione da fogli elettronici o database di una tabella di dati.

Database journalism: anche chiamato structured journalism, è il principio secondo cui il contenuto delle notizie è organizzato intorno a dati strutturati, al contrario dell'impostazione classica delle notizie storico-centriche. La differenza con il data-driven journalism sta nel fatto che il primo è un processo in base al quale i giornalisti costruiscono storie utilizzando i dati come materiale primario, mentre con database journalism ci si riferisce alla struttura organizzativa dei contenuti.

Data-driven journalism: spesso abbreviato in "ddj", è un termine usato per descrivere il processo giornalistico basato sull'analisi e il filtraggio di grandi insiemi di dati con lo scopo di creare una notizia. È il corrispondente italiano del giornalismo dei dati.

Data mining: processo di estrazione di informazione da dati strutturati tramite l'applicazione di algoritmi che individuano le associazioni nascoste tra le informazioni, rendendole visibili.

Data warehouse: magazzino centrale di dati di un'organizzazione, progettato per consentire all'utente di ottenere e lavorare grosse quantità di dati.

Dataset: un contenitore di dati che corrisponde al contenuto di una tabella di un unico database. I dataset possono essere uniti tra loro, per esempio con Google Fusion Tables, per ottenere storie interessanti.

Excel: programma prodotto da Microsoft, dedicato alla produzione e alla gestione dei fogli elettronici. Questi sono identificati dall'estensione .xls, .xlsx, ma può importare ed esportare dati in altri formati, come il formato .csv. Excel si presenta come un foglio di calcolo composto da colonne e da righe, la cui intersezione forma una cella all'interno della quale si può inserire un valore o una funzione.

Fact checking: espressione inglese che significa verifica dei fatti. Il fact checking deriva dalla tradizione del giornalismo professionale: è un metodo empirico per dimostrare se dichiarazioni e fatti sono attendibili, attraverso la raccolta e il confronto di fonti.

FOIA: il Freedom of Information Act è una legge che dà al cittadino il diritto di accedere agli archivi di Stato statunitensi, a documenti riservati e coperti da segreto di Stato, garantendo la trasparenza della pubblica amministrazione nei confronti del cittadino e il diritto di cronaca e la libertà di stampa dei giornalisti. La prima nazione ad adottare un FOIA è stata l'America, ma a oggi sempre più paesi si stanno muovendo in questa direzione.

Infografica: informazione proiettata in forma grafica e visuale. Nasce dall'incrocio dell'arte grafica con il giornalismo e l'informatica. Le immagini possono essere 2D o 3D, animate o fisse. Vi sono molti software disponibili gratuitamente sul web, come Picktochart, Infogr.am o Easel.ly.

Open Data: secondo la definizione fornita dalla Open Definition, i dati aperti sono “dati che possono essere liberamente utilizzati, riutilizzati e ridistribuiti da chiunque, soggetti eventualmente alla necessità di citarne la fonte e di dividerli con lo stesso tipo di licenza con cui sono stati originariamente rilasciati”.

Open Refine: creato da Google, è uno degli strumenti di pulizia dati più utilizzato. È utile soprattutto quando si ha a che fare con dati come i nomi standardizzati. Una buona alternativa è Data Wrangler

Python: software Open Source che permette l'elaborazione, l'analisi statistica e la creazione di applicazioni ad hoc per la visualizzazione dei dati.

Precision journalism: termine suggerito da Everette E. Dennis a Philip Meyer in occasione della pubblicazione del suo libro, per sottolineare il contrasto tra il metodo scientifico e l'approccio pseudoartistico del Nuovo Giornalismo. Indica un nuovo modo di realizzare inchieste, ovvero tramite gli strumenti della matematica, della statistica e delle scienze sociali e comportamentali applicate alla pratica del giornalismo.

QGIS: Quantum GIS è un'applicazione desktop GIS Geographic(al) Information System, un sistema progettato per catturare, immagazzinare, manipolare, analizzare, gestire e rappresentare dati di tipo geografico.

Scraping: il raschiamento è l'uso di un computer per catturare dati da fonti online, dispersi fra vari materiali documentari. Lo scraping permette anche di catturare tutti quei dati che le organizzazioni preferirebbero non condividere.

Spreadsheet: corrispondente inglese di foglio elettronico, o foglio di calcolo, ideato da Dan Bricklin. È un software che permette di effettuare calcoli, elaborare dati e tracciare rappresentazioni grafiche. Il principio su cui si basa il foglio di calcolo è semplice: fornire una tabella, o foglio di lavoro, formata da celle in cui si possono inserire dati, numeri o formule. Lo strumento più utilizzato in questa fase del processo è Google Spreadsheets.

SQL: lo Structured Query Language è un linguaggio di interrogazione che permette di creare e modificare schemi di database, ma anche inserire, interrogare e gestire i dati memorizzati. È utile quando si hanno più di due fogli di calcolo da condividere o un grosso numero di dati da cercare. Esempi di SQL sono Access SQLite, MySQL o PostgreSQL.

Visualizzazione: a differenza dell'infografica, la visualizzazione assume il punto di vista dell'utente, che esplora le informazioni visuali presentate dall'autore secondo un percorso di progressivo approfondimento.

BIBLIOGRAFIA

AGOSTINI ANGELO, *Giornalismi. Media e giornalisti in Italia*, Il Mulino, Gennaio 2010

BARBANO-SASSU, *Manuale di giornalismo*, Laterza 2012

BERNERS-LEE TIM, *L'architettura del nuovo web: dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*, Milano, Feltrinelli, 2001.

BOLTER JAY D., GRUSIN RICHARD, *Remediation. Competizione e integrazione tra media vecchi e nuovi*, Guerini, 2003

BOLTER-GRUSIN, *Remediation. Competizione e integrazione tra media vecchi e nuovi*, Guerini, Varese 2011

BOWMAN-WILLIS, *We media. How audiences are shaping the future of news and information*. The Media Center at the American Press Institute, 2003

BRADSHAW PAUL, *Scraping for journalists, How to grab information from hundreds of sources, put it in data you can interrogate - and still hit deadlines*, Leanpub 2013

CAIRO ALBERTO, *L'arte funzionale. Infografica e visualizzazione delle informazioni*, Pearson 2013

CARLINI FRANCO, *Parole di carta e di Web*, Einaudi, Torino, 2004

CAROTENUTO GENNARO, *Giornalismo partecipativo. Storia critica dell'informazione al tempo di Internet*, Nuovi Mondi, Modena, 2009

CASTELLS MANUEL, *La nascita della società in rete*, 2002

CIOTTI-RONCAGLIA, *Il mondo digitale. Introduzione ai nuovi media*, Laterza, Bari 2006

DADER JOSÉ LUIS, *Periodismo de precisión: la vía socioinformática de descubrir noticias*, Ed. Síntesis, Madrid 2002

- DE KERCKHOVE DERRICK, *Dall'alfabeto a Internet*, Mimesis, 2008
- DIRITTO DI SAPERE, *LegalLeaks. Media e democrazia come usare il diritto di accesso all'informazione*, aprile 2013
- FAMA ANDREA, *Open data - data journalism. Trasparenza e informazione al servizio delle società nell'era digitale*, Simplicissimus Book Farm srl, 2011
- FIDLER ROGER, *Mediamorfosi. Comprendere i nuovi media*, Guerini e associati, Varese 2002
- GALLI DIEGO, *Che cos'è il giornalismo partecipativo? Problemi dell'Informazione*, Il Mulino, n°3 Settembre, 2005
- GILLMOR DAN, *We the media, Grassroots Journalism by the People, for the People*, O' Reilly Media, Sebastopol, 2006
- GOZZINI GIOVANNI, *Storia del giornalismo*, Bruno Mondadori, Milano 2000
- GRANIERI GIUSEPPE, *Umanità accresciuta. Come la tecnologia ci sta cambiando*, Laterza, Bari 2009
- GRAY JONATHAN, CHAMBERS LUCY, BOUNEGRU LILIANA, *The data journalism handbook*, Oreilly & Associates Inc 2012
- JOHNSON STEVEN BERLIN, *Interface Culture, How Technology transform the way we create & communicate*, Basic Books, 1999
- LÉVY PIERRE, *Cyberdemocrazia*, Mimesis, 2008
- LORENZ MIRKO, *Data-driven journalism: What is there to learn?*, Amsterdam 2010
- MAISTRELLO SERGIO, *Giornalismo e nuovi media. L'informazione al tempo del citizen journalism*, Apogeo, Milano 2010
- MANOVICH LEV, *Il linguaggio dei nuovi media*, Olivares, Milano, 2005

MAYER-SCHÖNBERGER VIKTOR, CUKIER KENNETH, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, Boston, New York 2013

MCLUHAN MARSHALL, *La galassia Gutenberg. Nascita dell'uomo tipografico*, Armando, Roma 1976

MENDUNI ENRICO, *I media digitali, Tecnologie, linguaggi, usi sociali*, Laterza, Roma – Bari, 2007

MEYER PHILIP, *Giornalismo e metodo scientifico. Ovvero il giornalismo di precisione*, Armando editore 2006

MEYER PHILIP, *Paper Route: Finding My Way to Precision Journalism*, iUniverse, 2012

MONMONIER MARK, *How to lie with maps*, 1996

O'REILLY RADAR TEAM, *Big Data Now: Current Perspectives from O'Reilly Radar*, O'Reilly team, 2011

PRATELLESI MARCO, *New Journalism, Teorie e tecniche del Giornalismo Multimediale*, Mondadori, Milano, 2008

ROGERS SIMON, *Facts are sacred, the power of data*, 2013

RUSS-MOHL, *Fare Giornalismo*, Il Mulino, Urbino 2011

RUSSO MASSIMO, ZAMBARDINO VITTORIO, *Eretici digitali, la Rete è in pericolo, il Giornalismo pure. Come salvarsi con un tradimento e dieci tesi*, Apogeo, 2009

SABADIN VITTORIO, *L'ultima copia del "New York Times"*, Donzelli, Roma, 2007

STAGLIANÒ RICCARDO, *Giornalismo 2.0. Fare informazione al tempo di Internet*, Carocci, Roma 2002

STAGLIANÒ RICCARDO, *Giornalismo 2.0. Fare informazione al tempo di internet*, Carocci, Roma 2004

TUFTE EDWARD ROLF, *The visual display of quantitative information*, Graphics Press, 1983

VISCONTI GRAZIA, *Giornalista online, Dal Web Writer al Web Editor*. Franco Angeli, Milano, 2002

YAU NATHAN, *Visualize this: the flowing data guide to design, visualization and statistics*, 2011

SITI WEB

5stardata.info – pagina esplicativa dello schema 5 star

ahref.eu – sito ufficiale della Fondazione <ahref

blackwellreference.com – libreria online multimediale

blog.ouseful.info – blog di Ouseful.info, gestito da Tony Hirst

cafebabel.fr – rivista di attualità basata sul giornalismo partecipativo

careersatft.com – pagina del Financial Times dedicata alle offerte di lavoro

census.gov – sito web contenente i dati del censimento americano

ckan.net – catalogo online contenente dataset utili reperibili in Internet

corriere.it – news on line del quotidiano Il Corriere della sera

cottica.net – spazio personale dell'economista Alberto Cottica

daily.wired.it – notizie su Internet e sulla tecnologia della rivista Wired

datadrivenjournalism.net – corsi, eventi e risorse online dal team di DataDrivenJournalism

datajcrew.sudmediatika.it – notizie sul data journalism nel mondo

datajournalism.it – sito nato dall'omonima mailinglist italiana

datakind.org – sito web della comunità di appassionati dei dati

dataninja.it – blog sui dati di Alessio Cimarelli e Andrea Nelson Mauro

dati.istat.it – data warehouse dell'Istat

dirittodisapere.it – sito dell'omonima associazione nata per promuovere la trasparenza della pubblica amministrazione

[ejc.net](#) – sito web dell’European Journalism Centre, ente promotore del data journalism

[elmundo.es](#) – notizie e approfondimenti del giornale El Mundo

[factchecking.civiclincs.it](#) – piattaforma media civico promossa da [factcheck.it](#)

[flickr.com](#) – sito web multilingua per la condivisione di fotografie

[foia.it](#) – sito ufficiale dell’Iniziativa per l’adozione di un FOIA in Italia

[groups.google.com/d/forum/spaghettiopendata](#) – gruppo Google della comunità di Spaghetti Opendata

[groups.google.com/group/datajournalismitaly](#) – gruppo Google di Datajournalismitaly

[holovaty.com](#) – sito personale del giornalista Adrian Holovaty

[jplusplus.org](#) – sito web della comunità Journalism++

[knightlab.northwestern.edu](#) – sito del team di giornalisti della NWU

[lanacion.com.ar/data](#) – blog sugli Open Data de LA NACION

[lastampa.it/blogs/datablog](#) – blog del quotidiano La Stampa sui dati

[lefigaro.fr/infographie](#) – sezione dedicata alle infografiche de Le Figaro

[nieman.harvard.edu](#) – portale web della fondazione Nieman

[okfn.org](#) – spazio dedicato alla promozione di dati e contenuti aperti

[onlinejournalismblog.com](#) – blog del giornalista Paul Bradshaw

[open.blogs.nytimes.com](#) – blog sugli Open Data del New York Times

[opendata.zeit.de](#) – spazio dedicato agli Open Data del tedesco Zeit Online

[opendatablog.ilsole24ore.com](#) – blog sui dati de Il Sole 24 Ore

[owni.fr](#) – portale web indipendente specializzato in data journalism

powerreporting.com – sito web di banche dati ideato da Dedman

propublica.org – testata web non-profit indipendente fondata da Steiger

schoolofdata.org – progetto della OKFN per avvicinare le persone ai dati

simonrogers.net – blog dell'ex editor del Guardian Data Simon Rogers

slideshare.net – applicazione che raccoglie e condivide presentazioni

spaghettiopendata.org – sito web di riferimento di Spaghetti Opendata

theguardian.com/uk – portale online del quotidiano più letto d'Inghilterra

theguardian.com/data – sezione del Guardian dedicata al data journalism

uncp.edu – archivio online del Washington Post