

Università degli Studi di Torino  
Dipartimento di Culture, Politica e Società  
Corso di Laurea in Comunicazione, Ict e Media



Open e Linked Data.

Nuove frontiere per la comunicazione e la sociologia

Candidato: Marco Gritti

Relatore: Giuseppe Tiplado

Anno Accademico 2015/2016



# Indice

<b>Introduzione</b> .....	<b>3</b>
<b>Obiettivo della tesi</b> .....	<b>3</b>
<b>Struttura della tesi</b> .....	<b>5</b>
<b>Capitolo 1. Il <i>data journalism</i> e la <i>media logic</i></b> .....	<b>6</b>
<b>1.1. <i>Data journalism</i> tra metodo scientifico e strumenti</b> .....	<b>6</b>
1.1.1. Il <i>data journalism</i> ed i dati della pubblica amministrazione .....	15
<b>1.2. I <i>media studies</i></b> .....	<b>17</b>
<b>1.3. L'analisi del contenuto</b> .....	<b>21</b>
<b>Capitolo 2. Metodo e metodologia di ricerca</b> .....	<b>25</b>
<b>2.1. I dati nelle Pubbliche Amministrazioni</b> .....	<b>25</b>
2.1.1. <i>Open data</i> .....	26
2.1.2. <i>Open government</i> .....	29
<b>2.2. Agenzia per l'Italia Digitale</b> .....	<b>31</b>
<b>2.3. Il contesto giuridico</b> .....	<b>34</b>
Il <i>Freedom of Information Act</i> .....	37
<b>2.4. L'Indicatore di Tempestività dei Pagamenti</b> .....	<b>39</b>
<b>2.5. Materiale di analisi: Dati</b> .....	<b>42</b>
<b>2.6 Materiale di analisi: Testi</b> .....	<b>46</b>

Appendice. <i>I linked data</i> .....	48
<b>Capitolo 3. <i>Case Study</i>: la Pubblica Amministrazione tra dati e narrazione</b> .....	<b>54</b>
<b>3.1. I testi</b> .....	<b>54</b>
Corpus n.1: Decreto Trasparenza .....	59
Corpus n.2: <i>Freedom of Information Act</i> .....	66
Corpus n.3: <i>Open Government</i> .....	70
<b>3.2. I dati</b> .....	<b>75</b>
Appendice al <i>case study</i> : i testi .....	93
<b>Conclusioni</b> .....	<b>111</b>
<b>Ringraziamenti</b> .....	<b>117</b>
<b>Bibliografia</b> .....	<b>118</b>
<b>Sitografia</b> .....	<b>125</b>

# Introduzione

La mia è una tesi sperimentale, un progetto condotto in collaborazione con Synapta Srl<sup>1</sup>, start-up innovativa fondata a Torino che sviluppa software e progetti relativi alla tecnologia *linked data*. Il lavoro, durato sei mesi, si è articolato in due progetti di analisi condotti in successione temporale, ma legati dalla stessa domanda cognitiva. La prima parte, svolta direttamente, nei tempi e nei modi, con Synapta, è stata un lavoro di analisi di dati delle Pubbliche Amministrazioni, in particolare dell'Indicatore di Tempestività dei Pagamenti, e ha portato alla stesura di un report che, di fatto, rappresenta un esempio di *data journalism*. La seconda parte, invece, svolta sotto la diretta supervisione del dottor Tiplido, è un lavoro di analisi del contenuto di testi giornalistici.

Cercherò di chiarire nel primo capitolo gli aspetti più rilevanti, rispetto alla mia ricerca, del concetto di *data journalism*, e rimando al capitolo metodologico per la definizione dei corpora analizzati e per una trattazione della questione dei *linked data*. Ora mi sembra più dirimente esplicitare l'obiettivo della mia tesi e la domanda cognitiva che ne ha sotteso la stesura.

## Obiettivo della tesi

Quanto mi ha spinto a collaborare con Synapta Srl, prima nel corso del tirocinio curricolare previsto dal corso di laurea in Comunicazione, Ict e Media e poi proseguendo con la stesura della tesi in azienda, è stata innanzitutto la curiosità di spendermi in un progetto concreto di giornalismo con i dati. La start-up fondata da Federico Morando e Alessio Melandri, come accennato, si occupa di scrivere software per estrarre, ripulire, collegare e, in generale, curare i dati.

---

<sup>1</sup> <https://synapta.it/>, consultato il 20/04/2016.

Il lavoro che ho intrapreso mi ha portato, attraverso mesi di raccolta, analisi e interpretazione di dati sulle pubbliche amministrazioni, alla pubblicazione di un report, in cui ho analizzato l'Indicatore di Tempestività dei Pagamenti dei comuni piemontesi, un dato che le amministrazioni sono tenute a pubblicare sul proprio sito istituzionale e che indica il tempo medio impiegato dalle stesse amministrazioni per pagare i fornitori di prodotti e servizi.

L'obiettivo iniziale, pertanto, era quello di lavorare con i dati per estrapolarne un racconto giornalistico. La collaborazione con Synapta Srl mi ha permesso di entrare in contatto con una realtà lontana dal mondo giornalistico *tout court*, ma che si occupa di sviluppare software che si integrano perfettamente nell'idea di raccontare la realtà attraverso i dati. Assunta tale consapevolezza, ed insieme quella che neanche sei mesi di lavoro sulla tesi sarebbero stati sufficienti a portare ad uno stato di accettabile compiutezza un progetto di *linked data journalism*, la ricerca è evoluta in due direzioni. La prima riguardante l'analisi delle tematiche che, di fronte al giornalismo con i dati, si aprono a un dibattito. Da questo punto di vista, così, si è scelto di procedere alla definizione di un quadro per quanto possibile esaustivo della questione *open government*, strettamente correlato alle *issues open data* e trasparenza. Il *background* della mia ricerca è voluto essere il tema dell'accesso ai dati e del loro riutilizzo, i due paradigmi del mondo *open*. In tal senso si è reso utile ripercorrere i passi legislativi che nel corso degli anni hanno contribuito a dotare l'Italia di un impianto di (presunta) apertura dei dati, a cominciare dal Decreto Trasparenza del 2013, fino al Freedom of Information Act recentemente approvato.

L'altra direzione del mio lavoro di tesi, invece, ha prodotto una *content analysis* condotta attraverso corpora alimentati da articoli provenienti da quotidiani e settimanali, e selezionati sulla base di ricerche attraverso specifiche *keywords*, espressione dei temi di trasparenza e di riutilizzo.

Procedendo in tale maniera è stato possibile identificare innanzitutto le tematiche più attuali nel discorso sulle pubbliche amministrazioni, coniugarle in seno al discorso sul *data journalism*, analizzarle sui media tradizionali e provare ad indagarle con un lavoro di analisi sui dati.

## **Struttura della tesi**

Il capitolo iniziale è di natura introduttiva. In particolare si riprendono i concetti fondamentali del *data journalism*, se ne tratteggiano metodi e strumenti attraverso una ricostruzione storica dei principali esempi, e si prova a declinarli al racconto delle pubbliche amministrazioni con l'ausilio di una bibliografia, ad oggi, ancora poco corposa. Sempre nel primo capitolo si cerca di definire un quadro relativo alle teorie della *communication research* e dei *media studies* che possono risultare utili a definire un quadro bibliografico a supporto del mio lavoro. Il capitolo due rappresenta il cuore dell'analisi che, in quanto sperimentale, ha bisogno di un corposo bagaglio metodologico: lì trovano spazio, innanzitutto, gli approfondimenti sulle questioni di governo aperto, di dati open, di amministrazione trasparente. È sempre nel secondo capitolo, poi, che definisco il metodo, gli strumenti, i tempi e le ragioni dell'analisi condotta sul *case study* oggetto del capitolo tre. Il capitolo conclusivo, infine, riassumerà le evidenze emerse nel corso del lavoro e rappresenta il tentativo di tratteggiare attualità e direttrici di sviluppo di un discorso, l'*open data government*, da cui dipendono il futuro dell'affare pubblico e, in parte, del giornalismo.

# Capitolo 1

## *Il data journalism e la media logic*

In questo capitolo iniziale intendo affrontare alcune questioni preliminari utili a fornire un adeguato impianto teorico su cui andranno ad inserirsi le pagine metodologiche, oggetto del prossimo capitolo, e il caso di studio di cui al capitolo 3. In particolare, ora, ritengo importante affrontare il discorso su cosa sia il *data journalism*, che rappresenta il mezzo ed il fine della mia analisi, cercando di declinarlo sul piano del racconto delle pubbliche amministrazioni. Sarà quindi innanzitutto circoscritto, in un quadro temporale, il significato del termine *data journalism* ed il metodo utilizzato nella costruzione di una storia attraverso i dati. Si provvederà quindi a tracciare alcune direttrici di sviluppo del giornalismo con i dati da, e verso, le pubbliche amministrazioni, al fine di valutarne ambiti di applicazione e potenzialità, da un punto di vista giornalistico e sociale. Una seconda parte del capitolo verterà invece sull'introduzione all'analisi del contenuto e dei mass media, come da titolo del riferimento bibliografico essenziale in materia<sup>2</sup>, sulla base di alcune riflessioni in ambito di sociologia dei media. Verranno pertanto richiamati i principali contributi teorici dei *communication studies*.

### **1.1. *Data journalism* tra metodo scientifico e strumenti**

Per poter disporre di una adeguata definizione di *data journalism* è opportuno chiarire fin da ora cosa siano questi dati. Numeri, innanzitutto. Il giornalismo dei dati ha come materia grezza numeri, ma una tale definizione pecca di superficialità: in un mondo digitalizzato come quello attuale, infatti, ogni genere di informazione è descritta in termini binari, cioè attraverso l'utilizzo di una sequenza di zeri e di uno

---

<sup>2</sup> Tipaldo G., *L'analisi del contenuto e i mass media*, il Mulino, Bologna, 2014.

[Bounegru *et al.* 2012]. Non si tratta pertanto di cifre, almeno non solo: con il termine dati si intendono le informazioni contenute in database processabili attraverso l'utilizzo di software, spesso scritti appositamente per estrapolarne un valore<sup>3</sup>. Si parlerà diffusamente nel capitolo 2 di *open data*, ovvero di quei dati aperti, tipicamente prodotti dalle pubbliche amministrazioni, verso cui il giornalismo nutre una forte attrazione proprio perché rappresentano il materiale da cui partire per raccontare una storia. Come vedremo più avanti, una delle potenzialità dei dati, il valore aggiunto che il *data journalism* cerca di cogliere, è la capacità di trovare storie, intese in un vero senso giornalistico, che altrimenti non potrebbero essere raccontate [Fama 2011]. I dati rappresentano il punto di partenza per questo tipo di giornalismo. Tuttavia il successo di questo filone, le cui radici si trovano in Gran Bretagna e negli Stati Uniti<sup>4</sup>, toglie il velo ad una verità nota a molti professionisti, ma a cui è opportuno fare cenno ricorrendo alle parole di Mirko Lorenz. Il giornalismo di oggi è sotto assedio, afferma il giornalista del *Deutsche Welle*, messo in discussione da meccanismi di produzione ribaltati dall'avvento del web 2.0 e dall'atteggiamento di un pubblico generalmente più interessato, ma meno disposto a identificarsi come semplice utente. Se un tempo la stampa rappresentava la porta dell'informazione, il modo per mettersi in relazione con la realtà esterna, oggi viviamo in un'epoca caratterizzata da un flusso ininterrotto di notizie che si trasmettono da una moltitudine di fonti [Bounegru *et al.* 2012: 3-5], in cui il web la fa da padrona. Web che Mimmo Candito, storica firma de *La Stampa*, paragona ad una prateria, una distesa di informazioni che racchiude l'universo intero, ma al tempo stesso contiene la possibile falsificazione di tutto. Uno spazio straordinario in cui «diventa problematica la costruzione di una conoscenza credibile, attendibile, poiché il vero e il falso si propongono con una identica valenza semantica, cioè possono essere consumati allo stesso modo senza che se ne possa sapere la reale natura» [Candito 2009: 125, postfazione a Pulitzer 1904]. Senza volermi addentrare in un

---

<sup>3</sup> Cos'è il data journalism, su <http://www.ahref.eu/it/biblioteca/data-journalism-1/cose-il-data-journalism.html>, consultato il 09/08/2016.

<sup>4</sup> Le radici del modern *data journalism* possono essere intraviste nel CAR, *Computer Assisted Reporting*, il primo approccio sistematico all'utilizzo del computer per la raccolta e l'analisi dei dati. Vale la pena fare riferimento al caso delle elezioni statunitensi del 1952, quando la *Columbia Broadcasting System* decise di avvalersi dell'Univac, uno dei primi elaboratori sviluppati, per predire l'esito delle elezioni sulla base delle dichiarazioni di voto del 7% della popolazione.

terreno dagli squisiti contributi, e che meriterebbe una tesi di approfondimento a parte, sia sufficiente riportare un altro stralcio dello scritto di Candito che riprende le parole di Michael Bloomberg, fondatore del colosso Bloomberg L.P., multinazionale operante nel settore dei mass media che ne ha fatto, secondo Forbes, il 14° uomo più ricco al mondo al 2015. “Che c’entra internet con il giornalismo? Internet è uno straordinario archivio”, sentenza Bloomberg. Le parole dell’ex sindaco di New York City rappresentano, forse involontariamente, l’esatto punto da cui partire nella definizione del *data journalism*: internet rappresenta una fonte, straordinaria, inesauribile, di dati. Tuttavia il lavoro del giornalista non può, e non deve, esaurirsi alla consapevolezza di operare in un ambiente mediale in cui il digitale spadroneggia. Timothy Berners-Lee, inventore del world wide web, è convinto della necessità, da parte dei professionisti dell’informazione, di approfittare della mole di dati offerti dal web. Un tempo il giornalista raccoglieva le storie da raccontare nei bar, chiacchierando con la gente – ricorda – ma oggi deve studiare attentamente i dati ed estrapolarne storie interessanti: il futuro del giornalismo è nei dati<sup>5</sup>.

Se è probabilmente corretto affermare che il futuro del giornalismo sia nei dati, non è certamente sbagliato ricordare come, già nel passato, ottime inchieste giornalistiche abbiano preso corpo a partire proprio dai dati. Un breve *excursus* storico, in questo senso, può aiutare a comprendere quanto, del giornalismo *tout court*, rimanga nel giornalismo dei dati. I giornalisti di oggi hanno certamente a disposizione set di dati molto più ampi che in passato, nonché strumenti di analisi più rapidi nell’elaborazione e software grafici più potenti, ma dalla storia del *data journalism* è importante ereditarne il metodo. Fu Philip Meyer il primo a teorizzare il giornalismo dei dati, nel 1973, ed il suo *Precision journalism* rappresenta il punto di partenza della ricerca e della definizione di questa branca del giornalismo. Le riflessioni dell’allora quarantatreenne cronista statunitense nascono in contrapposizione a quel *new journalism* teorizzato da Tom Wolfe, secondo cui il giornalismo avrebbe dovuto accompagnarsi alla cura della forma, in nome della fusione tra letteratura e informazione [Johnson e Wolfe 1974]. Al contrario, Meyer invoca il rigore

---

<sup>5</sup> Arthur, C. «Journalists of the future need data skills, says Berners-Lee», su *The Guardian*, 19/11/2010 consultato il 02/08/2016.

scientifico, arrivando a paragonare il giornalista ad uno scienziato [Meyer 2006]. Un paragone non così azzardato, stando anche alle parole del fisico statunitense Lawrence Cranberg, secondo cui il giornalismo è in se stesso scienza e un giornalista adeguatamente preparato, responsabile, è uno scienziato di professione poiché, in maniera non dissimile, scienziati e giornalisti procedono nelle stesse fila e sono a servizio del comune bisogno dell'umanità di condividere conoscenze e intendimenti [Cranberg 1989]. Almeno nei progetti di Meyer, dunque, si tratta di applicare il metodo scientifico al giornalismo. A tal proposito Robert Niles, ex direttore di OJR e cronista del *Los Angeles Times*, definisce metodo scientifico un modello che permette di raccogliere le informazioni, metterle alla prova e condividerle<sup>6</sup>. Porre una domanda cognitiva circa l'ambito di interesse, informarsi, salendo "sulle spalle dei giganti" usando la fortuna espressione di Giovanni di Salisbury, formulare un'ipotesi e metterla alla prova, raccogliendo dati o effettuando esperimenti ed osservandone gli esiti, analizzando le evidenze che emergono, accettando o confutando l'ipotesi iniziale, ed infine rendere pubblici la metodologia seguita ed i dati ottenuti. Sono questi gli step fondamentali della ricerca condotta secondo il metodo scientifico, secondo Niles. In che modo giornalismo e scienza si assomigliano? Stando a quanto affermato da Matt Thompson<sup>7</sup> è la ricerca della verità ad avvicinare le due discipline.

Un saggio del 1983 firmato da Cole, Combs, Stevenson e Shaw individuava nella subalternità della notizia all'evento la principale critica al giornalismo. Più di trent'anni dopo tale lettura rimane attuale, benché l'informazione sia alle prese anche con altri problemi, primo tra tutti il complesso rapporto con il digitale e la sua tipica orizzontalità che rende produttori chi prima era consumatore<sup>8</sup>. Tornando alle parole

---

<sup>6</sup> Niles, R. «A journalist's guide to the scientific method, and why it's important», 23/08/2011, consultato il 10/08/2016.

<sup>7</sup> Thompson, M. «What Journalist can learn from scientist and the scientific method», su *Poynter*, 01/09/2011, consultato il 10/08/2016.

<sup>8</sup> Il termine *prosumer*, coniato da Alvin Toffler, è una crasi di *producer* e *consumer* ed indica un consumatore che è al tempo stesso produttore. Come si legge su Treccani alla pagina dedicata, il neologismo nasce per descrivere il comportamento dei consumatori nell'epoca che vede l'uscita dal classico modello di *mass communication*. Anni in cui la produzione abbandona la serialità per provare ad aderire in maniera più concreta ai diversificati gusti del pubblico, che indirettamente entra a far parte del processo produttivo. Il termine *prosumer* si riferisce però anche alla fase di rivoluzione digitale in corso da un quindicennio, indicativamente dopo il crollo delle *dotcom*, quando le piattaforme che andranno a profilare ciò che sarà chiamato web 2.0 prevedono la stretta interazione con gli utenti, intesi come produttori in maniera diretta.

di Cole e degli altri tre studiosi statunitensi, alla subalternità all'evento consegue una scarsa profondità di indagine, a cui i professionisti dell'informazione cercano di rimediare avvicinandosi ai dati, cioè all'analisi della società attraverso metodi precisi, quantitativi. Con il computer si rendono possibili analisi di variabili economiche, sociali e politiche ed il loro andamento nel tempo [Cole *et al.* 1983]. Il giornalismo di precisione, così lo chiamavano allora gli autori del saggio, può essere la risposta al bisogno di informazione più informativa che la società già allora aveva cominciato a percepire. Un giornalismo in grado di spiegare, oltre al chi, al cosa, al quando, al dove ed al come, il perché, di sviscerare con l'ausilio dei dati i legami tra una singola situazione e un insieme di eventi.

È stato questo il metodo che ha portato alla ribalta le prime inchieste con i dati, a cominciare dal report<sup>9</sup>, curato da Philip Meyer e altri cronisti del *Detroit Free Press*, sulla rivolta razziale dei neri a Detroit nel 1967, capace di vincere il premio Pulitzer in giornalismo locale, fino al lavoro<sup>10</sup> di Bill Dedman dell'*Atlanta Journal-Constitution* sui pregiudizi razziali delle banche, maggiormente restie a concedere prestiti ai cittadini di colore rispetto ai bianchi di Atlanta. Altra inchiesta in grado di vincere il Pulitzer adoperando numeri e metodi di analisi quantitativi è stata *What went wrong*<sup>11</sup>, approfondimento del 1992 di Stephen Doig sugli effetti dell'uragano Andrew abbattutosi sulle coste della Florida meridionale, sul sudovest della Louisiana ed il nordovest delle Bahamas. Incrociando quattro diverse basi di dati (i rapporti sugli accertamenti post uragano, le imposte patrimoniali del 1992 sugli edifici, il catasto della contea, e le licenze edilizie concesse sulle aree edificabili), Stephen Doig ed il suo team al *Miami Herald* furono in grado di scorgere una correlazione causale negli effetti dell'uragano: gli edifici più recenti, quelli costruiti dal 1979 in poi, sembravano aver subito i maggiori danni, e l'impressione che se ne ebbe fu quella di un progressivo allentarsi della fermezza del Regolamento Edilizio della Florida Meridionale dietro a pressioni di speculatori.

---

<sup>9</sup> Meyer *et al.*, «The People beyond 12th street: a survey of attitudes of Detroit Negroes after the riots of 1967», in *Detroit Free Press*, 1967. L'inchiesta dimostra, attraverso i numeri, quanto infondata fosse la teoria secondo cui i rivoltosi sarebbero stati in maggior parte provenienti dai ceti più disperati e meno istruiti della scala sociale statunitense.

<sup>10</sup> Dedman, B. «The Color of Money», in *Atlanta Journal-Constitution*, 1988.

<sup>11</sup> Doig, S. «What Went Wrong», in *Miami Herald*, 1992.

Ma come si articola in concreto un'inchiesta con i dati? Tipicamente il *data journalist* lavora seguendo tre momenti differenti. Dopo aver definito l'ambito di ricerca, si occupa innanzitutto di raccogliere i dati: in questa fase l'obiettivo è quello di avere a disposizione l'intero materiale su cui lavorare. È evidente come, in virtù dell'enorme varietà offerta dalla digitalizzazione dell'informazione, i dati possano presentarsi in maniera diversa: semplici numeri, come nel caso di studio proposto nel capitolo 3, coordinate geografiche, informazioni provenienti da censimenti o da questionari su abitudini o comportamenti, dati meteorologici o semplici orari, e via scorrendo. Tali dati possono dunque provenire da fonti predisposte alla loro pubblicazione, come nel caso di pubbliche amministrazioni o di enti di ricerca, oppure possono essere raccolti attraverso il cosiddetto processo di *scraping*, vale a dire l'estrazione dal web, come ad esempio nel caso in cui le informazioni siano trasformate in tabelle da pdf o html. La raccolta dei dati, qualunque essi siano, porta il professionista ad avere una tabella strutturata, ossia un dataset utilizzabile attraverso fogli di calcolo (come Microsoft Excel o LibreOffice Calc). In un secondo momento, immediatamente successivo alla raccolta, il giornalista si premura di pulire i dati, ovvero di ricondurli a un formato univoco che ne consenta l'operabilità. Le informazioni contenute nei fogli di calcolo devono avere tutte le stesse caratteristiche, essere espresse in uguali unità di misura ed essere memorizzate secondo gli stessi formati. La seconda fase del lavoro di *data journalism* richiede l'analisi od interpretazione dei dati: fondamentale è intanto l'approcciarsi ai dati raccolti in modo scettico, privo di preconcetti. Il giornalista, di fronte ai dati, deve preoccuparsi di farli esprimere in maniera naturale, prima di confrontarsi con le proprie aspettative e conoscenze pregresse. Michael Blastand, freelance scozzese collaboratore per la BBC, lo ribadisce con fermezza nel suo contributo al testo *The Data Journalism Handbook*, affermando che, con l'enorme mole di dati a disposizione, il *data journalism* sia in grado di apportare un limitato valore aggiunto al giornalismo tradizionale se non si dispone di una mentalità aperta, dal momento che i dati sono oggettivi solo nella misura in cui vengono trattati come tali [Bounegru *et al.* 2012: 152]. Inoltre è opportuno tenere a mente che un dataset è in grado di fornire risposte soltanto se è stato predisposto con adeguati record e variabili: in altre parole, prima di attendersi dai dati informazioni di un certo tipo, e quindi di

interrogarlo in un certo modo, è indispensabile comprendere quanto effettivamente in grado di svelare. Non diversamente da un testo, un database sembra sottostare alla valutazione dell'autenticità pragmatica, intesa secondo quanto teorizzato da Topolski, la quale impone di «circoscrivere l'insieme delle domande cognitive cui il documento è in grado di rispondere e con quale autorevolezza» [Cardano 2003: 68]. Si tratta perciò di definire l'area di autenticità del documento, ossia «la somma di quei problemi su cui quella data fonte è in grado di rispondere modo veridico» [Topolski 1973; trad. it. 1975, 501]. Sempre in seno al discorso sull'analisi e l'interpretazione dei dati può essere utile riportare quanto sostiene Simon Rogers, docente di *data journalism* all'università di Berkeley, San Francisco. Secondo l'autore di *Facts are Sacred* la preparazione dei giornalisti del futuro dovrà avvenire insegnando non tanto gli strumenti del giornalismo coi dati, destinati ad essere rimpiazzati da strumenti via via sempre più avanzati, quanto piuttosto un metodo, ovvero l'attitudine a ricercare storie nei numeri, in modo non dissimile da quanto avviene tradizionalmente con persone in carne ed ossa.

Un discorso a parte, perché articolato su almeno due piani, lo merita la terza e conclusiva fase del processo di costruzione di una notizia attraverso i dati. Una volta raccolti ed elaborati i dati giunge il momento di pubblicare la storia. Si tratta del momento maggiormente creativo perché, ed è il primo piano di cui sopra, il racconto passa attraverso la visualizzazione dei dati. Con l'ausilio di istogrammi, grafici, diagrammi a linea o a torta, mappe, ed ogni altro tipo di supporto, è opportuno far parlare i numeri. Si tratta di quel variegato apparato iconico che prende il nome di infografica, utile a rendere più avvincente la storia e a snellire il racconto testuale [Bounegru *et al.* 2012]. Una definizione pittoresca ma emblematica di infografica la fornisce David McCandless, autore di *Information is beautiful*. «Siamo sommersi dalle informazioni – sostiene il *data journalist* e *information designer* londinese – e ciò di cui abbiamo bisogno forse sono degli schemi colorati e ben fatti che ci aiutino a navigare nel mare di dati, una sorta di moderno atlante» [McCandless 2011]. E proprio il ricorso alla metafora dell'atlante mi porta ad introdurre il secondo piano relativo alla pubblicazione della notizia attraverso le immagini: visualizzare l'informazione, infatti, risulta spesso cruciale già nella fase di interpretazione e

analisi dei dati, svelando correlazioni, dipendenze e legami tra fenomeni che espressi in semplici sequenze di zero e di uno difficilmente potrebbero emergere. Il consiglio degli autori di *The Data Journalism Handbook*, ancora oggi il manuale di riferimento per chi intende avvicinarsi a questa branca del giornalismo, è infatti quello di ricorrere alla visualizzazione grafica dei dati nel corso dell'indagine, e non rimandarla ad una volta finito il pezzo, per poter scovare nuovi filoni degni di approfondimento.

Per quanto riguarda la pubblicazione della storia è naturalmente buona abitudine quella di non limitarsi all'utilizzo dell'apparato visivo, quanto piuttosto integrare il testo, redatto secondo i canoni del giornalismo tradizionale (e con questo naturalmente non si vuole intendere che testi giornalistici rispettino standard simili tra televisione, carta stampata e web, quanto piuttosto che indipendentemente dall'utilizzo dei dati la cura nella stesura dei testi è imprescindibile), con grafici e infografiche. Il racconto si disloca pertanto su diversi piani, anche grazie alle potenzialità offerte dagli strumenti utilizzati: a seconda che il supporto sia cartaceo piuttosto che digitale o televisivo è possibile utilizzare stratagemmi differenti, dall'utilizzo delle tecniche di impaginazione alla *user experience*, fino alla multimedialità e all'interazione con l'utente, o ancora ricorrere alla cosiddetta *gamification*<sup>12</sup>.

Un ulteriore aspetto da tenere in debita considerazione nella fase di pubblicazione della storia riguarda infine l'opportunità di rendere disponibili i dati utilizzati nel corso dell'analisi, cioè di offrire al lettore la possibilità di esplorare il materiale grezzo su cui il giornalista ha lavorato. Sono almeno quattro le ragioni che spingono in questa direzione. Per quanto riguarda il rapporto con i lettori, stando a quanto sostiene Cheryl Phillips del *Seattle Times*, i vantaggi del condividere i database e le informazioni trattate nel pezzo stanno innanzitutto in un sensibile rafforzamento della credibilità del giornalista e della redazione, ed in secondo luogo nella

---

<sup>12</sup> Con il termine inglese *gamification* si è soliti indicare l'applicazione e l'utilizzo di tecniche derivate dal gioco e dall'intrattenimento in ambiti esterni al gioco stesso. Tali meccaniche si declinano in particolare nel web design, dove la cura degli aspetti di ludicizzazione consentono una migliore esperienza d'uso all'utente. Nell'ambito di interesse del presente lavoro si può intendere *gamification* come le opportunità di utilizzo, navigazione e consultazione messe a disposizione dei fruitori dell'informazione.

possibilità, da parte dell'audience stessa, di interagire con la redazione, muovere critiche e dare consigli su ambiti da esplorare in futuro. A questo si aggiungono altre due ragioni, evidenziate da Steve Doig, docente della *Walter Cronkite School of Journalism* all'università dell'Arizona. La prima ha a che vedere con il metodo scientifico di cui ho parlato in precedenza: la condivisione dei dati, e la puntuale ricostruzione degli step seguiti, consentirebbe ad altri ricercatori e giornalisti di ripetere l'analisi [Bounegru *et al.* 2012: 179], aggiungendo nuove evidenze, confutando eventuali teorie, o confermandone la bontà, il tutto seguendo le orme tracciate da quanto succede abitualmente nelle scienze dure. Ne parlerò più diffusamente nel prossimo paragrafo, ora mi sembra più dirimente riportare la seconda ragione citata da Steve Doig, che fa riferimento ad un potenziale beneficio economico del pubblicare i dati analizzati sul sito web del giornale: coinvolgere i lettori, dice il docente statunitense, può generare traffico e alimentare le conseguenti ricadute economiche [Bounegru *et al.* 2012: 179].

Per concludere il breve discorso di inquadramento tracciato per il *data journalism* può rivelarsi utile aggiungere una considerazione ultima sul metodo, emersa nel corso di un recente lavoro condotto con il dottor Tiplado. «La corretta applicazione delle procedure di *data journalism* non può prescindere dal rispetto e dalla costante verifica di quelle prescrizioni metodologiche che attribuiscono ai risultati ottenuti i requisiti che la ricerca sociale esige in quanto ricerca scientifica» [Tiplado e Gritti 2015: 38]. In particolare si tratta delle nozioni di validità e di attendibilità. Secondo le definizioni offerte da Corbetta, validità si riferisce «al grado col quale una certa procedura di traduzione di un concetto in variabile effettivamente rileva il concetto che si intende misurare» [Corbetta 1999: 125], mentre il concetto di attendibilità ha a che vedere con «il grado con il quale una certa procedura di traduzione di un concetto in variabile produce gli stessi risultati in prove ripetute» [*ibidem*]. Considerata la facile identificazione del dato, inteso come numero, con il concetto di variabile, inteso come prodotto della traduzione dei concetti secondo lo schema di Lazarsfeld [Lazarsfeld 1967] risulta di cruciale importanza nel *data journalism* tenere a mente i concetti di validità ed attendibilità, onde evitare di portare l'analisi su un piano di inconsistenza teorica.

### 1.1.1. Il *data journalism* e i dati della Pubblica Amministrazione

Il presente paragrafo prende spunto da un articolo di Colin Porlezza, professore associato all'*Institute of mass communication and media research* dell'università di Zurigo, apparso su *Problemi dell'Informazione* lo scorso aprile, intitolato *Dall'open journalism all'open government? Il ruolo del data journalism nella trasparenza e nella partecipazione*. Il suo saggio ribalta, per certi versi, la questione tra *open government* e giornalismo con i dati. Abituati ad intendere l'apertura delle pubbliche amministrazioni come punto di partenza per un controllo del governo, Porlezza si chiede se, ed in che termini, il *data journalism* possa avere un impatto nel facilitare l'accesso ai dati della pubblica amministrazione. Secondo l'autore «l'idea che il *data journalism* possa creare e favorire una cultura di maggior apertura e trasparenza nella pubblica amministrazione si fonda sul concetto di *open journalism* [...] (inteso) come un nuovo concetto di giornalismo basato sull'apertura della redazione al pubblico e sulla collaborazione con lo stesso» [Porlezza 2016]. Di particolare interesse, a proposito del giornalismo aperto, è il decalogo<sup>13</sup> stilato da Alan Rusbridger nel 2012, da caporedattore del britannico *The Guardian*. Sono soprattutto gli ultimi due punti a rappresentare il collante con quanto scritto in precedenza relativamente al metodo scientifico e all'utilità di condividere il materiale su cui si è condotta l'analisi. Si legge, infatti, che «la pubblicazione può essere l'inizio del processo giornalistico piuttosto che la conclusione» e che l'*open journalism* «è trasparente e aperto alle sfide – comprese la correzione, i chiarimenti e le aggiunte». Naturalmente, ma è utile ricordarlo una volta in più, l'apertura del giornalismo significa non soltanto offrire al lettore la possibilità di partecipare al processo di costruzione della notizia, quanto anche rendere trasparente il proprio lavoro, pubblicare le proprie risorse e divulgare i diversi passi intrapresi [Porlezza 2016]. Applicare la cultura aperta al *data journalism* può rappresentare un paradigma nuovo dunque sul piano dei contenuti, in nome del *crowdsourcing*, come dimostrano recenti casi in nord Europa, tra Norvegia [Karlsen e Stavelin 2014], Svezia [Appelgren e

---

<sup>13</sup> Rusbridger, A. «Q&A with Alan Rusbridger: the future of open journalism», su *The Guardian*, 25/03/2012, consultato il 19/08/2016.

Nygren 2014] e anche Regno Unito [Daniel e Flew 2010], sebbene sia utile tenere a mente quanto, di diverso, vi sia nei modelli di giornalismo<sup>14</sup> in Europa [Hallin e Mancini 2004]. Se il *crowdsourcing*, e quindi la raccolta collaborativa di idee e di contenuti proposti dal pubblico e da sviluppare in seno ad una redazione giornalistica, può offrire un prezioso contributo nel rilancio del mondo editoriale, Porlezza sottolinea tuttavia quanto un giornalismo prodotto dai dati possa ampliare il divario tra professionisti e audience. Il *data journalism*, richiedendo, oltre alle consuete norme deontologiche della professione, competenze tecniche di natura tecnologica, rischierebbe di formare una élite di professionisti e di alimentare quel senso di distacco tra produttori e consumatori di informazione già in atto, come testimoniato ad esempio dai dati sulla vendita dei quotidiani.

A proposito della domanda cognitiva alla base del saggio, ovvero la capacità del *data journalism* nel facilitare l'accesso ai dati dell'amministrazione pubblica, Porlezza identifica in due differenti ambiti il grado di incisività di questa pratica. Il primo riguarda il pubblico: l'impatto del giornalismo con i dati è legato al successo in termini di audience. Un'ottima inchiesta che non incontra un pubblico adeguato diventa la migliore inchiesta non letta [Porlezza 2016]. Il secondo aspetto riguarda gli editori perché, se non si riesce a catalizzare la loro attenzione, le potenzialità di questo giornalismo rimarranno inespresse [*ibidem*] ed i modelli di sostenibilità economica un grande punto interrogativo.

L'intero impianto metodologico del saggio di Porlezza regge su interviste semi-strutturate a *data journalists* attivi in organizzazioni che si occupano di giornalismo con i dati, in particolare *Hacks/Hackers*<sup>15</sup>, *Open Polis*<sup>16</sup>, *DataJournalism.it*<sup>17</sup> e

---

<sup>14</sup> Daniel Hallin e Paolo Mancini teorizzano l'esistenza di tre modelli di sistema dei media. Uno lo definiscono pluralista-polarizzato, caratterizzato da un giornalismo politicamente schierato, orientato al commento e alimentato dai sussidi economici della politica, identificandolo nei paesi mediterranei, dal Portogallo alla Grecia, passando per Spagna, Francia ed Italia. Il modello che gli autori definiscono democratico-corporativo, tipico dei paesi centro-nord europei (dall'Austria alla Norvegia), sarebbe caratterizzato dalla tutela della libertà di stampa e dalla tendenza verso una stampa commerciale neutrale, mentre il modello cosiddetto liberale, espressione dei paesi anglosassoni, degli Stati Uniti e del Canada, presenterebbe caratteristiche di forte professionalizzazione e limitato intervento statale.

<sup>15</sup> <http://hackshackers.com/>, consultato il 24/07/2016.

<sup>16</sup> <http://www.openpolis.it/>, consultato il 24/07/2016.

<sup>17</sup> <http://www.datajournalism.it/>, consultato il 24/07/2016.

*Spaghetti Open Data*<sup>18</sup>. L'analisi delle conversazioni evidenzia, secondo l'autore, uno scarso impatto, da parte del *data journalism*, nel «favorire l'accesso ai dati pubblici e (nel) trasmettere alla pubblica amministrazione i valori di trasparenza e di partecipazione» [Porlezza 2016: 185]. Tuttavia, sottolinea l'autore, il giornalismo coi dati, attraverso inchieste e l'attivismo dei professionisti che ne auspicano la diffusione, è in grado di «inserire il problema dell'apertura nell'agenda pubblica e dunque dirigerli l'attenzione necessaria» [*ibidem*]. Da tale consapevolezza è uscito rafforzato il progetto della mia tesi, costruito sull'analisi dei prodotti mediali e sull'agenda dei media. Il paragrafo che segue intende offrire gli strumenti teorici indispensabili ad affrontare il caso di studio proposto nel terzo capitolo.

## **1.2 I *media studies***

Un paragrafo dedicato ai contributi che negli anni si sono moltiplicati nell'ambito dei media non può evidentemente prescindere da un inquadramento chiaro di cosa sia lecito intendere con tale parola, *media*. A partire dalla prima metà del XX secolo, scrive Paccagnella, nel panorama degli studi sulla comunicazione compare questo termine, plurale del latino *medium*, a cui abitualmente viene attribuito il significato di mezzo o di strumento, a cui si associa presto l'inglese *mass*, andando ad intendere i mezzi di comunicazione di massa [Paccagnella 2010]. Come primo, vero, *mass medium* l'autore identifica la radio, quello strumento così simile ad un telegrafo, almeno nelle intenzioni di utilizzo<sup>19</sup>, ma privo dei fili, in grado di diffondere il proprio segnale attraverso onde radio. Ciò che di fatto rende la radio un mezzo di comunicazione di massa è la natura della trasmissione del messaggio. Viaggiando attraverso l'etere, e non più via filo, il messaggio era potenzialmente ricevibile da chiunque, secondo quello che sarebbe stato poi chiamato modello di comunicazione

---

<sup>18</sup> <http://www.spaghettiopendata.org/>, consultato il 24/07/2016.

<sup>19</sup> La radio, proposta agli inizi del Novecento da Guglielmo Marconi in applicazione delle ricerche di Maxwell, Hertz, Edison e Volta, era stata inizialmente intesa come strumento di trasmissione di messaggi in ambito militare, specialmente per consentire la comunicazione con e tra le navi da guerra [Paccagnella 2010].

*broadcast*<sup>20</sup>. Una comunicazione pertanto priva di un destinatario preciso, come sarebbe stato, di lì a poco, la televisione che nacque tra la fine degli anni '20 e l'inizio dei '30. Intenzionato ad offrire un quadro intenso e sostenibile degli studi dei media, mi appoggio pertanto alla definizione del sociologo britannico Denis McQuail, secondo cui «nella comunicazione di massa gli emittenti sono quasi sempre [...] comunicatori di professione [...], il contenuto simbolico o messaggio è spesso fabbricato in modi standardizzati [...]» [McQuail 2000, trad. it. 2001: 55]. Da una tale definizione emerge immediatamente quanto uno degli interessi di chi si occupa di produrre informazione mass mediale sia la capacità del messaggio di rivelarsi inclusivo nei confronti del pubblico, capace cioè di indirizzarsi a fasce diversificate della popolazione. La cosiddetta teoria ipodermica<sup>21</sup> ha rappresentato il primo tentativo di approccio sistematico allo studio dei mass media [Paccagnella 2010], tracciando in un certo senso il solco su cui si sono inseriti i primi contributi rilevanti. Quanto può essere utile evidenziare fin da ora è il modello delle 5 W, elaborato da Harold Lasswell, che individua in cinque variabili (*who, what, whom, where, what effects*) le componenti fondamentali in cui è possibile scomporre l'oggetto di comunicazione [Lasswell 1927]. Da questa suddivisione sono nate le diverse correnti di studi sulla comunicazione di massa: alcune ricerche si sono concentrate sul *chi*, pertanto sull'emittenza, altre sul *cosa*, quindi sul messaggio, altre ancora sul *a chi*, focalizzandosi sui pubblici, fino all'analisi degli *effetti* dell'esposizione ai messaggi mediali. La parte di studi sul *come*, inteso come *con quale mezzo*, è il filone al momento più esplorato, anche in virtù di contributi come quello di Andrew Chadwick. Secondo l'autore, co-direttore del New Political Communication Unit del dipartimento di politica e relazioni internazionali alla Royal Holloway londinese, il sistema dei media si sta arricchendo in complessità ed articolazione, in una parola ibridando. La compenetrazione dei media nuovi e vecchi,

---

<sup>20</sup> Il modello *broadcasting*, nella comunicazione, identifica quel tipo di comunicazione, tipica della prima era della scarsità dominata da radio e televisione [Vellar 2015], in cui obiettivo delle emittenti era tipicamente il guadagnare pubblico e di conseguenza ricavi pubblicitari.

<sup>21</sup> La teoria dell'ago ipodermico risale ai primi decenni del novecento e sostiene che mezzi di comunicazione come radio, stampa e cinema fossero in grado di manipolare le persone. Tale visione, derivata in particolare dall'esperienza propagandistica della Prima Guerra Mondiale, identificava il messaggio come una sorta di proiettile, o di ago, in grado di «inocularsi sotto la pelle delle persone» [Paccagnella 2010: 98].

online ed offline, costituiscono il terreno privilegiato dell'analisi sui mezzi di comunicazione di oggi [Chadwick 2013].

Tornando alla teoria dell'ago ipodermico va detto che, pur nella limitatezza di un approccio ingenuo e semplificato, ha rappresentato la base per le ricerche successive. In particolare, sin dagli anni quaranta, alcuni studi hanno cercato di superare il modello stimolo-risposta come teorizzato da Lasswell, partendo dall'evidenza di reazioni diverse, da parte di persone diverse, all'esposizione dello stesso messaggio mediale. Carl Hovland ha il merito di essere stato il primo a prendere in considerazione l'intervento di variabili capaci di spiegare la varietà delle risposte individuali, costruendo il cosiddetto modello S-IV-R<sup>22</sup>, e sottolineando anche la capacità selettiva dei destinatari, in grado di scegliere a quali messaggi esporsi e quali parti del messaggio elaborare e memorizzare. Il fatto che Hovland fosse uno psicologo di professione è emblematico di quanto, gli studi sulla comunicazione e la ricerca sui mass media, abbia fin dal principio toccato discipline afferenti ad ambiti variegati, dalla sociologia sicuramente, alla semiotica, fino a filosofia, psicologia, storia e linguistica.

Se i primi studi conducono alla consapevolezza che «non è possibile analizzare né comprendere gli effetti dei mass media senza considerare il contesto sociale in cui agiscono» [Paccagnella 2010: 104], con la *two-step flow of communication* [Lazarsfeld *et al* 1944] si aggiunge un ulteriore tassello al quadro. Si coglie infatti l'esistenza di un doppio livello nella ricezione e nella fruizione del messaggio mediale, attraverso cui il pubblico recepisce il contenuto grazie all'intervento di leader di opinione, individui attivi a livello di partecipazione sociale, interessati agli argomenti di discussione e, soprattutto, percepiti come degni di fiducia. La teoria del flusso di comunicazione a due stadi deve il nome proprio al fatto che il messaggio raggiunge prima gli opinion leader, che lo interpretano ed elaborano, ed in secondo momento le persone comuni che dai leader di opinione traggono alcune chiavi di lettura e a cui si rivolgono per avere conferme. Quello di Lazarsfeld è il primo

---

<sup>22</sup> Il Modello S-IV-R, ovvero Stimolo-Variabili Intervenienti-Risposta presuppone che le variabili in gioco nella fruizione del messaggio mediano l'impatto dello stimolo. Tale lettura spiegherebbe la varietà delle risposte individuali, l'effettiva inefficacia in taluni casi, ed altre circostanze di effetto boomerang, con risvolti imprevisti ed indesiderati [Paccagnella 2010].

contribuito a rivedere, in una certa misura, la centralità del messaggio in quanto tale, e aprendo il varco a studi che si concentrano maggiormente sull'aspetto relazionale della comunicazione. In questo senso può essere valutata la piega verso una lettura orientata dallo struttural-funzionalismo statunitense, che conduce alla formulazione, da parte di Elihu Katz, della *uses and gratification theory* [Katz 1974], secondo cui l'uso che il pubblico fa dei mezzi di comunicazione di massa è orientato alla soddisfazione dei propri bisogni, per riceverne una gratificazione [Paccagnella 2010].

La ricerca europea sui media, nel frattempo, segue altri percorsi in particolare all'insegna della teoria critica elaborata dalla Scuola di Francoforte. Horkheimer e Adorno sono i due autori che danno sostanza alla *mass media research*, introducendo il concetto di industria culturale. Con questo termine i due autori identificano il complesso dei mezzi di comunicazione di massa, inizialmente cinema, radio e stampa, a cui negli anni cinquanta si aggiunge la televisione, concepito come una vera e propria industria, le cui forme di organizzazione e produzione e i cui prodotti sono «standardizzati e omologati alle necessità di un consumo culturale di massa» [Paccagnella 2010: 111]. Quelli che un tempo erano chiamati destinatari dei messaggi, che oggi chiameremmo utenti o magari *prosumer*, la teoria critica li definisce consumatori, niente altro che l'ultimo anello di una catena di montaggio che offre prodotti (messaggi mediali) preconfezionati, i cui standard sono determinati dalla capacità di raccolta pubblicitaria e di penetrazione sul mercato. È invece la scuola di Birmingham a costruire quel filone di ricerca definita amministrativa, all'interno del più ampio spettro dei *Cultural Studies*, secondo la quale la dimensione culturale non rappresenta un patrimonio immutabile, quanto piuttosto un insieme di processi sociali, frutto dell'agire e del pensare delle persone comuni. In questo senso, ovvero di cultura come opera in divenire, arricchita dal contributo dei singoli componenti della massa, vanno letti i principali contributi teorici alla ricerca sui media, che si articola in due filoni principali. Il primo riguardante la produzione, che gli autori riconoscono come espressione dell'ideologia dominante; il secondo riferito al consumo mediale, inteso come «parte della lotta delle classi subalterne per l'affermazione della propria specifica identità» [Paccagnella 2010: 115]. Questo

fecondo filone della ricerca amministrativa si è rivelato in grado di produrre modelli come l'*encoding-decoding* [Hall 1980] e le *audiences studies* [Morley 1980].

La breve carrellata offerta in queste pagine, inevitabilmente incompleta per ragioni di spazio e opportunità tematica, è stata pensata per garantire una sufficiente consapevolezza della complessità della questione relativa alla *mass media communication*. Con il medesimo intento di non affrontare in maniera ingenua il caso di studio presentato nel terzo capitolo, procedo ora a definire in maniera più articolata la questione sull'analisi del contenuto, conscio altresì della necessità di riprendere alcuni di questi argomenti proprio nel capitolo dedicato al caso di studio, restringendo pertanto il loro approfondimento in questa sede.

### **1.3 L'analisi del contenuto**

*«L'analisi del contenuto è un insieme ampio ed eterogeneo di tecniche manuali o assistite da computer di interpretazione contestualizzata di documenti provenienti da processi di comunicazione in senso proprio (testi) od di significazione (tracce e manufatti), avrei come obiettivo finale la produzione di inferenze valide e attendibili»*  
[Tipaldo 2014: 42]

Le parole di Tipaldo cercano di condensare una serie di definizioni, metodi, pratiche e strumenti [Markoff *et al.* 1975; Livolsi e Rositi 1988] spesso problematici già nell'identificare in maniera inequivocabile cosa sia da intendersi con il termine contenuto. Da qui è opportuno partire, appoggiandomi alla triplice distinzione offerta da Klaus Krippendorff nella definizione di *contenuto* rispetto al testo in analisi. In particolare l'autore tedesco propone un distinguo tra le definizioni che, negli anni, si sono premurate di considerare il contenuto come una componente interna del testo, come una proprietà della fonte del testo stesso, o come un dato emergente nel processo di analisi di un testo in relazione ad un determinato contesto [Krippendorff 2004, in Tipaldo 2014: 31]. La questione, lungi dall'essere una mera disquisizione terminologica, rappresenta l'inevitabile punto di partenza perché «da ciascuna classe discende una concettualizzazione del termine *contenuto* decisamente tipica [...] e

dunque si avranno procedure di costruzione della documentazione empirica significativamente differenti» [Tipaldo 2014: 31]. Ogni modo di intendere il contenuto determina un approccio differente alle tecniche ed agli strumenti di analisi. Vederlo come componente interna del testo, ad esempio, sottintende ad una analisi intesa come «controllo di determinate ipotesi su fatti di comunicazione [...] e che a tale scopo utilizza procedure di scomposizione analitica e di classificazione, normalmente a destinazione statistica» [Livolsi e Rositi 1988: 66]. Tale lettura, tuttavia, sembra peccare di realismo ingenuo nell'assumere «una scontata linearità tra i processi di messa in forma del messaggio da parte dell'emittente e di interpretazione a carico del destinatario» [Tipaldo 2014: 35]. La seconda classe di definizioni identificata da Krippendorff, come detto, associa il contenuto ad aspetti legati a stati o proprietà ascrivibili all'emittente, ovvero alla fonte che ha prodotto il testo, concependo la sua analisi come una tecnica che descrive le peculiarità dei testi rispetto (dicendolo *à la* Lasswell) al cosa viene comunicato a chi e come, consentendo al ricercatore di inferire conclusioni relative alla fonte e alle sue intenzioni comunicative, oltre alle eventuali conseguenze [Krippendorff 2004]. L'autore tedesco, tuttavia, propone una personale lettura del contenuto, comunemente definito come approccio epistemologico: in questo caso l'analisi del contenuto è definita sì come una tecnica, pertanto definita da procedure specialistiche indipendenti dall'autorità del ricercatore, ma viene chiarito come il testo non abbia qualità oggettive assolute, dal momento che il proprio patrimonio informativo non esiste indipendentemente dal ricercatore, definito come caso particolare di lettore<sup>23</sup>. Secondo Krippendorff, infatti, «analizzare un testo significa innanzitutto interpretarlo, ovvero attribuire significati alle parti di cui è composto, sottoponendo tali ipotesi di senso a un processo di confutazione testuale» [Tipaldo 2014: 39], dove

---

<sup>23</sup> A tal proposito vale senz'altro la pena investire alcune righe della presente ricerca per offrire il punto di vista di Umberto Eco. Il compianto semiologo alessandrino, ne *I limiti dell'interpretazione* [Eco 1990] propone di analizzare le intenzioni di un testo, fondandosi su una triplice distinzioni idealtipiche. In particolare Eco distingue tra *intentio auctoris*, espressione di quello che chiama Autore Empirico (o reale), di fatto insondabile ed inconoscibile, e *intentio operis*, termine col quale intende l'immagine di autore che prende forma nella mente del lettore, il quale cerca di ricostruire i significati dell'estensore originario. Infine Eco parla di *intentio lectoris*, strettamente legata al destinatario: in questo caso va tenuto in considerazione che i percorsi di senso che emergeranno dalla decodifica del messaggio potranno seguire strade anche profondamente discostanti da quanto immaginato, e previsto, dall'autore, poiché il lettore aggiungerà al consumo del messaggio il proprio bagaglio esperienziale.

con questo si termine si intende una verifica sistematica delle conclusioni cui il ricercatore giunge, riferendosi al testo come unico parametro di riferimento empirico. In questa sede tornano di attualità, al fine di valutare l'efficacia delle inferenze, i già richiamati concetti di validità e attendibilità, come definiti da Piergiorgio Corbetta. Dalla definizione di Krippendorff risalta anche l'importanza connessa al contesto di analisi: tale lettura matura dalla consapevolezza che ogni testo esiste a prescindere dal ricercatore che lo esamina, e che è stato originariamente indirizzato ad un insieme eterogeneo di pubblici, la cui fruizione risente evidentemente del contesto in cui avviene.

Sono tre, tipicamente, gli ambiti di applicazione dell'analisi del contenuto. Stando alla classificazione proposta da Holsti [1969], sulla base degli iniziali studi di Berelson [1952], gli obiettivi possono essere produrre inferenze sugli antecedenti della comunicazione, descrivere i tratti peculiari della comunicazione, e produrre inferenze sugli effetti della comunicazione. Se nel primo caso il focus è rivolto alla fonte e al processo di messa in codice del messaggio, il secondo filone riguarda piuttosto canale, messaggio e destinatario, mentre il terzo mi consente di aprire una breve parentesi su quelli che, nel corso dei *media studies*, sono stati i principali contributi allo studio degli effetti dei mezzi di comunicazione.

A tal proposito, con l'intento di chiudere il cerchio del precedente excursus di stampo strettamente sociologico, vale la pena soffermarmi su almeno tre formulazioni teoriche, a cominciare dalla più nota, l'*agenda setting* [McCombs e Shaw 1972]. Secondo gli autori i media sono in grado di influenzare l'agenda dell'opinione pubblica attraverso la definizione degli argomenti da trattare e l'enfasi che viene loro riservata. In particolare, scegliendo gli argomenti di cui parlare, adottando un determinato tipo di notiziabilità, i media sarebbero in grado di orientare, con grado variabile, l'attenzione del pubblico, evidenziando le informazioni a cui occorre prestare attenzione, suggerendo alla gente i temi attorno cui pensare. Diversa la questione per quanto riguarda l'*agenda framing*, dove con il termine *frame* Ervin Goffman [1974] intende il modo in cui una notizia è presentata al pubblico, sia da un punto di vista grafico/visivo, sia da un punto di vista di significato. Il concetto è una sorta di approfondimento dell'*agenda setting*, un suo

secondo livello, tale per cui nel modo in cui l'emittente tratta il messaggio sia contenuta la chiave di lettura per la sua decodifica. In tal modo, oltre che sul cosa pensare, i mass media sarebbero in grado di orientare la lettura, di influenzare il modo in cui pensare.

È opportuno offrire un richiamo, per completezza teorica, anche alla teoria dei differenziali di conoscenza (*knowledge gaps theory*), secondo cui le possibilità di accesso all'informazione, intesa come un vero e proprio bene economico, sono diverse per le differenti fasce della popolazione: in particolare, i segmenti sociali dotati di status socioeconomico di alto livello acquisiscono l'informazione in maniera più veloce rispetto alle fasce medio basse, alimentando un progressivo differenziale di conoscenza [Tichenor *et al.* 1970].

# Capitolo 2

## Metodo e metodologia di ricerca

Il presente capitolo, di natura metodologica, mira a descrivere la logica, le modalità e gli strumenti seguiti nel corso dell'indagine. Per quanto riguarda la logica sarà importante definire gli aspetti legislativi e legali che convergono a definire il quadro entro cui la mia tesi si inserisce. Verranno pertanto prese in considerazione, e delineati in maniera per quanto possibile esaustiva, le tematiche di *open data* e di *open government*, innanzitutto, utili a creare un background sul quale successivamente descrivere il contesto giuridico inerente la trasparenza delle pubbliche amministrazioni.

La seconda parte del capitolo verterà piuttosto sulla puntuale ricostruzione dei passi seguiti durante la mia analisi, da un punto di vista sia teorico che tecnico. Saranno così riprese le modalità di raccolta, analisi e descrizione dei dati sulle pubbliche amministrazioni, e le tecniche seguite nel corso dell'analisi dei testi adoperati come corpora.

Nell'appendice a secondo capitolo, infine, troveranno collocazione le pagine che consentiranno di entrare a contatto con il mondo *linked data*, la tecnologia di cui si occupa Synapta già più volte richiamata nelle pagine di questa tesi.

### 2.1. I dati nelle Pubbliche Amministrazioni

«Le pubbliche amministrazioni (PA) hanno una caratteristica comune: accumulano dati. Li accumulano per poter funzionare e per dimostrare di aver ben funzionato» [Morando 2013: 7]. L'avvento della tecnologia di massa, caratteristica degli ultimi anni, ha implicato che questi dati cominciassero a presentarsi in formato digitale.

Dati cartografici, catastali, informazioni meteorologiche, orari delle farmacie, banche dati dei tempi di percorrenza medi dei mezzi pubblici: sono tutti esempi di informazioni del settore pubblico o, in inglese, *public sector information* (PSI). Come sottolinea Morando in apertura del *Libro Bianco PSI*, spesso una singola informazione ha limitata utilità, ma il valore aggiunto può nascere dalla raccolta organizzata e dalla successiva combinazione di più informazioni con fonti diverse<sup>24</sup>.

Accesso e riutilizzo sono le parole chiave su cui insiste il presente lavoro di tesi. Appoggiandomi alle definizioni proposte nel testo curato da Morando, con riutilizzo si intende l'uso diverso e ulteriore, rispetto all'uso istituzionale, per cui il dato è stato raccolto o generato dalla PA [Morando *et al.* 2013: 8], mentre accesso, cioè la semplice disponibilità di consultare le informazioni, è soltanto la preconditione per poter riutilizzare i dati.

Volendo organizzare l'introduzione delle tematiche in maniera più chiara possibile, si è scelto di procedere andando a descrivere *open data*, *open government* e il quadro legislativo a livello europeo ed italiano in maniera separata. Si tratta tuttavia di argomenti profondamente intrecciati gli uni agli altri, in cui persino distinguere la *consecutio temporis* risulta un esercizio impegnativo: decreti legislativi, fenomeni sociali e trend tecnologici risultano spesso sovrapposti e di cui è difficile trovare radici solide. È pertanto opportuno tenere a mente l'inestricabile legame che unisce i prossimi paragrafi della tesi.

### **2.1.1. *Open data***

Volendo preparare il terreno concettuale che sta alla base del movimento *open government* è opportuno sgomberare il campo dalle parole e cercare fatti concreti.

---

<sup>24</sup> Morando porta un esempio tanto chiarificatore sulle potenzialità del riutilizzo dei dati pubblici da suggerirmi di inserirlo in nota. Si tratta degli orari dei mezzi pubblici: molte aziende municipalizzate realizzano siti che mostrano ai cittadini come spostarsi coi loro mezzi, ma spesso tali percorsi finiscono con la tratta di competenza della specifica municipalizzata, rendendo difficile l'utilizzo di questi dati da parte di un utente interessato a viaggiare tra comuni diversi con i mezzi pubblici gestiti da aziende diverse. Tuttavia un servizio come Google Maps, alimentato anche da dati e orari dei mezzi pubblici provenienti da fonti diverse, è in grado di rispondere ai bisogni dei viaggiatori mostrando percorsi che integrano informazioni diverse.

L'apertura del governo dipende in buona misura dai dati che mette a disposizione e dal modo in cui lo fa: si parla in questo caso di *open data*, o ancor meglio di *open government data*. Secondo la definizione offerta dalla *Open Definition*<sup>25</sup> “i dati aperti sono dati che possono essere liberamente utilizzati, riutilizzati e ridistribuiti da chiunque, soggetti eventualmente alla necessità di citare la fonte e di condividerli con lo stesso tipo di licenza con cui sono stati originariamente rilasciati”. Simone Aliprandi, fondatore e responsabile del Progetto Copyleft-italia.it<sup>26</sup> riporta, nel suo testo *Il fenomeno open data. Indicazioni e norme per un mondo di dati aperti*, gli aspetti più importanti che riguardano i dati aperti. Innanzitutto la loro disponibilità ed accesso: i dati devono essere disponibili nel loro complesso, per un prezzo non superiore ad un ragionevole costo di riproduzione, preferibilmente mediante scaricamento da internet, e soprattutto devono consentire la modifica, presentandosi in un formato utile a questa operazione. Per quanto riguarda il riutilizzo e la redistribuzione i dati devono consentire l'opportunità di combinarli con altre basi di dati. Un dato aperto deve poi ammettere la partecipazione universale, con ciò intendendo l'assenza di discriminazioni né di ambito di iniziativa né contro soggetti o gruppi. Aliprandi allinea il concetto di dato aperto con quello di interoperabilità, ossia la capacità di combinare una base di dati con un'altra [Aliprandi 2014].

Non avendo, il presente contributo, l'ambizione di voler offrire una trattazione esauriente degli aspetti relativi alle tecniche utili a rendere aperti i dati, sia sufficiente ricordare che, usando le parole di Aliprandi, gli *open data* devono essere aperti sia dal punto di vista tecnico che da quello legale. In particolare i dati devono essere disponibili in grandi quantità in un formato *machine-readable*. Ad esempio, con riferimento al formato PDF, comunemente utilizzato come tecnologia per la

---

<sup>25</sup> <http://opendefinition.org/>, consultato il 15/06/2016.

<sup>26</sup> L'espressione inglese copyleft è un gioco di parole su copyright e individua un modello di gestione dei diritti di autore basato su un sistema di licenze attraverso le quali l'autore indica ai fruitori dell'opera che essa può essere utilizzata, diffusa e modificata liberamente, nel rispetto di alcune condizioni essenziali. Tecnicamente funziona attraverso la scelta, da parte dell'autore in quanto detentore dei diritti di autore, di una delle licenze disponibili e la sua applicazione all'opera. Per quanto riguarda i contenuti, le licenze più diffuse sono le Creative Commons. Attualmente ne esistono sei: imprescindibile è l'attribuzione dell'opera all'autore originario, a cui si possono aggiungere la licenza che determina la condivisione dell'opera derivata secondo gli stessi standard, quella che impedisce la creazione di opere derivate e quella non commerciale. Per approfondimenti sia rimanda a <http://www.copyleft-italia.it/>, consultato il 24/07/2016.

pubblicazione di documenti delle pubbliche amministrazioni, lo studioso fa notare come file di questo tipo siano lì leggibili da essere umani, ma sia invece molto difficile renderle utilizzabili dai computer, limitando pesantemente la capacità da parte di altri di riutilizzare quei dati.

Nel 2007, poi, l'*Open Government Working Group* ha definito gli otto principi [Reggi 2012] che rendono aperti i dati prodotti dalle amministrazioni. L'informazione fornita deve essere:

- completa, tale per cui tutti i dati pubblici, ossia non soggetti a limitazioni di privacy, sicurezza o privilegio, devono essere disponibili.
- primaria, con cui si intende che i dati vengano raccolti e pubblicati alla fonte, senza essere aggregati o modificati.
- tempestiva e aggiornata, per preservarne il valore.
- accessibile alla più ampia gamma di utenti per il maggior numero di scopi.
- *Machine-processable*, ovvero strutturata in maniera da consentirne l'elaborazione automatizzata.
- non discriminatoria, che significa che i dati siano disponibili a tutti senza la necessità di registrarsi per averne accesso.
- non proprietaria, priva pertanto di controllo esclusivo
- *license-free*, non soggetta a copyright, brevetti o marchi di fabbrica.

Morando sottolinea un corollario importante alla definizione proposta dall'*Open Definition* che, pur non essendo esplicitata, merita di essere rilevata: «non è consentita nessuna discriminazione dei riutilizzatori o delle classi di riutilizzo. In particolare, le restrizioni di tipo “non-commerciale”<sup>27</sup> non sono ammesse. Ovvero, se il riutilizzo commerciale è vietato, i dati non sono veramente aperti» [Morando 2013: 23]. Si tratta di un aspetto secondario per quello che concerne il riutilizzo ai fini giornalistici di cui il presente lavoro intende occuparsi, ma non lo è affatto nella definizione degli standard operativi del mondo dei dati aperti.

---

<sup>27</sup> Come introdotto in nota 3, si tratta di una tipologia di licenza Creative Commons. La condizione “non commerciale” vieta di utilizzare il materiale coperto da questo tipo di licenza per il perseguimento di un vantaggio commerciale o di un compenso monetario. Per approfondimenti <https://creativecommons.org/licenses/by-nc-nd/2.5/it/>, consultato il 31/07/2016.

### 2.1.2. *Open government*

«Il paradigma *open data government* si sta rapidamente affermando come strumento di trasparenza delle politiche pubbliche per migliorare i processi decisionali, rendere i cittadini più informati e consapevoli, favorire la creazione di nuovi servizi» [Reggi 2012].

Definire con precisione il concetto di *open government*, lungi dall'essere un semplice esercizio di stile, ha in questo lavoro l'obiettivo di consentire di inscrivere l'analisi che segue in un quadro teorico solido. Per quanto varie siano le definizioni che negli anni sono state proposte, quella di Lathrop e Ruma nel loro *Open Government. Transparency, Collaboration and Participation in Practice* [Lathrop D., Ruma, L. 2010] ha il merito di seguire l'evolversi del concetto. Secondo gli autori, a livello più elementare la nozione di *open government* ha a che fare con il diritto, da parte delle persone, di accedere ai documenti ed al modo di agire dei governi. Un'idea, quella del diritto dei cittadini di partecipare all'affare governativo, che trova le radici nell'Illuminismo francese e che viene sancito nella Dichiarazione di Indipendenza degli Stati Uniti (1776) e nella successiva Costituzione Americana (1787). In tale generica definizione, teoricamente condivisa da tutti i paesi democratici del mondo, convergono risvolti che meritano di essere presi in considerazione, a cominciare dagli aspetti tecnici. Il concetto di *open government* infatti sta evolvendo, spinto dal movimento per l'*open source*, verso una dimensione di partecipazione nelle procedure governative. In maniera non dissimile da come, per l'*open source*, gli utenti sono in grado di integrare ed apportare modifiche al codice sorgente di un software, nel caso del governo aperto i cittadini non si limitano ad avere l'accesso ad informazioni e documenti prodotti dalle amministrazioni, ma entrano attivamente nelle procedure del settore pubblico. Già si rende evidente come, oltre all'aspetto tecnico e tecnologico, convergano attorno l'*open government* aspetti culturali, legali e produttivi che costituiscono il terreno, per la verità non ancora completamente pronto, in cui il *data journalism* prova a crescere.

Lathrop e Ruma offrono un interessante parallelismo, almeno ad un livello semantico, tra l'evoluzione del web verso il concetto di piattaforma e quello del

governo aperto. Allo stesso modo del web 2.0, infatti, si sta facendo largo la definizione, camaleontica stando alle parole degli autori, del governo 2.0. La definizione, per la verità, non sembra essere ancora del tutto condivisa: tra chi la legge come l'utilizzo dei social media da parte delle amministrazioni, chi la vede attraverso la lente della trasparenza intesa come concessione delle API<sup>28</sup>, e chi la interpreta come l'adozione di strumenti software, dal *cloud computing* alle wiki, dal *crowdsourcing* alle applicazioni mobile, non sono pochi a mancare il nocciolo della questione. Il governo 2.0, stando agli autori, è quanto rimane del governo sfrondata dalla propria apparenza, cioè il meccanismo di azione collettiva. Si traduce quindi nell'uso della tecnologia per la soluzione di problemi collettivi a livelli locale e globale [Lathrop D., Ruma, L. 2010: 11-12]. Identificare il governo come una piattaforma significa intendere l'informazione prodotta dai cittadini e per conto di essi come linfa vitale per l'economia e la nazione. Pertanto il governo ha il dovere di trattare l'informazione come un bene della nazione. I cittadini sono in grado di proporsi come attori dell'innovazione a livello pubblico sui temi di interesse, essendo oramai dotati di competenze oltre che spinti dall'interesse a farlo. Il governo diviene così propulsore delle *best practices* piuttosto che unico produttore [ibidem].

Tale processo rappresenta l'allontanamento da quel modello che Donald Kettl definisce "governo come distributore automatico" [Kettl 2008] tale per cui i cittadini, pagando le tasse, attendono passivamente la fornitura dei servizi e si limitano eventualmente a criticare l'operato del governo. Ma cosa succede se si ripensasse all'amministrazione come gestore di un supermercato piuttosto che come un distributore automatico? La lettura di Lathrop e Ruma suggerisce un approccio sul modello bazar, come inteso da Eric Steven Raymond in *La cattedrale ed il bazar*. Nel suo testo del 1997 Raymond analizza il fenomeno open source, partendo dal progetto Linux<sup>29</sup> di Linus Torvalds, paragonando lo sviluppo del sistema operativo

---

<sup>28</sup> L'API, acronimo inglese per Application Programming Interface, in italiano Interfaccia di programmazione di un'applicazione, è uno strumento utile a rendere disponibile ad altri programmatori le funzionalità di un programma. Tipicamente, un'applicazione web può offrire delle Open API, cioè API disponibili sul web, di cui un programmatore può servirsi, integrandole nel proprio programma.

<sup>29</sup> GNU/Linux è un sistema operativo libero, espressione del movimento *freeware*. La peculiarità di Linux, rispetto ad altri sistemi operativi proprietari come Microsoft Windows e Apple Mac OS, risiede nella licenza con il quale è distribuito, che consente a chiunque, oltre che di utilizzarlo, di

aperto a quanto succede in un grande bazar e confrontandolo con i software proprietari, sviluppati secondo un modello imposto dall'alto, "a cattedrale" per usare le parole dell'autore. Tornando a Lathrop e Ruma, intendere il governo come piattaforma significa vederlo come un bazar dove i servizi offerti non sono predeterminati, bensì come un luogo (virtuale, in questo caso) in cui la comunità si scambia idee e servizi.

Slancio importante per il governo aperto fu quello alimentato dal presidente statunitense Barack Obama, non appena insediato alla Casa Bianca nel gennaio del 2009. Il *Memorandum for the Heads of Executive Departments and Agencies*<sup>30</sup> sottolinea l'impegno a creare un livello di apertura nella sua amministrazione senza precedenti, puntando su trasparenza, per consentire ai cittadini di essere informati su cosa il governo stia facendo, partecipazione, per permettere al governo stesso di prendere decisioni migliori sulla base della conoscenza diffusa all'interno della società, e collaborazione tra i diversi livelli dell'amministrazione, i settori del no profit e quello privato, con l'obiettivo di condividere le informazioni in grado di identificare le migliori opportunità derivanti dalla cooperazione. Questi tre principi sono alla base della *Open Government Initiative* americana.

## **2.2. Agenzia per l'Italia Digitale**

Le Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico, pubblicate nel 2014 dall'Agenzia per l'Italia Digitale, rappresentano il punto da cui è più utile partire nell'analisi del fenomeno riguardante i dati aperti. Prima ancora, tuttavia, può essere utile richiamare brevemente gli step legislativi che hanno portato alla predisposizione dell'Agenzia stessa.

È l'articolo 19 del Decreto Legge 22 giugno 2012, n. 83 "Misure urgenti per la crescita del Paese", ad istituire l'Agenzia per l'Italia Digitale, "sottoposta alla

---

modificarlo , copiarlo e analizzarlo. Per approfondimenti <https://www.linux.it/>, consultato il 02/08/2016.

<sup>30</sup> [https://www.whitehouse.gov/the\\_press\\_office/TransparencyandOpenGovernment](https://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment), consultato il 24/07/2016.

vigilanza del Presidente del Consiglio dei Ministri o dal Ministro da lui delegato, del Ministro dell'economia e delle finanze, del Ministro per la pubblica amministrazione e la semplificazione, del Ministro dello sviluppo economico e del Ministro dell'istruzione, dell'università e della ricerca". Tale Agenzia è preposta alla realizzazione degli obiettivi dell'Agenda digitale italiana<sup>31</sup> in accordo con l'Agenda Digitale Europea. L'Agenda digitale italiana, in particolare, rappresenta l'insieme di azioni e norme per lo sviluppo di tecnologie, dell'innovazione e dell'economia digitale<sup>32</sup> in Italia, messe a punto in accordo con l'Agenda Digitale Europea<sup>33</sup>. È a livello europeo, infatti, che si trovano i primi richiami alla questione digitale. Europa 2020<sup>34</sup>, la strategia decennale per la crescita sviluppata dall'Unione Europea, organizza in sette punti gli obiettivi da raggiungere, uno dei quali è rappresentato proprio dall'agenda digitale, intesa come mercato unico digitale per condurre l'Europa verso una crescita intelligente, sostenibile ed inclusiva. Identificati gli ostacoli che minano l'agenda digitale, tra cui la frammentazione dei mercati digitali, la mancanza di interoperabilità e la mancanza di alfabetizzazione digitale e di competenze informatiche, la Commissione Europea ha indicato nell'aumento di interoperabilità di dispositivi, applicazioni, banche dati, servizi e reti una necessità primaria<sup>35</sup> da perseguire. L'Italia ha organizzato le misure da seguire a livello nazionale nell'Agenda Digitale Italiana, affidando il compito di garantirne la realizzazione all'Agenzia. Organizzazione<sup>36</sup> che, in particolare, "contribuisce alla diffusione dell'utilizzo delle tecnologie dell'informazione e della comunicazione, allo scopo di favorire l'innovazione e la crescita economica", anche elaborando "indirizzi, regole tecniche e linee guida in materia di omogeneità dei linguaggi, delle procedure e degli standard, anche di tipo aperto, per la piena interoperabilità e

---

<sup>31</sup> Per approfondimenti si rimanda all'articolo 47 del decreto-legge 9 febbraio 2012, n. 5, convertito in legge con modificazioni dalla legge 4 aprile 2012, n. 35.

<sup>32</sup> <http://www.agid.gov.it/agenda-digitale/agenda-digitale-italiana>, consultato il 24/07/2016

<sup>33</sup> <https://ec.europa.eu/digital-single-market/>, consultato il 24/07/2016

<sup>34</sup> Europa 2020 è il piano strategico decennale, messo a punto dall'Unione Europea nel 2010, per la crescita ed il lavoro. Si tratta di una serie di obiettivi da raggiungere in ambiti che spaziano dal mercato del lavoro all'inclusione sociale, passando per ambiente, ricerca e sviluppo ed educazione.

<sup>35</sup> <http://eur-lex.europa.eu/legal-content/IT/TXT/?uri=URISERV%3Aasi0016>, consultato il 24/07/2016

<sup>36</sup> L'organizzazione dell'ente è strutturata in Direttore Generale (Antonio Samaritani in carica da maggio 2015), Comitato di Indirizzo, composto da nove rappresentanti, e Collegio dei Revisori dei Conti.

cooperazione applicativa tra i sistemi informatici della pubblica amministrazione e tra questi e i sistemi dell'Unione Europea"<sup>37</sup>.

L'edizione 2014 delle Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico, elaborate dall'Agenzia per l'Italia Digitale rappresenta la definizione più recente degli "standard tecnici, compresa la determinazione delle ontologie dei servizi e dei dati, con l'obiettivo di rendere il processo omogeneo a livello nazionale, efficiente ed efficace" [AA.VV. 2014: 17]. Nel testo si fa presente come le pubbliche amministrazioni raccolgano, organizzino e gestiscano una vasta quantità di dati pubblici, ma che negli anni, l'autonomia delle singole amministrazioni e le modalità di gestione dei dati abbiano contribuito a creare «isole di informazioni, con scarsa visione sistemica», dati che se «fino a poco tempo fa avevano un ruolo funzionale al perseguimento dei compiti istituzionali delle amministrazioni, assumono (oggi) una differente valenza in termini di stimolo dell'economia digitale, sviluppo dell'innovazione e trasparenza amministrativa» [AA.VV. 2014: 6]. Nello stesso testo si fa presente come, compito delle pubbliche amministrazioni, debba essere anche quello di «favorire i processi di innovazione sociale in tutte le forme, creando presupposti di conoscenza, partecipazione ed innescando cicli virtuosi che possano produrre idee, servizi, modelli di gestione e di business innovativi ed alternativi», sottolineando come tali processi non possano scaturire in modo endogeno all'amministrazione, senza l'interazione con e stimoli dagli utenti.

Distinzione importante, al fine di sgomberare da equivoci il campo di quanto si dirà più avanti nella dissertazione, riguarda i concetti di condivisione di dati e di dati di tipo aperto. «Mentre la condivisione di dati riguarda solitamente contesti ristretti e può agire sulla base di un determinato scopo di condivisione e su un insieme di dati specifici, inclusi anche dati personali, i dati aperti sono invece tipicamente non riferibili a singole persone e disponibili gratuitamente per l'uso, il riutilizzo e la distribuzione da parte di chiunque anche per finalità commerciali, soggetti al massimo alla richiesta di indicare la fonte di provenienza dei dati e di riutilizzarli

---

<sup>37</sup> Decreto-legge 22 giugno 2012, n. 83, articolo 20 "Funzioni".

secondo gli stessi termini per cui sono stati licenziati originariamente. Nella pratica poi, si tende ancora spesso a confondere il rilascio dei dati aperti con la messa a disposizione online di dati pubblici in un portale, trascurando gli aspetti di disaggregazione, neutralità tecnologica, meta datazione completa, e legali richiesti per la conformità al paradigma *open data*» [*ibidem*: 20].

### **2.3. Il contesto giuridico**

Il contesto giuridico italiano relativo alle pubbliche amministrazioni di cui è opportuno tenere conto in questa sede si è sedimentato nel corso dell'ultimo quindicennio, in particolare a partire dalla legge 150 del 7 giugno del 2000 "Disciplina delle attività di informazione e di comunicazione delle pubbliche amministrazioni", pubblicata nella Gazzetta Ufficiale n. 136 del 13 giugno 2000<sup>38</sup>. Tale legge si inserisce in verità in un più complesso decennio, l'ultimo del ventesimo secolo, segnato dal dirompente sviluppo del world wide web con effetti evidentemente non limitati alle cosiddette *dotcom*. Dai primi anni Novanta, infatti, in Italia è cominciato un percorso che ha condotto al pieno riconoscimento della specificità della comunicazione pubblica rispetto ad altri ambiti della comunicazione, che partendo dalla legge 241 del 1990 ha portato alla definizione di strutture, strumenti e figure professionali dedicate alla cura e gestione dell'informazione e della comunicazione nella pubblica amministrazione [Ducci, G. 2016]. L'inseguita modernizzazione si è tradotta nel tentativo di perfezionare le pratiche amministrative all'insegna di trasparenza, semplificazione e snellimento della burocrazia, perseguendo l'efficacia e l'efficienza dei servizi attraverso il controllo e la valutazione del lavoro svolto. Il tutto riconoscendo la comunicazione come strumento indispensabile, risorsa necessaria per attuare il cambiamento e strategia per migliorare i servizi del settore pubblico [Rovinetti 2010 e 2014].

Il valore della Legge 150/2000, almeno stando a quanto maggiormente interessa mettere a fuoco nella presente analisi, risalta nel comma 4, laddove si distingue le

---

<sup>38</sup> <http://www.parlamento.it/parlam/leggi/001501.htm>, consultato il 25/07/2016.

attività di comunicazione da quelli di informazione. In particolare ai commi 6, 7, 8 e 9 si stabilisce l'esistenza di tre strutture – l'ufficio stampa, l'ufficio relazioni con il pubblico, l'ufficio del portavoce, distinte per compiti e per professionalità. Seppur ancora non completamente recepita [Pezzoli 2016: 37-38] la legge 150/2000 ha rappresentato il primo step di definizione delle pratiche cui le amministrazioni sono tenute ad attenersi.

Step legislativo di notevole rilievo è il decreto legge 179/2012, convertito in legge n. 221/2012. L'articolo 9, in particolare, riorganizza l'articolo 52 del CAD<sup>39</sup> sull'accesso telematico e sul riutilizzo dei dati delle pubbliche amministrazioni, stabilendo che PA, gestori di pubblici servizi e società partecipate a maggioranza pubblica, rendano noti “nel proprio sito web, all'interno della sezione Trasparenza, valutazione e merito, il catalogo dei dati, dei metadati e delle relative banche dati in loro possesso ed i regolamenti che ne disciplinano l'esercizio della facoltà di accesso telematico e il riutilizzo [...]”, e che “i dati e i documenti che le amministrazioni titolari pubblicano senza l'espressa adozione di una licenza si intendono rilasciati come dati di tipo aperto”, quindi secondo il principio dell'Open Data by default.

L'altro riferimento legislativo fondamentale è la legge 33 del 14 marzo 2013 “Riordino della disciplina riguardante gli obblighi di pubblicità, trasparenza e diffusione delle informazioni da parte delle pubbliche amministrazioni”, pubblicata nella Gazzetta Ufficiale del 5 aprile 2013, comunemente richiamata con il nome di Decreto Trasparenza. L'articolo 1 definisce trasparenza 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». Obiettivo, quello della trasparenza, da perseguire attraverso la pubblicazione «nei siti istituzionali delle pubbliche amministrazioni dei documenti, delle informazioni e dei dati concernenti l'organizzazione e l'attività delle pubbliche amministrazioni, cui corrisponde il diritto di chiunque di accedere ai siti direttamente ed immediatamente, senza

---

<sup>39</sup> Il CAD, Codice dell'amministrazione digitale, è il decreto legislativo 7 marzo 2005, n. 82. Nel testo si mettono a punto alcune disposizioni riguardanti l'utilizzo delle tecnologie dell'informazione e della comunicazione come strumenti privilegiati nei rapporti tra la pubblica amministrazione italiana ed i cittadini.

autenticazione ed identificazione» (art. 2). Per quanto riguarda la pubblicità ed il diritto alla conoscibilità, come espresso nell'articolo 3, «tutti i documenti, le informazioni ed i dati oggetto di pubblicazione obbligatoria [...] sono pubblici e chiunque ha il diritto di conoscerli, di fruirne gratuitamente e di utilizzarli e riutilizzarli».

Particolarmente rilevante, in riferimento al caso di studio di cui al prossimo paragrafo, è l'articolo 6 relativo alla qualità delle informazioni. Si legge che «le pubbliche amministrazioni garantiscono la qualità delle informazioni riportate nei siti istituzionali [...] assicurandone l'integrità, il costante aggiornamento, la completezza, la tempestività, la semplicità di consultazione, la comprensibilità, l'omogeneità, la facile accessibilità, nonché la conformità ai documenti originali in possesso dell'amministrazione [...]. L'articolo 9 determina le modalità concrete di pubblicazione dei dati, sancendo la predisposizione di un'area, all'interno dei siti delle pubbliche amministrazioni, denominata Amministrazione Trasparente, cui si rimanda all'appendice di questo capitolo per la descrizione delle singole sezioni (figura 2).

Nel corso del tirocinio curricolare svolto presso Synapta Srl si è scelto, di concerto con i responsabili della start-up, di seguire inizialmente un progetto relativo all'Indicatore di Tempestività dei Pagamenti, uno dei dati che le pubbliche amministrazioni sono tenute a pubblicare nella sezione Amministrazione Trasparente del proprio sito web. Ne parlerò diffusamente più avanti, ma già in questo paragrafo dedicato al quadro legislativo mi pare importante richiamare l'articolo 46 della legge 333/2013, Violazione degli obblighi di trasparenza – Sanzioni, secondo cui «l'inadempimento degli obblighi di pubblicazione previsti dalla normativa vigente o la mancata predisposizione del Programma triennale per la trasparenza e l'integrità costituiscono elemento di valutazione della responsabilità dirigenziale [...]».

In ambito europeo, infine, è importante richiamare le disposizioni contenute nella direttiva sull'informazione del settore pubblico (PSI), la cui prima versione risale al 2003, quindi diversi anni prima dell'affermazione del fenomeno *open data*. Tale direttiva, accolta in Italia con il decreto legislativo 36/2006, incoraggiava gli stati

membri a massimizzare il potenziale dell'informazione del settore pubblico attraverso il paradigma del riutilizzo. Anche la versione più recente della direttiva, la 2013/37/UE, ribadisce l'importanza di adottare standard che consentano il riuso dei dati, esprimendo in questo caso un carattere di obbligo.

Tutti gli step legislativi citati volgono a determinare un quadro caratterizzato, a livello continentale e nazionale, da apertura e trasparenza delle pubbliche amministrazioni, indispensabile incentivo al riutilizzo dei dati per offrire servizi, oltre che privilegiato terreno per i giornalisti che intendano adoperarsi al controllo del pubblico servizio.

### *Il Freedom of Information Act*

L'atto per la liberazione dell'informazione, come da traduzione letterale dall'originaria lingua inglese, è una legge statunitense sulla libertà di informazione, risalente al 4 luglio 1966, che sancisce il diritto pubblico di richiedere l'accesso ai documenti di tutte le agenzie federali degli States. Agenzie che sono tenute a pubblicare tali informazioni, a meno che non ricadano in uno dei nove ambiti di eccezioni, tra cui privacy personale e sicurezza nazionale<sup>40</sup>. Con l'emendamento del 1996, poi, il FOIA si è adeguato alla crescente digitalizzazione in corso nella società, predisponendo piattaforme online su cui pubblicare alcuni documenti. Pur aumentando, il numero di giorni in cui le agenzie federali devono evadere una risposta alle richieste di accesso rimangono al di sotto di un mese, venti giorni per la precisione [Henry C. 2003: 8]

Il Freedom of Information Act ha rappresentato per anni il modello di riferimento in ambito di trasparenza, spingendo oltre novanta paesi in tutto il mondo a dotarsene, alcuni con leggi dalle ampie possibilità, altri con maggiori limitazioni. Per quanto riguarda il caso italiano, la prima definizione legislativa in merito al diritto di accesso agli atti amministrativi risale alla legge 7 agosto 1990, n. 241, "Nuove norme in materia di procedimento amministrativo e di diritto di accesso ai documenti

---

<sup>40</sup> <https://www.foia.gov/about.html>, consultata il 01/08/2016.

amministrativi”. Si tratta, letto ora alla luce dei successivi sviluppi, di un primo tentativo di instaurare un rapporto collaborativo e paritario tra pubblica amministrazione e cittadini, caratterizzato dalla limitatezza di un codice che, ad esempio, sancisce all’articolo 24, comma 3, che “non sono ammissibili istanze di accesso preordinate ad un controllo generalizzato dell’operato delle pubbliche amministrazioni” e che prevede che la richiesta di accesso ai documenti debba essere motivata (articolo 25, comma 2). Un diritto di accesso agli atti, dunque, oggetto di notevoli limitazioni, tanto da far scaturire un acceso sentimento verso la necessità di legiferare anche in Italia sulla falsariga del FOIA. Negli ultimi anni Foia4Italy<sup>41</sup>, Foia.it<sup>42</sup>, la campagna su Riparte il Futuro<sup>43</sup> e la petizione su Change.org<sup>44</sup> hanno rappresentato alcune delle manifestazioni, da parte della società civile, a supporto di un disegno di legge che garantisca effettivo accesso ai documenti delle PA in Italia, a cui si sommano le voci dei mezzi di informazione.

Il decreto legislativo approvato dal Consiglio dei Ministri il 20 gennaio 2016 ha rappresentato il primo step verso l’approvazione del Freedom of Information Act italiano. Si trattava, nello specifico, della bozza del decreto attuativo della legge n. 124/2015, conosciuta come legge Madia, dal nome del Ministro per la semplificazione e la pubblica amministrazione. La proposta del governo, tuttavia, si è presto scontrata con richieste di modifiche per adeguarla al modello anglosassone. Tra le principali critiche<sup>45</sup> mosse al decreto legislativo: la possibilità, da parte delle pubbliche amministrazioni, di continuare ad applicare il cosiddetto silenzio-diniego, vale a dire il diritto di non evadere le richieste di accesso ai documenti (“decorsi inutilmente trenta giorni dalla richiesta, questa si intende respinta” recita l’articolo 6

---

<sup>41</sup> Foia4Italy è un progetto che mira ad ottenere anche in Italia strumenti già utilizzati all’estero per controllare l’operato della pubblica amministrazione. Obiettivo dichiarato è quello di “mettere ogni cittadino nelle condizioni di diventare agente di trasparenza”. Per approfondimenti <http://www.foia4italy.it/>, consultato il 20/07/2016.

<sup>42</sup> Foia.it è l’iniziativa di 62 promotori, guidati da Andrea Fama, per l’adozione di un Freedom of Information Act in Italia. Per approfondimenti <http://www.foia.it/>, consultato il 20/07/2016

<sup>43</sup> Riparte il Futuro è una comunità digitale apartitica, come si legge sul sito, fondata nel gennaio del 2013 ed alimentata da oltre un milione di persone con l’obiettivo di sconfiggere la corruzione promuovendo la trasparenza e la certezza del diritto. Per approfondimenti <https://www.riparteilfuturo.it/chi-siamo>, consultata il 20/07/2016.

<sup>44</sup> Change.org è una piattaforma di raccolta firme per petizioni su svariati argomenti. Per approfondimenti <https://www.change.org/about>, consultata il 20/07/2016.

<sup>45</sup> Non è un Freedom of Information Act, 12/02/2016, pubblicato su <http://www.foia4italy.it/news/non-e-un-freedom-of-information-act/>, consultato il 02/08/2016.

comma 5); la mancata predisposizione di adeguate sanzioni in caso di accesso illegittimamente negato o di mancata risposta; la indeterminazione delle eccezioni all'accesso; la mancata chiarezza nella definizione dei costi a carico del richiedente; la mancata predisposizione di rimedi stragiudiziali, stante la sola possibilità di ricorrere ai Tar, tribunali regionali spesso lenti ed onerosi; la mancata abrogazione del già citato articolo 24 comma 3 della legge 241/1990 che vieta il controllo generalizzato dell'operato della pubblica amministrazione.

La richiesta di apporre modifiche al decreto viene accolta dal governo anche in seguito alle indicazioni del Consiglio di Stato, e porta all'approvazione del FOIA, che si concretizza nel decreto legislativo 25 maggio 2016, n. 97, riguardante la revisione e la semplificazione delle disposizioni in materia di prevenzione della corruzione, pubblicità e trasparenza nella PA, attraverso le modifiche al decreto legislativo 14 marzo 2013, n. 33, il già citato Decreto Trasparenza. Si sancisce così tra le altre cose l'eliminazione del silenzio-diniego, il riconoscimento della gratuità dell'accesso in formato elettronico e cartaceo, ed in particolare si garantisce la possibilità di richiedere documenti, dati e atti delle pubbliche amministrazioni senza la necessità di dimostrare "un interesse diretto, concreto e attuale", come previsto dall'articolo 22, comma 1a della legge 241/1990.

Il FOIA, pubblicato in Gazzetta Ufficiale l'8 giugno 2016, è in vigore dal 23 giugno 2016 e le pubbliche amministrazioni hanno sei mesi di tempo per adeguarsi, periodo nel quale verranno rese note anche le Linee Guide dell'Anac, l'Autorità Nazionale Anticorruzione.

## **2.4. L'indicatore di tempestività dei pagamenti**

Le precedenti pagine hanno svolto la funzione di raccordo tra diverse tematiche la cui rilevanza è apparsa evidente nel corso del tirocinio a Synapta. Per quanto la startup torinese abbia individuato nella progettazione di software il proprio modello di business, le tematiche più rilevanti riguardano le pubbliche amministrazioni e i

contratti pubblici<sup>46</sup> e seguono percorsi integrati al lavoro di ricerca condotto dal Nexa Center for Internet & Society. Il primo progetto in cui mi sono speso nel corso dei mesi di stage curricolare prima, e di tesi in azienda poi, ha riguardato l'Indicatore di Tempestività dei Pagamenti.

Si tratta di un dato che le pubbliche amministrazioni sono tenute a pubblicare secondo quanto disposto dall'articolo 33 del Decreto Trasparenza (decreto legislativo 14 marzo 2013, n. 33). Il dato viene qui definito come l'indicatore dei "tempi medi di pagamento relativi agli acquisti di beni, servizi e forniture", da pubblicare con cadenza annuale. In verità il decreto, in quanto riordino della disciplina riguardante gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle pubbliche amministrazioni, determina solamente che l'indicatore dei pagamenti, già previsto dalla legge n. 69/2009, entri nello schema relativo alla cosiddetta sezione Amministrazione Trasparente che le PA italiane avrebbero dovuto adottare per i propri portali web. Già la citata legge 18 giugno 2009 n. 69, Disposizioni per lo sviluppo economico, la semplificazione, la competitività nonché in materia di processo civile, pubblicata in Gazzetta Ufficiale n. 140 del 19 giugno 2009, supplemento ordinario n. 95, prevedeva, all'articolo 23, comma 5, che "al fine di aumentare la trasparenza dei rapporti tra le amministrazioni pubbliche e gli utenti, a decorrere dal 1 gennaio 2009 ogni amministrazione pubblica determina (s)e e pubblica (s)e, con cadenza annuale, nel proprio sito internet o con altre forme idonee, un indicatore dei propri tempi medi di pagamento relativi agli acquisti di beni, servizi e forniture, denominato «indicatore di tempestività dei pagamenti»".

Gli step legislativi di cui è opportuno tenere conto per quanto riguarda l'indicatore di tempestività dei pagamenti sono pertanto innanzitutto la legge del 2009 e quella del 2013. Diversi sono gli schemi tipo e le modalità di calcolo utilizzate per definire l'indicatore: secondo la legge del 2009, il valore assunto dall'ITP indica il numero di giorni intercorrenti tra la data di ricezione della fattura e la data di pagamento dell'amministrazione al fornitore.

---

<sup>46</sup> <http://public-contracts.nexacenter.org/>, consultato il 20/04/2016.

Diverse sono le modalità attualmente applicate al calcolo dell'indicatore. L'articolo 8, comma 1c, del decreto legge del 24 aprile 2014, n. 66, Misure urgenti per la competitività e la giustizia sociale, introduce, a decorrere dall'anno 2015, un indicatore trimestrale di tempestività dei pagamenti che si va ad aggiungere ai dati annuali da pubblicare. La legge 23 giugno 2014, n. 89, converte in legge, con modificazioni, il suddetto decreto 66/2014, ribadisce l'istituzione dell'obbligo di pubblicare un indicatore di tempestività dei pagamenti annuale e quattro indicatori trimestrali all'anno, rimandando al decreto del Presidente del Consiglio dei Ministri la definizione delle modalità di calcolo e degli schemi di pubblicazione. Il DPCM 22 settembre 2014, pubblicato in Gazzetta Ufficiale n. 265, prende in considerazione l'ITP all'articolo 9, stabilendone il calcolo come "la somma, per ciascuna fattura emessa a titolo corrispettivo di una transazione commerciale, dei giorni effettivi intercorrenti tra la data di scadenza della fattura o richiesta equivalente di pagamento e la data di pagamento ai fornitori moltiplicata per l'importo dovuto, rapportata alla somma degli importi pagati nel periodo di riferimento". Con data di scadenza della fattura si intendono i termini previsti dall'articolo 4 del decreto legislativo 9 ottobre 2002, n. 231, come modificato dal decreto legislativo 9 novembre 2012, n. 192. In particolare, all'articolo 4, comma 2, si legge che il periodo di pagamento non può superare i trenta giorni dalla data di ricevimento da parte del debitore della fattura o di una richiesta di pagamento di contenuto equivalente, termine che può essere portato fino ad un massimo di sessanta giorni nel caso in cui debitori siano pubbliche amministrazioni, purché ciò sia "oggettivamente giustificato dalla natura particolare del contratto" (articolo 4).

L'ITP, così come previsto dalla legislatura corrente, è una media ponderata sulla base dell'importo delle fatture, ed indica il ritardo medio delle pubbliche amministrazioni nel pagamento delle fatture ai fornitori. Un valore pari a 0 significa che i pagamenti vengono saldati mediamente al trentesimo giorno (o ad altro giorno scelto dalle due parti come scadenza), valori negativi suggeriscono che le amministrazioni sono solite pagare prima della scadenza, valori positivi determinano un ritardo.

Un breve paragrafo merita di essere dedicato alle sanzioni previste in caso di ritardi nei pagamenti. Secondo quanto previsto dalla direttiva 2011/7/UE del Parlamento

Europeo e del Consiglio, ed in particolare dalla circolare del 16 febbraio 2011, il creditore ha diritto, senza che sia necessario un sollecito, agli interessi legali di mora<sup>47</sup>. Per quanto riguarda la legislazione italiana è opportuno fare riferimento all'articolo 41, comma 2, del decreto legge 66/2014, che fissa in novanta giorni per il 2014, ed in sessanta giorni a decorrere dal 2015, il limite massimo dei tempi medi di pagamento per le pubbliche amministrazioni. Qualora i tempi sforassero queste tempistiche, il decreto pone come sanzione il blocco delle assunzioni da parte dell'ente per l'anno successivo a quello in cui si è registrata la violazione dei termini.

## **2.5 Materiale d'analisi: Dati**

Il presente lavoro mira a mettere in discussione, attraverso un esempio pratico, quello riferito all'indicatore di tempestività dei pagamenti, il modello di trasparenza come sancito dal Decreto Trasparenza. Per quello che riguarda le successive modifiche introdotte dal Freedom of Information Act, la mia tesi si scontra con le tempistiche che non rendono possibile valutare su un piano effettivo la capacità, e la volontà, di recepimento delle direttive da parte delle pubbliche amministrazioni, stante il già ricordato termine di sei mesi durante i quali le amministrazioni possono adeguarsi. Va peraltro ricordato una volta in più come il FOIA italiano abbia ripercussioni di maggior portata per quello che riguarda le possibilità, da parte dei cittadini, di domandare documenti ed informazioni alle pubbliche amministrazioni, rispetto a quanto concerne la messa a disposizione, sui siti istituzionali delle PA, dei dati. Obiettivo, in questa fase, è stato indagare l'effettiva possibilità di ricavare da questi portali web dati utili ad analizzare l'andamento dell'affare pubblico. Dati, quelli cercati e raccolti, che come ricordato in precedenza dovrebbero essere pubblicati per legge, peraltro secondo standard chiaramente definiti.

La raccolta dei dati relativi all'indicatore di tempestività è partita inizialmente volgendosi ai capoluoghi di provincia italiani. Al 01/01/2016 il campione è composto

---

<sup>47</sup> L'interesse legale di mora è un interesse ad un tasso che è pari al tasso di riferimento, tipicamente il tasso di interesse applicato dalla Banca Centrale Europea alle sue più recenti operazioni di rifinanziamento, maggiorato di almeno otto punti percentuali.

da 117 comuni. Come detto in precedenza, l'ITP è un dato che le pubbliche amministrazioni sono tenute a pubblicare sul proprio sito internet istituzionale, e più precisamente all'interno della sezione Amministrazione Trasparente, nella sottosezione Pagamenti dell'Amministrazione.

La ricerca e raccolta dei dati è avvenuta in maniera completamente manuale, collegandomi volta per volta ai portali web dei comuni, navigandovi all'interno, trovando (quando presenti) i dati e copiandoli in una tabella Excel organizzata secondo i periodi di riferimento. Inizialmente la tabella prevedeva semplicemente una colonna per il dato ITP riferito al 2013, una per il dato del 2015, una per il 2015, quattro colonne per i quattro trimestri del 2015, una per il primo trimestre del 2016. Presto è emerso, durante la compilazione della tabella, che una tale organizzazione non rispecchiava i dati effettivamente pubblicati dalle amministrazioni. Per ogni periodo è stato necessario predisporre diverse colonne, al fine di salvare i dati in maniera direttamente confrontabile: spesso, infatti, per lo stesso periodo di riferimento, comuni diversi hanno pubblicato il dato relativo all'ITP secondo standard e attraverso calcoli differenti. Con riferimento all'anno 2013, ad esempio, è emerso che:

21 comuni hanno pubblicato il dato ITP secondo la legge 69/2009, intendo l'indicatore come la media dei giorni trascorsi dall'acquisizione della fattura al pagamento.

16 comuni hanno pubblicato il dato ITP relativamente ai codici SIOPE<sup>48</sup> indicati nella tabella A di cui al decreto legge n. 66/2014, convertito dalla legge n. 89/2014.

9 comuni hanno pubblicato il dato ITP secondo il decreto del Presidente del Consiglio dei Ministri del 22 settembre 2014, tale per cui l'indicatore è dato dalla media ponderata in base all'importo dei giorni di ritardo rispetto alla scadenza della fattura.

---

<sup>48</sup> Il SIOPE, Sistema informativo sulle operazioni degli enti pubblici, è un sistema di rilevazione telematica di incassi e pagamenti dei tesoriери di tutte le pubbliche amministrazioni italiane.

24 comuni hanno pubblicato un dato calcolato secondo modalità che non rispecchiano una di quelle esposte precedentemente precedenti, rendendo di fatto impossibile la determinazione effettiva dell'indicatore ed il confronto con i valori degli altri comuni. Alcuni dei motivi di indeterminazione del valore sono l'assenza di un numero preciso, la suddivisione nel calcolo dell'indicatore tra spese in conto corrente e spese in conto capitale, il mancato riferimento alle modalità di calcolo utilizzate e la specifica della legge applicata, o ancora il calcolo relativo ad un periodo diverso da quello che sarebbe dovuto essere (solo un trimestre invece dell'intero anno, ad esempio).

Dal 2014 in avanti si è assistito ad una maggiore standardizzazione dei dati, con sette amministrazioni comunali a pubblicare il dato secondo la legge del 2009 nel 2014, e il solo comune di Foggia ad utilizzare queste modalità di calcolo nel 2015. Rimane tuttavia alto seppur in diminuzione, negli stessi anni, il numero di amministrazioni che pubblicano l'indicatore di tempestività dei pagamenti in maniera non chiara (21 nel 2014, 11 nel 2015, 9 nel primo e 9 nel secondo trimestre del 2015, 7 nel terzo trimestre del 2015, 8 nel quarto trimestre del 2015, 6 nel primo trimestre del 2016).

Ciò che tuttavia desta la maggiore sorpresa è il numero di comuni che non pubblicano il dato relativo all'indicatore di tempestività dei pagamenti, sebbene le già ricordate disposizioni legislative in materia lo sanciscano come obbligo. La tabella che segue riassume, per ogni periodo analizzato, il numero di comuni a non aver pubblicato.

2013	2014	2015	2015 T1	2015 T2	2015 T3	2015 T4	2016 T1
50/117	8/117	21/117	20/117	22/117	27/117	30/117	41/117

Cominciata il 18 e terminata il 23 maggio 2016, la raccolta dei dati voleva inizialmente essere il principio di un semplice lavoro di analisi sulla falsariga del *data journalism*. Gli aspetti messi in luce in precedenza, relativi alle difficoltà di estrarre valore da dati non corrispondenti a quanto previsto dalla legge, e il paradosso che siano informazioni pubblicate proprio in nome della trasparenza amministrativa a non essere facilmente analizzabili, hanno spostato il focus del lavoro di tesi su un

secondo binario, quello delle questioni di *open government* e *open data* in ambito governativo, tenendo altresì in piedi il lavoro di analisi sui dati.

Un secondo lavoro portato avanti in Synapta, sempre concernente l'indicatore di tempestività dei pagamenti, mi ha visto impegnato in una simile analisi su un campione assai più corposo. Oggetto è stata infatti la collezione dei dati relativi ai comuni piemontesi con più di mille abitanti. Si tratta di 608 amministrazioni, per le quali si è scelto di procedere alla raccolta dei dati relativi al 2014, al 2015 (valori annuale e trimestrali) e al primo trimestre del 2016, escludendo quindi i dati del 2013, considerate le difficoltà incontrate a livello dei capoluoghi di provincia. La raccolta, cominciata il 3 giugno 2016 e terminata due settimane dopo, il 17 giugno, è avvenuta in maniera completamente manuale e ho proceduto secondo l'ordine decrescente in base al numero di abitanti. Le evidenze emerse sono confluite in due differenti lavori. Il primo, un ampio report cui si rimanda al capitolo 3 per la consultazione. Il secondo<sup>49</sup>, un blog post di sintesi del primo, pubblicato su ForumPa, sito internet di FPA s.r.l., società che organizza da oltre venticinque anni Forum Pa, evento che si tiene annualmente a Roma e rappresenta un momento di incontro tra pubbliche amministrazioni, imprese, mondo della ricerca e società civile.

In virtù delle dinamiche emerse durante il lavoro relativo all'indicatore di tempestività, in particolare il gap tra quanto previsto dalle leggi in vigore e l'effettiva pubblicità dei documenti delle pubbliche amministrazioni, si è scelto di proseguire il lavoro di tesi su un piano di analisi dei media. Per questa ragione si è proceduto raccogliendo una serie di articoli apparsi su alcune testate giornalistiche negli ultimi tre anni, con lo scopo di mettere in evidenza i *pattern* più rilevanti costruiti dall'agenda dei media.

---

<sup>49</sup> <http://www.forumpa.it/pa-digitale/dati-dati-aperti-sulla-tempestivita-dei-pagamenti-della-pa-alcuni-spunti-da-unindagine-volta-al-riutilizzo>

## 2.6 Materiale d'analisi: Testi

La scelta del materiale mediale da analizzare è iniziata definendo il periodo temporale cui far riferimento. In virtù delle tempistiche relative alle leggi più rilevanti ai fini del mio lavoro ho scelto di raccogliere gli articoli a partire dal 1 gennaio 2013, pochi mesi prima che il Decreto Trasparenza (33/2013) venisse approvato, fino all'ultimo giorno utile, il 13 luglio del 2016<sup>50</sup>. Ho proceduto poi a scegliere le testate da cui ricavare il materiale. In questo caso la scelta è ricaduta su Il Sole 24 ORE, quotidiano economico italiano per eccellenza, espressione di Confindustria, su La Stampa, quotidiano torinese scelto anche in virtù della vicinanza territoriale al lavoro svolto sull'indicatore di tempestività relativo ai comuni piemontesi, sul quotidiano romano la Repubblica, secondo per diffusione in Italia<sup>51</sup>, e sul settimanale l'Espresso, edito dal Gruppo Editoriale l'Espresso Spa, stesso gruppo de la Repubblica.

La scelta delle parole chiave attraverso cui svolgere l'analisi è stata guidata dalle tematiche emerse durante il lavoro sull'indicatore di tempestività. I concetti più spesso richiamati hanno rappresentato il fulcro attorno a cui procedere con la formulazione del corpus. In particolare ho scelto di cercare gli articoli, e di analizzarli poi, secondo le seguenti parole chiave:

“decreto trasparenza”, per un totale di 38 articoli, così articolati: sette articoli nel 2013, 9 nel 2014, 7 nel 2015 e 15 nei mesi del 2016 fino al 13 luglio.

“foia” o “freedom of information act”, per un totale di 108 articoli, di cui 15 risalenti al 2013, 15 al 2014, 28 al 2015 e 50 al 2016.

“open government”, per un totale di 37 articoli, di cui 12 risalenti al 2013, altrettanti al 2014, 2 al 2015, e 11 al 2016.

Il progetto iniziale prevedeva di costruire un corpus contenente articoli selezionati attraverso la *keyword* “indicatore di tempestività dei pagamenti”. Tuttavia la ricerca

---

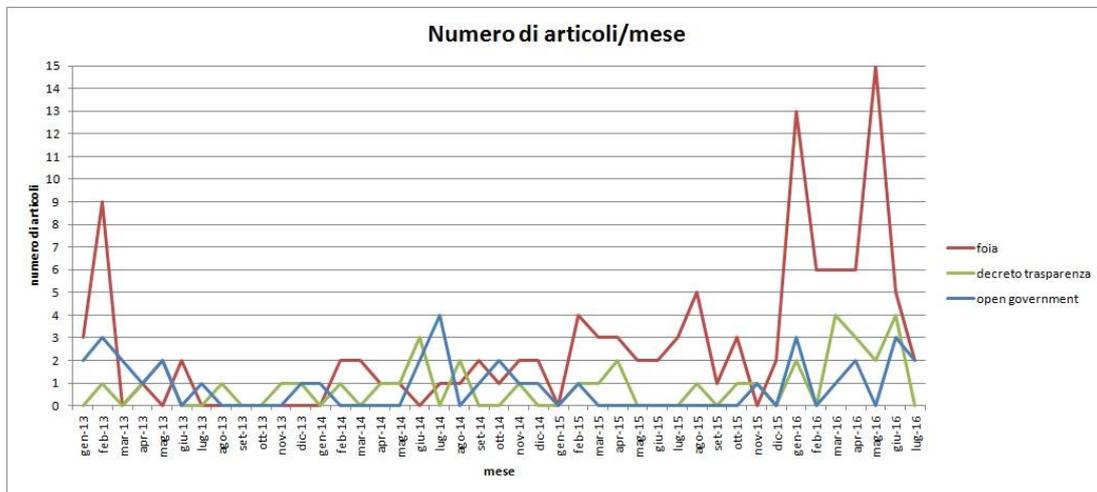
<sup>50</sup> La raccolta degli articoli e la composizione del corpus è infatti avvenuta in data 14 luglio 2016.

<sup>51</sup> Stando ai più recenti dati pubblicati da Ads (<http://www.adsnotizie.it/index.asp>) relativi al mese di agosto 2016, La Repubblica si colloca al secondo posto, per diffusione, tra i quotidiani italiani, con 248.007 copie vendute, dietro solo al milanese Corriere della Sera (254.805).

di tali articoli negli archivi digitali delle testate selezionate ha portato ad un numero troppo esiguo di tracce, 28 per la precisione, insufficienti ad un'analisi del contenuto metodologicamente solida.

L'analisi del materiale si è inizialmente rivolta ad un piano quantitativo, con l'obiettivo di evidenziare la distribuzione temporale degli articoli, per poi metterne in luce la contemporaneità ad eventi rilevanti.

Per quanto riguarda il numero di articoli pubblicati nei tre anni e mezzo presi in analisi, il grafico seguente (figura 1) ne descrive la distribuzione:



**Figura 1: Grafico di distribuzione temporale degli articoli**

Emerge immediatamente la curva mediatica segua ritmi solo parzialmente sovrapponibili. In particolare il termine “foia” ha visto tre picchi: uno a febbraio 2013, dove 9 articoli ne parlano, coincidente con la stesura del Decreto Trasparenza poi pubblicato a marzo, e con i programmi elettorali per le elezioni politiche del 24 e 25 febbraio. Un altro picco si nota a gennaio 2016, con 13 articoli che citano il Freedom of Information Act, periodo in cui si comincia a parlare insistentemente dei contenuti della riforma della pubblica amministrazione, di cui il Foia rappresenta uno

dei punti principali. L'ultimo picco coincide infine con maggio 2016, quando il Foia diventa legge.

Interessante è poi notare come la ricerca con parola chiave “decreto trasparenza” segua un andamento indipendente dall'approvazione della legge 33/2013. Si assiste infatti a dei picchi tra marzo e maggio 2016 quando, oltre che per l'imminente Foia, se ne parla in virtù di alcune inchieste, in particolare su la Repubblica, su stipendi e proprietà non dichiarate.

Queste sono solo alcune delle evidenze che emergono in maniera più immediata. Si rimanda al prossimo capitolo per l'analisi più profonda, e convergente su un piano qualitativo, del corpus di articoli raccolti.

## **Appendice. I *linked data***

La presente appendice intende offrire una breve panoramica sulla tecnologia *linked data*, ambito di lavoro di Synapta srl, la start up torinese nella quale ho investito i mesi di tirocinio curriculare e di stesura della presente tesi di laurea. Obiettivo è quello di chiarire di cosa si parla quando ci si riferisce ai dati collegati ed in che modo possono rappresentare un punto di svolta nel discorso sul *data journalism*.

Ho affrontato nel secondo capitolo la questione relativa agli *open data*, mettendo a più riprese in evidenza quanto il valore intrinseco maggiormente rilevante stia non tanto nella disponibilità dei dati, effetto della pubblicazione, quanto piuttosto nel loro riutilizzo. In tal senso vale la pena sottolineare una volta in più come «l'effettivo valore degli *open data* si sprigiona nel momento in cui dati di diversa provenienza - e quindi messi a disposizione da soggetti differenti - vengono combinati tra loro» [AA.VV. 2015: 7]. Tale combinazione di dati, che in gergo tecnico viene definita interoperabilità, si concretizza tipicamente nell'utilizzo congiunto di basi di dati differenti oppure, detto in termini meno astratti, nel poter confrontare, incrociare, visualizzare, in una parola interpretare le informazioni memorizzate in database differenti. L'interoperabilità si gioca evidentemente su piani differenti, a cominciare

da quello tecnico. La rimozione degli ostacoli che impediscono la fruizione delle differenti sorgenti informative comincia dalla definizione di standard tecnologici, definendo «l'insieme delle condizioni che consentono a sistemi diversi si scambiare automaticamente informazioni» [*ibidem*]. Altrettanto importante, e già brevemente richiamata nel secondo capitolo, è la questione sull'interoperabilità semantica, che si definisce sul piano del significato dei dati, e sull'interoperabilità giuridica, che abilita alla manipolazione dei dati nei modi desiderati senza il timore di incorrere in violazioni di diritti di proprietà intellettuali.

L'elaborazione teorica, nonché tecnica, dei *linked data* si deve a Tim Berners-Lee, colui che ad inizio anni '90 mise a punto, insieme a Robert Cailliau, il *world wide web*. Ed è proprio dal successo del web che è opportuno partire nell'affrontare il discorso sulla tecnologia *linked data*. Se è vero che, in virtù del rapido accumulo di contenuti, il web è divenuto un enorme contenitore di dati, è altrettanto vero che «l'informazione condivisa al suo interno [...] è costituita da documenti complessi intrecciati tra loro mediante collegamenti ipertestuali, che tuttavia non esplicitano la relazione tra le diverse pagine e concetti» [AA.VV 2015: 29]. Per dirla *à la* Kuhn [Corbetta 1999: 18, citato da Kuhn 1962; trad. it. 1969: 138], la tecnologia *linked data* si inserisce in un quadro di rivoluzione<sup>52</sup> del web, da una concezione che lo intende come raccolta di documenti a ciò che è comunemente definito come web semantico. Si tratta di un «processo volto a strutturare i contenuti messi a disposizione sul web, arricchendoli della loro connotazione semantica e collegandoli con altre informazioni, al fine di creare nuova conoscenza» [AA.VV. 2015: 30].

Risulta utile, in questa fase, fare riferimento al concetto di *open data* classificati secondo il modello a stelle<sup>53</sup> elaborato da Berners-Lee. Secondo l'inventore del web,

---

<sup>52</sup> Thomas Samuel Kuhn teorizza come cambio di paradigma nelle scienze. Con il termine paradigma, Kuhn intende non solo un modello di riferimento, ma anche alcune indicazioni indispensabili per costruirlo. Con l'adozione di un paradigma, uno studioso acquisisce teorie, metodi e criteri per svilupparlo [Corbetta 1999: 18, citato da Kuhn 1962; trad. it. 1969: 138]. In questo senso, ritengo fondato asserire che il *data journalism* operi un cambio di paradigma, quella che Kuhn chiamerebbe rivoluzione, all'interno dell'universo giornalistico. Con la definizione di giornalismo di precisione sono emersi sì, nel panorama della professione, strumenti differenti rispetto al passato, ma soprattutto una diversa concezione teorica, che affonda le proprie radici nel metodo scientifico, come ho cercato di dimostrare nel primo capitolo

<sup>53</sup> Tim Berners-Lee, *Linked Data*, <https://www.w3.org/DesignIssues/LinkedData.html>, 27/07/2006, consultato il 24/04/2016.

una stella è appannaggio dei dati aperti messi a disposizione sul web, “così come sono”. Due stelle identificano i dati pubblicati in maniera strutturata, ad esempio come tabella Excel, anziché come immagine scannerizzata; tre stelle le meritano i dati messi a disposizione secondo standard aperti (proseguendo sull’esempio tabellare, utilizzando CSV invece di Excel). La quarta stella caratterizza i dati caratterizzati dall’utilizzo di un URL, *Uniform Resource Locator*, per identificare gli oggetti pubblicati. La quinta stella, infine, indica che il dato pubblicato è collegato ad altri dati, quando sia cioè *linked*. Tale classificazione mostra come i dati collegati rappresentino, ad oggi, lo stato dell’arte per quanto riguarda gli *open data*. Va naturalmente evidenziato come pubblicare l’informazione in questi termini sia più dispendioso, in termini di tempo e di competenze da acquisire, piuttosto che la semplice condivisione di un dato “così com’è”, ma rispettare questo tipo di standard garantisce un elevato grado di riutilizzo.

A questo punto si dovrebbe aprire una parentesi di natura tecnica, con l’intento di chiarire le modalità di pubblicazione che fanno di un dato aperto un dato collegato. Tuttavia il presente lavoro lambisce solo tangenzialmente il discorso tecnologico strettamente inteso. Per questa ragione ci si limiterà, in questa sede, a riportare alcuni stralci utili a ricostruire il processo di pubblicazione, concentrandosi in maniera più approfondita sulle modalità di interazione con i database contenenti dati collegati.

## Le regole del W3C

Il livello 5 stars, come visto, prevede che i dati siano pubblicati in maniera collegata. Per raggiungere questo obiettivo il W3C<sup>54</sup>, il consorzio internazionale per il World Wide Web, ha stabilito quattro regole. La prima prevede di usare URIs (*Uniform Resource Identifiers*) per identificare le risorse pubblicate in maniera univoca; la seconda suggerisce di usare il protocollo http per raggiungere quella risorsa sul web; la terza stabilisce che siano RDF<sup>55</sup> e SPARQL<sup>56</sup> gli standard deputati alla descrizione

---

<sup>54</sup> <https://www.w3.org/>, consultato il 31/08/2016.

<sup>55</sup> RDF è il principale standard utilizzato per la modellazione, la codifica, lo scambio ed il riutilizzo di dati *linked*. Consapevole che una trattazione tecnica esula dalle tecnologie esula dall’obiettivo, e dalle

ed all'interrogazione delle risorse; la quarta suggerisce di includere collegamenti ad altre URI per arricchire il patrimonio di conoscenza.

Parlare di *linked open data* presuppone l'esistenza di «formati e modelli standard, interoperabili, *machine-readable* e ontologie di riferimento, al fine di rendere possibile la massima integrazione tra fonti diverse di conoscenza» [AA.VV. 2015: 33]. Per fare questo è opportuno far presente la necessità di adottare un adeguato vocabolario e di seguire una ragionevole ontologia. Con il primo termine si indicano «collezioni di termini e delle relative definizioni pensati per risolvere l'ambiguità tipica del linguaggio naturale», mentre l'ontologia è «una rappresentazione condivisa ed esplicita di un dominio di interesse» [*ibidem*] in grado di esprimere in maniera dettagliata le relazioni esistenti tra le entità che costituiscono il dominio.

### *Linked data* nella pubblica amministrazione

Tradurre dati prodotti dalle pubbliche amministrazioni in *linked data* è uno degli obiettivi che Synapta si è prefissata nel momento di proporsi sul mercato. Nel corso del tirocinio prima, e dello sviluppo del progetto di tesi poi, sono entrato a contatto con il progetto ContrattiPubblici.org<sup>57</sup>, un motore di ricerca alimentato ad oggi da più di cinque milioni e mezzo di contratti, stipulati tra quasi quindicimila enti pubblici italiani e oltre seicentocinquantamila aziende. Il sito consente all'utente di effettuare ricerche sui contratti stipulati dalle pubbliche amministrazioni, semplicemente digitando il nome dell'ente o il codice che la identifica. Il motore di ricerca è naturalmente alimentato da *linked data*: in particolare le triple, che come visto in nota 55 rappresentano il formato in cui sono espresse le informazioni, sono organizzate come entità articolate su proprietà come stazione appaltante,

---

competenze da me possedute, sia sufficiente ricordare come RDF non sia un formato, bensì un *data model*, un metodo di memorizzazione delle informazioni. Tali informazioni, in particolare, sono espresse in termini di triple, asserzioni caratterizzate da soggetto-predicato-oggetto. Soggetto e oggetto della tripla sono URIs che identificano risorse.

<sup>56</sup> SPARQL, acronimo di Sparql Protocol And Rdf Query Language, è un linguaggio di interrogazione per i dati rappresentati secondo le triple RDF. Le query, dunque le interrogazioni, vengono inviate da un client ad un server, detto *end-point* Sparql, attraverso l'utilizzo del protocollo http.

<sup>57</sup> <https://contrattipubblici.org/>, consultato il 29/08/2016

aggiudicatario, importo, durata del contratto (data di inizio e data di fine), prezzo concordato, tipo di procedura, oltre naturalmente al codice identificativo del contratto.

Le direttrici di sviluppo in ambito LOD (*linked open data*) sono naturalmente quelle verso la pubblicazione diretta, da parte delle pubbliche amministrazioni, dei dati sottoforma di triple. In tal senso alcune amministrazioni virtuose hanno già provveduto a pubblicare alcuni dati in formato collegato, quindi sottoforma di triple. Tra queste, la Camera dei Deputati, il Comune di Firenze e la Provincia di Carbonia-Iglesias. Altre, hanno incominciato a produrre dati in RDF, senza però offrire il punto di accesso a tali dati, il cosiddetto *end-point*.

Il valore dei *linked open data*, è utile rimarcarlo una volta in più, consiste nella semplicità di pubblicazione e di consultazione dei dati espressi in tale *data model*, e dalle possibilità di riutilizzo e di interoperabilità tra database differenti.

Di seguito il dettaglio delle sezioni previste dal Decreto Legislativo 14 marzo 2013, n. 33.

Denominazione sotto-sezione 1 livello	Denominazione sotto-sezione 2 livello	Contenuti (riferimento al decreto)
Disposizioni generali	Programma per la Trasparenza e l'integrità	Art. 10, c. 8, lett. a
	Atti generali	Art. 12, c. 1,2
	Oneri informativi per cittadini e imprese	Art. 34, c. 1,2
Organizzazione	Organi di indirizzo politico-amministrativo	Art. 13, c. 1, lett. a
		Art. 14
	Sanzioni per mancata comunicazione dei dati	Art. 47
	Rendiconti gruppi consiliari regionali/provinciali	Art. 28, c. 1
	Articolazione degli uffici	Art. 13, c. 1, lett. b, c
	Telefono e posta elettronica	Art. 13, c. 1, lett. d
Consulenti e collaboratori		Art. 15, c. 1,2
Personale	Incarichi amministrativi di vertice	Art. 15, c. 1,2
		Art. 41, c. 2, 3
	Dirigenti	Art. 10, c. 8, lett. d
		Art. 15, c. 1,2,5
		Art. 41, c. 2, 3
	Posizioni organizzative	Art. 10, c. 8, lett. d
	Dotazione organica	Art. 16, c. 1,2
	Personale non a tempo indeterminato	Art. 17, c. 1,2
	Tassi di assenza	Art. 16, c. 3
	Incarichi conferiti e autorizzati ai dipendenti	Art. 18, c. 1
	Contrattazione collettiva	Art. 21, c. 1
	Contrattazione integrativa	Art. 21, c. 2
	OIV	Art. 10, c. 8, lett. c
Bandi di concorso		Art. 19
Performance	Piano della Performance	Art. 10, c. 8, lett. b
	Relazione sulla Performance	Art. 10, c. 8, lett. b
	Ammontare complessivo dei premi	Art. 20, c. 1
	Dati relativi ai premi	Art. 20, c. 2
	Benessere organizzativo	Art. 20, c. 3
Enti controllati	Enti pubblici vigilati	Art. 22, c. 1, lett. a
		Art. 22, c. 2, 3
	Società partecipate	Art. 22, c. 1, lett. b
		Art. 22, c. 2, 3
	Enti di diritto privato controllati	Art. 22, c. 1, lett. c
		Art. 22, c. 2, 3
Attività e procedimenti	Rappresentazione grafica	Art. 22, c. 1, lett. d
	Dati aggregati attività amministrativa	Art. 24, c. 1
	Tipologie di procedimento	Art. 35, c. 1,2
	Monitoraggio tempi procedurali	Art. 24, c. 2
	Dichiarazioni sostitutive e acquisizione d'ufficio dei dati	Art. 35, c. 3
Provvedimenti	Provvedimenti organi indirizzo politico	Art. 23
	Provvedimenti dirigenti	Art. 23
Controlli sulle imprese		Art. 25
Bandi di gara e contratti		Art. 37, c. 1,2
Sovvenzioni, contributi, sussidi, vantaggi economici	Criteri e modalità	Art. 26, c. 1
	Atti di concessione	Art. 26, c. 2
		Art. 27
Bilanci	Bilancio preventivo e consuntivo	Art. 29, c. 1
	Piano degli indicatori e risultati attesi di bilancio	Art. 29, c. 2
Beni immobili e gestione patrimonio	Patrimonio immobiliare	Art. 30
	Canoni di locazione o affitto	Art. 30
Controlli e rilievi sull'amministrazione		Art. 31, c. 1
Servizi erogati	Carta dei servizi e standard di qualità	Art. 32, c. 1
	Costi contabilizzati	Art. 32, c. 2, lett. a
		Art. 10, c. 5
	Tempi medi di erogazione dei servizi	Art. 32, c. 2, lett. b
	Liste di attesa	Art. 41, c. 6
Pagamenti dell'amministrazione	Indicatore di tempestività dei pagamenti	Art. 33
	IBAN e pagamenti informatici	Art. 36
Opere pubbliche		Art. 38
Pianificazione e governo del territorio		Art. 39
Informazioni ambientali		Art. 40
Strutture sanitarie private accreditate		Art. 41, c. 4
Interventi straordinari e di emergenza		Art. 42
Altri contenuti		

Figura 2: Sezioni Amministrazione Trasparente secondo quanto previsto dal Dlgs 33/2013

# Capitolo 3

## *Case study. La Pubblica Amministrazione* **tra dati e narrazione**

Il presente capitolo presenta, nella prima parte, il caso di studio relativo all'analisi del contenuto e rappresenta la concretizzazione del lavoro di analisi dal versante mediologico mentre, come già accennato, un primo approccio relativo all'analisi dei dati in senso giornalistico sarà riportato nella seconda parte del capitolo.

### **3.1 I testi**

Nel primo capitolo si è cercato di offrire uno sguardo di insieme al discorso sull'analisi del contenuto, utile ad aprire la strada alle pagine che seguono. In particolare, questo terzo capitolo verterà inizialmente sulla descrizione del software utilizzato, andando a chiarire la logica che sostiene l'analisi, mentre oggetto della seconda parte sarà l'analisi vera e propria, intervallata da ulteriori chiarimenti tecnici che si renderanno necessari via via che le evidenze emergeranno.

### **Lo strumento e le fasi metodologiche**

Da un punto di vista metodologico, per il caso di studio in oggetto si è scelto di utilizzare T-lab 9.1.0.<sup>58</sup>, software che unisce strumenti linguistici, statistici e grafici sviluppato da Franco Lancia. Un ipotetico progetto di lavoro è costituito da una serie di operazioni su un medesimo corpus che possono essere ricondotte a sei successive

---

<sup>58</sup> <http://tlab.it/it/presentation.php>

fasi. Data per scontata la fase di raccolta dei testi, che nel nostro caso è avvenuta in data 14 luglio 2016 come riportato nella nota metodologica del secondo capitolo, l'analisi comincia con la preparazione del corpus. Si tratta della fase di trasformazione dei testi da analizzare in un file, detto corpus, che possa essere elaborato da T-lab. Tra i particolari requisiti vi sono la condizione di essere formato in solo testo con codifica ASCII/ANSI e la necessità di preparare il corpus in un formato .txt. Se, come nel caso della nostra analisi, il corpus è costituito da più testi e da più variabili, è necessario ricorrere alla definizione di un modello in grado di essere letto da T-lab. Da un punto di vista formale, infatti, ogni riga di codifica ha il seguente formato: inizia con una stringa di quattro asterischi (\*\*\*\*) seguita da uno spazio bianco (*blank*). Una stringa così formata viene interpretata dal software come un testo od una unità di contesto definita dall'utilizzatore. La riga di codifica prosegue poi con l'aggiunta di stringhe costituite da singoli asterischi (\*) ed etichette che individuano casi in maniera inequivocabile (IDnumber), variabili e modalità, e finisce con un ritorno di carrello, il volgarmente detto "a capo". Può essere a questo punto riportare una stringa utilizzata nel corpus analizzato, in maniera da chiarire questo passaggio formale.

```
**** *IDnumber_001 *Data_ago2013 *Testata_espres
```

Con tale stringa si indica in maniera univoca il frammento di testo (che verrà scritto nel rigo sottostante alla stringa), facendo presente che le due variabili sulle quali si articolerà l'analisi saranno la data, espressa in questo caso come mese e anno, e testata.

Successiva alle fase di preparazione del corpus è quella di importazione. Questa operazione trasforma il file di testo predisposto dall'utilizzatore in un insieme di tabelle integrate nel database di sessione. Il processo di importazione si articola in una serie di operazioni la cui prima, temporalmente, è la normalizzazione del testo, che consiste nella verifica del corpus con il duplice obiettivo di consentire un corretto riconoscimento delle parole come forme grafiche e di risolvere dal principio alcuni casi di ambiguità. Tale normalizzazione prevede, tipicamente, alcune trasformazione del file oggetto dell'analisi, come l'eliminazione di spazi vuoti in eccesso, la

marcatura degli apostrofi, la riduzione delle maiuscole. Operazione più importante delle altre e pertanto degna di nota è la marcatura di stringhe riconosciute come nomi propri di persone e di luoghi e la trasformazione delle forme grafiche riconosciute come locuzioni o *multiwords*<sup>59</sup> in stringhe unitarie da intendersi pertanto come tali nel processo di analisi. Altre operazioni che rientrano nel computo dell'importazione del corpus sono il riconoscimento di *stop-words*, ovvero parole che non veicolano un significato in maniera autonoma, come articoli, avverbi ed interiezioni, e la segmentazione del corpus, ovvero l'individuazione dei contesti elementari che possono essere di quattro tipi: frasi, vale a dire contesti elementari marcati da punteggiatura forte (come ad esempio i segni . ? !), frammenti, cioè sequenze di parole interrotte da ritorno di carrello e di dimensione inferiore a quattrocento caratteri, paragrafi, ossia contesti elementari arcati da punteggiatura forte e ritorno di carrello, oppure ancora testi brevi.

Più rilevante è il discorso riguardante la lemmatizzazione, altra operazione che rientra tra quelle di importazione. Facendo riferimento al concetto di lemma, identificabile come ogni singola voce di un dizionario, la lemmatizzazione riconduce le forme dei verbi all'infinito presente, quelle dei sostantivi e degli aggettivi al maschile singolare, quelle delle preposizioni articolate alla loro forma senza articolo e via discorrendo, consentendo così l'analisi unica dei lemmi riconducibili alla stessa radice.

Una volta importato il corpus, il software procede a domandare all'utilizzatore se desidera utilizzare le impostazioni automatiche o quelle personalizzate. La differenza consiste nel modo in cui il software andrà a selezionare le parole chiave. Con tale termine T-lab intende semplicemente tutte le unità lessicali (parole, lemmi, lessie e categorie) che man mano vengono incluse nelle tabelle da analizzare. L'inclusione od esclusione delle unità lessicali dall'elenco delle parole chiave dipende pertanto dalla scelta adoperata in sede di impostazione. Nel caso in cui si scelga di proseguire con le impostazioni predefinite T-lab seleziona le parole con valori di occorrenza

---

<sup>59</sup> In linguistica e semantica con il termine *multiwords* si è soliti intendere sequenze di parole che, a livello del significato, costituiscono una unità lessicale.

superiori alla soglia minima<sup>60</sup>, calcola i valori del TF-IDF<sup>61</sup> o applica il test del CHI quadrato<sup>62</sup> a tutti gli incroci di ogni parola selezionata per tutti i testi in analisi, e seleziona infine le parole con i valori maggiori nel metodo utilizzato, estrapolando così le parole che, all'interno del testo, sembrano far la differenza.

Se tale è il metodo di individuazione delle parole chiave di un testo, differente è il discorso attorno al vocabolario del corpus. Con questo termine, infatti, si intende la lista che include tutte le parole contenute nel corpus, i cosiddetti *word types*, le loro occorrenze, ossia i *word tokens*, e i corrispondenti lemmi. Tale strumento rappresenta un'utile cartina tornasole in grado di mostrare la ricchezza del vocabolario ed altri dati interessanti, quali l'*apax*, ovvero le occorrenze con frequenza 1, e la frequenza massima riscontrabile nel testo.

Tipicamente il lavoro svolto con l'ausilio di T-Lab prosegue lavorando con gli strumenti lessico. Si lavora pertanto sul vocabolario del corpus, eventualmente personalizzandolo secondo le esigenze del ricercatore, procedendo con le operazioni di disambiguazione e la definizione delle liste di *stop-words* e *multi-words*. Si rimanda alle prossime pagine la descrizione delle operazioni attuate nel caso specifico della mia tesi. È infatti più utile, per completezza metodologica, concludere ora la carrellata relativa alle diverse fasi in cui il lavoro su T-Lab si compone. La quarta fase riguarda la selezione delle parole chiave: si tratta di un momento particolarmente rilevante in quanto l'intervento del ricercatore in fase di selezione delle *keywords* inciderà sugli output prodotti dall'analisi, in particolare nel caso delle scelte adottate per il presente lavoro. In questa ricerca si è scelto infatti di procedere ad analisi di limitata portata, stante il disegno della ricerca volto sì ad indagare la rappresentazione mediale di alcuni temi riguardanti la pubblica amministrazione e

---

<sup>60</sup> La soglia di frequenza minima imposta da T-lab per l'inclusione di un termine all'interno dell'elenco delle parole chiave è quattro. Tuttavia, a seconda delle dimensioni del corpus si applicano calcoli differenti utili a definire il valore di soglia: tipicamente è incluso tra il primo ed il secondo decile del range.

<sup>61</sup> La misura definita TF-IDF consente di valutare l'importanza di un termine, ovvero di un'unità lessicale, all'interno di un documento, ossia un'unità di contesto. Si calcola moltiplicando il numero delle occorrenze del termine per il logaritmo del rapporto tra il numero complessivo dei documenti del corpus e del numero dei documenti del corpus che contengono il termine.

<sup>62</sup> Il CHI quadro è un test statistico utile a verificare se i valori di frequenza di un fenomeno, come quelli di un'unità lessicale, siano significativi diversi da quelli attesi secondo una normale distribuzione.

*l'open government*, ma in combinazione con un piano di analisi a più ampio raggio. Degli strumenti analisi, che rappresentano la quinta e più corposa fase di lavoro su T-Lab, si è pertanto scelto di adoperare in particolare l'analisi tematica dei contesti elementari. Si tratta di una tecnica afferente alla famiglia delle analisi delle corrispondenze lessicali, utile ad ottenere una rappresentazione grafica dei contenuti dei testi del corpus. In particolare tale analisi tematica elabora il materiale oggetto di analisi e restituisce una serie di cluster tematici. Anche in questo caso le scelte metodologiche sono appannaggio del ricercatore, il quale può scegliere di procedere secondo le modalità automatiche suggerite da T-Lab piuttosto che adoperando alcune modifiche in fase di impostazione. In particolare è consentito scegliere, a proposito del metodo di analisi, tra un *clustering* non supervisionato, basato sull'algoritmo *Bisecting K-means*, e un *clustering* supervisionato, basato sul dizionario delle categorie. Per ragioni di opportunità, in questa sede, si è proceduto a scegliere l'approccio non supervisionato, in grado di offrire una immagine grafica dei valori ottenuti dalla differenza tra le frequenze osservate e quelle attese<sup>63</sup>. Il processo non supervisionato adotta, infatti, un approccio di costruzione di una tabella dati unità di contesto x unità lessicali con valori del tipo presenza/assenza. In questo modo il software procede verificando l'eventuale presenza di determinate parole (unità

---

<sup>63</sup> È utile introdurre ora in maniera più completa alcuni aspetti tecnici propri del software T-Lab. L'analisi tematica dei contesti elementari segue un modello che ricalca in buona misura quello messo a punto da Charles Osgood nel 1959 chiamato analisi delle co-occorrenze. In questo caso il concetto che sta alla base delle analisi prende origine dal tradizionale approccio quantitativo all'analisi del contenuto, muovendo dall'assunto che, se alcune parti di un messaggio co-occorrono per un numero di volte superiore a quello presupposto dal semplice caso, questo è in qualche misura indicativo delle associazioni che informano il modo di pensare del mittente [Osgood 1959].

Concretamente, si tratta di costruire una matrice *casi x variabili*, con in riga le unità di analisi testuali ed in colonna le categorie entro cui sono state classificate. In ognuna delle celle risultanti dall'incrocio riga per colonna si procede a segnare con un + l'eventuale presenza della categoria all'interno dell'unità di analisi, e con il segno – l'eventuale assenza.

Ulteriori recenti sviluppi dell'analisi del contenuto prevedono di trasformare il testo in vettori, cioè in grandezze matematiche la cui descrizione richiede di prendere in considerazione più informazioni contemporaneamente. Procedendo in questo modo il testo diviene operativizzabile in una stringa di valori la cui «elaborazione statistica consente di far emergere similarità semantiche tra coppie di valori o tra segmenti, cioè sequenze di occorrenze di lunghezza prestabilita dal ricercatore» [Tipaldo 2014: 73].

lessicali) all'interno di determinate frasi (unità di contesto), procedendo con una tipica analisi delle co-occorrenze<sup>64</sup> prima, e con un'analisi di tipo comparativo<sup>65</sup> poi.

Tornando all'algoritmo *Bisecting K-means*, è opportuno spendere ora alcune parole utili alla sua definizione, prima di procedere all'analisi del caso in questione. L'algoritmo divide il testo e lo elabora proponendolo al ricercatore come un grafico a dispersione in cui coppie di coordinate individuano punti in un piano cartesiano. La rappresentazione grafica dei vettori, ovvero dei temi che attraversano il testo, facilita la comprensione dei *pattern* del testo stesso ed è propedeutica all'interpretazione degli output, operazione in capo all'analista, che rappresenta la sesta ed ultima fase prima dell'elaborazione del report.

È giunto ora il momento di procedere all'esposizione dei primi, elementari, risultati delle analisi del caso di studio. Per chiarezza espositiva procederò presentando i corpora in maniera distinta, rimandando al capitolo conclusivo il compito di rappresentare una summa del lavoro.

## Corpus n.1. Decreto Trasparenza

Come introdotto nelle scorse pagine, la prima operazione da svolgere sui corpus è rappresentata dalla verifica del vocabolario attraverso la disambiguazione e la selezione delle parole chiave che andranno a rappresentare le unità lessicali utili alla narrazione della storia. Si tratta, a ben vedere, di un'operazione delicata e a cui è opportuno dedicare il tempo necessario, dal momento che la scelta adottata in sede di parole chiave è in grado di influenzare gli esiti dell'intera ricerca. Per questa ragione il ruolo del ricercatore deve limitarsi, per quanto possibile, alla semplice pulizia dei

---

<sup>64</sup> Parlare di co-occorrenze significa conteggiare il numero di volte in cui due o più unità lessicali sono contemporaneamente presenti all'interno degli stessi contesti elementari.

<sup>65</sup> Le analisi comparative comprendono, in T-Lab, una serie di operazioni tra cui l'analisi delle specificità, l'analisi delle corrispondenze, la cluster analysis e le tabelle di contingenza. Di particolare rilievo è l'analisi delle specificità, uno strumento che permette di evidenziare quali forme grafiche sono tipiche all'interno di un testo oppure in una qualsiasi delle sue parti definita da una variabile categoriale [Lancia 2012].

dati, in maniera non dissimile dal compito che attende il *data journalist* nel momento di approcciarsi al proprio *dataset*.

Nel mio caso, ed in particolare con riguardo al corpus composto dagli articoli selezionati tramite la ricerca d'archivio dei quotidiani attraverso la *keyword* “decreto trasparenza”, ho scelto di procedere con una disambiguazione utile a ricondurre i termini *amministrazione*, ed il relativo plurale, e il termine *PA*, al termine *pubblica amministrazione*. È stata poi mia premura quella di ricondurre al medesimo lemma sostantivi portatori del medesimo significato ma presenti in forme diverse, tipicamente singolare e plurale. Ho così disambiguato i termini *anno-anni*, *bilancio-bilanci*, *caso-casi*, *cittadino-cittadini*, *compenso-compensi*, *obbligo-obblighi*, oltre al caso *ente-ente locale*.

Successiva alla disambiguazione è stata la costruzione personalizzata del vocabolario attraverso la definizione delle parole chiave. Anche in questo caso si è scelto di procedere nell'ottica di non inficiare l'analisi con un intervento massivo da parte del ricercatore. Si sono pertanto scelte 510 parole chiave utili successivamente in fase di analisi tematica.

Da un punto di vista analitico, il vocabolario è stato esplorato in maniera statistica, ricorrendo all'analisi delle occorrenze, utili a verificare, ad un livello elementare, l'emergere dei termini maggiormente ricorrenti. Su tale piano si collocano pertanto evidenze di scarsa profondità metodologica, in grado tuttavia di definire un primo quadro relativo alle già esposte tesi di *agenda setting* e di *agenda framing*. Sono d'altronde il ricorso ripetitivo di termini e la loro collocazione all'interno del testo gli aspetti più immediati e in grado di stabilire una connessione semantica tra testo e lettore. Nel caso in analisi, scegliendo di escludere i termini *trasparenza* e *decreto* per ovvie ragioni di opportunità<sup>66</sup>, il termine *pubblica amministrazione* è risultato quello con la maggiore occorrenza, presente in 130 occasioni nelle diverse declinazioni a cui ho fatto riferimento in precedenza. Seguono *euro*, con 58

---

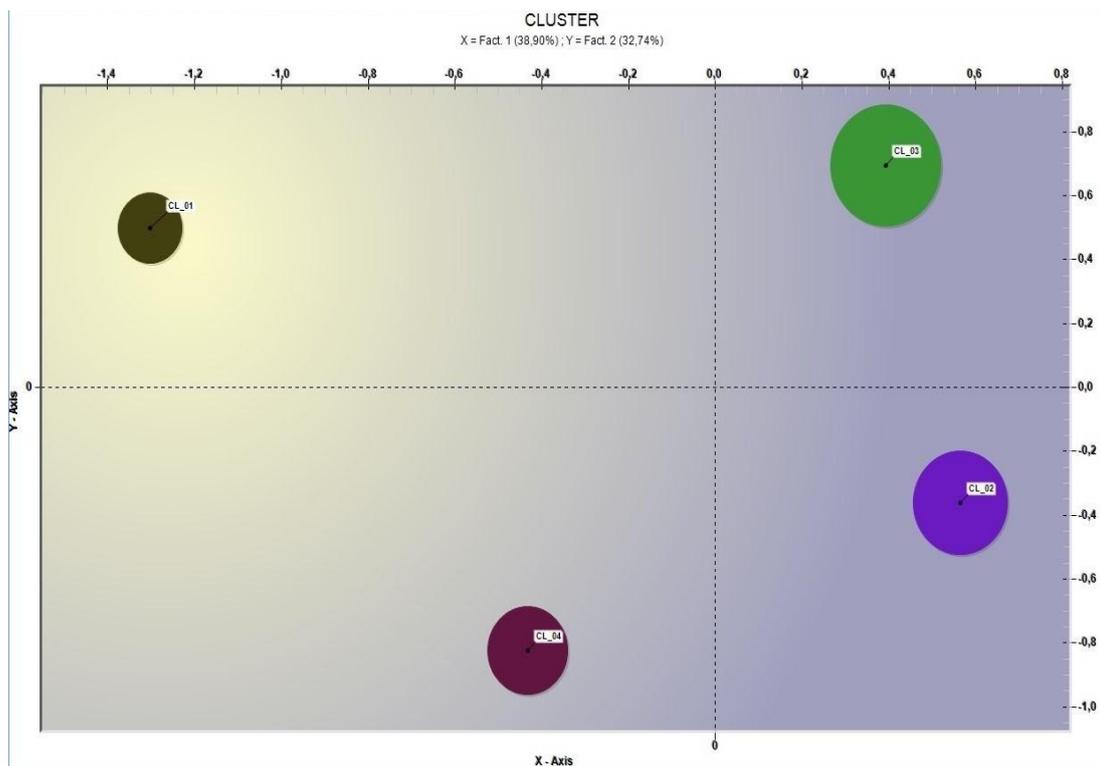
<sup>66</sup> La scelta di escludere tali termini dall'analisi è dovuta al fatto che la selezione degli articoli è proceduta attraverso la ricerca, negli archivi dei quotidiani, proprio dei termini *decreto* e *trasparenza*. La stessa logica ha naturalmente sotteso l'incedere nel caso degli altri tre corpora, costruiti attraverso la ricerca, nei medesimi archivi, dei termini *foia*, *open government* e *itp*.

occorrenze, *pubblicare* (55), *dati* (50) e *legge* (48). Si riporta in appendice al capitolo (tabella 1) l'elenco delle occorrenze calcolate da T-Lab.

Per leggere con sufficiente maturità analitica l'elenco delle occorrenze è opportuno richiamare brevemente quanto contenuto nel cosiddetto Decreto Trasparenza, il decreto legislativo 14 marzo 2013, n.33/2013. Se ne è parlato diffusamente nel corso del secondo capitolo, qua può essere utile ricordare come, tale passo legislativo, abbia rappresentato l'avvio a quella rivoluzione digitale delle pubbliche amministrazioni in nome della trasparenza e della lotta a corruzione e sprechi di risorse pubbliche. Uno degli aspetti più noti del decreto trasparenza è, nei fatti, l'istituzione della sezione Amministrazione Trasparente all'interno dei portali web delle singole amministrazioni. Nelle ambizioni originarie tale sezione avrebbe dovuto rappresentare il canale privilegiato, per i cittadini, verso una sorta di controllo dell'affare pubblico, attraverso la disponibilità di una serie di dati resi aperti e consultabili da chiunque. Tuttavia, come ha evidenziato il lavoro svolto in collaborazione con Synapta circa l'Indicatore di Tempestività dei Pagamenti, la effettiva natura *open* di tali dati è messa in discussione da una serie di vincoli tecnologici e di pubblicità.

Tornando alla questione relativa al vocabolario, si osserva una netta preponderanza di termini strettamente legati alla questione tecnica della legge. Tra questi, *pubblicare, dati, legge, articolo, norma, sanzioni, pagamento*. La settima unità lessicale, tuttavia, apre un versante che è quello in cui il presente lavoro di ricerca si inserisce. Con 43 occorrenze, il termine *informazione* rappresenta la prova di quanto intensa, già nel momento della pubblicazione in Gazzetta Ufficiale del testo di legge, fosse la questione relativa alle potenzialità giornalistiche offerte dalla (presunta) apertura delle pubbliche amministrazioni.

Passando all'analisi tematica dei contesti elementari, tale è la rappresentazione grafica dei cluster (figura 3).

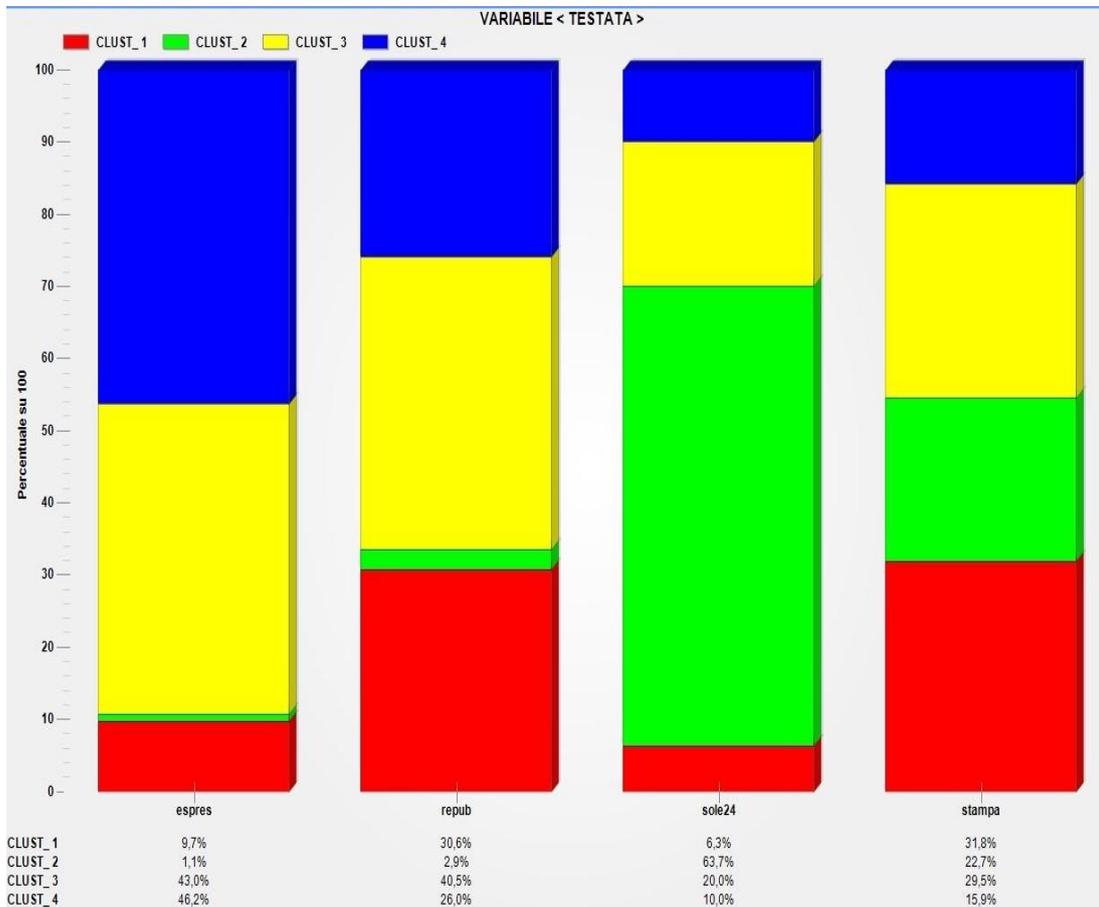


**Figura 3: Corpus "Decreto Trasparenza". Rappresentazione grafica dei cluster tematici**

Utile alla definizione dei temi come offerti da T-Lab è l'analisi dei cluster attraverso la rilettura dei frammenti di testo. La ricerca dei fattori più rilevanti avviene tipicamente attraverso l'analisi dei *contesti significativi*, da cui è possibile tornare sul corpus importato originariamente e servirsi del supporto del software, che in questo caso evidenzia, frammento dopo frammento, gli stralci maggiormente caratteristici.

Da un primo sguardo al grafico, integrato ad una lettura intensiva dei frammenti del corpus, è possibile intendere la disposizione dei cluster in relazione a due direttrici tematiche rappresentate dagli assi cartesiani. Il cluster 1 appare in evidente contrapposizione rispetto al cluster 2, così come il cluster 3 con il cluster 4. Prima di elaborare ipotesi di senso relative ai fattori che attraversano il corpus può essere utile soffermarsi sulla suddivisione dei fattori in rapporto alla variabile testata, andando cioè a valutare l'incidenza di un determinato tema per ciascuna delle testate prese in considerazione. Nel caso del corpus elaborato sulla base della *keyword* decreto

trasparenza, risulta chiara la contrapposizione dei temi all'interno di giornali diversi, come dimostra il grafico (figura 4).



**Figura 4: Corpus "Decreto Trasparenza". Percentuale di articoli di ogni testata per ogni cluster tematico**

Il cluster 1, che rappresenta l'insieme di un certo numero di articoli caratterizzati dagli stessi temi, è popolato in gran parte da frammenti provenienti da La Stampa e La Repubblica, mentre il cluster 2 è chiaramente abitato in maniera preponderante dai testi de Il Sole 24 Ore. Tale aspetto, senza timore di condurre a conclusioni affrettate, aiuta nella definizione dei fattori presenti nel corpus, in virtù delle notevoli differenze editoriali che caratterizzano i tre giornali. Se La Stampa e La Repubblica rappresentano la voce dell'informazione quotidiana<sup>67</sup>, il tratto caratterizzante de Il

<sup>67</sup> Tale affermazione va intesa al netto dell'esclusione de Il Corriere della Sera dal corpus, oltre che dal recente processo di fusione tra il Gruppo Espresso e Itedi, le cui implicazioni editoriali sono ancora lontane dall'essere chiare.

Sole 24 Ore è evidentemente il taglio economico del proprio giornalismo. Alla luce di questo necessario, e non scontato, avviso, è possibile intendere l'asse delle ascisse proprio come il binomio "informazione-legislazione", dal momento che il cluster 1 sembra ricondurre il tema del decreto trasparenza ai casi politici in cui, la trasparenza prevista dalla legge, non ha trovato riscontri. In particolare si fa riferimento ai casi in cui alcuni protagonisti della scena politica hanno fatto ricorso a scudi per mantenere celata l'entità dei propri redditi, o ancora la questione sul bilancio di Expo 2015. Il cluster 1, quello composto in massima parte da articoli di La Stampa e La Repubblica, tornano così quasi quotidianamente sulla battaglia elettorale tra Sala e Parisi, per la poltrona di sindaco di Milano alle elezioni di giugno 2016, e sul dibattito romano tra Virginia Raggi e Roberto Giachetti, tra gli incarichi di Civitavecchia non dichiarati e le consulenze rimaste occultate. Spiccano così termini quali *accusa, attaccare, polemica*, oltre ad una serie di tematiche legate ai tristemente noti episodi di corruzione attorno all'Esposizione Universale di Milano 2015, con tanto di puntuali riferimenti ai termini *vicenda, procura, esposto*. Dall'altro lato, il cluster 2 è caratterizzato da un tratto maggiormente legato alle disposizioni in materia di trasparenza, dall'*obbligo della pubblicità dei dati ai bilanci*, fino alle *sanzioni*. Il corpus di articoli ricondotti al cluster 2, quasi tutti provenienti da Il Sole 24 Ore, rappresenta tipicamente il materiale di stampo economico-informativo, scevro di ogni componente di colore.

Emerge così in maniera piuttosto evidente il differente taglio offerto dai quotidiani ai temi che afferiscono alla stessa sfera, quella del decreto trasparenza.

Dall'altro lato, i cluster 3 e 4 rappresentano l'articolazione del discorso mediale attorno al tema dell'accesso ai dati. Più che un'antitesi tra due universi tematici, i due cluster sembrano rappresentare due facce della stessa medaglia. Se il cluster numero 3 è caratterizzato dalla ricorsività dei temi di (*diritto di*) *accesso, freedom of information act, informazione, Italia*, il cluster numero 4 rappresenta la concretizzazione di tali tratti in una serie di inchieste condotte da L'Espresso. Anche in questo caso, infatti, l'identificazione della provenienza degli articoli che compongono i diversi cluster rivela la natura degli stessi. Se, da una parte, L'Espresso rappresenta i valori dell'odierno giornalismo d'inchiesta, una testata

come La Repubblica adotta un approccio evidentemente diverso, pur facendo parte della stessa proprietà editoriale. Un cluster, quello numero 4, caratterizzato pertanto da lemmi quali *euro*, *ministro-moglie* (utile ad identificare la questione dell'intestazione, da parte di esponenti politici, di beni e redditi alle proprie consorti per sfuggire al decreto trasparenza che li obbliga alla pubblicità), e ancora a termini quali *reddito*, *stipendio* e *guadagnare*.

Anche in questo caso il corpus rivela differenze su un piano metodologico più che concettuale. Il tema dell'accesso ai dati, come garantito dal Decreto Trasparenza, rappresenta un tema caldo sull'agenda di tutti i giornali. È la natura del discorso a cambiare, proprio in virtù di caratteristiche quali il mezzo (supporto e tempistiche di pubblicazione) ed il pubblico di riferimento. Una testata come Il Sole 24 Ore sembra offrire ai propri lettori uno strumento utile all'azione, dal momento che si rivolge ad una platea di professionisti spesso e volentieri coinvolti da tali tematiche. Testate come La Repubblica e La Stampa sembrano piuttosto costruire il proprio racconto sulla quotidianità, alla ricerca di una chiave di lettura da offrire ai propri lettori e di maturare pertanto un rapporto fiduciario in grado di protrarsi nel corso del tempo. Il discorso, infatti, pare articolarsi tipicamente sulla cronaca, affidando ad esempi concreti la ricostruzione di un universo tematico più complesso. L'Espresso rappresenta evidentemente un *unicum* di questo corpus, a cominciare dalla frequenza di pubblicazione. La cadenza settimanale della rivista garantisce una maggiore capacità di tematizzazione e di approfondimento. Non si leggeranno pertanto notizie di cronaca: le sue pagine saranno piuttosto l'occasione per capovolgere l'annosa questione della subalternità del giornalismo alla notizia. In tal senso L'Espresso rappresenta uno dei possibili tipi di concretizzazione del discorso giornalistico condotto attraverso l'*open data*. Ciò che tuttavia sembra mancare è un discorso quantitativo, fondato sul *fact checking* dei dati offerti dalle pubbliche amministrazioni, dal momento che la narrazione offerta dalla rivista sembra collocarsi, su un piano stilistico, come inferenziale-esplicativa. Tale è il caso di testi costruiti con la finalità di attribuire ai fatti narrati la prospettiva interpretativa dell'autore [Tipaldo 2014], mentre sul piano del tono utilizzato si possono rintracciare alcune forme che contribuiscono a definire L'Espresso come

dimostrativo, cioè che segue la successione logica di ipotesi, dimostrazione, verifica e tesi.

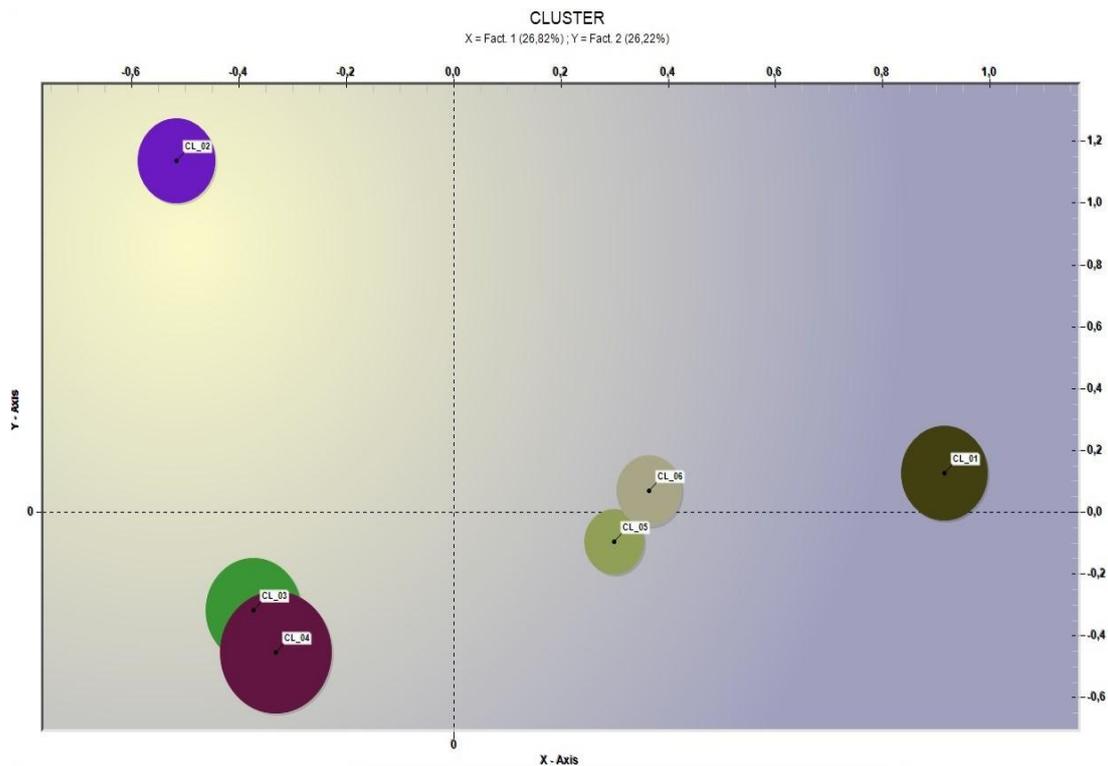
## Corpus n.2. Freedom of Information Act

Prima di approcciarsi al discorso relativo all'analisi del contenuto del corpus composto dai testi contenenti le *keywords* “*freedom of information act*” è opportuno richiamare alcuni aspetti introduttivi a tale corpus. Si tratta, intanto, di 108 articoli pubblicati tra il 1 gennaio 2013 ed il 13 luglio 2016. Cinquanta di questi centotto articoli risalgono al 2016, anno in cui il discorso attorno l'opportunità, o meglio la necessità, di una legge sul diritto all'accesso si è acceso ed ha portato alla promulgazione della legge stessa.

L'analisi del vocabolario ha messo in evidenza i termini occorsi il maggior numero di volte. La *multi word* “pubblica amministrazione” ricorre 571 volte, seguita da “trasparenza” (in 249 occasioni), “accesso” (233), “dati” (204) e “legge” (201).

In appendice (tabella 6) la tabella del vocabolario riferito al corpus “*freedom of information act*”.

Come nel caso del corpus relativo a *decreto trasparenza* si è scelto di procedere con l'analisi tematica dei contesti. L'elaborazione dei dati mostra l'esistenza di sei macrotemi espressi in altrettanti cluster tematici, come si può osservare dall'immagine (figura 5).



**Figura 5: Corpus "Freedom of Information Act". Rappresentazione grafica dei cluster tematici**

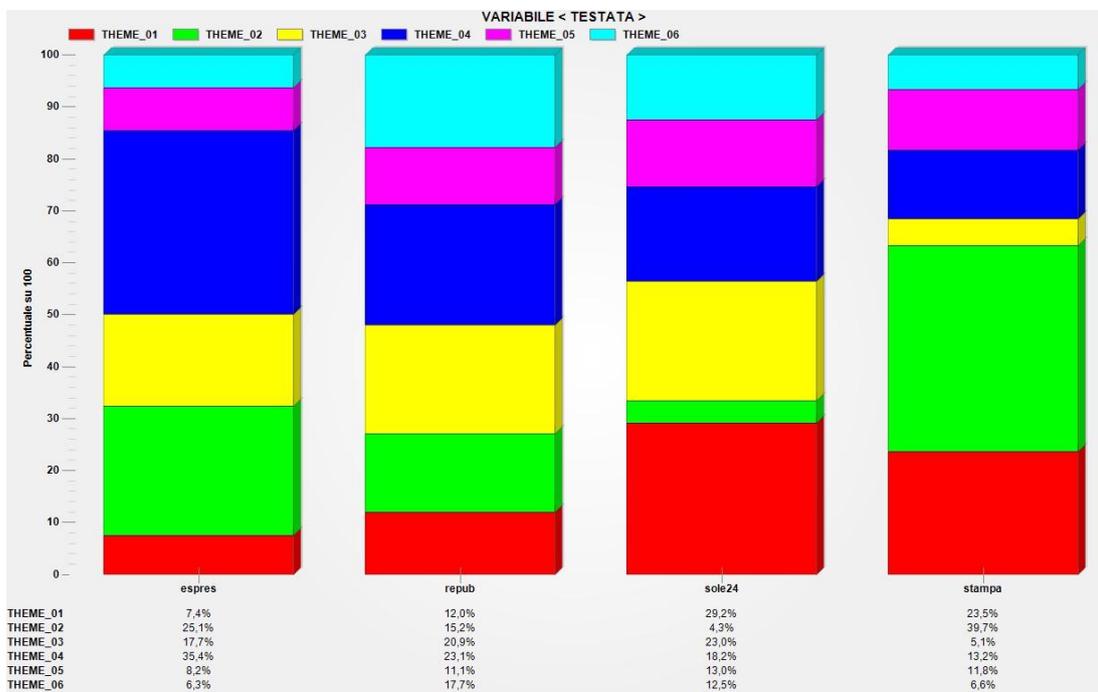
Si rimanda all'appendice del presente capitolo per l'esplorazione approfondita dei vocabolari tipici di ogni singolo cluster (tabelle da 7 a 12) . In questa sede ritengo più importante estrapolare un significato dalle analisi condotte sul corpus.

Un elementare approccio visivo ai cluster suggerisce una lettura orientata su quattro grandi categorie tematiche, e su una forte contrapposizione tra il cluster 1 ed il cluster 2. Il grafico sembra organizzarsi sui contenuti geografici dei due corpora: se nel cluster 1 tornano sovente termini afferenti alla sfera politica italiana, da *Grillo* a *Bersani*, da *Imu* al *Movimento 5 Stelle*, il cluster 2 si articola sulla ricorsività di personaggi appartenenti alla cronaca estera, da *Julian Assange* a *Hillary Clinton*, passando per *Carlo* principe di Galles, fino a *Edward Snowden*. Il discorso si articola spesso sulla contrapposizione, riguardante l'accesso ai dati e le possibilità di riutilizzo, tra la situazione italiana e l'estero. Si è già sottolineato negli scorsi capitoli il ritardo con cui il governo italiano ha accolto le richieste della società civile, in

particolare il movimento Foia4Italy, nel preparare una riforma sulle pubbliche amministrazioni e la loro trasparenza.

I cluster 3 e 4, che graficamente mostrano una certa vicinanza se non addirittura similarità tematica, si rivelano prossimi nei contenuti anche di fronte all'analisi dei contesti elementari. In tal senso è lecito guardare alle due collezioni di testi come due facciate dello stesso edificio. Fuor di metafora, il cluster numero 3 colleziona stralci relativi al processo di definizione della legge sulla libertà di accesso agli atti, tracciando un profilo storico della *Riforma Madia*, gli step seguiti dal *governo Renzi* verso l'*approvazione* del testo, in accordo con i pareri della *Camera* e della società civile (in corsivo i lemmi più ricorrenti). Il cluster 4, dal canto suo, racchiude i frammenti dei testi che esplicano le principali disposizioni previste dal Freedom of Information Act.

Anche in questo caso può essere utile, al fine di appoggiarsi su evidenze statistiche nell'elaborazione di ipotesi, a verificare la composizione dei diversi cluster sulla base della variabile testata (figura 6).



**Figura 6:** Corpus "Freedom of Information Act". Percentuale di articoli di ogni testata per ogni cluster tematico

L'Espresso colloca la maggior parte (35,4%) dei propri articoli nel cluster 4, quello in cui convergono i frammenti utili a ricostruire il cosa prevede la legge. Tuttavia proprio l'analisi dei contesti elementari, limitata agli articoli del cluster 4 proveniente da L'Espresso, evidenziano un rapporto di stretta correlazione tra le disposizioni previste e una serie di termini quali *tradimento*, *troppo poco*, *difetto*, lemmi che suggeriscono di leggere tale cluster in maniera più critica, in un senso di inchiesta giornalistica come effettivamente la rivista romana è abituata a fare. Il quotidiano La Repubblica si ritrova nel 20,9% dei casi all'interno del cluster 3, e nel 23,1% dei casi nel cluster 4 che, come abbiamo visto, rappresentano il discorso sul Freedom of Information Act condotto attraverso il percorso di approvazione e l'esplorazione dei contenuti. Dato interessante è quello de Il Sole 24 Ore che offre ai propri lettori un *frame* interpretativo spesso caratterizzato dalle implicazioni sul piano dei possibili investimenti pubblici da sbloccare come conseguenza della trasparenza delle PA prevista dal FOIA. Per quanto riguarda La Stampa, la maggior parte degli articoli rientrano nel cluster 2, quello caratterizzato dagli argomenti di cronaca estera, mentre i cenni al processo di approvazione del Freedom of Information Act italiano si limitano a testi più brevi, strettamente descrittivi degli step seguiti nel corso della discussione in aula del testo del ministro Madia.

Da quanto affermato finora emerge una certa marginalità dei cluster 5 e 6. Si tratta di un raggruppamenti di scarso rilievo quantitativo, ma caratterizzati dalla ricorsività di termini espressione della presunta svolta digitale in atto nella società e nella politica italiana. Con particolare riguardo al cluster 6, si può affermare che i temi riguardino in particolare le potenzialità dell'*open government*, della trasparenza dell'affare pubblico, ma letti in relazione con le possibile derive che l'accessibilità comporta. La questione della privacy e della tutela dei dati sensibili rappresentano il focus di questa collezione di materiale. Si tratta quindi di un ambito tutt'altro che marginale nel discorso sul *data journalism* e sulle implicazioni deontologiche già richiamate nei precedenti capitoli. L'analisi di dati, la curiosità nell'indagare storie e nel trovare chiavi di lettura inedite nel materiale prodotto dalle pubbliche amministrazioni non può prescindere dalla consapevolezza della necessità di tutelare le fonti, uno dei pilastri del mestiere del giornalista. In questo senso si apre un ampio scenario in cui

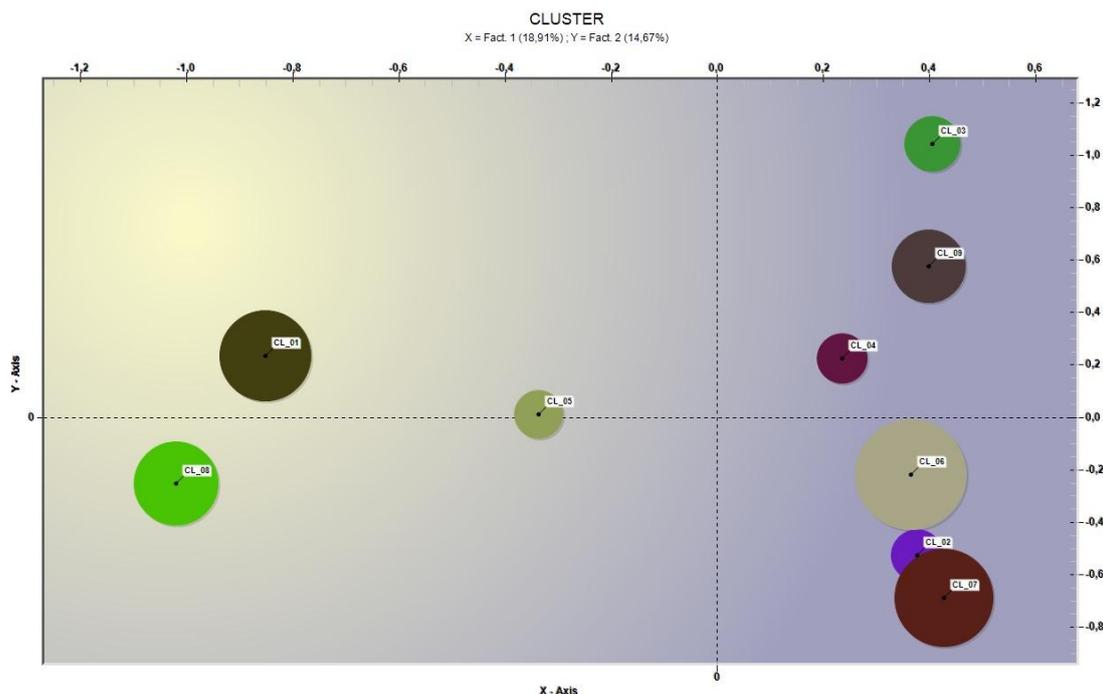
il lavoro del *data journalism* non può limitarsi all'interpretazione dei dataset. Le implicazioni sociali, e legali, necessitano di essere lette e considerate in maniera approfondita, onde evitare di trasformare uno strumento, i dati al servizio dell'informazione, in un'arma. La professionalità del giornalista, in questo senso, deve uscire rafforzata, e non scalfita, dal processo di innovazione in corso nelle redazioni.

### Corpus n.3. Open Government

Il corpus numero 3 è composto da 37 articoli provenienti da tre dei quattro giornali presi in considerazione per questo lavoro di ricerca. Attraverso lo strumento di archivio digitale, infatti, non è stato possibile risalire agli articoli, contenenti le *keywords* "open government", apparsi sul quotidiano La Stampa. Nel caso del presente corpus, di conseguenza, l'analisi sarà condotta solo sui testi provenienti dai quotidiani La Repubblica e il Sole 24 Ore, oltre che sul settimanale L'Espresso.

Come sempre, le prime operazioni condotte sono state volte alla normalizzazione del corpus e alla definizione del vocabolario attraverso l'individuazione delle parole chiave. Secondo tale approccio, ad emergere come lemmi dalla maggiore occorrenza sono *cittadini*, presente in 86 occasioni, *data* (80), la maggior parte delle volte (72) in abbinamento all'aggettivo *open*, seguiti da *digitale* (63) ed *Italia* (61). Numericamente meno frequenti, ma di notevole rilevanza analitica già da questa elementare verifica, sono i termini *nuovo* (occorrente 43 volte) e *innovazione* (40), capaci di veicolare il senso di novità legato alle tematiche di governo aperto di cui si è approfonditamente discusso nel capitolo 2.

Rimandando all'appendice del capitolo per la tabella di occorrenza (tabella 13), e per le tabelle utili a ricostruire i vettori tematici del testo (tabelle da 14 a 18), è utile anche in questo caso esplorare il corpus attraverso l'analisi tematica dei contesti elementari, sempre secondo il metodo di clustering non supervisionato. Il grafico risultante secondo l'algoritmo *Bisecting K-means* è quello riportato in figura 7.



**Figura 7: Corpus "Open Government". Rappresentazione grafica dei cluster tematici**

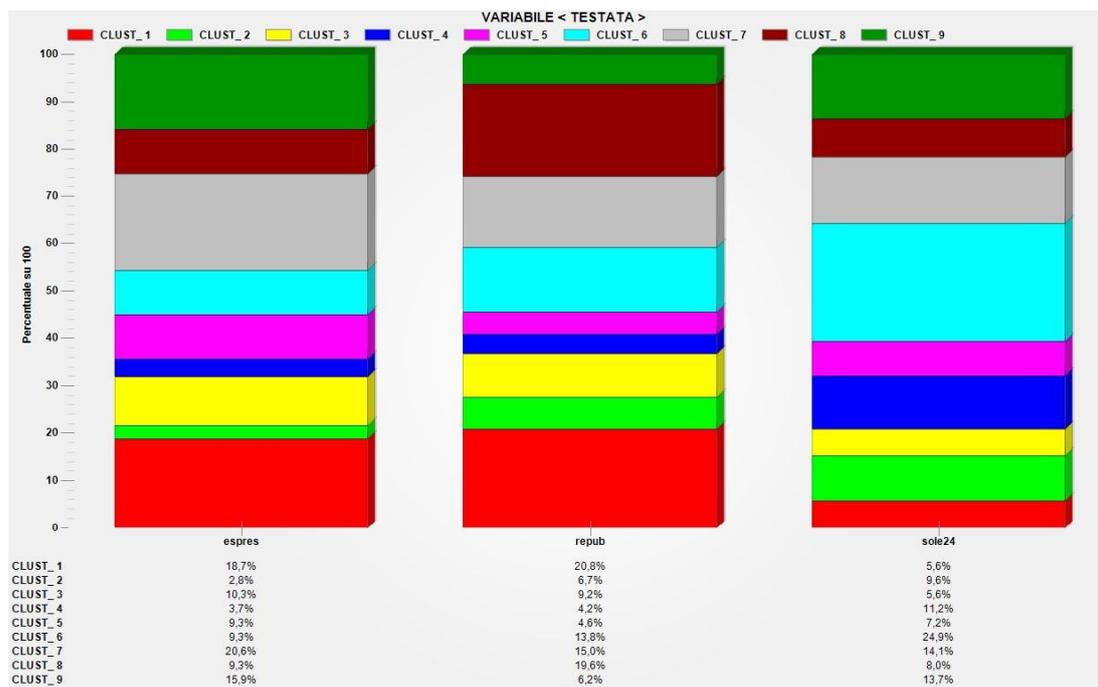
La disposizione dei cluster tematici suggerisce una lettura orientata su una chiara contrapposizione tra due gruppi, disposti sull'asse delle ascisse. L'analisi delle occorrenze dei termini inclusi nel cluster 1, uno dei due disposti sulla sinistra del grafico, vede il lemma *aula* al primo posto, con un valore del chi quadrato di oltre 84 punti, seguito da altri termini di chiaro stampo politico quali *sinistra*, *destra*, *movimento*, *maggioranza*, *governo*, *giunta* e *Pd*. Situazione non particolarmente dissimile per quello che riguarda il cluster 8 che, oltre ad una vicinanza grafica suggerita dal grafico, rivela congruenze sul piano del contenuto. A emergere sono altri termini legati alla cronaca politica, in questo caso però riferiti strettamente alla questione romana: *sindaco*, *Ama*, *rifiuti*, *ambiente*, *assessore*, *Raggi*, *Paola Muraro* (Assessore all'Ambiente del Comune di Roma), *Roma*, *Fortini* (Presidente di Ama), *Luca Bergamo* (Assessore alla cultura del comune di Roma, *trasporti*, *Atac*, solo per riportarne alcuni. La narrazione mediale della questione sull'*open government* dunque rivela fin da subito una stretta relazione con la cronaca, in particolare per quello che riguarda le vicende politiche nazionali. Si fa pertanto riferito al dibattito in aula (cluster 1), e alla questione romana (cluster 8). Proprio quest'ultima rappresenta

un punto di svolta nell'interpretazione testuale, dal momento che il Movimento 5 Stelle ha spesso fatto ricorso ai temi dell'apertura dei dati e all'integrazione dell'affare pubblico tra amministrazioni e cittadini come emblema programmatico della propria offerta politica. È stata però la stessa dinamica di *open government* ad incrinare la solidità della giunta Raggi non appena insediatasi in Campidoglio, in un cortocircuito di cause e conseguenze su un piano politico e sociale.

Esposte le principali evidenze dei cluster 1 e 8, passiamo al lato destro del grafico. In questo caso si è scelto di concentrarsi sui cluster quantitativamente più rilevanti, ossia il 6, il 7 e il 9.

*Dataset, Freedom of information act, riuso, formato aperto, rilasciare*, sono i termini più ricorrenti nel cluster 6. *Informazione, tecnologia, banche dati, riutilizzare, statistica*, sono alcuni dei lemmi di maggior rilievo nel cluster 7. *Programma, digitale, digitalizzazione, rete, banda larga* sono invece i termini più ricorrenti nel cluster 9. Ricorrendo ad un'analisi congiunta di tali dati si può, con un accettabile grado di consapevolezza, intendere il grafico come orientato sul vettore di "attualità-prospettiva", o ancora di "oggi-domani", intendendo in questo senso la contrapposizione di cronaca, dello status quo, dell'attualità, con quanto concerne la sfera delle potenzialità, del futuro del governo dell'affare pubblico, indissolubilmente legata al tema dell'alfabetizzazione digitale della società, proprio nell'ottica di una collaborazione amministrazione-cittadini, emblema del governo aperto.

Può essere utile cercare una conferma della lettura proposta nelle righe precedenti nella composizione dei diversi cluster secondo la testata (figura 8).



**Figura 8: Corpus "Open Government". Percentuale di articoli di ogni testata per ogni cluster tematico**

Da questo punto di vista, emerge di fatto che la maggior parte degli articoli de La Repubblica rientrano nei cluster 1 e 8, quelli identificati come il racconto della cronaca, nazionale e romana, mentre il Sole 24 Ore si colloca in netta preponderanza nel cluster 6 (24,9% dei propri articoli), il cui pattern di riferimento è in effetti la natura tecnica del processo di *open government*. L'Espresso, infine, trova la maggior collocazione nei cluster 7 e 9, a conferma di una certa predisposizione nella tematizzazione dei fatti e nella collocazione degli avvenimenti in un'ottica di insieme, tipica del processo di settimanalizzazione dell'informazione. Emerge, tuttavia, un parallelismo tra l'Espresso e La Repubblica nell'analisi del cluster 8, quello che tratta la questione romana. Le ragioni sono piuttosto evidenti, a cominciare dall'appartenenza allo stesso gruppo editoriale, basato peraltro proprio nella Capitale. Da un punto di vista contenutistico, poi, la natura degli articoli riferiti alla città oggi guidata dal Movimento 5 Stelle ha un stampo chiaramente di inchiesta, segno anche di un certo schieramento in fatto di linea editoriale. Superfluo richiamare in questa sede il profilo storico del gruppo editoriale l'Espresso; sia sufficiente ricordare come la questione Mondadori-Espresso a cavallo degli anni '90

abbia acuito una polarizzazione in fatto di linea editoriale, già storicamente tracciata dalla rivista settimanale l'Espresso.

Tornando ancora sulla questione di vocabolario, interessante è notare come ricorrano con una certa frequenza lemmi riconducibili alla questione sanitaria: *paziente, dati sanitari, sanità, health*, sono il sintomo di uno degli utilizzi di *open data* a cui la società civile maggiormente tiene. La mappatura dei servizi di assistenza sanitaria, la verifica delle modalità di somministrazione di tali servizi, i relativi costi, sono aspetti caldi nell'agenda dei media e in quella pubblica; altrettanto importanti sono le potenzialità a cui l'integrazione di differenti dataset contenenti dati sanitari potrebbe condurre, e anche in tal senso la questione di *linked data* consentirebbe l'esplorazione di nuove frontiere, nazionali o globali, nella ricerca di nuove cure e di soluzioni a emergenze sanitarie.

## 3.2 I dati

Quello che segue è il report relativo ai dati sull'Indicatore di Tempestività dei Pagamenti pubblicati dai comuni piemontesi secondo la metodologia esposta nel secondo capitolo.

### **Piemonte in costante ritardo sui pagamenti ai fornitori. E va sempre più lento**

L'indicatore di tempestività dei pagamenti esprime il numero di giorni in cui i comuni piemontesi pagano le fatture ai fornitori. Nei comuni sopra i 15 mila abitanti i giorni di ritardo sono in media più di 20. A Torino addirittura 33. I dati sorprendenti arrivano dai comuni piccoli, dove si paga con maggiore puntualità.

I dati sono riferiti ai comuni del Piemonte con più di mille abitanti e svelano chi paga in tempo e chi in ritardo. Non sono pochi, però, i casi in cui le pubbliche amministrazioni mancano di pubblicare i dati sulla tempestività dei pagamenti, nonostante il Decreto Trasparenza del 2013 lo sancisca come obbligo.

### **Cos'è l'ITP e come viene calcolato?**

Parliamo dei pagamenti dei comuni piemontesi ai fornitori di merci o servizi. L'ambito è delicato ed il quadro complesso. Ma dalla confusione emergono alcuni trend che i dati raccolti da Synapta consentono di raccontare con precisione.

*I comuni piemontesi pagano in ritardo*, oggi ed in generale negli ultimi due anni. I dati, pochi per la verità, riferiti ai pagamenti del primo trimestre 2016, dicono che sono poco più di 6, in media, i giorni di ritardo con cui i comuni saldano i propri debiti. Attenzione, però, il dato va letto nel modo corretto, prima di balzare a conclusioni affrettate: l'indicatore di tempestività dei pagamenti (d'ora in poi ITP), quel numero che le pubbliche amministrazioni hanno l'obbligo di pubblicare

trimestralmente nella sezione Amministrazione Trasparente del proprio sito web, rappresenta il ritardo medio dei pagamenti. Tale ritardo esprime il numero di giorni oltre la scadenza delle fatture impiegati per pagare: tipicamente, secondo quanto previsto dal decreto legislativo 192/2012 del 9 novembre 2012, le pubbliche amministrazioni hanno trenta giorni per saldare i propri debiti con i fornitori. L'ITP conteggia i giorni che eccedono questa scadenza. Il calcolo è, per la verità, ancora più complicato: il decreto del Presidente del Consiglio dei Ministri del 22 settembre 2014 (pubblicato in G.U. n.265 del 14 novembre 2014) prevede che il calcolo dell'ITP sia ponderato sulla base dell'importo delle fatture. Di seguito un esempio:

*Si consideri un caso ipotetico in cui nell'anno 2014 siano state pagate tre fatture.*

*Fattura A: scadenza 4 aprile 2014; pagamento avvenuto il 16 settembre 2014 per un importo dovuto al fornitore pari a 54 mila euro;*

*Fattura B: scadenza 30 novembre 2014; pagamento avvenuto il 2 novembre 2014 per un importo dovuto al fornitore di 60 mila euro;*

*Fattura C: scadenza 30 novembre 2014; pagamento avvenuto il 4 dicembre 2014 per un importo dovuto al fornitore di 120 mila euro.*

*Il calcolo prevede le seguenti elaborazioni:*

	importo in euro (a)	data emissione fattura	data scadenza (b)	data pagamento (c)	differenza in giorni effettivi tra il pagamento e la scadenza (d) = (c)-(b)	ritardo ponderato (a) * (d)
fattura a	54.000,00	05/03/2014	04/04/2014	16/09/2014	165,00	8.910.000,00
fattura b	60.000,00	31/10/2014	30/11/2014	02/11/2014	-28,00	- 1.680.000,00
fattura c	120.000,00	31/10/2014	30/11/2014	04/12/2014	4,00	480.000,00
<b>somma</b>	<b>234.000,00</b>					<b>7.710.000,00</b>
<b>indicatore di tempestività dei pagamenti (2) / (1)</b>						<b>32,95</b>

*L'indicatore di tempestività dei pagamenti è in questo caso pari a 32,95.*

### **ITP. La situazione: quando pagano le pubbliche amministrazioni?**

Chiarite le modalità di calcolo dell'ITP, possiamo ad analizzare i dati.

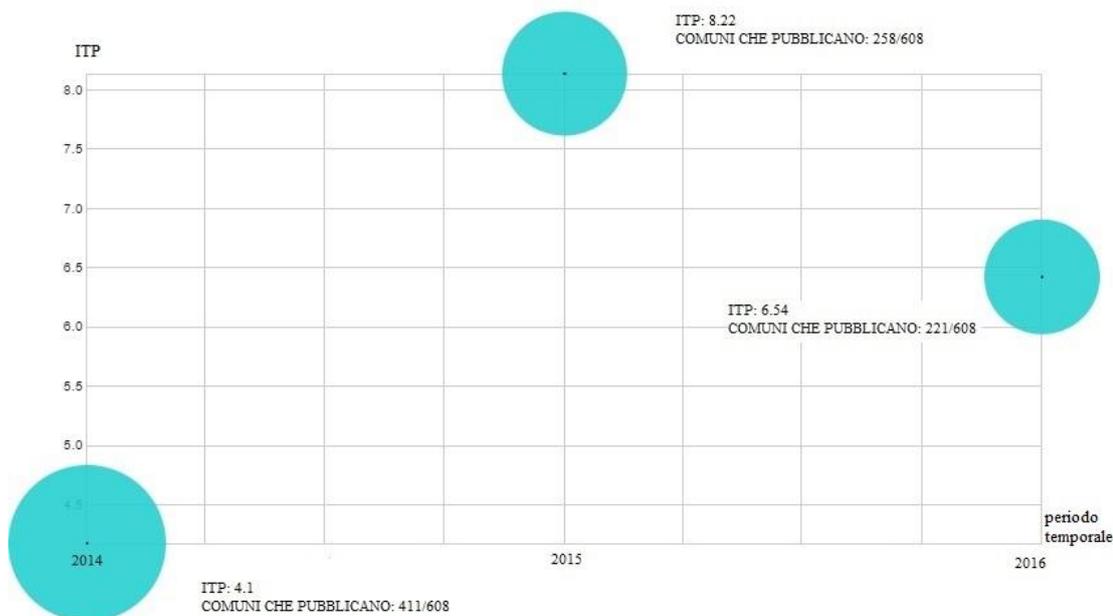
La nostra inchiesta ha riguardato i comuni piemontesi con più di mille abitanti ed il campione è composto da 608 comuni. L'analisi ha messo in luce le tempistiche di pagamenti di tre periodi: l'anno 2014, l'anno 2015 ed il primo trimestre 2016.

La tabella che segue riguarda le medie complessive dell'ITP dei 608 comuni analizzati:

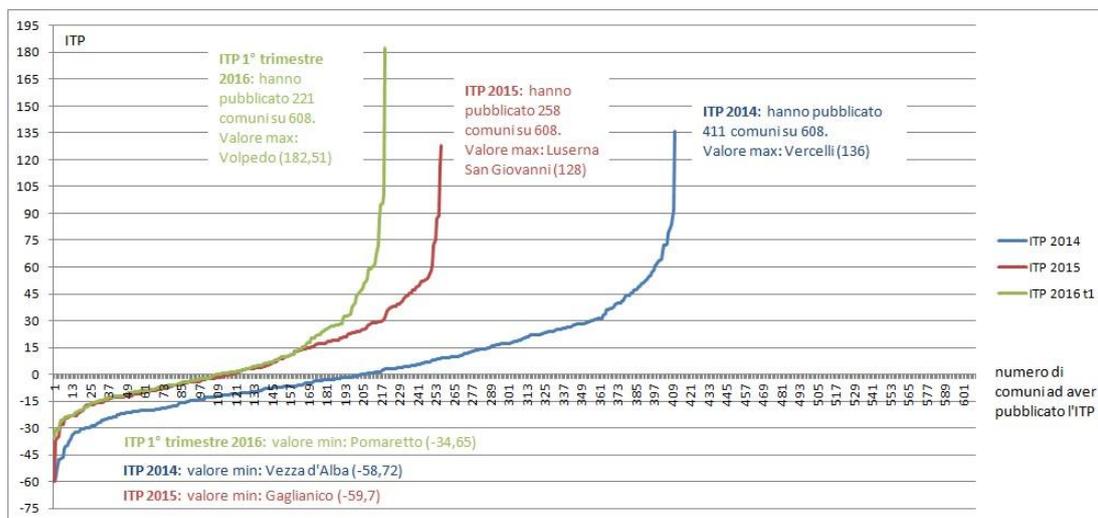
Anno 2014	Anno 2015	1° trimestre 2016
4.1	8.22	6.54

In tutti e tre i periodi presi come riferimento, dunque, la media dei tempi di pagamento mostra un ritardo. Considerando la scadenza media delle fatture di trenta giorni, come ricordato prima, i comuni dovrebbero avere idealmente ITP compresi tra -30 e 0, anche se la ponderazione del calcolo secondo l'importo delle fatture determina una certa varianza, facendo "pesare" di più le fatture con importi elevati rispetto a quelle maggiormente economiche. In ogni caso un valore positivo dell'ITP rileva un ritardo medio nei tempi di pagamento.

Su piano cartesiano i valori medi complessivi di tutti i 608 comuni piemontesi riportati nella tabella precedente seguono questo schema. La dimensione dei punti dipende dal numero di comuni che hanno pubblicato i dati. Emerge immediatamente che il numero di comuni ad aver pubblicato i dati stia diminuendo in maniera sensibile. Se nel 2014, primo anno di applicazione del Decreto Trasparenza (D.Lgs 33/2013 del 14 marzo 2013), erano stati 411 comuni piemontesi su 608 a pubblicare il dato, nel 2015 lo avevano fatto solo 258, e nel primo trimestre del 2016 solo 221.



Simile è il grafico che illustra, per i tre periodi, il range di valori assunti dall'ITP dei comuni.

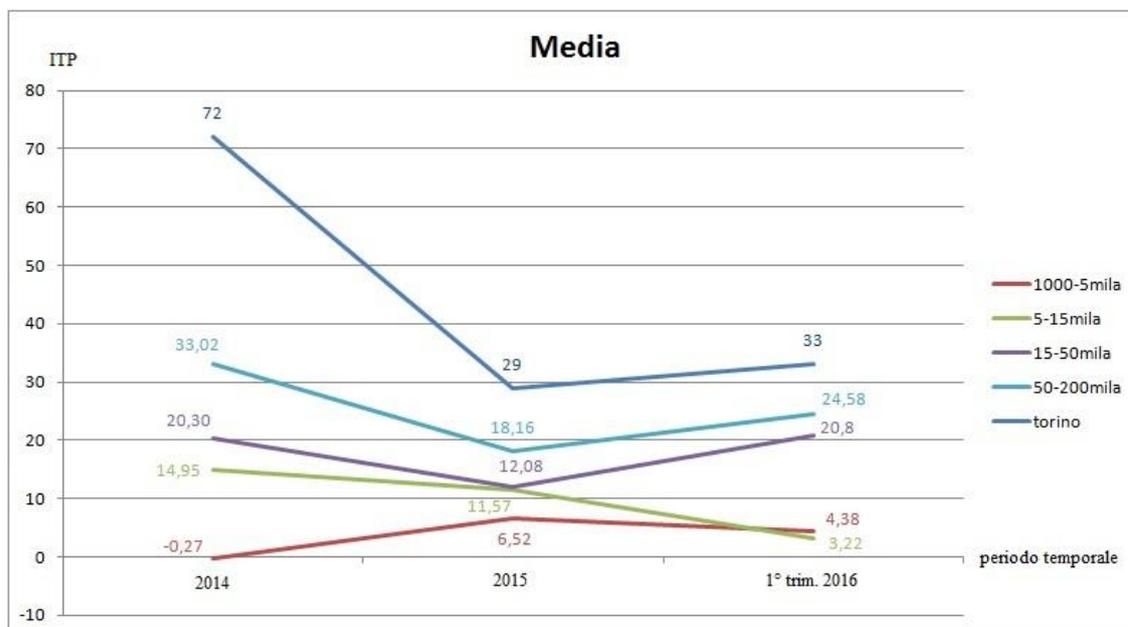


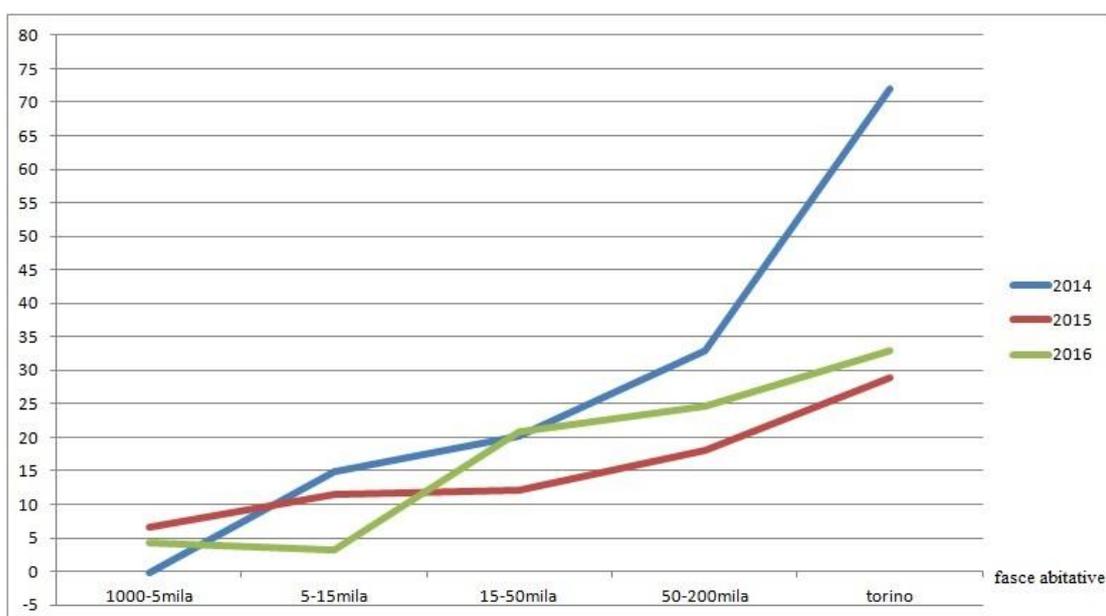
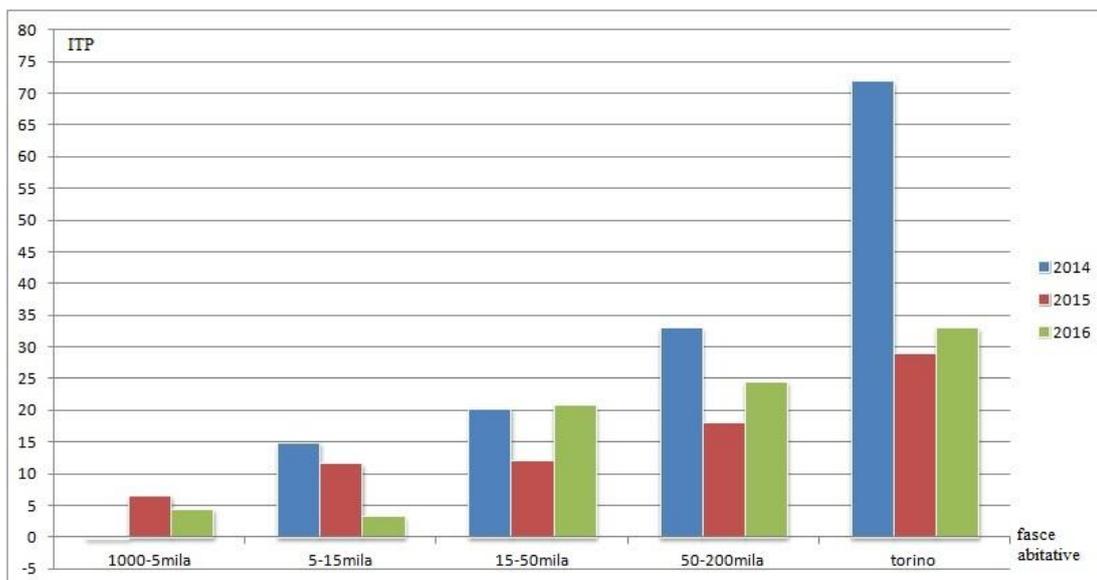
## Il problema. Si paga in ritardo

Più significative sono le analisi condotte segmentando i comuni secondo il numero di abitanti. Abbiamo proceduto utilizzando cinque fasce.

abitanti	n. comuni	media itp 2014	media itp 2015	media itp 1°trim 2016
1.000 – 4.999	475	-0.27	6.53	4.38
5.000 – 14.999	87	14.95	11.57	3.22
15.000 – 49.999	40	20.3	12.08	20.8
50.000 – 200.000	5	33.02	18.16	24.58
> 200.000 (Torino)	1	72	29	33

Dalla tabella emerge come i comuni piemontesi siano abituati a pagare in ritardo. Solo nel 2014, e limitatamente ai comuni piccoli, è accaduto che i debiti venissero pagati entro la scadenza.





I due grafici mostrano come la media dei tempi di pagamento aumentino all'aumentare del numero di abitanti. In altre parole, i comuni grandi pagano molto più lentamente che i paesi piccoli, fenomeno interessante perché riflesso di almeno due aspetti. I comuni piccoli affrontano evidentemente minori spese ai fornitori, ma

devono probabilmente far fronte anche a minore disponibilità di risorse. La bilancia tra entrate e uscite, tuttavia, sembra essere favorevole.

### **Pagare in ritardo è un'abitudine. Le sanzioni**

Sono previste sanzioni ai comuni che pagano in ritardo?

Sì, secondo quanto previsto dalla direttiva 2011/7/UE del Parlamento Europeo e del Consiglio. In particolare la circolare del 16 febbraio 2011, relativa alla lotta contro i ritardi di pagamento nelle transazioni commerciali, prevede che “il creditore abbia diritto, senza che sia necessario un sollecito, agli interessi legali di mora - vale a dire interessi ad un tasso che è pari al tasso di riferimento (tasso di interesse applicato dalla BCE alle sue più recenti operazioni di rifinanziamento) - maggiorato di almeno otto punti percentuali.”

L'articolo 41, comma 2, del decreto legge 66/2014 introduce ulteriori specifiche. In particolare la norma fissa in 90 giorni per il 2014, ed in 60 giorni a decorrere dal 2015, il limite massimo dei tempi medi di pagamento per le pubbliche amministrazioni. Ciò significa che tali termini si applicano alla media dei tempi di pagamento del periodo preso a riferimento e non alla singola fattura. Il decreto legge 66/2014, poi, pone come sanzione il blocco delle assunzioni da parte dell'ente per l'anno successivo a quello in cui si è registrata la violazione dei termini (90 giorni per il 2014 e 60 giorni dal 2015). Una sanzione severa che, colpendo indistintamente ogni violazione ai tempi di pagamento, può investire amministrazioni che abbiano registrato ritardi per motivi a loro non imputabili come ad esempio i limiti del Patto di Stabilità e la disponibilità economica delle singole amministrazioni. A questo proposito è possibile che i piccoli comuni, in modo particolare, abbiano scarse entrate dirette e debbano rimettersi ai tempi di trasferimento di risorse da parte degli enti maggiori, con conseguenti allungamenti dei tempi.

Ciò che forse è più assurdo, però, è la mancata predisposizione di sanzioni a carico dei comuni che non pubblicano il dato relativo all'ITP. Il Decreto Trasparenza 33/2013, infatti, sancisce solo che il responsabile per la prevenzione della corruzione,

come previsto all'articolo 1, comma 7, della legge 6 novembre 2012, n.190, svolga le funzioni di responsabile per la trasparenza, controllando l'adempimento degli obblighi di pubblicazione e segnalando all'organo di indirizzo politico, all'Organismo indipendente di valutazione (OIV), all'Autorità nazionale anticorruzione e, nei casi più gravi, all'ufficio di disciplina i casi di mancato o ritardato adempimento degli obblighi di pubblicazione.

All'articolo 46, comma 1, poi, viene sancito che “l'inadempimento degli obblighi di pubblicazione previsti dalla normativa vigente o la mancata predisposizione del Programma triennale per la trasparenza e l'integrità costituiscono elemento di valutazione della responsabilità dirigenziale”: l'eventuale responsabilità sarà dunque a carico esclusivo del dirigente di riferimento ed interverrà sulla corresponsione della retribuzione.

### **Il secondo problema. Chi non pubblica**

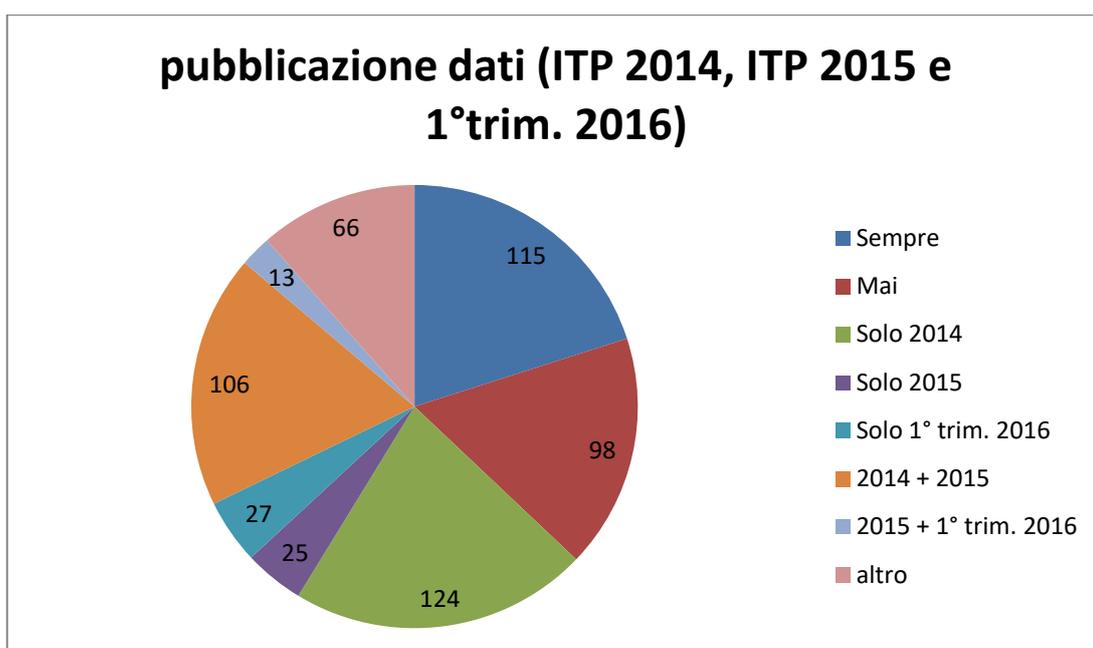
Dal 2014 al 1° trimestre 2016, come già detto, i comuni piemontesi ad aver pubblicato il dato ITP si è quasi dimezzato, passando da 411 a 221 (-46,23%).

Sempre con riferimento al periodo dal 2014 al 1° trimestre 2016, sono 98 i comuni piemontesi sui 608 analizzati a non aver mai pubblicato il dato ITP secondo quanto previsto dalla legge. Quasi un comune su sei è abituato a ignorare le disposizioni di legge che obbligano le amministrazioni a pubblicare i propri dati in nome di quella trasparenza che dovrebbe “favorire forme diffuse di controllo sul perseguimento delle funzioni istituzionali e sull'utilizzo delle risorse pubbliche” (art. 1, comma 1, del Decreto Trasparenza 33/2013 del 14 marzo 2013).

Evitando di prendere in considerazione i valori trimestrali del 2015, di cui parliamo al prossimo paragrafo, abbiamo scelto di rappresentare graficamente la situazione di pubblicazione dell'ITP. In particolare abbiamo calcolato il numero di comuni che abbiano soddisfatto le seguenti condizioni:

- mai pubblicato un dato: 98/608

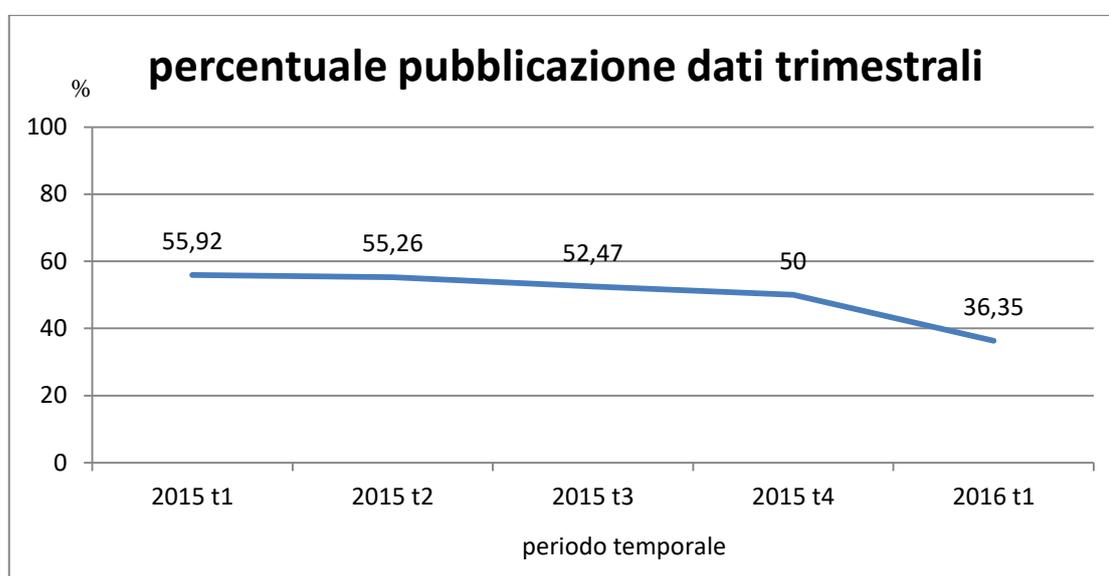
- sempre pubblicato (ITP annuale 2014, ITP annuale 2015, ITP 1° trimestre 2016): 115/608
- pubblicato solo ITP annuale 2014: 124/608
- pubblicato solo ITP annuale 2015: 25/608
- pubblicato solo ITP 1° trimestre 2016: 27/608
- pubblicato ITP annuale 2014 e ITP annuale 2015: 106/608
- pubblicato ITP annuale 2015 e 1° trimestre 2016: 13/608
- altro: 66/608



### **I dati trimestrali sulla pubblicazione**

Proviamo ora a seguire l'andamento della pubblicazione dell'ITP su base trimestrale, in modo da avere periodi temporali confrontabili con maggior rigore. In questo caso i dati a disposizione riguardano i quattro trimestri del 2015 ed il primo del 2016.

	2015 t1	2015 t2	2015 t3	2015 t4	2016 t1
numero di comuni che pubblicano dati	340	336	319	304	221
percentuale	55,92%	55,26%	52,47%	50%	36,35%

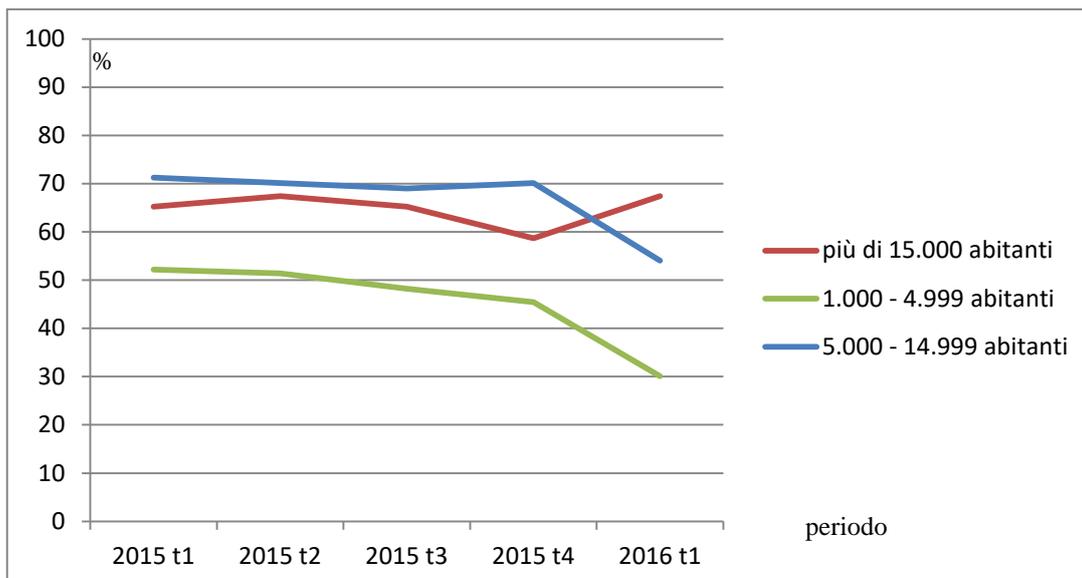


La percentuale di pubblicazione del dato ITP trimestrale decresce costantemente, seppur in misura meno accentuata. Dal dato del 1° trimestre 2016 è opportuno rilevare che i comuni piemontesi sembrano essere lenti non soltanto nel pagare le proprie fatture, ma anche nel pubblicare i dati.

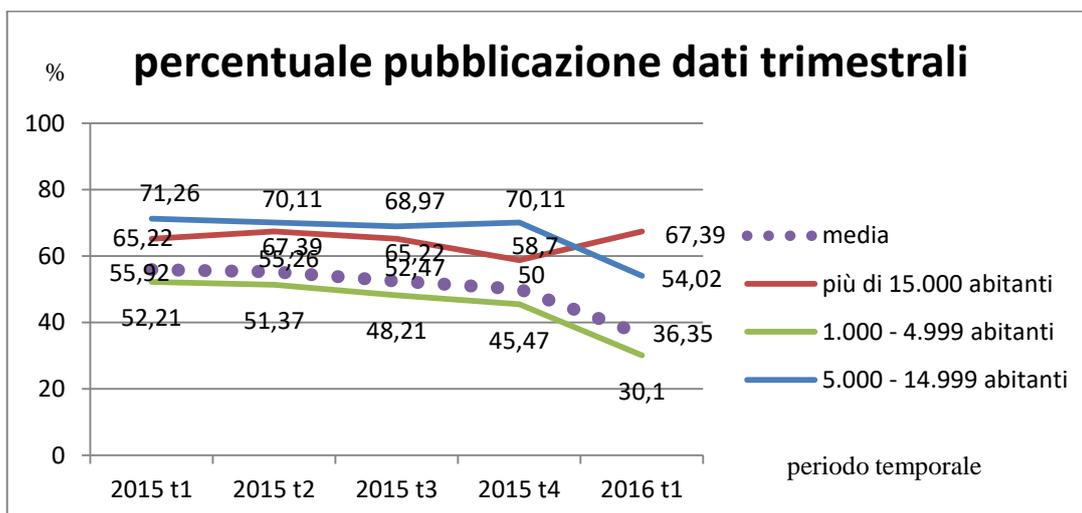
Proviamo ora a segmentare la percentuale di pubblicazione dei dati trimestrali secondo le fasce abitative, per capire se i trend riflettono le stesse evidenze emerse nelle precedenti analisi.

Si è deciso di unire i dati dei comuni facenti parte delle fasce abitative 15.000 – 49.999, 50.000 – 200.000 e oltre 200.000, in quanto queste ultime due rappresentano un universo troppo ristretto (rispettivamente 5 ed 1 comuni) per poter essere statisticamente rilevanti.

	2015 T1	2015 T2	2015 T3	2015 T4	2016 T1
1.000 – 4.999 abitanti: numero di comuni che pubblicano dati	248	244	229	216	143
<i>percentuale</i>	52.21%	51.37%	48.21%	45.47%	30.1%
5.000 – 14.999 abitanti: numero di comuni che pubblicano dati	62	61	60	61	47
<i>percentuale</i>	71.26%	70.11%	68.97%	70.11%	54.02%
oltre 15.000: numero di comuni che pubblicano dati	30	31	30	27	31
<i>percentuale</i>	65.22	67.39	65.22	58.7	67.39

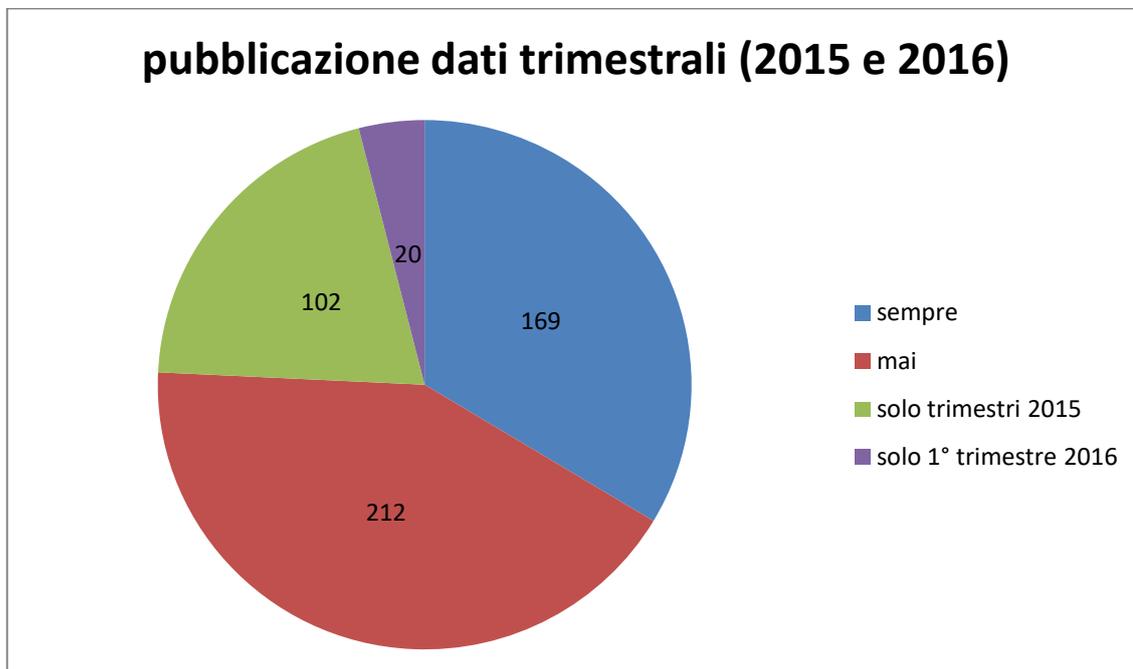


Incrociando i dati elaborati otteniamo il seguente grafico che mostra l'andamento della percentuale di pubblicazione dei dati trimestrali dell'ITP suddivisa secondo le fasce abitative dei comuni piemontesi, con l'immediato confronto con la media complessiva della percentuale di pubblicazione degli stessi dati.



Emerge immediatamente come la maggiore base rappresentata dai comuni piccoli influenzi in maniera decisa l'andamento complessivo del dato medio. Tuttavia lo stesso trend è riscontrabile nelle altre fasce abitative (quelle dei comuni con più di mille abitanti) fino al 2014. L'unica anomalia è rappresentata dalla percentuale di pubblicazione dei dati da parte dei comuni con più di 15.000 abitanti che, lo ricordiamo, in questo grafico unisce le fasce 15.000 -49.999, 50.000 – 200.000, e la città di Torino che è l'unica in Piemonte con più di 200.000 abitanti.

- pubblicato sempre: 169/608
- pubblicato mai: 212/608
- pubblicato solo i trimestri 2015: 102/608
- pubblicato solo il 1° trimestre 2016: 20/608



Quanto possiamo dedurre da questi dati conferma quanto affermato in precedenza, ovvero la recente tendenza a pubblicare meno dati. In particolare nel 1° semestre 2016 67 comuni hanno smesso di pubblicare il dato ITP, a fronte di solo 20 che hanno incominciato a farlo.

Secondo quanto disposto dal già citato Decreto del Presidente del Consiglio dei Ministri del 22 settembre 2014, le pubbliche amministrazioni hanno l'obbligo di pubblicare i dati trimestrali entro il trentesimo giorno dalla conclusione del trimestre cui si riferisce. Il dato relativo al 1° trimestre 2016, pertanto, sarebbe dovuto essere pubblicato entro il 30 aprile 2016. È opportuno ricordare che la raccolta dei dati, avvenuta manualmente, è incominciata in data 3 giugno 2016 e si è conclusa in data 17 giugno 2016, procedendo in ordine decrescente secondo il numero di abitanti.

### **Problema di standard. I dati sporchi**

Abbiamo già chiarito i criteri seguiti per la raccolta dei dati dai siti delle pubbliche amministrazioni. Sono infatti stati presi in considerazione solo quelli pubblicati secondo il DPCM 22 settembre 2014, lo standard legislativo attualmente utilizzato. Altra questione riguarda i dati sì pubblicati, ma privi delle necessarie informazioni. Ad esempio:

Anno 2014	10,06
-----------	-------

**Figura 9: dati pubblicati senza la specifica della modalità di calcolo utilizzata** (<http://www.comune.savigliano.cn.it/servizi/gestionedocumentale/visualizzadocumento.aspx?ID=10193>)

## INDICATORE DI TEMPESTIVITÀ DEI PAGAMENTI

Sezione relativa all'indicatore di tempestività dei pagamenti, come indicato all'art. 33 del d.lgs. 33/2013.

- 21.44 (ultimo trimestre 2014)
- 24.46 (primo trimestre 2015)
- 24,80 (secondo trimestre 2015)
- 7,45 (terzo trimestre 2015)
- 4.55 (ultimo trimestre 2015)
- 8.54 (primo trimestre 2016)

Figura 10: dati di cui si fatica a cogliere il segno + o -

(<http://www.comune.cortemilia.cn.it/Home/Menu/tabid/23636/Default.aspx?IDDettaglioPagina=28079>)

### COMUNE DI CONZANO Provincia di ALESSANDRIA

#### Tempestività dei pagamenti delle pubbliche amministrazioni Art. 9 - D.P.C.M. 22/09/2014

##### Indicatore di tempestività dei pagamenti

Nel periodo **30.04.2015 - 31.07.2015** l'indicatore globale è pari a **7,82 giorni**

*Tale indicatore è ottenuto elaborando tutte le fatture pagate nel periodo considerato.  
La data di scadenza delle fatture in cui tale data non è stata indicata è calcolata automaticamente il 30° giorno successivo la data di registrazione.*

*Secondo quanto previsto dal D.P.C.M. 22/09/2014, art.9, l'indicatore è ottenuto sommando il prodotto tra numero di giorni e importo dovuto di ogni fattura e rapportando tale somma all'importo dei pagamenti di fatture avvenuti nel periodo considerato.*

Figura 11: trimestri calcolati sui mesi sbagliati

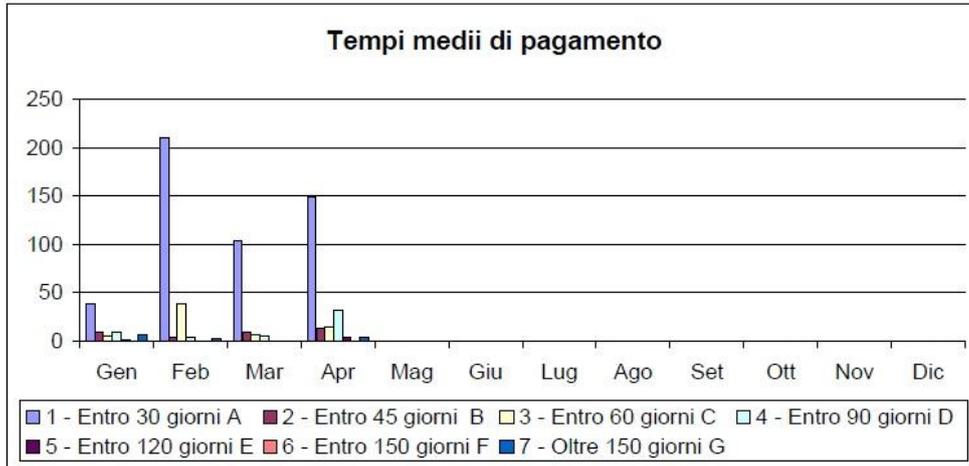
(<http://www.studiok.it/trasparenza/conzano/attachments/article/128/TEMPESTIVITA'%20PAGAMENTI%20FATTURE%20II%20TRIMESTRE%202015.pdf>)



Tempi medi pagamento fatture Esercizio 2016

Comune di Canelli

	Gen	Feb	Mar	Apr	Mag	Giu	Lug	Ago	Set	Ott	Nov	Dic	TOTALE	%
1 - Entro 30 giorni A	38	210	104	149	0	0	0	0	0	0	0	0	501	75,34
2 - Entro 45 giorni B	9	4	9	13	0	0	0	0	0	0	0	0	35	5,26
3 - Entro 60 giorni C	5	38	7	14	0	0	0	0	0	0	0	0	64	9,62
4 - Entro 90 giorni D	9	4	5	31	0	0	0	0	0	0	0	0	49	7,37
5 - Entro 120 giorni E	1	0	0	4	0	0	0	0	0	0	0	0	5	0,75
6 - Entro 150 giorni F	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7 - Oltre 150 giorni G	6	2	0	3	0	0	0	0	0	0	0	0	11	1,65
Tempo Medio	55,09	22,95	22,24	32,91	0	0	0	0	0	0	0	0	28,30977444	



**Figura 12: dati espressi come tempi medi di pagamento anziché come giorni di ritardo**  
 ([http://www.comune.canelli.at.it/servizi/funzioni/download.aspx?ID=4393&IDc=784&nomeFile=TempiMedii\\_784\\_4393.pdf](http://www.comune.canelli.at.it/servizi/funzioni/download.aspx?ID=4393&IDc=784&nomeFile=TempiMedii_784_4393.pdf))



Da un lato quantitativo, ad emergere sono le problematiche economiche: i comuni pagano in ritardo i fornitori, con conseguenze facilmente immaginabili sulla qualità dei servizi offerti, che con ogni probabilità andrà peggiorando, sugli investimenti in innovazione e sui pagamenti successivi degli stessi fornitori, in probabile diminuzione, sul gettito fiscale degli stessi comuni che, non potendo contare sulle tasse pagate dai fornitori finiscono per autoalimentare un circolo regressivo. Naturalmente va osservato come la lentezza nei pagamenti possa non dipendere dalla sola cattiva gestione delle attività economiche del singolo comune, quanto piuttosto anche da una generale lentezza nella redistribuzione delle risorse dal centro.

Da un lato qualitativo, invece, l'analisi mostra la diffusa inosservanza delle leggi. La pubblicazione dei dati IPT è, almeno stando al 1° trimestre 2016, una circostanza rara (36.35% dei comuni). La sezione Amministrazione Trasparente dei siti internet delle pubbliche amministrazioni, sebbene siano passati tre anni dal Decreto 33/2013 che ne ha sancito l'organizzazione, non rappresenta uno standard sufficiente a garantire una reale trasparenza. Le modalità di pubblicazione non sono rispettate in diversi casi, con vizi di natura diversa.

La difficoltà nel reperire i dati utili all'inchiesta sull'ITP suggerisce la necessità di adottare un sistema unico di pubblicazione dei dati. La soluzione potrebbe essere l'adozione di un portale unico a livello regionale, se non nazionale, dove i responsabili della trasparenza dei comuni possano inserire il valore ITP relativo ai differenti periodi.

## Appendice al *case study*: i testi

Si riportano qui le tabelle utilizzate nell'elaborazione e nell'analisi dei dati su T-Lab.

LEMMA	OCCORRENZE
PUBBLICHE_AMMINISTRAZIONI	130
EURO	58
PUBBLICARE	55
DATI	50
LEGGE	48
ARTICOLO	45
INFORMAZIONE	43
GIORNI	42
COMUNE	40
ENTE	40
DECRETO_TRASPARENZA	38
MINISTRO	33
NORMA	33
FOIA	31
GOVERNO	31
CITTADINI	30
SOCIETÀ	30
SANZIONI	30
PAGAMENTO	29
PREVISTO	28
ATTO	26
VEDERE	24
POLITICO	23
CHIEDERE	23
INCARICHI	23
REDDITO	22
TEMPO	22
ANNO	22
RISPETTO	21
RENDERE	21
PUBBLICI	21
PRESIDENTE	21
RESPONSABILE	20
PREVEDERE	20

**Tabella 1: Vocabolario dei lemmi, e relative occorrenze, del corpus "Decreto Trasparenza"**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
Expo	103,792	19	19
Raggi	91,262	18	19
candidato	81,877	15	15
Pd	58,602	12	13
Sala	47,778	10	11
consigliere	42,18	10	12
accusa	38,149	7	7
attaccare	38,149	7	7
polemico	38,149	7	7
volta	38,149	7	7
credito	32,693	6	6
depositare	32,693	6	6
Giuseppe	32,693	6	6
Roberto	32,693	6	6
stella	32,693	6	6
vicenda	32,693	6	6
Giachetti	31,657	7	8
sindaco	27,511	8	11
Asl_di_Civitavecchia	27,238	5	5
autocertificazione	27,238	5	5
campagna_elettorale	27,238	5	5
Campidoglio	27,238	5	5
confronto	27,238	5	5
io	27,238	5	5
M5S	27,238	5	5
Virginia_Raggi	27,238	5	5
incarico	25,457	9	14
campagna	21,786	4	4
conferire	21,786	4	4
dem	21,786	4	4
esposto	21,786	4	4
procura	21,786	4	4
quartiere	21,786	4	4
mandato	21,065	5	6
uomo	21,065	5	6
partire	21,046	8	13
assessore	17,985	6	9
eletto	17,985	6	9
ricevere	17,985	6	9
bilancio	16,812	7	12
Asl	16,708	5	7

**Tabella 2: Corpus “Decreto Trasparenza”. Cluster 1, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
ente	40,791	31	40
ente_locale	38,477	19	20
organismo	37,991	17	17
pagamento	36,559	24	29
giorni	32,392	30	42
partecipare	27,394	17	20
vedere	26,177	19	24
Dlgs	26,143	15	17
bilanci	24,553	11	11
responsabile	22,557	16	20
soggetto	21,812	13	15
sanzioni	21,48	21	30
società	21,48	21	30
obblighi_di_publicazione	20,081	9	9
scattare	18,511	10	11
società_partecipate	18,511	10	11
riduzione	17,847	8	8
Sole	17,847	8	8
fattura	16,901	12	15
relativo	15,624	23	38
capoluogo	15,613	7	7
ritardi	15,613	7	7
violazione	15,613	7	7
previsto	14,6	18	28
indicatore	14,136	8	9
controllato	13,38	6	6
fondazione	13,38	6	6
pubblico_interesse	13,38	6	6
tagli	13,38	6	6
adempimento	13,32	9	11
scadenza	13,32	9	11
comune	13,254	23	40
indice	12,861	10	13
articolo	12,817	25	45
obblighi	12,616	11	15
attività	12,498	13	19
comprendere	11,967	7	8
DI	11,967	7	8
adozione	11,147	5	5
commi	11,147	5	5
consuntivo	11,147	5	5

**Tabella 3: Corpus “Decreto Trasparenza”. Cluster 2, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
Foia	68,458	31	31
pubbliche_amministrazioni	48,4	77	130
accesso	41,859	19	19
freedom_of_information_act	39,648	18	18
testo	35,228	16	16
informazione	33,56	31	43
Italia	33,495	17	18
cittadini	33,284	24	30
legge	28,186	32	48
cittadino	26,4	12	12
paese	21,992	10	10
approvare	20,384	11	12
costi	17,586	8	8
rendere	15,791	15	21
diritto_di_accesso	15,385	7	7
Foia4Italy	15,385	7	7
forma	15,385	7	7
sicurezza	15,385	7	7
eccezione	14,589	11	14
riforma	13,905	8	9
accesso_civico	13,185	6	6
libertà	13,185	6	6
motivazione	13,185	6	6
vero	13,185	6	6
conoscere	11,769	7	8
garantire	11,769	7	8
mondo	11,769	7	8
interessato	10,985	5	5
linea	10,985	5	5
punti	10,985	5	5
regola	10,985	5	5
rifiuto	10,985	5	5
segreto	10,985	5	5
tutela	10,985	5	5
consiglio_dei_ministri	10,682	9	12
documenti	10,682	9	12
ufficio	10,682	9	12
richiesta	10,507	12	18
possesso	9,652	6	7
questione	9,652	6	7
ricorso	9,652	6	7

**Tabella 4: Corpus “Decreto Trasparenza”. Cluster 3, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
euro	124,643	48	58
ministro	68,392	27	33
moglie	52,749	15	15
libera_professione	49,222	14	14
stipendio	42,174	12	12
politico	42,066	18	23
reddito	38,824	17	22
medico	35,131	10	10
parlamentare	29,104	13	17
Catricalà	28,094	8	8
guadagnare	28,094	8	8
lordo	28,094	8	8
parente	26,681	9	10
patrimoniale	24,577	7	7
universitario	24,577	7	7
leggere	23,228	8	9
coniuge	21,062	6	6
patrono	21,062	6	6
titolare	21,062	6	6
anno_scorso	19,792	7	8
consenso	19,792	7	8
figli	17,548	5	5
intramoenia	17,548	5	5
ospedale	17,548	5	5
paperoni	17,548	5	5
auto	16,381	6	7
grazie	16,381	6	7
proprietà	16,381	6	7
pubblicare	14,809	24	55
compensi	14,406	9	14
Angelino_Alfano	14,036	4	4
chirurgia	14,036	4	4
genitore	14,036	4	4
Giannini	14,036	4	4
Monti	14,036	4	4
on_line	14,036	4	4
Orsola	14,036	4	4
ospedaliero	14,036	4	4
Rizzoli	14,036	4	4
sorella	14,036	4	4
grado	13,263	7	10

**Tabella 5: Corpus “Decreto Trasparenza”. Cluster 4, Caratteristiche dei lemmi**

LEMMA	OCCORRENZA
PUBBLICA_AMMINISTRAZIONE	571
TRASPARENZA	249
ACCESSO	233
DATI	204
LEGGE	201
INFORMAZIONE	197
CITTADINI	191
PUBBLICO	190
GOVERNO	176
NUOVO	175
DECRETO	167
DIGITALE	157
DIRITTO	156
RIFORMA	139
ITALIA	138
DOCUMENTI	133
ATTO	129
PUBBLICI	129
TESTO	125
PAESE	123
ANNI	103
CHIEDERE	102
POLITICO	101
RENDERE	101
RICHIESTA	99
GIORNI	90
ITALIANO	90
NOSTRO	86
ARRIVARE	83
MADIA	83
PUBBLICARE	81
UNICO	80
MINISTRO	79
CITTADINO	78
TEMPO	77
PREVEDERE	77
PUBBLICAZIONE	76
NORMA	75
CORRUZIONE	73
LAVORO	69
PUBBLICA	68
SOCIETÀ	67
PROPRIO	67
ULTIMO	65
METTERE	65

**Tabella 6: Vocabolario dei lemmi, e relative occorrenze, del corpus "Freedom of Information Act"**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
grillo	192,898	40	40
miliardo	186,252	40	41
fiscale	160,683	36	38
politico	126,884	60	101
investimento	120,231	31	36
reddito	92,408	24	28
corruzione	78,07	41	73
lotta	70,678	16	17
declino	67,44	14	14
locale	61,954	24	36
imprese	60,842	32	57
movimento	57,068	19	26
Bersani	52,982	11	11
imposta	52,982	11	11
riduzione	50,516	19	28
incarichi	50,04	21	33
Imu	48,163	10	10
debito	46,752	11	12
fermare	46,42	14	18
evasione	46,177	12	14
euro	45,09	21	35
fondi	44,332	16	23
Ingroia	43,345	9	9
regionale	42,602	14	19
patrimoniale	41,968	13	17
banda	41,049	18	29
intervento	41,049	18	29
fiscal	38,527	8	8
stella	38,492	16	25
settore	38,296	13	18
largo	37,275	17	28
Asl	37,232	9	10
ridurre	35,033	13	19
abitazione	33,71	7	7
Beppe	33,71	7	7
compact	33,71	7	7
Monti	33,344	14	22
proporre	32,856	17	30
partire	32,646	27	60
aumentare	32,579	10	13
tagli	32,579	10	13
programma	31,408	16	28

**Tabella 7: Corpus “Freedom of Information Act”. Cluster 1, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
Assange	270,201	44	44
svedese	220,999	36	36
Ny	141,116	23	23
Carlo	126,18	23	25
e-mail	110,099	26	34
lettera	106,781	27	37
dipartimento	101,872	19	21
Julian	98,139	16	16
scrivere	96,625	33	55
procuratore	85,864	14	14
Prosecution	85,864	14	14
Marianne	79,728	13	13
interrogare	77,691	15	17
caso	72,795	30	56
Clinton	71,676	14	16
Close	67,456	11	11
Ecuador	67,456	11	11
interrogatorio	67,456	11	11
trono	67,456	11	11
principe	61,321	10	10
accusa	60	13	16
Svezia	60	13	16
americano	59,917	26	50
segreto	57,766	24	45
ambasciata	57,433	15	21
Fbi	57,433	15	21
inglese	57,433	15	21
sicurezza	55,726	25	49
erede	55,187	9	9
estradizione	55,187	9	9
Londra	55,01	13	17
rilasciare	53,945	10	11
WikiLeaks	53,945	10	11
Stoccolma	49,053	8	8
Nsa	46,656	13	19
ufficiale	45,134	12	17
indagine	44,992	18	33
magistrato	43,865	15	25
compromettere	42,919	7	7
Crown	42,919	7	7
Swedish	42,919	7	7

**Tabella 8: Corpus “Freedom of Information Act”. Cluster 2, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
riforma	294,389	111	139
Madia	203,272	70	83
ministro	152,468	61	79
Renzi	134,571	50	62
testo	131,33	78	125
decreti	124,306	44	53
Consiglio	120,31	46	58
Marianna	110,145	29	29
attuativo	110,063	39	47
decreto	95,608	86	167
camera	93,593	42	58
approvare	88,542	43	62
delega	88,406	34	43
definitivo	85,619	29	34
consiglio_dei_ministri	79,586	41	61
conferenza	64,619	21	24
Matteo	53,545	18	21
attendere	43,931	21	30
pareri	42,582	15	18
governo	40,752	71	176
premier	39,15	24	39
Palazzo_Chigi	37,501	19	28
senato	35,536	15	20
esame	35,365	13	16
Scia	30,358	8	8
commissioni	28,424	12	16
giorni	27,633	39	90
addetti_ai_lavori	26,562	7	7
poste	26,562	7	7
primo	26,467	28	58
conferenza_stamp	25,237	8	9
furbetti	25,237	8	9
varo	25,237	8	9
codice	25,229	15	24
insediamento	25,031	11	15
pubblica_amministrazione	24,918	167	571
iniziale	24,768	9	11
commercio	24,761	10	13
venerdì	24,761	10	13
indicazione	24,404	14	22
arrivare	24,008	37	88
legge	23,943	70	201

**Tabella 9: Corpus “Freedom of Information Act”. Cluster 3, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
accesso	437,256	196	233
diritto	309,62	134	156
atto	168,346	96	129
cittadino	117,75	61	78
richiesta	74,749	62	99
informazione	72,715	101	197
eccezione	71,965	39	51
pubblica_amministrazione	70,605	229	571
cittadini	59,869	94	191
possesso	58,127	21	22
documenti	57,174	71	133
richiedere	53,912	35	50
Tar	52,813	22	25
motivazione	49,176	18	19
interessi	46,778	33	49
accedere	45,84	22	27
rifiuto	43,218	16	17
sanzioni	40,354	26	37
ricorso	39,921	22	29
negare	36,17	15	17
civico	35,301	32	53
dati	33,925	87	204
pubblicazione	33,902	41	76
Foia	32,957	78	179
diniego	32,923	11	11
riproduzione	32,923	11	11
obbligo	32,822	33	57
prevedere	32,402	53	109
anticorruzione	30,478	22	33
casi	29,924	25	40
risposta	28,962	23	36
garantire	27,783	27	46
rifiutare	27,454	12	14
trasparenza	27,424	98	249
motivare	26,67	14	18
limiti	26,655	16	22
previsto	25,116	25	43
specifico	24,135	12	15
conoscere	23,011	28	52
archivi	22,475	9	10
motivato	22,475	9	10
silenzio-diniego	22,475	9	10

**Tabella 10: Corpus “Freedom of Information Act”. Cluster 4, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
dati	197,63	84	204
sito	131,647	35	63
stipendi	89,855	13	15
istituzionale	86,24	19	30
strumento	53,847	22	52
personali	53,563	10	14
open	48,773	18	40
democrazia	48,733	10	15
economico	41,596	18	44
medio	40,873	6	7
livello	37,813	11	21
esigenza	37,799	9	15
protezione	36,472	7	10
ruolo	34,989	13	29
servizi	34,159	17	45
utilizzare	32,574	15	38
qualità	31,961	7	11
rapporto	29,991	16	44
azione	25,601	8	16
fornire	25,17	12	31
media	20,779	9	22
compiti	19,486	4	6
pubblicare	19,379	21	81
paese	18,456	28	123
curriculum	18,091	7	16
porre	17,829	9	24
attività	17,24	17	63
andamento	17,152	3	4
elettore	17,152	3	4
riuso	17,152	3	4
specifica	17,152	3	4
supporto	17,152	3	4
sanitario	16,956	10	29
formato	16,919	6	13
posto	16,919	6	13
bando	15,997	5	10
twitter	15,997	5	10
on-line	15,78	16	60
analizzare	15,646	4	7
cifra	15,646	4	7
dimensione	15,646	4	7
efficacia	15,646	4	7

**Tabella 11: Corpus “Freedom of Information Act”. Cluster 5, Caratteristiche dei lemmi**

internet	96,304	31	57
risorse	70,788	19	31
mettere	57,459	33	85
startup	56,66	16	27
utente	51,608	14	23
funzioni	45,461	14	25
scuola	45,461	14	25
Swartz	44,368	11	17
diffuso	40,194	12	21
diffondere	38,533	15	31
piattaforma	37,586	11	19
perseguimento	36,595	7	9
software	35,953	9	14
identità	35,004	10	17
utilizzo	32,064	10	18
mondo	31,76	18	46
noi	30,823	16	39
Piacentini	29,835	6	8
organizzazione	29,818	11	22
creativo	29,121	4	4
navigare	29,121	4	4
istituzione	27,703	11	23
controllo	27,211	19	54
serve	25,776	11	24
civile	25,619	16	43
on-line	25,589	20	60
favorire	25	10	21
web	24,383	15	40
progetto	22,051	12	30
Aaron	21,725	4	5
repubblicano	21,725	4	5
restituire	21,725	4	5
Samaritani	21,725	4	5
digitalizzazione	21,634	6	10
forma	19,54	11	28
Tim	19,158	5	8
carabiniere	18,969	7	14
mercato	18,75	9	21
lanciato	18,689	6	11
banda	18,274	11	29
repubblica	18,274	11	29
raccogliere	16,909	7	15
figlio	16,84	4	6

**Tabella 12: Corpus “Freedom of Information Act”. Cluster 6, Caratteristiche dei lemmi**

LEMMA	OCCORRENZA
CITTADINI	86
DATA	80
OPEN	72
DIGITALE	63
ITALIA	61
NUOVO	43
GOVERNO	42
INNOVAZIONE	40
PUBBLICA_AMMINISTRAZIONE	38
PAESE	37
INFORMAZIONE	36
COMUNE	35
RETE	35
POLITICO	32
REGIONE	32
SITO	32
PUBBLICO	31
PA	30
DIVERSO	30
INTERNET	30
OGGI	30
PROGRAMMA	30
ON-LINE	29
PIATTAFORMA	29
EUROPEO	28
SERVIZI	28
TRASPARENZA	28
ITALIANO	27
STARTUP	27
MINISTRO	27
RAPPORTO	27
STRUMENTO	26
SANITARIO	26
SISTEMA	26
WEB	26
RENDERE	26
RICERCA	25
ACCESSO	25
CHIEDERE	24
PROGETTO	24
PRIMO	24
AMMINISTRAZIONE	24

**Tabella 13: Vocabolario dei lemmi, e relative occorrenze, del corpus "Open Government"**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
aula	84,26	13	13
sinistra	51,815	8	8
annunciare	50,717	10	12
movimento	38,321	8	10
destra	32,37	5	5
Giulio_Cesare	32,37	5	5
maggioranza	32,37	5	5
governo	31,646	18	42
valutare	31,611	6	7
giunta	31,341	9	14
restare	29,427	8	12
certo	27,679	7	10
romano	27,679	7	10
cittadini	27,597	28	86
contratti	25,337	5	6
E-democracy	25,337	5	6
incarichi	25,337	5	6
Pd	22,992	9	17
consigliere	22,061	6	9
Campidoglio	20,686	8	15
scranni	19,147	4	5
politica	18,972	9	19
stella	18,764	6	10
questione	16,66	5	8
M5S	16,217	7	14
città	15,011	8	18
Anac	13,102	3	4
porre	12,749	7	16
parlamento	12,059	6	13
posto	12,059	6	13
squadra	11,577	5	10
giornalista	11,569	4	7
Parole	11,569	4	7
riguardare	10,506	6	14
ultimo	10,121	7	18
prima	10,047	8	22
mancare	9,77	5	11
noi	9,77	5	11
affiancare	9,375	3	5
governo_Monti	9,375	3	5
leggere	9,375	3	5
mano	9,375	3	5

**Tabella 14: Corpus “Open Government”. Cluster 1, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
paziente	39,587	10	11
sezione	35,091	9	10
dataset	34,783	11	14
Act	30,326	9	11
decreto	26,016	8	10
Freedom_of_Information	21,762	7	9
riuso	21,716	6	7
formato_aperto	18,192	4	4
rilasciare	18,192	4	4
rivelare	18,192	4	4
santé	18,192	4	4
data	18,171	29	80
trasparente	17,585	6	8
centri_di_ricerca	17,324	5	6
agenzia	16,678	8	13
associazione	15,502	7	11
stati_uniti	14,42	6	9
utile	14,42	6	9
open	13,71	25	72
applicare	13,516	5	7
dati_sanitari	13,208	7	12
mancaza	12,999	4	5
oggetto	12,999	4	5
sviluppato	12,999	4	5
voce	12,999	4	5
processo	12,477	9	18
salute	11,087	8	16
Italia_digitale	10,714	5	8
raccolta	9,937	6	11
prevedere	9,699	7	14
Sviluppo	9,699	7	14
americano	9,61	4	6
atto	9,61	4	6
attuazione	9,61	4	6
azione	9,61	4	6
crowdfunding	9,61	4	6
stimolare	9,61	4	6
amministrazione	9,101	10	24
Agenda_digitale_italiana	8,787	3	4
istituire	8,787	3	4
ricerca_scientifica	8,787	3	4
Stati_Generali	8,787	3	4

**Tabella 15: Corpus “Open Government”. Cluster 6, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
informazione	135,152	30	36
tecnologia	50,479	11	13
diritto	35,778	10	14
Cnr	28,877	5	5
Ministero_della_Salute	28,877	5	5
banche_dati	28,019	6	7
pubblico	27,959	15	31
rapporto	23,993	13	27
detenere	23,098	4	4
Inail	23,098	4	4
organizzazione	23,098	4	4
diritti	23,083	6	8
Istat	22,428	5	6
prodotto	22,428	5	6
Fabrizio	18,125	7	12
verificare	17,871	5	7
comunicazione	16,915	4	5
edizione	16,915	4	5
festival	16,915	4	5
Inps	16,915	4	5
rilevare	16,915	4	5
riutilizzare	16,915	4	5
statistica	16,915	4	5
attività	16,277	6	10
verso	16,277	6	10
mondo	15,814	8	16
corruzione	12,85	4	6
difficile	12,85	4	6
stabilire	12,85	4	6
paese	12,258	13	37
capire	12,151	7	15
commerciale	11,535	3	4
digital_divide	11,535	3	4
of	11,535	3	4
spaziale	11,535	3	4
ente	11,311	8	19
banda_larga	10,701	7	16
pagamento	10,192	6	13
accedere	9,996	4	7
operato	9,996	4	7
quantità	9,996	4	7
registrare	9,996	4	7

**Tabella 16: Corpus “Open Government”. Cluster 7, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
sindaco	73,41	13	16
ama	59,175	8	8
rifiuti	58,143	9	10
ambiente	51,263	9	11
assessore	51,263	9	11
Raggi	49,325	11	16
Paola_Muraro	44,369	6	6
Roma	39,072	10	16
Fortini	36,969	5	5
Luca_Bergamo	36,969	5	5
Trasporti	36,339	6	7
domanda	33,094	9	15
Andrea_Lo_Cicero	29,571	4	4
comitato	29,166	5	6
vertice	29,166	5	6
Atac	22,086	4	5
incontrare	19,514	5	8
nomina	19,514	5	8
nostro	19,064	6	11
Marino	17,141	4	6
partire	16,531	8	19
bilancio	16,341	5	9
partecipare	16,271	9	23
capo_di_gabinetto	15,167	3	4
emergenza	15,167	3	4
lancio	15,167	3	4
milioni_di_euro	15,167	3	4
riunione	15,167	3	4
stimare	15,167	3	4
parlare	15,061	8	20
aspettare	13,647	4	7
notizia	13,647	4	7
raggiungere	13,647	4	7
regionale	13,647	4	7
ultimo	12,5	7	18
Campidoglio	11,287	6	15
buona	11,015	3	5
hacker	11,015	3	5
palazzo_Senatorio	11,015	3	5
prima_riunione	11,015	3	5
quartiere	11,015	3	5
questionario	11,015	3	5

**Tabella 17: Corpus “Open Government”. Cluster 8, Caratteristiche dei lemmi**

LEMMAS & VARIABLES	CHI <sup>2</sup>	E.C. IN CLUSTER	E.C. IN TOTAL
lacono	72,792	10	10
programma	65,043	18	30
digitale	50,906	26	63
incentivi	35,723	6	7
citare	28,667	5	6
digitalizzazione	28,667	5	6
rete	25,768	14	35
mettere	25,005	10	21
banda_larga	21,675	8	16
alfabetizzazione	16,823	4	6
istruzione	16,823	4	6
punta	16,823	4	6
infrastruttura	16,012	5	9
startup	15,863	10	27
temi	15,117	7	16
agenda	14,898	3	4
conti	14,898	3	4
denaro	14,898	3	4
e-commerce	14,898	3	4
gruppi	14,898	3	4
Rangone	14,898	3	4
risparmiare	14,898	3	4
on-line	13,731	10	29
investimento	13,558	7	17
uso	13,535	5	10
capitolo	13,376	4	7
rivoluzione	13,376	4	7
internet	12,785	10	30
imprese	12,188	7	18
bisognare	10,976	7	19
formazione	10,825	4	8
approccio	10,802	3	5
competenze	10,802	3	5
diffusione	10,802	3	5
Facebook	10,802	3	5
largo	10,802	3	5
miliardi_di_euro	10,802	3	5
presente	10,802	3	5
supporto	10,802	3	5
ufficiale	10,802	3	5
ricerca	9,352	8	25
Pa	9,087	9	30

**Tabella 18: Corpus “Open Government”. Cluster 9, Caratteristiche dei lemmi**

# Conclusioni

Nel presente capitolo conclusivo mi appresto a condensare le principali evidenze emerse nel corso dell'analisi, partendo dal caso di studio e leggendone gli esiti alla luce dei primi due capitoli. La natura composita della mia tesi e i diversi territori esplorati consentono di offrire un quadro piuttosto nitido, ma articolato su tematiche diverse, il che rende necessaria un'organizzazione del discorso in diversi paragrafi. È tuttavia opportuno approcciarsi alle presenti conclusioni con la consapevolezza dell'inestricabile legame che unisce le diverse parti del discorso.

Seguendo la successione temporale che ha caratterizzato la preparazione della tesi di laurea, l'esposizione comincerà dal progetto di *data journalism* relativo all'indicatore di tempestività dei pagamenti, conscio che una prima lettura dei dati è già stata offerta in seno al report (pagina 75). Per tale ragione, più che una lettura dei dati, in questa sede mi preme sottolineare alcuni evidenze relative al metodo, organizzandone la spiegazione in tre momenti distinti, i tre momenti in cui tipicamente si articola un progetto di *data journalism*.

Uno dei primi nodi critici affrontati nel corso dell'inchiesta sui dati ITP delle pubbliche amministrazioni piemontesi è stato infatti riscontrato durante la raccolta dei dati. Sebbene le disposizioni in materia di pubblicità dei dati delle pubbliche amministrazioni siano chiare e in atto da oramai tre anni, la percentuale di pubblicazione dell'ITP si attesta, almeno per quello che riguarda il 2015 e il primo trimestre 2016, su soglie lontane persino dal 50%. Al 17 giugno 2016, può essere utile ricordarlo, solo 221 comuni su 608 avevano pubblicato un dato che sarebbe dovuto essere disponibile dal 30 aprile, come termine ultimo. Anche in riferimento al 2015 è emersa una notevole disattenzione nei confronti della legge, dal momento che, sempre al 17 giugno 2016, erano 258 su 608 le amministrazioni ad aver pubblicato un dato che sarebbe dovuto essere stato reso pubblico entro il 31 gennaio. Sempre in riferimento alla raccolta dei dati è significativo il fatto che si sia dovuto

procedere in maniera totalmente manuale, dal momento che i portali web delle singole amministrazioni presentano un grado di omogeneità e di similarità basso, che impedisce un processo di *scraping* automatizzato.

Se la raccolta ha presentato alcune difficoltà tecniche, la fase di interpretazione dei dati si è rivelata complicata, e altresì stimolante, proprio per l'emergere di due distinti piani di analisi. Il primo quantitativo, che almeno nelle intenzioni sarebbe voluto essere quello principale, se non l'unico. Da questo punto di vista, infatti, la preoccupazione principale è stata quella di far esprimere i dati in maniera naturale, approcciando il dataset con operazioni statistiche elementari e con la visualizzazione dei dati stessi, ed è la terza fase, tramite grafici. Decisamente più impegnativa è stata la lettura dei dati nel momento in cui mi sono reso conto che, anche da un punto di vista qualitativo, sarebbe stato interessante costruire un dataset. In quest'ottica si è così concentrata l'interpretazione dei dati relativi alle percentuali di pubblicazione, e in particolare alle serie storiche. Il valore dell'inchiesta, da un punto di vista giornalistico, si è così spostato dall'analisi del valore assunto dall'ITP all'analisi della mancata osservanza del Decreto Trasparenza.

Con riguardo al lavoro di *data journalism*, pertanto, la criticità maggiore è stata l'iniziale difficoltà nel trovare una chiave di interpretazione ai dati. Scoprire che proprio nell'assenza dei dati si celava la più succosa evidenza di un lavoro di giornalismo *data-driven* è una lezione che non dimenticherò facilmente e che mi aiuterà nell'approcciarmi ai futuri progetti in maniera meno ingenua. Un dataset, d'altronde, è in grado di fornire risposte soltanto se è stato predisposto con adeguati record e variabili [Bounegru *et al.* 2012]. In termini di ricerca sociale, si tratta di definire e cogliere l'autenticità pragmatica [Topolski 1973] del foglio di calcolo e dei dati lì immagazzinati.

Dall'analisi del contenuto è opportuno innanzitutto richiamare le principali direttrici tematiche svelate dall'analisi dei contesti elementari.

Il primo corpus, quello costruito sulla *keyword* "Decreto Trasparenza", si articola su una doppia direttrice tematica. La prima mette in evidenza un distanziamento tra come quotidiani di informazione pura, La Repubblica e La Stampa, e informazione

specialistica, il Sole 24 Ore, approcciano il tema della trasparenza. Dall'altra, il corpus sembra suddiviso tra una serie di articoli di stampo teorico-prescrittivo, intendendo con questo termine una predisposizione a trattare la *issue* sotto la lente del cosa preveda la legge, e una serie di frammenti che rappresentano, invece, il tentativo di concretizzare in inchieste il lavoro su quanto, le pubbliche amministrazioni, rendono pubblico. È però chiarire, fin da ora, la natura di tali inchieste, approfondendo il discorso iniziato nel capitolo 3, all'interno del paragrafo sul corpus "Decreto Trasparenza". Già allora si era sottolineato come, tale approccio, fosse appannaggio quasi esclusivo de l'Espresso, giornale settimanale di approfondimento. La natura degli articoli, tuttavia, rivela come, in tali articoli, manchi nettamente il ricorso ad un'analisi dei dati. In altre parole, il tipo di informazione proposta dal settimanale si colloca nel solco della tradizione della carta stampata, dove il testo la fa da padrone e l'argomentazione trova ampio spazio. Un tipo di informazione di assoluta qualità, ma che fa risaltare l'assenza di un'indagine giornalistica fondata sui dati. Va tuttavia rimarcato come, la generale catalizzazione dello spazio verso un discorso sul metodo più che sull'indagine numerica, non possa essere letta semplicemente come scelta editoriale, quanto piuttosto anche come conseguenza di un mezzo, la carta stampata, che concede limitati slanci al *data journalism*. Ulteriore nota degna di evidenza è la consapevolezza che il Freedom of Information Act, la legge sull'accesso ai dati e ai documenti prodotti dalle pubbliche amministrazioni, ha in Italia una storia assolutamente recente, essendo stato approvato lo scorso maggio, e che le amministrazioni stanno vivendo, proprio in questi mesi, il periodo di adozione che scadrà a dicembre.

Il discorso attorno al processo, attualmente in corso, di adeguamento legislativo introduce altri due evidenze emerse nell'analisi sul contenuto dei corpora. La prima, relativa al consueto ricorso, da parte degli organi di informazione, all'antitesi tra Italia ed estero, un tratto particolarmente marcato nel corpus "Freedom of Information Act". Il confronto tra la situazione italiana e quella estera è spesso tratteggiato in termini di ritardo: negli Stati Uniti la legge sulla libertà di informazione risale al 1966 ed è la più nota, sebbene ve ne siano di più antiche. La Svezia, ad esempio, ha regolamentato l'accesso ai documenti prodotti dagli enti

pubblici nel 1776 e da allora, prima dell'Italia, in tutto il mondo sono stati più di novanta i paesi a dotarsene. Ancora più significativa mi sembra la seconda evidenza di cui sopra, riguardante la dislocazione del discorso sui contrapposti piani di *oggi e domani*, presente e futuro. Si tratta della principale conclusione a cui si giunge dall'analisi del corpus "Open Government", da cui si ricava anche una sorta di sovrapposizione semantica tra il concetto di futuro e quello di digitalizzazione. Dietro ad entrambe le antitesi sopra esposte sembra pertanto vivere l'idea che la digitalizzazione rappresenti la svolta verso l'*open government*.

Attenzione, però, a non cadere in un'ingenua e onirica attesa del futuro. La svolta digitale rappresenta evidentemente, soltanto un mezzo e non il fine. Inoltre, anche nel caso di un'eventuale piega digitale intesa in senso di pubblicazione degli atti, il processo sarebbe solo ad uno stadio iniziale. Come dice Morando, infatti, l'accesso, ovvero la semplice disponibilità di consultare le informazioni, è soltanto la preconditione per poter riutilizzare i dati [Morando *et al.* 2013]. È pertanto sul piano del riutilizzo che si gioca la partita del governo aperto e, di riflesso, del giornalismo investigativo con i dati. Inevitabilmente, il grado di apertura del governo dipende dai dati che mette a disposizione e dalla loro qualità, come ad esempio essere strutturati in maniera tale da risultare processabili ed elaborati in maniera automatizzata, ma anche da un comune intendimento volto ad accrescere il processo collaborativo tra pubblica amministrazione e società civile, unite nell'intento di trovare soluzioni a problemi collettivi.

Volendo trovare una lettura conclusiva al lavoro svolto, alla luce delle evidenze sopra esposte, mi sembra legittimo affermare che la questione relativa all'*open data government* veda uno scollamento tra le diverse componenti prese in analisi. In particolare, ritengo che la questione sull'*open government* come inteso nel presente lavoro di tesi possa inscrivere in un quadrato i cui lati sono rappresentati dai seguenti elementi: *open data*, *data journalism*, giornalismo tradizionale, aspetti legislativi. Le righe che seguono mirano a spiegare le ragioni di quanto ho poco fa definito scollamento.

«Le pubbliche amministrazioni (PA) hanno una caratteristica comune: accumulano dati. Li accumulano per poter funzionare e per dimostrare di aver ben funzionato» [Morando 2013: 7], e tale è il processo di produzione di dati. Come a più riprese evidenziato, il concetto di governo aperto sta evolvendo verso una dimensione collaborativa e di partecipazione nelle procedure di amministrazione: in un governo aperto i cittadini non si limitano ad avere accesso ai documenti, ma collaborano col governo stesso che diviene pertanto una piattaforma, il suolo fertile su cui far fiorire *best practices* collaborative [Lathrop D., Ruma, L. 2010] e creare presupposti di conoscenza. È tuttavia proprio sul piano di trasformazione del governo aperto da concetto a realtà che si incontrano le prime difficoltà: l'autonomia delle singole amministrazioni e le modalità di gestione dei dati hanno contribuito a creare «isole di informazioni, con scarsa visione sistemica» [AA.VV. 2014: 6] che, se da un lato rallentano la definizione di *open government* come descritto poco fa, dall'altro inibiscono, almeno parzialmente, anche lo sviluppo di un giornalismo con i dati cui viene a mancare la materia prima.

Accanto a questi due aspetti vanno presi in considerazione altre due realtà: il quadro legislativo e la situazione del giornalismo contemporaneo. Con riguardo a quest'ultimo, si è sviscerato nelle pagine precedenti il contenuto ed il framing editoriale adottato dalle diverse testate. Spesso gli organi di informazione, almeno in riferimento al periodo preso in considerazione nell'analisi del contenuto del terzo capitolo, hanno proposto ai propri pubblici una lettura orientata al racconto dei processi legislativi e alla cronaca politica più che all'investigazione dei processi in atto nelle pubbliche amministrazioni. Il racconto giornalistico, in altre parole, sembra inseguire la propria essenza, in una sorta di ritardo che si concretizza nel parlare di *open data* più che nel cercare storie attraverso tali dati aperti.

Per completezza metodologica, va tuttavia tenuto a mente il momento storico in cui tale lavoro di laurea si iscrive, a cavallo dell'adozione del Freedom of Information Act italiano. Lungi dal voler rappresentare un'attenuante all'inevitabile ritardo nella definizione di standard legislativi all'altezza, i prossimi mesi rappresenteranno l'occasione, per le pubbliche amministrazioni, per scrollarsi di dosso la scomoda etichetta di sprechi e mala gestione di risorse e, per il governo, di proporsi come

propulsore di innovazione, in grado di trattare l'informazione come un bene della nazione.

# Ringraziamenti

Nei sei mesi di lavoro sulla tesi di laurea si sono alternati periodi di entusiasmo a momenti di difficoltà, dovuti in particolare al carattere innovativo del mio lavoro. Proprio in quei frangenti è emersa l'importanza di soggetti il cui supporto è risultato imprescindibile al raggiungimento dell'obiettivo che mi ero preposto. Per tale ragione sento di voler spendere alcune righe di ringraziamento a chi è risultato decisivo nella stesura della tesi di laurea, a cominciare dal mio relatore, dott. Giuseppe Tipaldo, in grado di dimostrarmi fiducia lasciandomi ampia libertà di scelta ma guidandomi nei decisivi snodi metodologici del lavoro.

Allo stesso modo un sincero ringraziamento va a Federico Morando e ad Alessio Melandri, fondatori di Synapta Srl, senza le cui disponibilità a seguirmi nello stage ogni sforzo sarebbe risultato vano.

Un grazie anche a Franco Carcillo e Guido Vernerò importanti, l'uno, nell'aiutarmi a trovare il giusto ambito in cui spendermi e, l'altro, nell'offrirmi delucidazioni in alcuni passaggi di ricerca sull'Indicatore di tempestività dei pagamenti.

Un grazie, infine, a Nicola, prezioso nell'introdurmi velocemente all'utilizzo di strumenti tecnici che avrebbero altrimenti rallentato il lavoro di ricerca.

# Bibliografia

AA.VV.

2015 *Gli open data in ambito parlamentare*, Nexa Center for Internet & Society.

Aliprandi S.,

2014 *Il fenomeno open data*, progetto copyleft-italia.it. Ledizioni.

Appelgren, E. e Nygren, G.

2014 *Data journalism in Sweden*, in «Digital Journalism», 2(3), pp. 394-405.

Berelson, B.

1952 *Content analysis in communication research*, Glencoe, Free Press.

Bounegru, L., Chambers, L. e Gray, J. (a cura di)

2012 *The Data Journalism Handbook*, O'Reilly Media.

Cardano, M.

2003 *Tecniche di ricerca qualitativa*, Roma, Carocci.

Chadwick A.

2013 *The hybrid media system*, Oxford, Oxford University Press.

Cole, Combs, M.E., R.R., Shaw, D.L., Stevenson, R.L.

1983 *Il giornalismo di precisione*, in «Problemi dell'informazione», anno VIII n.1, pp. 89-105.

Corbetta, P.

1999 *Metodologia e tecniche della ricerca sociale*, Bologna, Il Mulino.

Cranberg, L.

1989 *Plea for Recognition of Scientific Character of Journalism*, in «Journalism Educator», Winter, pp. 46-49.

Daniel A. e Flew T.

2010 *The Guardian reportage of the UK MP expenses scandal: a case study of computational journalism*, in «Communications Policy and Research Forum 2010», pp. 15-16, November 2010, Sydney.

de Sola Pool, I. (a cura di)

1959 *Trends in Content Analysis*, Urbana, University of Illinois Press.

Ducci G.

2016 «*Lavori in corso*» nella PA connessa. *Il ruolo delle strutture di comunicazione nella gestione dei social media e lo sportello polifunzionale 3.0*, «Problemi dell'informazione», anno XLI, n.1, pp. 113-136, Bologna, Il Mulino.

Eco, U.

1990 *I limiti dell'interpretazione*, Milano, Bompiani.

Fama, A.

2011 *Open Data - Data Journalism, Trasparenza e informazione al servizio delle società nell'era digitale*, Edizione elettronica a cura di Simplicissimus Book Farm.

Ferrer Conill, R. e Karlsson, M.

2015 *The gamification of journalism*, in Davis, D.Z. e Gangadharbatla, H., *Emerging research and trends in gamification*, pp. 356-383, IGI Global.

Goffman, E.

1974 *Frame analysis. An essay on the organization of experience*, Boston, Northeastern University Press.

Hall, S.

1980 *Encoding/decoding*, in Hall, S., Hobson, D., Lowe, A. e Willis, P. (a cura di), *Culture, Media, Language*, London, Hutchinson.

Hallin, D. e Mancini, P.

2004 *Modelli di giornalismo*, Roma, Bari, Laterza.

Henry, C.

2003 *Freedom of information act*, Nova Publishers, Novinka Books, New York.

Holsti, O.R.

1969 *Content analysis for the social sciences and humanities*, Reading, Massachusetts, Addison-Wesley.

Hovland, C., Lumsdaine, A. e Sheffiel, F.

1949 *Experiments on mass communication*, Princeton, N.J., Princeton University Press.

Johnson, E.W. e Wolfe, T.

1974 *The New Journalism: with an anthology edited by Tom Wolfe and E. W. Johnson*, Londra, Pan Books.

Karlsen, J. e Stavelin, E.

2014 *Computational journalism in Norwegian newsrooms*, in «Digital Journalism», 8 (1), pp. 34-48.

Katz, E., Blumler, J. E Gurevitch, M.

1974 *The use of mass communication*, Beverly Hills, California, Sage.

Kettl, W.

2008 *The Next Government of the United States: Why Our Institutions Fail Us and How to Fix Them*, Donald, W. Norton & Company.

Krippendorff, K.

2004 *Content analysis: an introduction to its methodology*, II ed., Thousand Oaks, California, Sage.

Kuhn, T.S.

1962 *The structure of scientific revolution*, Chicago, University of Chicago Press; trad. it. *La struttura delle rivoluzioni scientifiche* Torino, Einaudi, 1969.

Lancia, F.

2012 *Manuale utente T-Lab 8.1*.

Lasswell, H. D.

1927 *Propaganda Technique in the World War*, London, Kegan Paul.

Lathrop D. e Ruma, L., (a cura di),

2010 *Open government. Trasparenza, Collaborazione e Partecipazione in Pratica*, New York, O'Reilly,

Lazarsfeld, P. F.

1967 *Metodologia e ricerca sociologica*, Bologna, Il Mulino.

Lazarsfeld, P., Berelson, B. e Gaudet, H.

1944 *The people's choice. How the voter makes up his mind in a presidential campaign*, New York, Columbia University Press.

Livolsi, M. e Rositi, F.

1988 *La ricerca sull'industria culturale*, Roma, La Nuova Italia Scientifica.

Markoff, J., Shapiro, G. e Weitman, S. R.

1975 *Toward the integration of content analysis and general methodology*, in Heise, D. R. (a cura di), *Sociological Methodology*, San Francisco, California, Jossey-Bass.

McCandless, D.

2011 *Information is Beautiful, Capire il mondo al primo sguardo*, Milano, BUR Extra.

McCombs, M.E. e Shaw, D.L.

1972 *The agenda-setting function of mass media*, in «Public Opinion Quarterly», 36, pp. 176-187, Oxford, Oxford University Press.

McQuail, D.

2000 *Mass communication theory*, London, Sage, trad.it. *Sociologia dei media*, Bologna, Il Mulino.

Meyer, P.

1967 *The people beyond 12th street: a survey of attitudes of Detroit Negroes after the riot of 1967*, Detroit, Detroit Urban League, Detroit Free Press.

Meyer, P.

2006 *Giornalismo e metodo scientifico, ovvero il giornalismo di precisione*, Roma, Armando Editore.

Morando, F. (a cura di)

2013 *Libro bianco per il riutilizzo dell'informazione del settore pubblico* (1° ed. 2012), EVPSI, edizione digitale.

Morley, D.

1980 *The Nationwide audience. Structure and decoding*, London, British Dilm Institute.

Osgood, C.E.

1959 *The representation model and relevant research methods*, in de Sola Pool [1958, 33-38].

Paccagnella, L.

2010 *Sociologia della comunicazione*, Bologna, Il Mulino.

Porlezza, C.

2016 *Dall'open journalism all'open government? Il ruolo del data journalism nella trasparenza e nella partecipazione*, in «Problemi dell'informazione», anno XLI n.1, pp. 167-194, Bologna, Il Mulino.

Pulitzer, J.

1904 *Sul giornalismo*, Torino, Bollati e Boringhieri.

Raymond, E.S.

1997 *La cattedrale e il bazaar*, trad. it Bernardo Parrella (1999), O'Reilly Media.

Reggi, L.

2012 *Transparency on the beneficiaries of structural funds in Italy and Europe*, Ministero dello Sviluppo Economico, Dipartimento per lo Sviluppo e la Coesione economica, Unità di valutazione degli investimenti pubblici, Issue 27, Collana materiali UVAL.

Rogers, S.

2013 *Facts are sacred*, London, Faber & Faber Non Fiction.

Rovinetti, A.

2010 *Comunicazione pubblica. Sapere & fare*, Il Sole 24 Ore, Milano

Rovinetti, A.

2014 *Professione: comunicatore pubblico*, Aras Edizioni, Fano.

Savaresi, S.M. e Boley, D.L.

2004 *A comparative analysis on the bisecting K-means and the PDDP clustering algorithms*, in «Intelligent Data Analysis», 8, pp. 345-362.

Tichenor, P., Donohue, G. e Olien, C.

1970 *Mass media and differential growth in knowledge*, in «Public Opinion Quarterly», Oxford, Oxford University Press.

Tipaldo G.

2014 *L'analisi del contenuto e i mass media*, Bologna, Il Mulino.

Tipaldo G. e Gritti, M.

2015 *Il giornalista nel pallone. Il racconto del calcio nell'era dei Big data*, in «Problemi dell'Informazione», anno XL, n.1, pp. 35-56, Bologna, Il Mulino.

Topolski, J.

1973 *Metodologia historii*, Warsaw, PWN; trad. it. *Metodologia della ricerca storica*, Bologna, Il Mulino.

Vellar, A.

2015 *Le industrie culturali e i pubblici partecipativi: dalle comunità di fan ai social media*, Roma, Aracne Editrice.

# Bibliografia

<http://www.adsnotizie.it/index.asp>  
<http://www.agid.gov.it/agenda-digitale/agenda-digitale-italiana>  
<http://www.ahref.eu/it/biblioteca/data-journalism-1/cose-il-data-journalism.html>  
<https://www.change.org/about>  
<https://contrattipubblici.org/>  
<http://www.copyleft-italia.it/>  
<https://creativecommons.org/licenses/by-nc-nd/2.5/it/>  
<http://www.datajournalism.it/>  
<http://www.derivativeworks.com/2008/09/complete-scans.html>  
<https://ec.europa.eu/digital-single-market/>  
<http://eur-lex.europa.eu/legal-content/IT/TXT/?uri=URISERV%3Asi0016>  
<http://www.foia4italy.it/>  
<https://www.foia.gov/about.html>  
<http://hackshackers.com/>  
<https://www.linux.it/>  
<http://www.ojr.org/a-journalists-guide-to-the-scientific-method-and-why-its-important/>  
<http://opendefinition.org/>  
<http://www.openpolis.it/>  
<http://www.parlamento.it/parlam/leggi/>  
<http://www.poynter.org/2011/what-journalists-can-learn-from-scientists-the-scientific-method/144581/>  
<http://powerreporting.com/color/>  
<http://public-contracts.nexacenter.org/>  
<https://www.riparteilfuturo.it/>  
<http://www.spaghettiopendata.org/>  
<https://synapta.it/>  
<https://www.theguardian.com/commentisfree/2012/mar/25/alan-rusbridger-open-journalism>  
<https://www.theguardian.com/technology/organgrinder/2010/nov/19/berners-lee-journalism-data>

<https://www.w3.org/>

[https://www.whitehouse.gov/the\\_press\\_office/TransparencyandOpenGovernment](https://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment)