

Politecnico di Milano
School of Architecture, Planning and Construction Engineering

Urban Planning and Policy Design
Academic year 2018/2019

Data Literacy.
Un'indagine sull'utilizzo dei big data nei Comuni Italiani

Master thesis

Mentor: Prof. Grazia Concilio

Student: Francesco Fagiani (matricola 862535)

Indice

0. Introduzione	6
1. Il valore dei dati	7
1.1. Big Data.....	7
1.2. Open Data	17
1.3. Data Literacy	19
2. I Big Data nella Pubblica Amministrazione.....	22
2.1. I dati comunali	22
2.2. Data-driven city (o comuni).....	24
2.2. Database comunali.....	26
2.3. Risorse	29
3. Metodo analitico e statistico.....	32
3.1. Questionario online.....	33
3.2. Criteri per la definizione del campione.....	33
3.3. Campione omogeneo.....	36
4. Risultati dell'analisi	40
5. Conclusioni	68
Bibliography.....	69
Allegato 1 - Questionario.....	71
Allegato 2 - Percorso delle domande	78
Allegato 3 - Dettaglio comuni per invio questionario.....	79
Elenco dei comuni	80

Indice delle figure

Figura 1 - Numero di comuni, Popolazione e Superficie in base ai cinque gruppi di popolazione	35
Figura 2 - Mappa della penisola italiana con i 259 comuni che hanno risposto al questionario (aggiornamento 23/03/2019).....	40
Figura 3 - Provenienza risposte per Gruppo d'appartenenza.....	43
Figura 4 - Provenienza risposte per macroregione ISTAT.....	43
Figura 5 - Ruolo all'interno del Comune. Totale comuni risposte.....	43
Figura 6 - Ruolo all'interno del Comune. Campione omogeneo	43
Figura 7 - Settore comunale. Campione omogeneo.	45
Figura 8 - Settore comunale. Totale risposte.....	45
Figura 9 - Conosci la definizione di big data?.....	45
Figura 10 - Conosci la definizione di "big data"?	47
Figura 11 - Conosci la definizione di "big data"?	47
Figura 12 - Volume - grande quantità di dati	47
Figura 13 - Volume - grande quantità di dati	47
Figura 14 - Velocità - rapidità di acquisizione /	47
Figura 15 - Velocità - rapidità di acquisizione /	47
Figura 16 - Varietà - diverse tipologie di dati	48
Figura 17 - Varietà - diverse tipologie di dati	48
Figura 18 - Procedure - capacità di gestione e analisi	48
Figura 19 - Procedure - capacità di gestione e analisi.....	48
Figura 20 - Pubblicità - condivisione dei dati	48
Figura 21 - Pubblicità - condivisione dei dati	48
Figura 22 - Fonti - da dove vengono acquisiti i dati	49
Figura 23 - Fonti - da dove vengono acquisiti i dati	49
Figura 24 - Utilizzo dei big data.....	49
Figura 25 - Utilizzo dei big data.....	49
Figura 26 - Produzione interna di big data.....	49
Figura 27 - Produzione interna di big data.....	49
Figura 28 - Dove lavori vengono UTILIZZATI dei "big data"?.....	50
Figura 29 - Dove lavori vengono PRODOTTI.....	51
Figura 30 - Settore di provenienza	51
Figura 31 - Formato	51
Figura 32 - Fornitura dati a terzi	51
Figura 33 - Organizzazioni a cui vengono forniti i dati	52
Figura 34 - Formato dei dati forniti a organizzazioni terze	52
Figura 35 - Utilizzo big data di terze organizzazioni	53
Figura 36 - Organizzazione da cui derivano i dati	53
Figura 37 - Ragioni utilizzo dati terze organizzazioni	53
Figura 38 - Nella pubblica amministrazione in cui lavori vengono utilizzati dati prodotti da altre organizzazioni?	54
Figura 39 - Nella pubblica amministrazione in cui lavori vengono utilizzati dati prodotti da altre organizzazioni?	54
Figura 40- Secondo te i "big data" potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?.....	54
Figura 41 - Secondo te i "big data" potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?.....	54

Figura 42 - Secondo te "big data" di altre organizzazioni potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?	55
Figura 43 - Secondo te "big data" di altre organizzazioni potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?	55
Figura 44 - Da quanto tempo la pubblica amministrazione in cui lavori utilizza "big data"?.....	55
Figura 45 - Da quanto tempo la pubblica amministrazione in cui lavori utilizza "big data"?.....	55
Figura 46 - I "big data" hanno prodotto maggiore efficienza.....	55
Figura 47 - I "big data" hanno prodotto maggiore efficienza.....	55
Figura 48 - I "big data" hanno migliorato i processi decisionali.....	56
Figura 49 - I "big data" hanno migliorato i processi decisionali.....	56
Figura 50 - I "big data" hanno aiutato a migliorare i servizi al cittadino	56
Figura 51 - I "big data" hanno aiutato a migliorare i servizi al cittadino	56
Figura 52 - Credi che l'amministrazione pubblica per cui lavori utilizzi a pieno le potenzialità dei "big data"?.....	56
Figura 53 - Credi che l'amministrazione pubblica per cui lavori utilizzi a pieno le potenzialità dei "big data"?.....	56
Figura 54 - Quali risultati si cercano dall'analisi dei "big data"?	57
Figura 55 - Da quanto tempo la pubblica amministrazione in cui lavori utilizza "big data"?.....	57
Figura 56 - Chi si occupa dei "big data" dove lavori?.....	59
Figura 57 - Esiste una strategia definita per l'utilizzo degli OD?.....	59
Figura 58 - Esiste una struttura per il trattamento dei dati confidenziali?.....	59
Figura 59 - Chi si occupa dei "big data" dove lavori?.....	61
Figura 60 - Chi si occupa dei "big data" dove lavori?.....	61
Figura 61 - Esiste una strategia definita per l'utilizzo degli OD?.....	61
Figura 62 - Esiste una strategia definita per l'utilizzo degli OD?.....	61
Figura 63 - Esiste una struttura per il trattamento dei dati confidenziali e della privacy?	61
Figura 64 - Esiste una struttura per il trattamento dei dati confidenziali e della privacy?	61
Figura 65 - Conoscenza di cosa è un dato, come si raccoglie, analizza, visualizza e condivide. 62	
Figura 66 - Esiste una struttura di valutazione delle fonti, dei processi e dei risultati dei "big data"	62
Figura 67 - Esiste una struttura di valutazione delle fonti, dei processi e dei risultati dei "big data"	62
Figura 68 - Capacità di valutare l'affidabilità e l'imparzialità delle fonti.....	62
Figura 69 - Conoscenza di cosa è un dato, come si raccoglie, analizza, visualizza e condivide. 62	
Figura 70 - Capacità di valutare l'affidabilità e l'imparzialità delle fonti.....	62
Figura 71 - Capacità di interpretare le informazioni derivate dai dati	63
Figura 72 - Capacità di interpretare le informazioni derivate dai dati	63
Figura 73 - Capacità di usare i dati per prendere decisioni, risolvere problemi correnti	63
Figura 74 - Capacità di usare i dati per prendere decisioni, risolvere problemi correnti	63
Figura 75 - Capacità di usare strumenti tecnici avanzati nelle diverse fasi del processo.....	63
Figura 76 - Capacità di usare strumenti tecnici avanzati nelle diverse fasi del processo.....	63
Figura 77 - Conoscenza di come trattare la privacy e la sicurezza dei dati.....	64
Figura 78 - Conoscenza di come trattare la privacy e la sicurezza dei dati.....	64
Figura 79 - La pubblica amministrazione dove lavori ha organizzato corsi di formazione per facilitare l'utilizzo dei "big data"?	64
Figura 80 - La pubblica amministrazione dove lavori ha organizzato corsi di formazione per facilitare l'utilizzo dei "big data"?	64
Figura 81 - La pubblica amministrazione dove lavori ha organizzato corsi di formazione per facilitare l'utilizzo dei "big data"?	65
Figura 82 - Argomento del corso	65

Figura 83 - Secondo te perchè la pubblica amministrazione in cui lavori non utilizza "big data"?	66
Figura 84 - Secondo te i "big data" potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?	66
Figura 85 - Secondo te perchè la pubblica amministrazione in cui lavori non utilizza dati di altre organizzazioni?	67
Figura 86 - Secondo te "big data" di altre organizzazioni potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?	67

0. Introduzione

Lo sviluppo tecnologico che investe tutti sta cambiando il modo di vivere, di lavorare e di relazionarsi con gli altri. Le persone sono iper-connesse, le abitudini e i comportamenti vengono tracciati costantemente, collezionando ed aggregando informazioni che costituiscono *big data*. Questi dati vengono considerati una fonte di conoscenza a cui finora non era stato possibile accedere. Sono molteplici i settori a cui si applicano, così come l'utilizzo che se ne può fare, tra cui il settore pubblico in particolare legato alla definizione delle *policy*.

La ricerca scientifica è molto attiva in questo campo ed è orientata alle potenzialità ed agli utilizzi possibili dei *big data*. Tutto ciò è imprescindibile senza una adeguata *data literacy*. Questo termine viene definito come la capacità di utilizzare i dati in maniera critica, in funzione della presa di decisioni e dell'offerta di servizi. Tale capacità richiede una conoscenza dei dati, di cosa sono, di come essi vanno utilizzati e gestiti, conoscere quale contributo possono offrire all'organizzazione nel quale si opera, ma anche i limiti che si pongono nel loro utilizzo.

Questa tesi si ispira ad una indagine svolta da PoliVisu nel 2018 all'interno di una ricerca più ampia relativa ai dati e alle innovazioni tecnologiche.¹ La ricerca analizza l'utilizzo di *big data* nel settore della mobilità e dei trasporti, inoltre prova a stabilire il livello di *data literacy* delle pubbliche amministrazioni europee attraverso un questionario indirizzato ai soggetti che operano all'interno di organizzazioni pubbliche.

L'indagine ha riguardato gli aspetti principali che riguardano la *data literacy*, dalla sua definizione, alle competenze, fino alle strutture e alle applicazioni utilizzate nell'elaborazione dei dati.

La tesi utilizza il medesimo metodo d'indagine della ricerca citata. Il questionario ha avuto il compito di collezionare informazioni sulla *data literacy* nei comuni italiani. Si è prestata particolare attenzione affinché il campione risultasse rappresentativo dell'intera penisola. Questa operazione ha considerato un criterio geografico – funzionale all'ottenimento di risposte da tutte le regioni - ed uno basato sulla popolazione residente nel comune.

La tesi è strutturata in quattro capitoli, un primo capitolo che analizza i dati e il loro valore nella società attuale, provando a restituire una definizione dei principali termini presenti nel questionario. Un secondo capitolo che cerca di spiegare perché è necessario focalizzare l'attenzione sulle pubbliche amministrazioni, con una particolare attenzione ai comuni, oggetto dell'analisi. Il terzo capitolo stabilisce il metodo utilizzato nell'analisi del campione e nella restituzione dei risultati che saranno oggetto del quarto capitolo.

¹ (Data & Technical innovation deliverables, 2019)

1. Il valore dei dati

1.1. Big Data

Di Big Data si sente sempre più spesso parlare, sono numerose le ricerche e gli articoli che affrontano questo argomento e che dimostrano un interesse sempre crescente verso questo termine. Utilizzando Google Trends per verificare la popolarità delle ricerche correlate alla voce big data, si può notare un importante incremento a partire dal 2012 che solo recentemente tende a stabilizzarsi. Vengono riportate di seguito alcune testimonianze di ricercatori ed esperti di IT e mondo digitale, nel tentativo di mostrare da dove e perché scaturisce l'interesse per questo settore in costante evoluzione. Per prima cosa va analizzato il loro valore, infatti l'aumento d'interesse è fortemente legato alla creazione di un valore economico. Nel capitalismo del futuro, ormai prossimo, il "capitale" stesso sarà costituito dai dati. Questa è la tesi sostenuta da Vinton Mayer-Schonberger (docente di internet governance and regulation all'Oxford Internet Institute) uno dei massimi esperti mondiali di big data, secondo cui i dati rappresenteranno l'elemento strutturale dei mercati. L'esperto prefigura la creazione di un sistema economico data-rich, attraverso cui i mercati saranno resi più equi, meno centralizzati e più aperti alla competizione grazie alla equilibrata diffusione dell'informazione. Quindi valore e equità.

"I mercati funzionano solo se c'è un buon flusso di informazioni: se questo flusso migliora, allora si avrà un mercato più efficiente, quello che io chiamo il data rich market".²

Sembra difficile riportare questo discorso, che pare riguardare soltanto le grandi e potenti aziende della silicon valley, ad un livello concreto e fare un parallelo con le Pubbliche amministrazioni italiane. È chiaro però che in un'economia che si regge sempre più e in modo fondamentale sull'utilizzo dei dati, le istituzioni pubbliche siano chiamate a rimanere al passo, costantemente aggiornate e pronte al cambiamento che si prospetta loro. L'utilizzo dei dati oltre a un punto di vista prettamente economico investe le sfere della trasparenza e del bene comune, migliorabili anche attraverso l'utilizzo appropriato dei dati. Per questo motivo analizzare l'intreccio tra dati, economia, bene comune e istituzioni è doveroso.

Chi riprende il tema dei dati come moneta (data as currency) è Michelle Dennedy, responsabile della privacy di Cisco, in particolare studia i processi di privacy engineering, perché per poter estrarre valore dai dati si crea un evidente problema di privacy. L'esperta ribadisce il passaggio da una visione di *data as oil* (dati come petrolio, come valore) ovvero da estrarre e raffinare, a un approccio di *data as currency*, i dati come moneta. Ma aggiunge una questione cruciale in questo campo, quella della privacy.

"I dati possono generare valore perché di fondo c'è un accordo sociale e una fiducia diffusa sul fatto che vengano raffinati e analizzati nel rispetto della privacy".³

Il passo più urgente è quello di giungere a una regolamentazione da parte dei governi sulle modalità con cui i dati vengono raccolti, controllati e strutturati dalle grandi aziende che ora controllano i mercati in modo monopolistico. Solo a quel punto sarà possibile lasciarsi alle spalle mercati basati sul capitale e aprirci a una nuova economia basata sui dati. Perciò è importante considerare il valore della privacy sin dall'inizio del processo ovvero dalla raccolta dei dati; è attraverso la privacy si costruiscono prodotti e servizi. Al cuore della questione privacy c'è un

² (Mayer-Schonberger, 2018)

³ (Dennedy, 2018)

approccio di fondo che riguarda tutti, a partire proprio dalle pubbliche amministrazioni. Secondo la Denedy, oggi le pubbliche amministrazioni lavorano con un unico enorme *data lake* nazionale, ma si tratta di un errore. Secondo l'opinione dell'esperta questo è esattamente l'approccio di *data as oil* che genera errori e fughe di informazioni.

Il valore intrinseco dei dati costituisce una forma di potere da parte di chi li possiede, lo sostiene Christopher Wylie, ex consulente di Cambridge Analytica e whistleblower che ha portato alla luce lo scandalo che ha investito la società presso cui lavorava per il presunto possesso non autorizzato dei dati personali di 87 milioni di account Facebook.

*“I dati sono come gli atomi di uranio: puoi costruire e distruggere una città, o una società. (...) Le nostre identità come terra di conquista dei nuovi conquistadores, gli attuali signori della rete”.*⁴

Oggi l'economia di dati esiste, Cambridge Analytica offre un esempio di come i dati vengono raccolti, analizzati e sfruttati senza che gli utenti ne abbiano consapevolezza e nemmeno un reale controllo sulle informazioni che loro stessi hanno fornito. Per questo è necessario che le piattaforme siano regolamentate in maniera molto stringente.

Massimiliano Barawitzka parlando di Dapp (decentralized app) prefigura una rivoluzione degli attuali modelli di business (dei GAFA – Google, Amazon, Facebook, Apple, i big buddies del web) che sono basati sostanzialmente sullo sfruttamento dei dati personali, costringendoli a cambi di rotta importanti. Attraverso la piattaforma Solid, l'inventore di Internet Tim Berners-Lee, prova a reinventare la sua precedente invenzione, il www (world wide web). L'obiettivo è ridare controllo dei dati alle persone cui appartengono, togliendoli ai big della tecnologia.

Proprio come negli esempi precedenti la privacy e un utilizzo sicuro dei propri dati e della propria identità digitale sono la modalità con cui reinventare il web. Ci si è resi conto che la libertà che si gode online non è così reale come si crede, soprattutto per quanto riguarda il possesso e l'utilizzo dei dati che forniamo, anche senza rendercene conto.

⁴ (Wylie, 2018)

1.1.1. Big Data: una definizione

Nella stesura di questo capitolo si è fatto riferimento alla recente letteratura che riguarda i big data cominciando dal paper al quale questa tesi si è ispirata: “*Experimental dimension of policy making*”⁵ redatto a gennaio 2018 nel corso della ricerca condotta sulla *data literacy* delle pubbliche amministrazioni europee sotto l’egida Polivisu.

Questo report è stato utile prima di tutto per capire con più precisione di cosa si parla esattamente quando si utilizzano termini come *Big Data* e *Smart city*, ma anche a selezionare con maggiore precisione le fonti e le ricerche da cui attingere alle informazioni utili. Perciò per prima cosa si cercherà di dare una definizione al termine che definisce il perimetro d’indagine del questionario, “*Big Data*”.

*“Big data is, essentially, everything captured or recorded digitally by modern information and communications technologies such as networked sensors, ‘smart’ objects and devices, the web and social media”*⁶

La traduzione letterale del termine tende a far intendere big data soltanto in funzione del loro volume, ovvero come grande mole di dati. Una definizione molto esplicativa a tal proposito appartiene a Michel Batty “*any data that cannot fit into an Excel spreadsheet*”⁷, qualsiasi dato che non può essere contenuto in un foglio Excel. In realtà anche il report di Batty prosegue aggiungendo caratteristiche alla definizione appena riportata. Si capisce che non è un’operazione così semplice e che alla prima definizione vanno aggiunte altre informazioni. Eaton, il primo autore citato, prosegue indicando le fonti da cui attingere questi dati:

*“It can take the form of text, web data, tweets, sensor data, audio, video, click streams, log files, banking and purchase data, social network chatter, traffic flow sensors, mobile phone GPS trails and smart energy meters”*⁸

Risulta evidente che il termine *big data* è ampio e abbraccia non solo i dati in senso stretto, ma la loro modalità di raccolta, la loro gestione e visualizzazione. Di seguito vengono individuate e spiegate le caratteristiche principali, come suggerito all’interno del primo report rilasciato all’interno della ricerca PoliVisu.

Se si fa riferimento alla definizione fornita dall’IT glossary Gartner i *big data* vengono considerati come

“high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation”.⁹

Una definizione che può sembrare generica ma che elenca tutte le caratteristiche che devono appartenere a questo tipo di dati.

- *High volume*: una grande mole di dati costituita da terabytes o addirittura petabytes.
- *High velocity*: questa caratteristica si riferisce alla raccolta, che deve essere rapida, cosa che, con l’utilizzo delle moderne tecnologie, permette di ottenere dati in tempo quasi reale.
- *High variety*: strutturati e destrutturati, spesso definiti nello spazio e nel tempo.

⁵ (Paola Pucci, Giovanni Vecchio, Grazia Concilio, 2018)

⁶ (Eaton C., Deroos D., Deutsch T., Lapis G., Zikopoulos P., 2012)

⁷ (Batty, 2013)

⁸ (Eaton C., Deroos D., Deutsch T., Lapis G., Zikopoulos P., 2012)

⁹ (Big Data, 2019)

- *Exhaustive*: la sfida è quella di ottenere campioni totali della popolazione analizzata. Comunque ci si riferisce a campioni significativamente più ampi delle tradizionali ricerche statistiche.
- *Fine-grained*: viene ricercata la granularità del dato e il massimo dettaglio possibile (sorge un problema di privacy, che viene rispettata grazie all'aggregazione e all'anonimato)
- *Relational*: presenza di campi comuni che permettano la relazione e integrazione di diversi dataset.
- *Flexible*: in questo senso i dati devono mantenere le caratteristiche di estensibili (aggiunta di nuovi campi) e scalabili (rapido cambio delle dimensioni)

A smentire la credenza secondo cui con il termine *big data* si definisce solo sulla base della dimensione dei dati viene a supporto il report di Boyd e Crawford pubblicato nel 2012¹⁰ in cui ribadiscono come principali caratteristiche dei *big data* la velocità di raccolta, il volume di dati e la loro varietà (diverse fonti), ma specificano che i *big data* non sono tali solo in base al loro volume “*big*”, ma in base alla possibilità che vengano collegati ad altri dati provenienti da fonti diverse. Quest'ultima definizione richiama molto l'idea di *Linked data* e di web semantico (Web 3.0) che a sua volta si ricollega al termine di *open data*, fondamentale quando si parla di pubbliche amministrazioni come avviene in questa tesi.

Prima di parlare di altre tipologie di dati è bene fornire una categorizzazione dei *big data* sulla base della tipologia delle fonti. All'interno del primo report a supporto della ricerca Polivisu viene ripresa la classificazione di Kitchin¹¹, che nel 2014 individua tre categorie di *big data*:

- *Directed Data*: la fonte diretta dei dati, sono chiamati anche *tracking data*, monitorano i movimenti individuali delle persone o di oggetti ad esse associati. Basti pensare ai dati sulla localizzazione forniti dagli smartphone che ci portiamo costantemente appresso. Sono generati attraverso una forma di sorveglianza “*form of surveillance*”, determinano luoghi di frequentazione e rivoluzionano il modo di valutare gli spostamenti delle persone che finora si basavano sulla matrice origine destinazione.
- *Automated Data*: vengono prodotti in modo automatizzato da dispositivi digitali come smartphone, clickstream (attraverso cui si monita il percorso degli utenti all'interno di un sito o una app), *IoT data* generati da sensori vari presenti sul territorio, dati di immagini comprese le immagini aerofotogrammetriche.
- *Volunteered data*: questi dati derivano dalle ricerche effettuate online, sui social network, in questo caso gli utenti stessi forniscono i dati, fungendo essi stessi da sensori, contribuendo a collezionare informazioni geografiche e non solo. L'interazione con i social attraverso la pubblicazione di un post o il commento di un'immagine e il caricamento di fotografie gli utenti forniscono dati e opinioni spesso personali. Esiste una componente volontaria, è l'utente che fornisce volontariamente questi dati, anche se spesso non conosce l'utilizzo che ne viene fatto.

Nella categoria dei dati volontariamente forniti (spesso anche in modo inconsapevole) dagli utenti, emerge una problematica importante di rispetto della privacy e della sicurezza dei dati, in particolare di quelli personali. Verrà citato in seguito il caso di Facebook e dell'utilizzo improprio di dati provenienti da milioni di utenti da parte di un'impresa privata, Cambridge Analitica. Affrontare il problema di sicurezza dei dati personali forniti da apparecchi individuali, in

¹⁰ (Boyd D., Crawford C., 2012)

¹¹ (Kitchin, 2014)

particolare da smartphone, è una questione delicata ma cruciale che verrà trattata nel paragrafo successivo.

La classificazione appena fornita è diversa da quella della *United Nations Economic Commission for Europe*, sviluppata durante la *Conference of European Statisticians* (2013) che si compone delle seguenti voci:

- *Administrative*: questa voce comprende tutti i dati raccolti all'interno di un programma amministrativo sia esso governativo oppure non governativo. Alcuni esempi comprendono cartelle cliniche elettroniche, visite mediche ospedaliere, registri delle assicurazioni, delle banche ecc.
- *Commercial or transactional*: derivano dalle transazioni effettuate tra due entità, come i dati delle carte di credito, i dati delle compravendite online, ecc.
- *Sensors*: i dati derivati dai sensori
- *Traking*: i dati che tracciano la posizione e gli spostamenti degli utenti. Dati GPS e derivati dagli smartphone e applicazioni, ecc.
- *Behavioural*: dati derivati dal comportamento degli utenti, *clickstram*, rispetto a un servizio a un prodotto, una applicazione o dalla visualizzazione di una pagina web.
- *Opinion*: esempio emblematico di questa categoria sono i commenti sui social media.

Una conseguenza importante che deriva dalla classificazione appena fornita, ma che è ancora più chiara se si confronta con la prima classificazione tratta da Kitchin¹², riguarda le implicazioni che si hanno, in base alla fonte, nell'utilizzo, nelle potenzialità e nei limiti ai *big data*. La modalità con cui essi sono generati è il prodotto di scelte che influiscono anche sull'utilizzo che se ne può fare nelle fasi successiva alla raccolta. Attraverso la ricerca PoliVisu è stato possibile individuare nel metodo *Knowledge Discovery form Databases* (KDD) le modalità di elaborazione dei *big data*. Dalle caratteristiche fornite fin qui è evidente la natura destrutturata dei big data. Essi sono provenienti da diverse fonti, contenuti in diversi database, spesso incompleti e quasi certamente non "noisefree". La traduzione letterale del termine *noisefree* è silenzioso, che si riferisce alla necessità di eliminare dalle banche dati i rumori, ovvero i dati non strettamente necessari all'analisi e mantenere soltanto il segnale. In questo caso con segnale si deve intendere l'informazione importante che si ottiene attraverso la strutturazione e l'organizzazione dei dati e che può comportare il loro dimensionamento. Dai dati di partenza, diversi a seconda della provenienza, non esiste un approccio automatico per raggiungere delle soluzioni e degli scenari. L'approccio *KDD* passa generalmente attraverso queste fasi: *selection; pre-processing, transformation; data mining; interpretation/evaluation*. Tre di queste fasi sono imprescindibili nell'applicazione del metodo e sono le seguenti:

- *Pre-processing*: questa fase riguarda le azioni di pulizia, la realizzazione di dati "noisefree", la definizione di un target preciso in funzione di una migliore analisi nelle fasi successive. La pulizia riguarda anche i dati mancanti o incompleti.
- *Data mining*: questa fase è composta a sua volta da sei classi;
 - *Anomaly detection (outlier/change/deviation detection)*. Vengono identificate eventuali anomalie presenti nella banca dati. Dati inusuali, valori inusuali, ecc.
 - *Association rule learning (dependency modelling)*. In questo passaggio vengono analizzate possibili relazioni tra diverse variabili. Ad esempio, riguarda i prodotti dei supermercati, l'interdipendenza tra variabili permette di determinare quali prodotti più di frequente sono comprati insieme, usando questa informazione per rilasciare offerte e fare marketing.

¹² (Kitchin, 2014)

- *Clustering*. Ricercare gruppi o strutture dei dati per similitudine, cominciando a delinearne una possibile strutturazione.
 - *Classification*. Tentare l'applicazione di strutture esistenti ai nuovi dati. Ad esempio, un software e-mail che prova a classificare la nuova posta come posta in arrivo oppure come spam.
 - *Regression*. Tentativo di applicare una funzione ai dati, una modellazione dei dati che garantisca il minimo errore, per stimare relazioni tra dati o set di dati.
 - *Summarization*. Elaborare una rappresentazione compatta e chiara del dataset, è la fase cruciale di visualizzazione e della creazione di report.
- *Results validation*: le operazioni di *data mining* potrebbero dare risultati che appaiono significativi, ma che in realtà non possono essere applicati ad un campione o prevedere comportamenti futuri. Un problema di questo genere viene denominato *overfitting*, ed è particolarmente significativo nel capo della *machine learning*.

Gli aspetti che caratterizzano i big data sono rilevanti per le pubbliche amministrazioni e nelle scelte politiche e di governo del territorio. Sorgono allo stesso tempo problemi derivati dalla complessità intrinseca di questi dati. L'aumento del livello di complessità non può comportare l'utilizzo di questi dati soltanto da parte di imprese private tecnologicamente avanzate. La pubblica amministrazione ha l'obbligo di dotarsi degli strumenti necessari per l'utilizzo di questi dati. Bruun e Givoni¹³ nel 2015 scrivono a tal proposito scrivono quanto segue:

(è necessario) “capire le implicazioni politiche di questi dati e di come le istituzioni dovrebbero evolvere per poter soddisfare questa richiesta. Le metodologie e gli strumenti sono fondamentali per sviluppare politiche efficaci, in modo strategico, superando le barriere attuative, come l'accettazione pubblica e politica”.

L'analisi della ricerca PoliVisu, focalizzata sullo studio dei big data nel campo della mobilità, da una parte riguarda un campo definito che quindi applica risposte e considerazioni specifiche, dall'altra il maggiore sviluppo che ha interessato l'utilizzo di big data nella mobilità rispetto ad altri campi, permette di anticipare problematiche che potrebbero verificarsi anche in altri settori in modo trasversale. È stato perciò molto utile investigare i big data nelle politiche urbane sulla mobilità, seguendo lo schema proposto dalla ricerca PoliVisu.

La necessità di utilizzo dei big data nelle pubbliche amministrazioni è dovuta ovviamente alle grandi opportunità in termini decisionali e di valutazione delle politiche. La ricerca PoliVisu che si focalizza sulla mobilità propone altre due altre ragioni che ne dimostrano la rilevanza.

La prima ragione è la nuova opportunità messa a disposizione di pianificatori, tecnici e politici dalle risorse digitali e tecnologiche sparse su tutto il territorio e soprattutto agganciate ad ogni individuo che possiede uno smartphone. Le tecnologie collegate in rete producono dati costantemente in maniera automatica, definendo i dati nello spazio e nel tempo con una risoluzione impossibile da raggiungere con le ricerche statistiche tradizionali. Basti pensare a cosa significa ottenere gli spostamenti delle persone tramite i dati telefonici. È possibile stabilire dinamiche della città in tempo quasi reale.

La seconda riguarda il cambiamento dei modelli di studio che deriva da questa disponibilità di dati. Ad oggi gli studi sulla mobilità si basavano su costose e lunghe ricerche statistiche che determinavano una matrice di origine destinazione e la relativa modalità di trasporto sulla quale venivano programmate tutte le attività correlate alla pianificazione dei trasporti. È evidente che i nuovi dati modificano profondamente, migliorandola, l'analisi e la programmazione sia a lungo, che a breve termine.

¹³ (Bruun E., Givoni M., 2015)

Si tratta di dati ampiamente disponibili, che non colgono soltanto un campione parziale di individui grazie alla diffusione capillare della tecnologia, di dati originariamente digitali che diminuisce al minimo i margini di errore, e dati di alta qualità da cui si possono determinare i comportamenti degli utenti.

È proprio nel campo della mobilità che si ha il più fertile utilizzo di big data, grazie alla disponibilità dei dati infatti, in questo campo si possono programmare, gestire e valutare cambiamenti e politiche pubbliche sui trasporti.

Non soltanto dagli “*Automated data*” possono derivare le analisi (vedi la classificazione proposta da Kitchin e riportata in precedenza). Politiche di trasporto possono essere analizzate e valutate anche sulla base dei “*Volunteered data*” generate dagli utenti in maniera volontaria ed estrapolate dai social network. Alcune ricerche hanno utilizzato questo tipo di dati per ottenere informazioni qualitative sull’offerta di trasporto, comfort, gestione delle problematiche o miglioramento dell’offerta.

1.1.2. Gestione dei Big Data

La gestione e il governo dei big data e dei dati in generale, è una condizione cruciale per lo sviluppo di politiche pubbliche guidate dall'analisi dei big data. Per questo, come già anticipato, le fonti di dati sono determinanti, perché proprio a partire da esse viene a determinarsi il processo di analisi. Di conseguenza gli attori che posseggono questi dati nella fase iniziale (dalla raccolta) vanno trattati con altrettanta attenzione. Anche in questa parte del capitolo la tesi si confronta con la ricerca di PoliVisu, utilizzando la stessa classificazione fornita dal report. Come nei casi precedenti la ricerca appena citata, volge lo sguardo al settore della mobilità e dei trasporti, valgono nuovamente le considerazioni fatte in precedenza sulla trasversalità di alcune questioni che emergono, perciò la classificazione non cambia. I soggetti che producono e gestiscono i dati rientrano in uno di questi casi.

Il primo fa riferimento alle istituzioni pubbliche e alle pubbliche amministrazioni. Con particolare riferimento ai comuni, si può dire che questa categoria possiede tutte le informazioni derivate da eventuali sensori posti nel territorio. I comuni hanno la facoltà di installare sensori del traffico, della lettura di targhe o sull'inquinamento, distribuirli all'interno dei confini comunali e utilizzare i dati senza limitazioni. Oltre a questo tipo di dati le pubbliche amministrazioni hanno a disposizione un archivio (non sempre del tutto digitalizzato) contenente pratiche, contratti, informazioni riguardo a moltissimi aspetti della vita pubblica comunale. La cartografia e l'urbanistica fanno parte di questi. Dagli uffici comunali potrebbero emergere moltissime informazioni, che se opportunamente raccolte e gestite costituirebbero una base di dati fondamentale. La creazione di dataset nei vari settori comunali a partire dalle pratiche presentate dai cittadini, magari in forma telematica tale da garantire una raccolta efficace in termini di tempo, determina un terreno fertile alla gestione e analisi di big data. Il loro valore è determinato anche dalla loro ampiezza, questi dati infatti vengono raccolti in tutti o quasi i campi, ottenendo così una conoscenza approfondita e a 360 gradi del comune in questione. Se si considera poi la possibilità di unire in modo organizzato i dataset di diversi comuni, il valore e le possibilità che si aprono sono ancora maggiori. Queste operazioni in parte sono già cominciate, lo stato, ma soprattutto le regioni, stabiliscono obblighi e adempimenti burocratici nella gestione e nella fornitura dei dati, per poi riorganizzarli in maniera aggregata. Ogni regione si organizza per conto proprio, determinando probabili squilibri, che il questionario cercherà di valutare.

Il secondo caso proposto riguarda le Aziende pubbliche. In questo caso la mobilità e i trasporti calzano in maniera particolare. Sono moltissime le aziende pubbliche o partecipate, controllate dallo Stato o da altre Istituzioni che giocano un ruolo fondamentale nel settore dei trasporti. Aziende che si occupano di trasporto pubblico, aziende che gestiscono il patrimonio infrastrutturale, che gestiscono informazioni legate alla mobilità di merci e persone. Anche la varietà di fonti è ampia, dai sistemi di sorveglianza, ai sensori di traffico diffusi sulle strade, nei parcheggi, sistemi di accesso alle infrastrutture, applicazioni smartphone, ecc. E questi esempi sono solo quelli legati al mondo dei trasporti, ma molti altri settori sono oggetto di questa categoria, le aziende ospedaliere, quelle legate alla gestione dei rifiuti o all'energia elettrica. Per non parlare di quelle che gestiscono le reti di telecomunicazione. Aziende che gestiscono il patrimonio di edilizia pubblica e tante altre ancora. Il report *Experimental dimension of policy*¹⁴ anche in questo caso, pur rimanendo nel settore specifico legato ai trasporti, fa emergere un problema trasversale che interessa il rapporto tra Istituzioni e aziende pubbliche o partecipate. Infatti, la disponibilità di dati da parte di questi gestori non si traduce automaticamente in una disponibilità da parte dell'Istituzione pubblica che affida il servizio. Esistono seri problemi di proprietà e di condivisione di questi dati, l'integrazione e la fornitura dei dati tra azienda e

¹⁴ (Paola Pucci, Giovanni Vecchio, Grazia Concilio, 2018)

istituzione non avviene sempre in maniera fluida come dovrebbe essere. A riguardo, si aggiunge un appunto che riguarda l'Agenzia delle Entrate e del Territorio. Si tratta di un ente pubblico vigilato sottoposto all'indirizzo e alla vigilanza del Ministero dell'Economia e delle Finanze. Questo esempio mette in luce un'ulteriore questione, quella della fruibilità dei dati, oltre che tra società partecipate e istituzioni, tra queste e le aziende private e i cittadini. I dati dovrebbero essere liberamente consultabili con licenza aperta da parte di tutti. Penso ad esempio alle banche dati catastali, che opportunamente ripuliti da intestazioni e altri dati sensibili che potrebbero far emergere problemi di privacy, potrebbero essere liberamente consultabili, ma che risultano difficilmente accessibili. La cartografia catastale che in altri paesi europei è fruibile e riutilizzabile sia da utenti che da macchine in Italia è soggetta a un macchinoso sistema di controllo e di blocco delle macchine (inserimento di *capcha*). Quest'ultimo argomento si ricollega alla questione degli open data.

Il terzo soggetto indicato nella produzione dei big data è rappresentato dalle Aziende private. Questo punto apre veramente infinite possibilità per l'utilizzo di dati e lo scambio di informazioni. Ogni settore, ogni tipo di azienda ormai raccoglie gestisce e scambia dati. Dalle aziende che si occupano di applicativi per web e smartphone, a quelle che gestiscono le reti di telecomunicazioni, ma anche aziende che controllano i propri processi tramite IoT in modo digitalizzato. Abbiamo inoltre visto come molte delle imprese private fanno dei dati il fulcro del loro business, la nuova moneta di scambio e anche la principale. Molte aziende offrono servizi in maniera gratuita in cambio dell'accesso ai nostri dati. In questa prospettiva, la gestione delle informazioni, la privacy dei dati e la sicurezza acquistano un ruolo fondamentale. La proprietà dei dati e la consapevolezza dei loro utilizzi possibili prerogativa di ogni cittadino e della sua identità digitale.

Infine, il soggetto che ha rivoluzionato il mondo dei dati diventando esso stesso un produttore, è rappresentato dai Privati cittadini. I privati cittadini possono agire come veri e propri sensori e questo li rende una fonte importantissima. Sorge, collegandosi al punto precedente, una questione cruciale che riguarda la volontarietà o meno di fornire questo tipo di dati a aziende che li riusano e rivendono. Attraverso l'utilizzo dello smartphone, di applicazioni web, dei social network o anche semplicemente usufruendo di servizi pubblici i privati cittadini offrono un'enorme mole di dati alle categorie precedentemente rappresentate. A partire da questa categoria, dai loro spostamenti, dalle loro opinioni, dall'utilizzo che fanno dei servizi, che vengono generati big data.

1.1.3. La Privacy e GDPR

L'utilizzo dei big data pone dei problemi rispetto al loro utilizzo effettivo. Se l'obiettivo è quello di un'influenza dei big data tale per cui i dati "parlino da soli" cioè la ricerca di un "*automatic smart city understanding*", è ancora necessario applicare un approccio basato sulla conoscenza dell'uomo per dare senso alle informazioni derivate dai dati. La capacità di selezionare le informazioni, soprattutto la capacità di dare loro un senso:

*"only those with the wherewithal to make sense of the information (through data mining or other means of investigation) were able to increase their use of the new, digitised data".*¹⁵

Solo coloro che utilizzano questi mezzi per dare un senso alle informazioni (...) sono stati in grado di incrementare l'uso dei nuovi dati digitali. Per questo è necessario "*make big data small*"; questa apparente contraddizione in termini, che è già stata affrontata all'interno del procedimento KDD, rende inevitabile una integrazione dei dati con altre forme di conoscenza, questa questione è oggetto di un importante dibattito accademico.

Inoltre, rimangono delle questioni aperte che riguardano in particolare la mobilità, come spiega la ricerca PoliVisu, ma che sono applicabili anche ad altri campi. Una prima questione riguarda la privacy, se proviamo a pensare ai *traking data* che derivano dagli smartphone essi si rivelano tanto precisi da permettere una profilazione degli utenti a partire dai dati con un conseguente problema di controllo dei cittadini. Nessuno può rimanere nascosto in un tale sistema, generando questioni di controllo e mancanza di libertà nel caso in cui venisse meno l'anonimato in un contesto di utilizzo dei dati lontano da quello accademico e orientato al controllo. Altro problema che si sono posti Song, Qu, Blumm e Barabási¹⁶ riguarda alcuni limiti tecnici all'utilizzo dei dati nella previsione di comportamenti umani, che in quanto tali mantengono un grado di non prevedibilità. Ma forse insieme alla privacy, il problema che caratterizza maggiormente i big data riguarda la rappresentazione incompleta di un problema o di un fenomeno. Questo potrebbe generare omissioni o esclusioni di intere popolazioni (statistiche) dalle banche dati. Ad esempio, la *digital divide* non permette a una popolazione omogenea di essere rappresentata: gli spostamenti e i luoghi frequentati dagli anziani, che generalmente non posseggono uno smartphone, non potranno essere considerati semplicemente perché non raccolti. Se non venissero considerati questi possibili errori si genererebbe una concentrazione del potere dei dati, in favore della popolazione che li fornisce a discapito di chi non rientra nel campione.

Nel testo del questionario non viene fatto specifico riferimento al GDPR (*General Data Protection Regulation*), ma sono presenti domande che indagano la conoscenza di tematiche relative alla privacy. Diversi partecipanti hanno indicato a proposito della privacy, l'adempimento richiesto dal GDPR riguardo il trattamento e la circolazione dei dati personali. Le PA sono obbligate a adeguarsi alla normativa entro 25 maggio 2018, data in cui sarà direttamente applicabile il regolamento UE 2016/679. Il Piano Industria 4.0 ha individuato tre priorità operative. La prima riguarda la designazione di un responsabile per la protezione dei dati. La seconda, l'istituzione del registro delle attività di trattamento, la terza, la notifica dei *data breach*.¹⁷ Sebbene non avvenga in maniera automatica, la presenza di una normativa, obbliga le PA a adeguarsi dotandosi del necessario. Anche se risulta un adempimento burocratico, viene imposto al responsabile un livello minimo di conoscenza del tema privacy e trattamento dei dati, ovvero una certa *data literacy*.

¹⁵ (Rabary C., Stoper M., 2014)

¹⁶ (Chaoming Song, Zehui Qu, Nicholas Blumm, Albert-László Barabási, 2010)

¹⁷ (A. Longo, R. Natale, 2018)

1.2. Open Data

Il riferimento a questa tipologia di dati è stato frequente nel capitolo precedente. Diversi sono i collegamenti tra i termini open data e big data. È necessario provvedere a dare una definizione anche agli open data che vada oltre la traduzione letterale di dati aperti.

*“Open data is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control”*¹⁸

Open Data è l'idea che alcuni dati dovrebbero essere disponibili liberamente a chiunque da utilizzare e pubblicare a proprio piacimento, senza restrizioni di *copyright*, brevetti o altri meccanismi di controllo

Gli obiettivi che si prefigge il movimento open data sono molto simili a quelli legati all'open source. Il termine open data è recente, anche se meno di big data. Il secondo infatti è diventato realtà con l'avvento di tecnologie che ne hanno reso possibile la raccolta e l'archiviazione in grandi quantità, cosa che è avvenuta soltanto recentemente. Mentre la filosofia alla base degli open data possiede una tradizione scientifica più “antica”, anche se l'interesse reale si è manifestato con l'avvento di Internet e del *World Wide Web* ormai trent'anni fa. Quindi il concetto di open data non è recente come si potrebbe pensare, anche se la formalizzazione del termine lo ha reso tale.

*“A piece of data is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike.”*¹⁹

“Un dato si definisce aperto se chiunque può usarlo, riutilizzarlo e redistribuirlo – ed è soggetto solo, al limite, alla attribuzione delle stesse caratteristiche di condivisione.”

La definizione appena proposta sembra essere indicata da molti report e siti web come la più completa ed attendibile. Questa definizione può essere applicata a tutti i dati, perciò tutti i dati, di tutte le tipologie, possono essere resi open. Big data, small data, altri tipi di contenuto, come immagini, testi, musica.

*“The Definition was based directly on the Open Source Definition from the Open Source Initiative and we were able to reuse most of these well-established principles and practices that the free and open source community had developed for software, and apply them to data and content”*²⁰

Sempre secondo la pagina web di Open Knowledge, a cui molte altre fonti si ispirano, “Open knowledge is any content, information or data that people are free to use, re-use and redistribute — without any legal, technological or social restriction”. Le caratteristiche chiave di questa definizione riguardano tre principi determinanti nel fare emergere le potenzialità:

- *Availability and access.* Chiaramente la disponibilità e l'accesso ai dati sono centrali nella definizione di un dato aperto. Questo significa che essi devono essere presenti nel loro insieme, preferibilmente scaricabili online. Inoltre, i dati devono essere presenti in una forma comoda e modificabile. Per riprendere l'esempio dei dati catastali forniti dall'Agenzia delle Entrate, pare chiaro che questa caratteristica sia del tutto assente.

¹⁸ (S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, 2007)

¹⁹ (Open Definition 2.1, 2019)

²⁰ (Open Definition 2.1, 2019)

- *Reuse and redistribution.* I dati devono essere forniti sotto termini che ne permettano il riuso e la riproduzione, inclusa l'integrazione con altri dataset. I dati devono essere *machine-readable*.
- *Universal participation.* Chiunque deve essere in grado di accedere a questi dati, senza discriminazione. Per questo non sono previste nemmeno restrizioni di tipo "non-commerciale" che escludano un utilizzo commerciale o limitato a certi scopi (es. educazione).

Gli elementi che definiscono questa apertura sono due, un'apertura di tipo legale ed una tecnica. L'apertura legale dipende dalle licenze applicate al dato che ne permettono un completo utilizzo. L'apertura tecnica riguarda la necessità di evitare barriere tecniche nell'utilizzo dei dati. Questo significa evitare report scannerizzati oppure tabelle in formato pdf, caratteristiche che determinano un difficile riutilizzo delle informazioni.

*"The information should be digital, preferably available by downloading through the internet, and easily processed by a computer too (otherwise users can't fully exploit the power of data – that it can be combined together to create new insights)"*²¹

Una recente ricerca sviluppata in Italia, curata da GovLab e dalla Fondazione Bruno Kessler ha dimostrato l'importanza degli open data per la realizzazione di prodotti e servizi da parte di imprese private. In questo caso si parla appunto della creazione di valore e sviluppo di servizi grazie a dati, che se rimanessero "chiusi", non permetterebbero di farlo. Queste aziende lavorano grazie ai dati messi a disposizione dalla pubblica amministrazione. Scopo della ricerca è stato proprio quello di censire e provare a raccontare queste realtà. L'iniziativa prende spunto da un'analogia ricerca OD500 che censiva le aziende americane che utilizzano i dati aperti. OD200 è il nome della ricerca sul territorio italiano. La ricerca è la dimostrazione di come i dati, in particolare quelli messi a disposizione dalle Pubbliche amministrazioni permettano di generare valore e lavoro. Ovviamente gli open data da soli, come del resto i big data che non vengono elaborati, non costituiscono valore aggiunto, del resto gli open data sono accessibili liberamente da tutti, come spiegato nelle definizioni. Il valore aggiunto si crea attraverso la loro rielaborazione e con l'aggregazione con altri dati, anche privati, come spiega la ricercatrice che ha realizzato la ricerca, Francesca De Chiara. I risultati raccontano che il 47 % delle imprese offre i servizi ad altre imprese, oltre il 30 % ha come cliente la pubblica amministrazione stessa e solo il 17% delle imprese censite offre servizi B2C (business to client). Molto interessante, visto che questa tesi si sviluppa nella facoltà di urban planning, è il fatto che i dataset maggiormente richiesti riguardano dati geografici.

1.2.1. Linked data

Non tutti gli open data sono uguali, esistono diverse sfumature di apertura che dipendono dalle caratteristiche mostrate nel paragrafo precedente. Tim Berners-Lee nel 2006 pubblica un report dal titolo "Linked Data" in cui definisce questo tipo di dati e propone una classificazione con 5 gradi per stabilire se i tuoi dati sono definibili linked.

"The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data."

²¹ (Open Definition 2.1, 2019)

★

Available on the web (whatever format) but with an open licence, to be Open Data

★★

Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★

As two stars plus non-proprietary format (e.g. CSV instead of excel)

★★★★

All the above plus, use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★

All the above, plus: Link your data to other people's data to provide context

1.3. Data Literacy

Questa sezione del deliverable *Experimental dimension of the policy* è stata molto d'aiuto nelle prime fasi di questa ricerca, durante la definizione degli obiettivi, ma anche per iniziare ad avere un'idea generale del fenomeno *data literacy* applicato alle pubbliche amministrazioni. Curiosando su internet infatti è facile trovare definizioni generiche o ricerche che applichino questo tema al settore delle imprese private o a quello scolastico (alfabetizzazione dei dati come primo approccio al loro utilizzo e studio). Molto più complesso è avere una panoramica del tema applicato al mondo della pubblica amministrazione, come propone il report appena citato. Esso si compone di tre paragrafi, un primo riguarda una definizione accademica del termine *data literacy*, un secondo che analizza le ricerche esistenti e infine un terzo che restituisce una definizione costruita a margine della ricerca PoliVisu.

Secondo la ricerca appena citata, nel particolare contesto a cui si applica, questa è la definizione di *data literacy*:

*“the ability to critically use data for developing policy, taking decisions and deliver services, wherever possible; this ability implies knowing what data are, how they are managed and used, what contributions they can provide to an organization as well as the limitations they are prone to”*²²

Ovvero, l'abilità di saper usare in modo critico i dati per sviluppare politiche, prendere decisioni e offrire servizi. Questa abilità implica la conoscenza dei dati, la loro gestione e il loro utilizzo; quale contributo essi possono offrire ad un'organizzazione e le limitazioni che si pongono in questo utilizzo. A queste definizioni ho scelto di aggiungere la definizione tratta dalla pagina web di Wikipedia, per mostrare come generalmente viene approcciato il tema. La *data literacy* è considerata come l'abilità di leggere, capire, creare e comunicare i dati come informazioni. La *data literacy*, più dell'alfabetizzazione come concetto generale, si concentra sulle competenze

²² (Paola Pucci, Giovanni Vecchio, Grazia Concilio, 2018)

necessarie nel lavorare con i dati. Si conclude con l'importanza di avere questo tipo di alfabetizzazione da parte di studenti, cittadini e lettori.

Nel report di Tibor Koltay, una delle fonti più citate quando si ricerca questo termine, si parla addirittura di diversi approcci alla *data literacy*.²³

Quindi non soltanto diverse definizioni, ma diversi approcci scovati tra le definizioni utilizzate da autori in precedenti ricerche. Koltay elenca e descrive gli autori che a partire dal 2004 hanno iniziato a parlare di *data literacy*, e ne esplora le differenze e le evoluzioni di questo termine.

La prima citazione appartiene a Shield (2004) secondo cui l'alfabetizzazione concerne la capacità di accedere, valutare, maneggiare, sintetizzare e presentare i dati. L'autore compara questo tipo di alfabetizzazione con un'altra, l'*information literacy*, che presuppone l'abilità di pensare in maniera critica rispetto concetti, obiettivi e argomenti che riguardano i dati.

Nel 2010 e nel 2013 altre due ricerche utilizzano il verbo "*to understand*" nella loro definizione; in particolare Qin e D'Ignazio parlano di abilità di capire, usare e gestire dati scientifici, mentre Mandinach e Gummer di capire e utilizzare i dati per prendere decisioni. La seconda come avrete notato aggiunge un elemento fondamentale, quello della funzione ultima nella *data literacy*, ovvero la presa di decisioni. Questi due autori oltre specifiche tecnico-conoscitive di chi utilizza i dati, si soffermano sulla loro capacità di trasformarli in "*actionable knowledge*". Questo riguarda una parte finora trascurata, che riguarda i risultati e il fine dei dati: sviluppo di ipotesi, identificazione dei problemi, interpretazione, pianificazione, monitoraggio delle azioni.

Calzada Prado e Marzal nel 2013 utilizzano una definizione che Koltay descrive come comprensiva delle precedenti:

*"data literacy enables individuals to access, interpret, critically assess, manage, handle, and ethically use data. Managing, that appears in this definition, comprises preservation and curation, and this definition is much more comprehensive than those above"*²⁴

Prado e Marzal aggiungono un nuovo elemento, l'utilizzo etico dei dati.

Koltay inoltre prova a categorizzare le definizioni sulla base dei diversi concetti che esprimono, individuando così la *data information literacy*,²⁵ *science data literacy*²⁶ e *research data literacy*.²⁷ Nella ricerca si sottolinea in particolare il fatto che a questi diversi concetti che sottendono tre diverse definizioni, corrispondono competenze e abilità diverse (non soltanto tecniche).

Nel suo report Koltay presenta anche una propria definizione di *data literacy*, che è la seguente:

*"data literacy as a specific skill set and knowledge base, which empowers individuals to transform data into information and into actionable knowledge by enabling them to access, interpret, critically assess, manage, and ethically use data"*²⁸

La qualità della ricerca sta anche nell'investigare la relazione tra *information literacy* e *data literacy*, individuando l'importanza di quest'ultima, al di là del fatto che essa venga considerata indipendente o meno rispetto alla *information literacy*. Inoltre, proseguendo altri concetti emergono da ulteriori definizioni, ad esempio qualità dei dati, scoperta, acquisizione, repository, sintesi, combinazione, produzione e riuso e ancora qualità. Tracciamento della provenienza dei dati, contesto, autenticità e affidabilità.

²³ (Koltay, 2017)

²⁴ (Calzada Prado J, Marzal M. A., 2013)

²⁵ (Carlson J, Fosmire M, Miller CC, et al., 2011)

²⁶ (Qin J, D'Ignazio J, 2010)

²⁷ (Shneider, 2013)

²⁸ (Koltay, 2017)

*“A comprehensive definition was provided that emphasized the importance of accessing, interpreting, critically assessing, managing, handling, and ethically using data”*²⁹

Anche un altro autore ha scelto di collezionare tutte le definizioni di data literacy a partire dal 2004, si tratta di Crusoe. La definizione che trae al termine della sua ricerca è la seguente:

*“the knowledge of what data are, how they are collected, analysed, visualized and shared, and is the understanding of how data are applied for benefit or detriment, within the cultural context of security and privacy”*³⁰

Ovvero la conoscenza di quello che i dati sono, di come vengono raccolti, analizzati, visualizzati e condivisi. È la comprensione di come i dati vengono applicati a beneficio o a scapito, all'interno del contesto culturale di sicurezza e privacy.

Anche questo autore sceglie una definizione omnicomprensiva del termine. La gestione di questi dati vale quanto le loro caratteristiche e il loro utilizzo. A partire da quest'ultima definizione la ricerca PoliVisu trae uno spunto importante, che determina alcune scelte. Crusoe sottolinea la differenziazione che può derivare dalla materia, dalle problematiche e dall'impostazione che si prendono in considerazione. Questo significa che il valore dei *big data* dipende dallo specifico problema che si prende in considerazione. Nel caso del report *Experimental dimension of policy*³¹, la questione riguarda la pianificazione della mobilità pubblica, ma anche i soggetti di un eventuale progetto: cittadini, imprese e istituzioni pubbliche. È a partire da questi tre gruppi e dalla loro prospettiva che la letteratura accademica ha definito la data literacy.

In particolare, la definizione che riguarda le istituzioni pubbliche, la categoria sulla quale si concentra questa tesi, è stata fornita nel 2017 dall'*OECD* che ha sviluppato una ricerca sulle competenze significative per un *“high performing civil service”* (servizio civico ad alto rendimento). Secondo l'*OECD* dalla prospettiva pubblica, *data literacy* significa:

*“means that, wherever possible, decisions should be based on data not hunches or guesses. Data isn't just for 'geeks', non-specialists must understand its importance”*³²

Ovunque sia possibile le decisioni dovrebbero essere basate su dati, non intuite o supposte. I dati non sono soltanto per *Geeks*, i non specialisti devono comprendere la loro importanza. Questo punto di vista determina un risvolto importantissimo sulla mia ricerca. I risultati del questionario dovranno essere valutati tenendo conto di questo presupposto: chiunque all'interno di un comune o di una pubblica amministrazione deve comprendere l'importanza dei dati, cosa che non è affatto scontato in un contesto lavorativo pubblico.

²⁹ (Koltay, 2017)

³⁰ (Crusoe, 2016)

³¹ (Paola Pucci, Giovanni Vecchio, Grazia Concilio, 2018)

³² (OECD, 2017)

2. I Big Data nella Pubblica Amministrazione

Il valore dei dati sottolineato nel primo capitolo è tale anche quando si parla di dati delle pubbliche amministrazioni, di comuni e di governo del territorio. Negli ultimi anni si è utilizzato molto il termine *smart city*, in particolare dal 2015 l'interesse per il termine è cresciuto notevolmente. Questo termine si riferisce da un lato all'utilizzo dei dati e delle applicazioni digitali e dall'altro a un territorio, quello urbano della città. La città è il banco di prova dove riutilizzare i dati per migliorare concretamente la vita dei cittadini. La dimensione locale è fondamentale per la generazione di valore, non solo con fini economici, nell'intento di rendere la città più vivibile. Nelle metropoli il tentativo di trasformazione in *smart city* è già cominciato, con diverse modalità e velocità il processo è avviato e continua a progredire. Nei comuni minori, questo processo è più difficile. Sembra chiaro però che la scala migliore in cui ricercare questo fenomeno è quella comunale. Questo capitolo cerca di indagare le motivazioni per cui si effettua la ricerca al livello istituzionale dei comuni.

2.1. I dati comunali

La mole di dati che tutti i giorni transita attraverso i server di un qualsiasi comune, è enorme. Tutti i dati raccolti durante la presentazione delle pratiche, da quelle edilizie fino a quelle anagrafiche, servizi sociali, erariali, tributarie, durante lo svolgimento di qualsiasi attività al servizio di cittadini e professionisti. Dati dei cittadini che entrano nei comuni, e che potrebbero essere elaborati funzionalmente alla creazione di valore. Le grandi aziende raccolgono dati e li riutilizzano per la creazione di valore o addirittura li rivendono. Le pubbliche amministrazioni posseggono moltissimi dati dei cittadini che hanno il dovere non solo di custodire ma anche di utilizzare correttamente nel prendere decisioni politiche, in tutte le sfere dell'agire di un comune o di un ente pubblico, ma non solo. Proprio per la natura pubblica di questi dati, e per la natura stessa delle pubbliche amministrazioni, al servizio dei cittadini e del bene comune, questi dati (opportunamente ripuliti in funzione delle normative sulla privacy) devono essere pubblicati in un formato aperto; come *open data* appunto. Tutto questo permetterebbe la generazione del valore latente dei dati, anche da parte di imprese private esterne, che li utilizzano per la creazione di applicazioni, per il miglioramento dei servizi per la creazione di valore, oltre che per rendere più trasparenti i processi amministrativi. Migliorare la fruibilità dei servizi, l'interoperabilità delle Istituzioni, la burocrazia, la trasparenza stessa di tutto ciò che avviene in un comune e che influenza fortemente le nostre vite di cittadini.

Senza la disponibilità e la capacità dei comuni, soprattutto dei piccoli comuni, a riutilizzare i dati nei processi interni, senza una appropriata conoscenza dell'argomento, la *data literacy* appunto, non è possibile liberare quel valore nascosto dei dati posseduti dalle pubbliche amministrazioni. Ogni aspetto relativo alla pianificazione urbanistica è strettamente connesso con l'apertura dei dati, in particolare di quelli riguardanti il piano urbanistico come ovvio, ma non solo. Tutte le informazioni presenti all'interno di pratiche edilizie, permessi di costruire, appalti di opere pubbliche, varianti di piani, ma non solo, dinamiche sociali difficilmente rilevabili da chi sviluppa un piano se non attraverso i dati posseduti e raccolti quotidianamente dal comune interessato e che permettono, almeno in parte, di conoscere dinamiche dei territori. Più spesso succede che i dati presenti nei bilanci nei DUP, nei documenti e nelle pratiche archiviate dai comuni, rimangano ad appannaggio dei tecnici comunali, che, spesso oberati di lavoro, nemmeno li utilizzano. Ma le potenzialità di utilizzo dei dati si sono moltiplicate oltremodo e non si limitano all'utilizzo in funzione della predisposizione di un piano urbanistico (cosa che si conciliava solo con una visione statica del futuro).

Le vere protagoniste di questa ricerca sono le Pubbliche Amministrazioni con particolare riferimento ai comuni italiani. Essi rappresentano le istituzioni pubbliche più vicine al cittadino, all'interno delle quali tra l'altro, l'urbanistica e il governo del territorio vengono sviluppate, e che quindi vanno considerate con estrema importanza dalla disciplina della nostra facoltà. Essi agiscono spesso con difficoltà, ma al tempo stesso con maggiore incidenza (e possibilità di controllo) rispetto a tutti gli altri livelli deputati alla pianificazione e al governo del territorio in Italia.

Il livello locale dei comuni offre uno sguardo ravvicinato dei processi pubblici, delle comunità locali, delle associazioni presenti sul territorio e dei cittadini stessi. Per questo ha una fondamentale importanza leggere il fenomeno dei *big data* e la *data literacy* a questo livello. Nei comuni i *big data* possono contribuire concretamente alla ridefinizione dei servizi per poter rispondere in modo efficiente ed efficace alle esigenze dei cittadini. Il comune ha la possibilità di coinvolgere i cittadini o le associazioni direttamente nella definizione delle politiche e dei progetti locali, e ciò può farlo anche basandosi sui dati. L'innovazione a questo livello è molto differente rispetto a quella portata avanti da regioni e stato, e in questo il contatto diretto con la popolazione è fondamentale. Anzi, in questo caso la scala del comune consente diverse opportunità, in comuni piccoli questo è ancora più vero. L'innovazione dei comuni potrebbe generare un nuovo rapporto tra residenti cittadini e amministratori locali, che oggi purtroppo si limita ad uno scambio diretto di accuse o plausi tramite i social network.

Il dibattito vero interno ai comuni è venuto meno, insieme all'anti-politica e allo scarso interesse della disciplina pubblica si è innescato un meccanismo che limita il dibattito cittadino. Parlo in questo caso con riferimento diretto al piccolo comune da cui provengo. Scarso interesse e scarsa partecipazione da parte dei cittadini, non solo dei giovani. L'innovazione dovrebbe riguardare anche i processi di generazione dell'agenda politica e della presa delle decisioni, a sua volta supportate dall'utilizzo dei dati, meglio big data. Partecipazione e collaborazione dei cittadini, è a questo che bisogna puntare. Un processo innovativo dovrebbe puntare a incorporare le idee provenienti dai cittadini e dalle associazioni che vivono il territorio, all'interno di un processo che insieme ai dati permetta di costruire politiche pubbliche adeguate, indirizzando sui binari corretti l'agenda politica in funzione di risolvere problemi di pubblica utilità. La *data literacy* in questo orizzonte si amplia, oltre a quella dei comuni e quindi dei dipendenti e degli amministratori, a quella dei cittadini. Ovviamente se tutti quanti sono consapevoli del valore reale che è possibile generare dai dati, politiche digitali adeguate saranno ben volute e implementate di conseguenza. Se l'interesse viceversa non esiste, allora quale scopo nel generare politiche *data-driven*, quando nessuno controlla, nessuno ha interesse che questo valore emerga. Possiamo continuare a riempirci la bocca di termini digitali inglesi, in particolare di smart city (termine presente nella maggior parte dei programmi politici cittadini) senza un reale interesse ad utilizzare i dati.

Lavorare con le pubbliche amministrazioni potrebbe significare lavorare con e per il bene comune, per conto e insieme ai cittadini e per il nostro territorio. L'apertura dei dati, la trasparenza delle procedure, la tensione al progresso e all'innovazione tecnologica (insite negli argomenti stessi) potrebbero già di per sé dare un significato, all'utilizzo concreto dei big data. In tal caso però il testo dovrebbe riguardare maggiormente gli effetti benefici dell'apertura dei dati, piuttosto che la *data literacy* delle pubbliche amministrazioni. Di fatto queste due tematiche sono fortemente in relazione tra di loro. Ricerche che affrontano l'effetto e le potenzialità dell'apertura dei dati della pubblica amministrazione sono numerose, non sarà questo l'oggetto delle analisi successive. Per tanto l'accento cade sulla *data literacy* dei comuni italiani, questione niente affatto secondaria. La loro capacità di raccogliere, elaborare, analizzare e restituire i dati in forma di politiche territoriali è cruciale, come si può intuire dai discorsi riportati nel primo capitolo.

I servizi offerti da un comune sono quelli più tangibili per un residente, incidono direttamente sulla vita, riguardano aspetti concreti e pratici della vita di tutti i giorni. Per questo l'utilizzo di big data nella definizione dei servizi si applica direttamente sulla vita dei cittadini.

2.2. Data-driven city (o comuni)

Il libro “*Beyond Transparency*”, offre questo esempio chiaro dell'uso dei dati per impostare una politica in collaborazione con cittadini e associazioni locali. Gli abitanti del quartiere Tanderloin insieme all'associazione GAFFTA hanno installato sensori di rilevamento dei rumori per raccogliere dati a supporto della loro richiesta di abbassamento del livello di rumore nel quartiere. Solo grazie ai dati raccolti i cittadini hanno potuto dimostrare che i rumori superavano il livello ammissibile, e questo ha permesso al comune di San Francisco di rilasciare permessi di costruire speciali, che limitassero gli orari di lavoro delle imprese di costruzioni che erano presenti nell'area. Questo utilizzo dei dati ha una evidente implicazione diretta sulle abitudini e sulla qualità della vita dei cittadini. I dati possono quindi essere utilizzati per sostenere decisioni politiche.

Ovviamente questo esempio riguardava un dataset raccolto e utilizzato direttamente dagli stessi cittadini. La pubblica amministrazione possiede, raccoglie dati, che anche se non resi pubblici possono essere utilizzati internamente nella definizione di politiche. Ovviamente in questo caso mancherebbe una evidenza pubblica, ma per ora manterrò separata la versione open data dall'utilizzo generale dei big data nei comuni. Certamente l'apertura aumenta notevolmente le possibilità di riutilizzo.

Come ammesso in precedenza, la presente tesi si sofferma maggiormente sulle implicazioni interne ai comuni e non sulle svariate possibilità che gli open data aprono anche al resto della società. La *data literacy* interna è un discorso ampio affrontato già da solo. Spero di rendere il più chiaro possibile i motivi per cui questa ricerca si focalizza sulle pubbliche amministrazioni e in particolare i comuni. Considero i comuni lo snodo fondamentale, il punto d'incontro tra la definizione delle politiche e le esigenze pubbliche, istanze dei cittadini, delle associazioni ma anche delle imprese private locali e non. Perciò l'utilizzo dei dati a livello comunale è una questione cruciale, sia per i cittadini che per il mondo della pubblica amministrazione. Anche perché una mole incredibile di informazioni (big data passano esattamente da qui, ed è all'interno dei comuni che andrebbero raccolte e collezionate prima di tutto.

Da un lato l'apertura dei dati per permettere lo sviluppo e la creazione di valore da parte di terzi, dall'altro (lato di cui si occupa questa tesi) la possibilità di sfruttare questi dati internamente per migliorare i processi interni, decisionali, le politiche e l'allocazione delle risorse.

*“Being a data-driven city is about more efficiently and effectively delivering the core services of the city. Being data-driven is not primarily a challenge of technology; it is a challenge of direction and organization leadership”.*³³

Secondo quanto spiega Michel Flowers capo dell'ufficio di analisi della città di New York, nel capitolo “*Data-driven city*” del libro *Beyond Transparency*, una caratteristica aggiuntiva dei big data è la geolocalizzazione dei dati. Inoltre, spiega che esiste una grossa differenza tra utilizzare i big data intesi soltanto in funzione della loro quantità (volume) ed essere *data-driven*. Collezionare i dati sul traffico in un file CSV non è utile di per sé. Questi dati possono diventare utili se utilizzati dai pianificatori, che attraverso la loro rielaborazione e conversione in mappe li utilizzano per ridisegnare il traffico. L'effettivo valore esiste se si hanno dei risultati. Il futuro è

³³ (Michael Flowers, Beth Blauer, 2013)

data-driven, questo è chiaro e ampiamente documentato dal primo capitolo, ma per esserlo è necessario che i comuni abbiano un alto livello di *data literacy*. Secondo Michel Flowers le istituzioni *data-driven* per eccellenza sono rappresentate dalle grandi metropoli americane. A proposito di questo si pone un tema, approfondito nella sezione di statistica, che riguarda l'utilizzo dei dati da parte delle grandi città e il confronto con la maggior parte dei comuni, piccoli o molto piccoli italiani, in cui più difficilmente si potrà parlare di *data-driven*. Questo coincide con un periodo in cui le città (e questo è ancora più vero in Italia) hanno sempre meno risorse a disposizione per offrire servizi adeguati ai propri cittadini. È indispensabile in questa circostanza trovare il modo per migliorare l'allocazione delle risorse e migliorare processi e sistemi decisionali dei comuni.

La definizione di una città (o nel nostro caso di un comune) *data-driven* ci è fornita sempre da Michel Flowers ed è la seguente:

*“A data-driven city is a city that intelligently uses data to better deliver critical services. Transparency, open data, and innovation are all important parts of the modern civic identity, [...] being a data-driven city is really about more efficiently and effectively delivering the core services of the city: smarter, risk-based resource allocation, better sharing of information agency-to-agency to facilitate smart decision-making, and using the data in a way that integrates in the established day-to-day patterns of city agency front line workers. Being data-driven is not primarily a challenge of technology; it is a challenge of direction and organizational leadership”.*³⁴

Quello che serve in un comune per poter risolvere dei problemi oppure per adottare politiche funzionali al cittadino, insomma per essere *data-driven*, sono i dati. Da soli però non sono abbastanza come abbiamo visto. Ottenere i dati riguarda solo la primissima fase del processo, ovvero la raccolta. Le modalità, gli strumenti e le possibilità di raccolta sono state ampiamente documentate nel primo capitolo. Possedere i dati per i comuni è reso sempre più semplice, così come la raccolta. L'archiviazione è fondamentale per permettere un corretto e rapido utilizzo nelle fasi successive. Altri dati provenienti da terzi sono certamente utili comunque.

I dati oltre ad essere collezionati hanno bisogno di essere collegati. Il collegamento dei dati è estremamente funzionale all'interno di una pubblica amministrazione. La strutturazione in uffici e settori permette la raccolta di dati differenti, ma questo non è sinonimo di un'integrazione dei settori e dei dati relativi tra loro. È necessaria una nomenclatura comune che permetta la perfetta integrazione, ovvero il collegamento di dati tra loro differenti. Spesso è impossibile effettuare questa operazione infatti i settori nel tempo si sono specializzati fornendo specifici nomi e modalità di collezione dei dati. In questi casi esistono dei dati di collegamento che permettono l'integrazione di diversi data set, come ad esempio i dati geografici. Spesso questi ultimi servono da tramite rispetto a istanze diverse riguardo diversi uffici e competenze.

L'integrazione dei dati è molto importante perché permette una maggiore incidenza nella risoluzione dei problemi, che non sono mai mono settoriali. La prima creazione di valore riguarda proprio l'integrazione dei dati a partire dai diversi dataset collezionati.

Per integrare quanto detto fin ora inseriamo un paragrafo che studia la presenza di dataset all'interno dei vari comuni. Il numero dei dati sarà influenzato dalla popolazione o piuttosto dalla regione di provenienza? L'approfondimento è stato proseguito nel capitolo di statistica.

“Analytics is not magic, and it's not necessarily complicated. Analytics really means intelligence, and intelligence is better information that helps us make better

³⁴ (Alissa Black, Rachel Burstain, 2013)

decisions. To the extent that we can automate that information gathering and analysis, for instance, in automatically of the approach. The most important thing to remember, however, is that we are not changing the approach.”

“An effective analytics project is one that gets in and gets out sight unseen. Let the results speak for the project.”³⁵

2.2. Database comunali.

Ogni comune italiano deve dotarsi di sistemi digitali adeguati all’esercizio dei servizi che ha il compito di garantire. L’agenda digitale ha stabilito l’obbligo di presentazione delle pratiche telematicamente. A proposito dei dati posseduti dai comuni, è stato utilizzato il servizio offerto dal nuovo catalogo “Basi Dati della pubblica amministrazione” creato da AGID, per provare a fare alcuni ragionamenti sugli effettivi dati che transitano da un comune. In questo catalogo recentemente pubblicato, vengono raccolti tutte le basi di dati (e gli applicativi che le utilizzano) della pubblica amministrazione. L’operazione è stata svolta in attuazione dell’art. 24-quater, comma 2, del D.L. n.90/2014, che prevedeva:

“Entro trenta giorni dalla data di entrata in vigore della legge di conversione del presente decreto, i soggetti di cui all’art. 2, comma 2, del codice di cui al decreto legislativo 7 marzo 2005, n. 82, e successive modificazioni, comunicano all’Agenzia per l’Italia digitale, esclusivamente per via telematica, l’elenco delle basi di dati in loro gestione e degli applicativi che le utilizzano.”

Quindi il catalogo contiene le informazioni comunicate dalle stesse amministrazioni, aggiornate al 25/05/2015. Attualmente il catalogo è composto dai dati comunicati da 13.822 amministrazioni, per un totale di 159.725 dataset.³⁶ Il maggior numero di basi di dati appartiene proprio ai comuni e loro consorzi e associazioni, voce che comprende 81.118 basi di dati, oltre la metà di tutti i database della pubblica amministrazione.

Questa statistica contiene il numero di DB di 6.676 comuni italiani. Ai primi posti troviamo le principali metropoli, in particolare elencherò i comuni con oltre mille basi di dati a partire dal primo classificato: Roma, Torino, Milano, Palermo, Napoli, Firenze, Genova. Da sottolineare il fatto che l’archivio è disponibile in formato aperto, sia come testo che come database in formato open source.

Ovviamente non è detto che tutte le informazioni contenute nei vari database comunali vengano utilizzate nei processi interni, o nella definizione delle politiche, o nei vari altri modi che sono stati presentati. Certamente alcuni di questi sono messi a disposizione in formato open data all’interno di portali, molto spesso questo accade tramite le regioni. Per questo un ulteriore passo affrontato è stato quello di provare a dare un livello alla struttura informatica delle regioni Italiane. Il grado di “informatizzazione” della regione è stato applicato solo in base ad un criterio. Purtroppo, non esiste una ricerca a livello italiano che prova a vagliare la capacità delle regioni di raccogliere, creare banche dati, analizzare i dati. Questo dovrebbe essere fatto in una forma simile alla ricerca di *Open Data Maturity* effettuata tutti gli anni dall’European data Portal.³⁷ Basandosi su quattro criteri il report fornisce un livello annuale per ogni stato dell’unione. I quattro criteri utilizzati sono i seguenti: *Policy*, che valuta la presenza di un quadro politico, un coordinamento nazionale, norme che regolano le licenze; *Impact*, l’impatto dal punto di vista strategico, politico,

³⁵ (Michael Flowers, Beth Blauer, 2013)

³⁶ (AGID, 2019)

³⁷ (Cecconi Gianfranco, Radu Cosmina, 2018)

ambientale, sociale ed economico; *Portal*, che analizza le caratteristiche, l'utilizzo e la sostenibilità dei portali; *Quality*.

Una ricerca più approfondita dovrebbe basarsi sugli stessi criteri utilizzati dal report *Open Data Maturity*, per mancanza di tempo è stato analizzato solamente il criterio "Portal" analizzando i portali open data delle regioni italiane, individuando il numero di dataset pubblicati e il formato con cui sono caricati i dati. Di seguito viene riportata la tabella riassuntiva con le regioni e il numero di dataset pubblicati. Trentino la regione con maggiore disponibilità di dati, seguita da Lombardia e Toscana. Fondo classifica ospita Calabria e Molise, di cui non si riesce a scovare nemmeno un portale open data.

#	Ente regionale	Linee Guida	Geoportale	Numero Dataset pubblicati
1	Trentino	Yes	Yes	6.690
2	Lombardia	Yes	Yes	5.117
3	Toscana	Yes	Yes	3.517
4	Emilia-Romagna	Yes	Yes	1.263
5	Piemonte	Yes	Yes	922
6	Sardegna	Yes	Yes	635
7	Liguria		Yes	555
8	Veneto		Yes	505
9	Lazio		Yes	371
10	Umbria	Yes	Yes	362
11	Friuli-Venezia Giulia	Yes	Yes	356
12	Basilicata	Yes	Yes	191
13	Abruzzo		Yes	162
14	Puglia	Yes	Yes	111
15	Campania	Yes	Yes	73
16	Sicilia		Yes	38
17	Marche		Yes	32
18	Valle d'Aosta	Yes	Yes	15
19	Calabria			0
20	Molise			0

Tabella 1 - Regioni per numero di dataset pubblicati nei portali OD

A *data-driven city* (o comune più in generale) si affianca il termine *data-driven government*. L'innovazione civica necessariamente dovrà passare tramite questa fase di utilizzo dei dati nel processo di indirizzo delle scelte politiche. Certo tutto ciò deve avere un metodo per poter funzionare. Secondo Beth Blauer un buon metodo che permette l'utilizzo dei dati per prendere decisioni migliori e coinvolgere la comunità è il seguente: Curare i dati e informare la popolazione a proposito del governo, permettere ai developers l'accesso ai dati, creare un ecosistema florido, creare un ambiente collaborativo in cui gli analisti parlano con governo e sviluppatori facendo comprendere a tutti la situazione reale per poter selezionare obiettivi concreti.

Nel primo capitolo è stato dimostrato l'importanza che i dati e in particolare i big data hanno assunto oggi, con possibili benefici a pioggia in tutte le discipline, non soltanto in quelle strettamente legate al digitale. Certo, per liberare le potenzialità e il valore da questi dati è

necessarie elaborazioni, integrazioni, monitoraggio, insomma, tutte quelle operazioni che sono state spiegate nel precedente capitolo.

Questo secondo capitolo cerca di applicare quello che è stato detto fin ora su tutti gli aspetti che regolano i comuni. Il comune è l'istituzione più vicina al cittadino, il livello più concreto e anche pratico, dove a mio parere è necessario (ma forse anche più utile) incidere attraverso l'uso di big data. Questi dati sono sempre più a disposizione, anche dei piccoli comuni, è sempre meno difficile disporne, raccogliarli attraverso sensori e utilizzarli nei propri processi politici e amministrativi. Il questionario cerca di indagare anche se esiste un interesse a proposito, o se spesso si traduce puramente adempimenti burocratici.

I servizi che i comuni offrono ai cittadini sono molteplici e comprendono moltissimi settori. Infrastrutture, acquedotti, sistema fognario, strade, opere pubbliche, edilizia privata, urbanistica, trasporto pubblico, educazione, servizi sociali, servizi medici di base, servizi agli anziani, sicurezza (estremamente attuale come tema), servizi pubblici al cittadino in generale.

Oltre all'elenco dei servizi ai quali è possibile estrarre valore applicare l'uso dei big data è necessario ripercorrere le fasi del KDD per capire come i big data possono essere integrati nei processi comunali, quali parti sono già funzionanti e avanzate per i comuni, quali sono i processi in cui sono più arretrati, dove è possibile e più importante agire per creare valore.

Il paper "*Data mining applications in public Organization*" i servizi principali che offre la pubblica amministrazione vengono raccolti in sette branche: economia e finanza, salute, criminalità e giustizia e difesa, lavoro e welfare sociale, E-government, educazione, trasporti.

Molto evidente se si parla di grandi città, dove il termine smart cities diventa onnipresente, in cui l'idea di data driven cities è consolidata. Nelle metropoli è alquanto evidente, l'uso di big data è imprescindibile per coloro che ricercano o che affermano di volere una smart city. Data driven city è un concetto più ampio. Architettura, ingegneria, urbanistica tutte queste discipline interessate. La presa delle decisioni

Nonostante ricerche importanti europee vedano l'Italia nei primi posti per quanto riguarda il livello di open data, l'apertura dei dati – con particolare riferimento al portale dati.gov.it – la sensazione che comune in Italia, è quella di lavorare con amministrazioni poco all'avanguardia dal punto di vista tecnologico e informatico e ciò è lampante nelle piccole realtà, cioè la maggioranza. Con dipendenti pubblici dall'età media molto elevata, carichi di esperienza, ma scarsamente abituati a lavorare con tecnologie informatiche, pare che tale piazzamento sia dovuto a politiche AGID sostenute a livello statale oppure a livello regionale. Parentesi rispetto alla presenza di dipendenti "con esperienza" all'interno dei nostri comuni, ai quali non saranno affiancati giovani colleghi (come avveniva in passato) e che facilmente andranno in pensione senza trasmettere il bagaglio di esperienze pratiche, che manca ai giovani e che più facilmente siamo in grado di adeguarci al contesto tecnologico che investe questo mondo e che inevitabilmente deve investire anche le amministrazioni pubbliche. Questa tendenza del resto si registra anche in aziende private, in Italia nel 2019 poca fiducia nei giovani. Insomma, le capacità informatiche quasi del tutto assenti nei comuni di medio-piccole dimensioni (la maggior parte) sono le più necessarie per una alfabetizzazione digitale e dei dati, appunto la data literacy. Questo pare ovvio, ma non lo è. Ricerche come la precedentemente citata "*Experimental dimension of the policy*", non affidano così tanta importanza alle capacità tecniche dei singoli, o meglio, credono che l'assenza di tecnici non determini automaticamente l'assenza di *data literacy* nelle PA, come viceversa la presenza di tecnici informatici preparati non garantisce alti livelli di *data literacy*.

Ad esempio, le amministrazioni che affidano i propri servizi digitali ad aziende private, oppure Regioni (che spesso assumono il ruolo di gestore dei servizi digitali) con dipartimenti in grado di gestire e fornire applicazioni complete che il dipendente utilizza quotidianamente e dalle quali

estrae, in maniera ordinata e quasi automatica, i dati che interessano la propria mansione e le attività dell'intera PA.

Per questo motivo è stato importante studiare i portali open data delle singole regioni, per determinare un livello, quantomeno generale, dei servizi informatici offerti ai propri comuni. Dal numero di dataset pubblicati è stato possibile determinare, grosso modo, quanti e quali dati i comuni raccolgono per conto della regione. Certamente si tratta di una valutazione generale e forse anche parziale, che non tiene conto della normativa regionale di riferimento e delle reali pratiche e regole che i comuni devono seguire nell'archiviazione dei dati. Anche le applicazioni fornite dalla regione potrebbero determinare un fattore importante per determinare il loro grado di digitalizzazione. Fattore importante ma non determinante, infatti ogni comune (o unioni di comuni) può stipulare convenzioni con imprese private, software house all'avanguardia dal punto di vista IT, per la dotazione di applicativi dedicati. Ma l'introduzione di società esterne complica ulteriormente il quadro, le possibilità e le varianti da tenere in considerazione. Perciò, ritengo i portali open data, insieme alla qualità e alla quantità dei dataset pubblicati un importante indicatore, esempio calzante in riferimento ai *data as oil*. La pubblicazione di questi dati infatti, si presuppone posta al termine di quelle operazioni che hanno raffinato il petrolio grezzo, i dati, per permetter loro di acquistare valore ed essere riutilizzabili. Ovviamente stiamo parlando di quel valore intrinseco che è stato delineato in precedenza. Concludendo, credo che la valutazione del livello regionale di servizi digitali offerti al comune, possa essere approfondita ulteriormente e che questo sia soltanto un primo tentativo. Non sono state trovate informazioni più dettagliate, di quelle presentate nel relativo capitolo, raggruppate in una sola ricerca di sintesi (evitando lo studio di una singola regione alla volta).

2.3. Risorse

Le risorse interne dei comuni vanno intese da due punti di vista: economico e della competenza. In questo capitolo il tema delle risorse viene associato alla potenzialità nell'utilizzo di big data e dei dati più in generale.

Risorse importanti perché permettono l'elaborazione delle informazioni e la trasformazione in politiche. Emerge chiaramente la differenza tra le grandi metropoli e i piccoli comuni. Sostanziale differenza il tema delle competenze.

Per questa analisi viene fatto particolare riferimento al capitolo "Public sector skills in the search of public value" del report dell'OECD sulle competenze dei dirigenti e dei dipendenti pubblici.

*"Professional civil services are as important as ever to respond to complex challenges and to deliver public value."*³⁸

I dipendenti pubblici sono più importanti che mai per rispondere adeguatamente alle sfide complesse e per restituire valore al pubblico. Il report si focalizza su tre caratteristiche fondamentali del dipendente o del dirigente pubblico. La professionalità è una qualità necessaria ma non sufficiente, è utile acquisire una visione strategica soprattutto a livello gestionale. Queste due caratteristiche hanno un limite causato dai vincoli legali del sistema organizzativo pubblico, per questo la terza caratteristica è l'innovazione.

Le competenze del settore pubblico per la creazione di valore a partire dai dati sono riassunte in quattro punti, che corrispondono a quattro aree, ciascuna rappresentante una mansione o una competenza specifica.

³⁸ (OECD, 2017)

- *Develop policy.* Lo sviluppo di policy basate sui dati e perciò basate sull'evidenza, significa avere la capacità di gestire problemi complessi. Se da un lato la politica e le amministrazioni elette determinano l'indirizzo politico da seguire, dall'altro anche uffici e personale hanno il compito di saper bilanciare obiettivi e determinare evidenze per le scelte politiche da attuare.
- *Commission and contract.* Punto di contatto tra chi fornisce il servizio e il cittadino.
- *Work with citizens.* Collaborare con i cittadini
- *Collaborate in Network.* Collaborare in rete, insieme ad altre amministrazioni, condivisione.

Ai dipendenti pubblici viene conferito un ruolo cruciale in questo report, proprio per questo, in virtù dei grandi cambiamenti “digitali” che investono le pubbliche amministrazioni è necessario un rapido adattamento e un continuo aggiornamento delle conoscenze.

*Identifying the skills needed is a first step towards developing a fit-for-purpose civil service for the twenty-first century. Building this civil service requires a new look at the way people are managed; one that recognises that public employees are neither homogenous nor mutually interchangeable.*³⁹

La sfida secondo l'OECD è anche quella di determinare quali saranno realmente le competenze richieste da un dirigente o dipendente pubblico nel prossimo futuro, quali sono le mancanze principali della pubblica amministrazione di oggi. In questo caso la definizione della data literacy potrebbe fornire una risposta. Attrarre e selezionare le giuste figure professionali con competenze innovative. Non solo attrarre dall'esterno, ma avere la capacità di coltivare le competenze interne, formare figure professionali adatte alle sfide che si pongono. Infine, si ragiona in funzione della definizione di un'organizzazione pubblica capace sia di formare, ma anche di ricollocarle nel giusto ruolo.

Un'altra ricerca importante nel campo delle competenze nella pubblica amministrazione è stata pubblicata da Sopra Steria per due anni consecutivi che riguarda i dipendenti pubblici del Regno Unito. La ricerca propone un'indagine del *ternd* digitale della pubblica amministrazione britannica ed è svolta attraverso un questionario compilato dai dipendenti pubblici britannici. Nel 2016 le risposte sono state 1.235 di cui 313 di dipendenti direttamente coinvolti in mansioni di trasformazione digitale. In particolare, quello che emerge da questo ampio campione di risposte viene riassunto nei seguenti punti.⁴⁰

Per il 75% degli intervistati ammette che le trasformazioni digitali hanno influenzato il loro modo di lavorare. Il 71% pensa che il mondo del digitale stia determinando un cambiamento in base a come i servizi sono forniti. Il 64% di loro ha fiducia che l'organizzazione in cui lavorano abbia una buona conoscenza dei servizi all'utente.

Poco più della metà (il 53%) dice di non avere competenze digitali adeguate e di non ricevere corsi di formazione per colmare questa lacuna. Il 43% di loro utilizza dati sul comportamento degli utenti per sviluppare servizi adeguati. La questione della competenza e della mancanza di formazione rispetto alle tematiche digitali risulta essere la barriera più importante alla trasformazione digitale. Seconda risulta essere la mancanza di risorse economiche a supporto della trasformazione.

In questa analisi delle principali ricerche che hanno riguardato le competenze di dirigenti e dipendenti delle pubbliche amministrazioni si affianca una ricerca (citata anche dal report sulla *data literacy* di PoliVisu) che riguarda le competenze degli operatori statistici, focalizzata sulle capacità tecnico-informatiche degli addetti.⁴¹ Il report è stato sviluppato da *UN Global Working*

³⁹ (OECD, 2017)

⁴⁰ (Sopra Steria, 2016)

⁴¹ (Task Team on Skills, Training and Capacity Building, 2015)

Group on Big Data e fa parte di una grande ricerca che ha toccato molte sfere d'applicazione dei big data. Da essa emerge la mancanza di competenza tecniche adeguate alla gestione dei big data e all'analisi statistica. Le tre competenze più citate all'interno della ricerca sono competenze metodologiche, di *Data scientist*, Modellatori matematici e statistici. In tutti e 86 gli stati interessati da questa ricerca svolta nel 2014, emerge la necessità di “*Skills and training for Big Data*” ovvero di competenza e formazione all'utilizzo dei big data. Questa ricerca considera in particolare gli uffici di statistica nazionale, vengono tralasciati i dipendenti pubblici o comunali. Anche nei comuni la necessità di formazione e competenza è cruciale. Le due voci che seguono riguardano la necessità della qualità delle strutture big data e l'accesso alle fonti di big data (che spesso coincide con un fattore economico).

3. Metodo analitico e statistico

La domanda a cui il presente capitolo cerca di fornire una risposta è la seguente: come determinare il livello di alfabetizzazione dei dati dei comuni Italiani? Questa operazione è stata svolta in conformità e in continuità con la ricerca europea di PoliVisu, già citata nei precedenti capitoli. La modalità utilizzata si basa sull'invio di un questionario a cui hanno risposto dipendenti, responsabili o membri eletti dei comuni. Questa ricerca non entra nei particolari della selezione dei partecipanti, anzi rileva una grossa difficoltà in fase di reperimento delle risposte, dovuta alla scarsa partecipazione ottenuta. Oltre alle risposte al questionario, il deliverable "*Data literacy survey report*"⁴² ha effettuato delle interviste approfondite su alcuni dipendenti che avevano dato la disponibilità. Una differenza sostanziale rispetto a questa tesi è rappresentata dal target utilizzato. Infatti, questa tesi come già spiegato, intende focalizzarsi sui comuni, mentre il report compreso nella ricerca PoliVisu comprendeva tre diversi livelli di pubblica amministrazione, il primo comprende stato, regioni, città metropolitane; il secondo agenzie pubbliche, aziende partecipate; il terzo gli utenti dei servizi di trasporto e i dipendenti delle aziende di trasporto. Inoltre, il metodo utilizzato dalla precedente ricerca considera una popolazione, di 40 mila componenti, che comprende i tre gruppi a cui il questionario è stato inviato, stimando 100 risposte per considerarsi rappresentativo. La definizione del campione si basa su tre condizioni principale: campione selezionato in modo casuale, margine di errore dell'8% e una *confidence rate* del 90%. Il metodo di selezione del campione proposto è stato riutilizzato per la definizione del campione di comuni della presente tesi. Le limitazioni previste dal *report survey* della ricerca PoliVisu riguardano altre tre condizioni: campione selezionato inappropriato (i partecipanti al sondaggio si comportano diversamente dal target selezionato); dimensione del campione inadeguato (le 100 risposte sperate potrebbero essere insufficienti, troppo poche rispetto alla popolazione di 40.000); risposte mancanti o inappropriate ("*false positive*" partecipanti che non costituiscono una reale rappresentazione della situazione nella loro organizzazione).

Il questionario sull'utilizzo dei big data è stato inviato online, ma rispetto alla mia tesi, esiste una seconda fase che riguarda un approfondimento attraverso intervista che tenta di fare emergere significati nascosti rispetto al questionario. L'intervista telefonica di circa trenta minuti ha coinvolto 17 partecipanti, ed è stata impostata in modo simile alla struttura del questionario. Questo ha permesso di far emergere posizioni personali o specifici argomenti. La presente tesi non ha utilizzato questo metodo nella rielaborazione delle risposte, la domanda sulla disponibilità all'approfondimento tramite un'intervista telefonica è comunque presente all'interno del questionario. Eventuali sviluppi futuri dell'analisi dovrebbero tenere conto anche delle risposte affermative a quest'ultima domanda, valutando l'ipotesi di contattare chi si è reso disponibile all'approfondimento.

⁴² (Giovanni Vecchio, Pavel Kogut, 2018)

3.1. Questionario online.

La forma del questionario, la sua struttura e le domande da cui è composto sono molto simili e in molti casi identiche al questionario presentato all'interno della ricerca PoliVisu e riportato nel deliverable "*Data literacy survey report*"⁴³. Si è cercato di rendere il più leggero, semplice e veloce lo scorrere delle domande, snellendo il testo della domanda, ponendo le domande in seconda persona, togliendo la possibilità di selezionare alcune risposte. Alcune domande sono state tolte completamente, mentre non è stata aggiunta alcuna domanda ulteriore. A seconda delle risposte date il partecipante intraprendeva una diversa corrente di domande. Per studiare le soluzioni adottate nel dettaglio si rimanda al testo del questionario utilizzato nella tesi⁴⁴, oppure allo schema riassuntivo che non riporta le possibili risposte, ma soltanto le domande e il percorso alternativo da seguire in base alla risposta delle domande binarie (si/no).⁴⁵

Dopo il riadattamento, il questionario è stato caricato sulla piattaforma *Google form*, applicazione che permette di costruire un sondaggio online e parteciparvi tramite l'invio del link. Il questionario è stato denominato "La data literacy nelle pubbliche amministrazioni italiane". Ai comuni è stata inviata una mail con allegato il link di accesso al questionario on line. Le risposte ricevute sono state scaricate in formato CSV dalla piattaforma *Google form* e rielaborate in locale. Il link fornito ai comuni non sarà messo a disposizione per evitare la partecipazione di persone non facenti parte delle categorie analizzate. Se si desidera ottenere l'accesso contattare l'autore della tesi. Di seguito verrà messo a disposizione il link della copia del questionario:

<https://forms.gle/Emg4PVG4fdBNEzvg6>

Il questionario a cui si accede da questo link è una copia dell'originale contenente le risposte dei comuni. Pertanto, è possibile partecipare e inviare le risposte, non influiranno sul campione originale. Il testo del questionario è riportato al termine della presente tesi come allegato I. Per partecipare in qualità di dipendente, consulente, responsabile al questionario, contattare l'autore della tesi o alla relatrice.

3.2. Criteri per la definizione del campione

L'utilizzo di questi due criteri ha il suo fondamento nell'ipotesi che essi abbiano una influenza importante sui risultati del questionario. La selezione del campione omogeneo si è basata su di essi, di seguito vengono analizzati i due criteri. Il primo è stato la suddivisione dei comuni sulla base della popolazione residente al loro interno. Il secondo criterio si è basato su una suddivisione geografica relativa alle regioni italiane. Questo determina la suddivisione della penisola geograficamente e ottenere un risultato omogeneo e non concentrato in determinate regioni, cosa che potrebbe falsare i risultati o non renderli applicabili a tutta l'Italia, quindi parziali. Inoltre, permette una suddivisione tra comuni di grandi dimensioni, di cui fanno parte i capoluoghi regionali e le città metropolitane, e le realtà minori che rappresentano la maggioranza dei comuni italiani.

La popolazione comunale è stato il primo criterio ad essere analizzato. La questione fondamentale è stata raggruppare la popolazione in classi in modo da rendere i risultati gestibili e omogenei. L'esercizio, in fase di definizione, è stato svolto su diverse fasce di popolazione al fine di

⁴³ (Paola Pucci, Giovanni Vecchio, Grazia Concilio, 2018)

⁴⁴ Allegato I.

⁴⁵ Allegato II.

analizzare quali classe si adattassero in modo migliore alla statistica sui big data. L'ipotesi fondamentale nella definizione delle classi riguarda due aspetti: competenza e risorse interne ai comuni. Indagare la *data literacy* in comuni di poche centinaia di abitanti rispetto a una metropoli come Milano ha uno scarto fondamentale fortemente influenzato dai due aspetti appena citati. L'utilizzo dei big data è imprescindibile rispetto a una conoscenza del tema, a competenze adeguate (spesso a disposizione solo di grandi comuni) e al permesso di dedicare tempo e risorse economiche a questi temi.

L'indagine sui dipendenti delle pubbliche amministrazioni e dei comuni in particolare ha influenzato la definizione delle fasce di popolazione.⁴⁶ Questa analisi è stata integrata dai dati messi a disposizione dal MEF che riguardano i dipendenti a tempo indeterminato per comune.⁴⁷ Inoltre si è fatto specifico riferimento alla normativa, art.14, comma 27 del D.L. n.78/2010 prevede quanto segue:

L'esercizio obbligatorio in forma associata delle funzioni fondamentali da parte dei comuni con popolazione fino a 5.000 abitanti ovvero fino a 3.000 abitanti se appartengono o sono appartenuti a comunità montane avviene, alternativamente, mediante:

- unione di comuni
- convenzione*

** Le convenzioni hanno durata minima di tre anni e devono raggiungere i livelli di efficacia e di efficienza individuati con il decreto del Ministro dell'Interno 11 settembre 2013*

In questo caso i piccoli comuni in collaborazione potrebbero avere a disposizione più risorse e competenze per l'utilizzo dei big data.

I gruppi, così sono stati denominate le fasce di popolazione, sono cinque. Gli scaglioni di popolazione sono stati determinati come riportato dalla seguente tabella.

	Fascia di popolazione	Numero di comuni	Popolazione	Superficie [kmq]
Gruppo 1	0 – 3.000	4437	5.630.269	119.800
Gruppo 2	3.000 – 7.000	1.724	7.963.880	68.969
Gruppo 3	7.000 – 15.000	1.046	10.563.868	47.587
Gruppo 4	15.000 – 30.000	441	9.094.385	65.710
Gruppo 5	> 30.000	306	27.231.571	38.262
Totale	-	7954	60.483.973	340.328

Tabella 2 - Definizione dei "Gruppi" per scaglioni di popolazione

Il raggruppamento scelto permette anche di mantenere una crescita progressiva fino ai 30.000 abitanti. Oltre questa soglia le classificazioni si interrompono. Non sono stati determinati ulteriori gruppi, infatti nel gruppo cinque rientra una piccola minoranza di comuni, che se ulteriormente suddivisa non avrebbe ottenuto rappresentanza nel campione e quindi durante l'analisi. Questo avrebbe determinato un grave squilibrio nei risultati, come è facilmente intuibile analizzando i grafici che seguono e che mostrano la proporzione dei comuni in base ai gruppi appena stabiliti. Se si considera la popolazione, nel mantenere un campione omogeneo e proporzionale a questo criterio, non si otterrebbero risultati (o quasi) dal gruppo 5, come dimostrato nel capitolo 3.3, rischiando di non considerare i comuni in cui risiede quasi la metà della popolazione italiana. Il grafico sulla superficie fornisce una ulteriore statistica, ininfluenza nella definizione dei cinque gruppi.

⁴⁶ (W. Tortorella, G. Marinuzzi, F. Chiovoloni, T. Ulivieri, 2015)

⁴⁷ https://bdap-opendata.mef.gov.it/opendata/spd_pca_org_ann_minis_01_2017

Il criterio geografico chiaramente mira a rendere omogeneo il campione evitando concentrazioni di risposte in determinate aree della penisola. In fase di definizione del parametro si è tentato di capire il giusto livello istituzionale a cui applicare la suddivisione. Infine, questo livello è stato definito quello regionale.

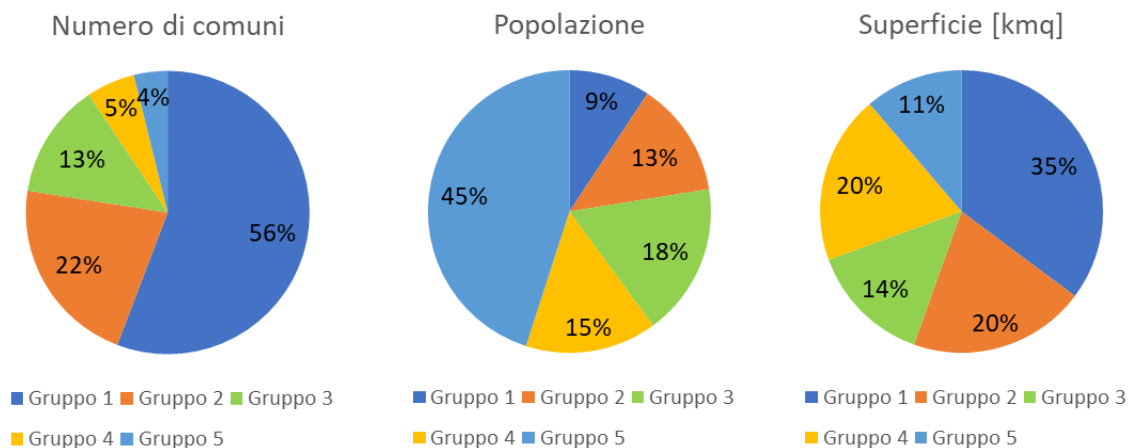


Figura 1 - Numero di comuni, Popolazione e Superficie in base ai cinque gruppi di popolazione

Utilizzare le regioni come suddivisione geografica è stata la scelta più ragionevole. In tal senso la suddivisione della penisola in maniera omogenea è risultata secondaria. Molto più significativo è il fatto che siano le regioni ad avere ampio margine di manovra (anche dal punto di vista normativo) sulle tematiche digitali. Certo il riferimento normativo generale resta di livello Nazionale, ma la raccolta dei dati, la creazione di portali, la gestione delle informazioni è una pratica consolidata dalle regioni e che rimane in campo a questa istituzione. Lo stesso portale nazionale dati.gov raccoglie in maggioranza i dati messi a disposizione dalle regioni. Solo in rari casi le province coincidono con maggior dettaglio anche da questo punto di vista tematico: sono questi i casi di Trento e Bolzano o delle città metropolitane. La cosiddetta Legge Del Rio, la n.56 del 7 aprile 2014 “Disposizioni sulle città metropolitane, sulle province, sulle unioni e fusioni di comuni”, entrata in vigore l’8 aprile 2014, ridisegna organi e competenze dell’amministrazione locale determinando lo svuotamento di poteri delle Province che si traduce in un “dettaglio eccessivo”. Per questo motivo è stato valutato il grado digitale delle Regioni e non di altri livelli territoriali.

Dalla definizione di questi due parametri è stata generata la seguente matrice, che individua il numero di comuni totale per regione e gruppo d’appartenenza. Oltre alla matrice inserita di seguito, si allega al termine (Allegato 3) una matrice che mostra in modo dettagliato popolazione, numero di comuni e numero di questionari necessari per completare il campione omogeneo.

Regioni	Gruppo 1 0 - 3 mila	Gruppo 2 3 - 7 mila	Gruppo 3 7 - 15 mila	Gruppo 4 15 - 30 mila	Gruppo 5 > 30 mila	Totale
Abruzzo	220	49	18	10	8	305
Basilicata	82	33	11	3	2	131
Calabria	256	96	33	11	8	404
Campania	264	112	89	44	41	550
Emilia-Romagna	82	102	89	36	22	331
Friuli-Venezia Giulia	131	45	28	7	4	215
Lazio	205	70	49	27	27	378
Liguria	164	34	25	6	5	234
Lombardia	787	405	211	74	39	1.516
Marche	125	50	31	11	12	229
Molise	116	14	3	1	2	136
Piemonte	957	142	50	29	19	1.197
Puglia	57	66	63	44	28	258
Sardegna	272	59	30	9	7	377
Sicilia	136	103	84	34	33	390
Toscana	78	67	74	33	22	274
Trentino-Alto Adige/Südtirol	207	64	11	6	4	292
Umbria	46	23	7	10	6	92
Valle d'Aosta/Vallée d'Aoste	65	8			1	74
Veneto	187	182	140	46	16	571
Totale	4.437	1.724	1.046	441	306	7.954

Tabella 3 - Matrice regione - gruppi dei comuni italiani

3.3. Campione omogeneo.

La selezione dei comuni per l'invio del questionario ha riguardato due criteri e alcune ipotesi sviluppate in una fase preliminare della tesi. Questa operazione è stata svolta utilizzando un metodo statistico tale per cui i risultati ai questionari di una piccola percentuale di comuni, opportunamente scelti, possa considerarsi rappresentativa di tutti i comuni italiani. Questo ha significato l'utilizzo di due criteri: la posizione geografica e la popolazione. I questionari sono stati inviati proporzionalmente evitando concentrazioni geografiche, individuandone il numero opportuno in ogni regione sulla base della loro popolazione.

A tal proposito, ho ottenuto l'opportunità di partecipare alle prime due lezioni del master di *Machine Learning* organizzato dal Cefriel in collaborazione con il Politecnico di Milano. L'argomento delle lezioni è stato l'introduzione al calcolo statistico e alle ipotesi di verifica, tenute dai prof. Simone Vantini e Francesca Ieva ricercatori presso il MOX.

Nonostante alcune difficoltà legate alla mancanza di solide basi nelle discipline di statistica e matematica – nell'intero ciclo di studi in pianificazione si frequenta soltanto un corso da quattro crediti – ho trovato le lezioni molto utili. L'utilità effettiva delle nozioni di statistica nella prima fase di presentazione del questionario è stata relativa. Credo infatti, alla luce di quanto appreso, che nella presentazione di un questionario si faccia riferimento a una statistica di tipo sociale, visto anche la tipologia di domande contenute in esso ma che certamente utile potrà essere nella fase di elaborazione dei risultati.

La "statistica sociale" è stata utile nella presentazione del questionario, che si è basata su indagini campionarie e modalità di somministrazione dei questionari. La qualità e la credibilità del risultato

è strettamente legata alla metodologia che si adotta nella selezione del campione, che a sua volta presuppone una conoscenza approfondita della totalità del fenomeno e dell'insieme da cui il campione deriva. Il campione deve essere una sintesi precisa della realtà, nel mio caso specifico il campione di comuni deve essere rappresentativo della totalità dei comuni. Per determinare tutto questo non si può prescindere quindi sia dal fenomeno che stiamo analizzando, la *data literacy* e i big data, che dai soggetti di questa analisi, i comuni italiani.

Per stabilire il numero di questionari da ricevere perché tutti gli esiti della tesi possano essere considerati validi, la tesi si avvale di un calcolatore di campioni statistici basato sulle seguenti formule. Una ricerca statistica standard utilizza un livello di confidenza del 95% con un margine di errore del 5%. Applicato al caso dei comuni questo significa che ho una popolazione di circa 8.000 unità. Il 95% di confidenza (*confidence level*) significa che se conducessi la stessa ricerca venti volte, diciannove (19/20) il mio risultato potrebbe differire al massimo del margine di errore che ho stabilito. Il 5% di errore significa che se venissero intervistati tutti e 8.000 i comuni, il risultato potrebbe differire al massimo tra più 5% o meno 5%.

Il calcolo applicato al presente questionario è il seguente.⁴⁸

1. *Sample Size:*

$$(*Distribution\ of\ 50\%) / ((Margin\ of\ Error\% / **Confidence\ Level\ Score) ^ Squared)$$

*Distribution 50%: la distribuzione riflette la differenza dei partecipanti su un certo argomento. Si mantiene una distribuzione del 50% che è la più conservativa.

**Confidence level score: è la deviazione standard che deriva dal livello di confidenza, per il 95% è pari a 1,96.

2. *Finite population correction:*

$$(Sample\ Size\ X\ Population) / (Sample\ Size + Population - 1)$$

Considerando quindi i valori standard utilizzati a livello statistico, ovvero livello di confidenza del 95% e margine di errore del 5%, si ottiene un numero di questionari valido pari a 367 (4,6% del totale dei comuni italiani).

Ci sono ulteriori considerazioni utili fare chiarezza nella determinazione del campione d'indagine tra i comuni italiani, manteniamo fedeli i calcoli sul campione omogeneo appena determinati. In principio la selezione riguardava circa il 12% dei comuni, ovvero all'incirca 1.000 questionari, da ripartire in modo proporzionale tra regioni e fasce di popolazione. Di fronte alla difficoltà di reperire tante risposte in così poco tempo (che tra l'altro superano le 367 risposte necessarie), ho diminuito la percentuale iniziale in modo da ottenere un campione il più possibile omogeneo con un numero ragionevole di comuni interessati (economicità dell'indagine). Passando ad esempio a percentuali dell'1,2% il campione si restringerebbe a 100 comuni, ma ben 4 regioni su 20 non avrebbero rappresentanza nel campione di comuni del gruppo uno (meno di 5 mila abitanti). Che salirebbe a 6 regioni non rappresentate nel gruppo successivo, il gruppo due; 4 sole regioni rappresentate nei comuni del gruppo 3 e senza alcun comune interessato oltre i 30.000 abitanti. Questo sta a significare che il comune di residenza di 27.231.571 abitanti, cioè quasi la metà dell'intera popolazione italiana, non rientrerebbe nel campione selezionato. Provando a superare

⁴⁸ (FluidSurveys, 2014)

superare il record precedente ottenuto dalla ricerca di PoliVisu, ovvero 127 questionari, la percentuale viene alzata al 2%: 160 comuni italiani interessati. In questo caso almeno un comune appartenente al gruppo 1 per regione entrerebbe a far parte del campione. Prima di proseguire nell'analisi è bene fare alcune precisazioni.

È evidente che la selezione del gruppo è fondamentale. Prima di stabilire tutto questo ho individuato quattro gruppi sulla base delle fasce di popolazione (i cui criteri di scelta ho fornito in precedenza), ma ho sempre determinato il numero di questionari riferendomi alla singola fascia e non al gruppo. In questo modo mantenevo il dettaglio, ma con un numero troppo elevato di questionari. Diminuendo o aumentando il numero di fasce, diminuendo o aumentando la popolazione in ciascuna fascia, il numero di comuni che ne fanno parte cambia. Il campione statistico raggiunge tutte le fasce con pochi questionari solo se le fasce sono poche, ma questo significa che il grado di dettaglio è minore.

Lo stesso ragionamento potrebbe essere fatto utilizzando come discriminante il criterio geografico, ora rappresentato dalle regioni, ma che si potrebbe configurare in macroregioni, oppure addirittura in Province. Diminuendo o aumentando il grado di dettaglio geografico che intendiamo ottenere.

In questo caso la selezione del grado di dettaglio geografico è più semplice rispetto alla popolazione. Utilizzare le regioni come suddivisione geografica è stata la scelta più ragionevole. L'equa suddivisione della penisola è un dettaglio trascurabile. Molto più significativo è il fatto che siano le regioni ad avere ampio margine di manovra (anche dal punto di vista normativo) sulle tematiche digitali. Certo il riferimento normativo generale resta di livello Nazionale, ma la raccolta dei dati, la creazione di portali, la gestione delle informazioni è una pratica consolidata dalle regioni e che rimane in campo a questa istituzione. Lo stesso portale nazionale dati.gov raccoglie in maggioranza i dati messi a disposizione dalle regioni. Solo in rari casi le province coincidono con maggior dettaglio anche da questo punto di vista tematico: sono questi i casi di Trento e Bolzano o delle città metropolitane. La cosiddetta Legge Del Rio, la n.56 del 7 aprile 2014 "Disposizioni sulle città metropolitane, sulle province, sulle unioni e fusioni di comuni", entrata in vigore l'8 aprile 2014, ridisegna organi e competenze dell'amministrazione locale determinando lo svuotamento di poteri delle Province che si traduce in un "dettaglio eccessivo". Per questo motivo è stato valutato il grado digitale delle Regioni e non di altri livelli territoriali. La scala di dettaglio del criterio geografico e adeguati strumenti digitali si incontrano proprio a livello della Regione. Di conseguenza la variabile sulla quale lavorare, tra i due criteri, rimane la popolazione e la sua suddivisione in fasce.

Anche nel caso del criterio della popolazione dobbiamo ricordarci che stiamo facendo riferimento ad una istituzione, ovvero il comune. È perciò necessario considerare una discrezionalità politica nel calcolo. Intendo dire che la suddivisione in fasce di popolazione non può essere effettuata soltanto sulla base di criteri matematici, perché essi potrebbero non tenere in considerazione effetti normativi: la popolazione di un comune determina o meno l'applicazione di una determinata norma, o l'applicazione della stessa norma con diverse modalità. Se prendiamo come esempio la già citata Legge Del Rio essa prevede "l'esercizio obbligatorio in forma associata delle funzioni fondamentali da parte dei comuni con popolazione fino a 5.000 abitanti". Capite bene quanto questo possa influire sui servizi digitali di due comuni che adottano una convenzione per la loro gestione associata. Un'altra soglia importante è quella dei 15.000 abitanti, oltre i quali variano non solo le modalità di elezione del Sindaco, ma anche gli strumenti di programmazione e gestione.

Certamente fondamentale per l'oggetto della tesi e nella determinazione del campione d'indagine sono i dipendenti comunali. Essi, secondo i dati elaborati dall'IFEL risultano pressoché omogenei, ma sotto i 5 mila abitanti ed oltre i 60 mila, aumentano passando da una media di 5.8

dipendenti ogni mille abitanti, a ben oltre i 7. Il Dm. interno del 10/04/2017 prevede per il triennio 2017-2019 dei rapporti medi dipendenti – popolazione validi per gli enti dissestati.

Dai primi tentativi di determinarle fino alla decisione finale, i raggruppamenti ed un'unica grande fascia per i comuni oltre i 30.000 abitanti. Lo scaglione ultimo. La prospettiva iniziale dei 1.000 questionari sarebbe stata assai allettante, ma decisamente poco economica in termini di tempo, senza considerare che ottenere tutte quelle risposte non sarebbe affatto scontato. Come già raccontato, una percentuale intorno al 2,5% sarebbe molto più sostenibile. Riducendo i questionari tra i 150 (1,9%) e i 250 (3,2%) e suddividendoli secondo le opportune fasce demografiche il numero diventa più gestibile, il campione rimane omogeneo, ma alcune correzioni sono necessarie.

Perciò non è possibile dividere soltanto in maniera progressiva, oltre una certa popolazione i gruppi determinati sulla base delle fasce di popolazione perdono importanza.



Figura 2 - Mappa della penisola italiana con i 259 comuni che hanno risposto al questionario (aggiornamento 23/03/2019)

4. Risultati dell'analisi

In questo capitolo si tenterà di riassumere e rielaborare le risposte al questionario ricevute. In particolare, si specifica che i risultati sono stati elaborati sulle risposte ricevute alla data 23/03/2019, tutti i dati analizzati di seguito fanno riferimento alle risposte ricevute fino ad allora, nuove risposte al questionario non vengono considerate in questa tesi.

L'invio delle mail ai comuni per la partecipazione al questionario è cominciato il primo febbraio 2019 tramite la mail ufficiale da studente del Politecnico. Sono state inviate 250 mail al campione di comuni selezionato secondo i criteri presentati nel precedente capitolo. Le mail dei comuni erano state collezionate dal sito ufficiale di ogni comune, erano state utilizzate le mail ordinarie, e dopo due settimane le risposte erano ferme a 30 (ovvero il 12%). Per ottenere un campione più indicativo è stato necessario comprare un account di posta elettronica certificata pec (francesco.fagiani@pec.it), nella maggior parte dei casi infatti, i comuni permettono di inviare mail alle proprie pec soltanto attraverso un altro indirizzo pec. Grazie all'open data di IndicePA

che mette a disposizione tutti gli indirizzi pec delle pubbliche amministrazioni italiane per la fatturazione economica, sono stati facilmente individuate le pec di tutti i comuni italiani. Sono state rinviate le mail ai comuni già contattati che non avevano ancora dato risposta (220) oltre ad altre 1.000 mail a nuovi comuni. Nella selezione di quest'ultimi è stata mantenuta la proporzione valida per mantenere il campione omogeneo invariato. Nelle nuove mail è stato aggiunto il numero di telefono personale, già alcuni comuni infatti avevano palesato il rischio di aprire un link sconosciuto (quello del questionario) all'interno della rete comunale. Tramite il numero di telefono i comuni avrebbero potuto accertarsi, o quasi, della sicurezza del link contattandomi direttamente. Il vantaggio della pec, oltre al fatto di poter contattare direttamente i comuni utilizzando lista degli indirizzi di IndicePA, è stato quello della certezza dell'avvenuta protocollazione.

Grazie ad una semplice macro VBA di Excel è stato possibile mandare un grande numero di mail tutte insieme. Ogni riga corrispondeva ad un comune, nella prima colonna era presente il testo della mail (in cui variavano il nome del comune e la regione di appartenenza) con richiesta di attenzione da parte dell'amministrazione, del segretario e degli uffici; nella seconda colonna era presente l'indirizzo pec principale del comune. Oltre alla prima tranches di 1.220 mail, sono state inviate altre mail in funzione di completare il campione omogeneo. Alla data del 23/03/2019 sono state inviate un totale di 1.878 mail tramite pec e 250 tramite l'indirizzo di posta elettronica ordinaria del politecnico. Quindi i comuni raggiunti dal link del questionario sono stati 1.908.

Il questionario ha ottenuto 281 risposte da parte di 259 comuni⁴⁹. Da alcuni comuni sono pervenute più di una domanda, in questo caso al questionario hanno partecipato diversi uffici. Due risposte sono risultate inutilizzabili in fase di elaborazione dei risultati, non avendo riportato una voce corretta alla domanda sul comune di provenienza del questionario. Il totale di risposte elaborate è 279. La tabella riporta la suddivisione delle risposte in base alla regione di provenienza e della fascia di popolazione a cui appartiene il comune rispondente. La tabella, come il resto delle statistiche è aggiornata al 23/03/2019.

⁴⁹ Data di riferimento 23/03/2019

Regioni	Gruppo 1 0 - 3 mila	Gruppo 2 3 - 7 mila	Gruppo 3 7 - 15 mila	Gruppo 4 15 - 30 mila	Gruppo 5 > 30 mila	Totale
Abruzzo	5	1	1	-	-	7
Basilicata	3	2	-	-	-	5
Calabria	6	2	2	-	-	10
Campania	5	1	2	2	4	14
Emilia-Romagna	3	2	6	4	2	17
Friuli-Venezia Giulia	3	1	3	-	-	7
Lazio	5	1	2	1	2	11
Liguria	5	8	3	-	-	16
Lombardia	27	16	7	2	4	56
Marche	3	1	1	-	-	5
Molise	2	-	-	-	-	2
Piemonte	23	5	3	3	6	40
Puglia	2	2	1	1	1	7
Sardegna	11	1	2	-	-	14
Sicilia	2	3	3	2	4	14
Toscana	2	2	1	9	2	16
Trentino-Alto Adige/Südtirol	7	2	-	-	-	9
Umbria	2	1	-	-	-	3
Valle d'Aosta/Vallée d'Aoste	2	-	-	-	-	2
Veneto	8	8	4	4	-	24
Totale	126	59	41	28	25	279

Tabella 4 - Matrice regioni - gruppi dei comuni da cui è stata ricevuta almeno una risposta

4.1.1. Analisi del campione omogeneo.

Per rappresentare l'Italia in maniera omogenea e ottenere dei risultati validi e omogenei per tutta la penisola, alcune elaborazioni dei risultati sono state effettuate su un campione omogeneo di comuni. Per fare ciò sono state selezionate 158 risposte in maniera casuale, ma allo stesso tempo rappresentative dell'Italia. I criteri utilizzati per la selezione del campione sono stati definiti nel capitolo precedente. I riferimenti al "campione omogeneo" che verranno fatti nei paragrafi successivi vanno intesi per quanto appena spiegato, utilizzando un campione di 158 risposte sulle 279 correttamente caricate. Per altri tipi di elaborazione, la definizione del campione è superflua, perché non è necessaria una valutazione unitaria, anzi è bene confrontare le risposte totali tra loro, prima in base al criterio della popolazione e poi sulla base del criterio geografico, facendo emergere eventuali incongruenze e marcate differenze.

4.1.2. Tipologia dei comuni.

La tabella 4 costituisce la matrice delle risposte in funzione della regione di provenienza (criterio geografico) e del gruppo di appartenenza, che rappresenta la fascia di popolazione. Le stesse informazioni presenti in tabella sono state inserite in forma di grafico, suddividendo i due criteri utilizzati nella formazione del campione: regione e abitanti. Una grossa fetta delle risposte arriva dal nord Italia, in particolare dal Nord-Ovest (41%) con 114 risposte. Dalle isole sono pervenuti solamente 28 questionari.

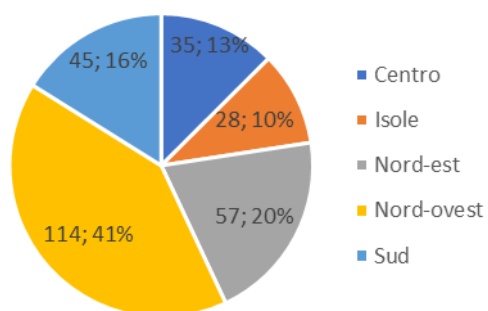


Figura 4 - Provenienza risposte per macroregione ISTAT

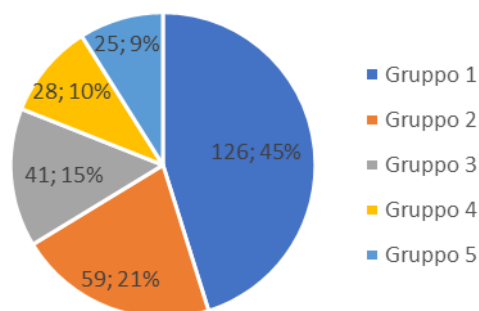


Figura 3 - Provenienza risposte per Gruppo d'appartenenza

Per quanto riguarda la provenienza in base ai gruppi, il 45% delle risposte proviene da comuni del Gruppo 1 (fascia di popolazione 0 – 3.000 abitanti). Come mostrato nel capitolo sulla statistica, di questa fascia fa parte la maggioranza dei comuni italiani.

4.1.3. Ruolo all'interno del comune.

Alcune domande, tra cui quella sul ruolo del rispondente all'interno della pubblica amministrazione, sono state omogenizzate. La risposta "altro" permetteva di specificare il proprio ruolo qualora non fosse presente tra le opzioni di scelta. La tipologia di risposte si è ampliata enormemente (nel caso della domanda sul proprio ruolo nel comune, quasi 80 differenti, dovute anche a piccole differenze nei caratteri – es. Tutte e TUTTE formano due tipologie), per questo in fase di rielaborazione le risposte sono state valutate singolarmente e riassegnate a una categoria

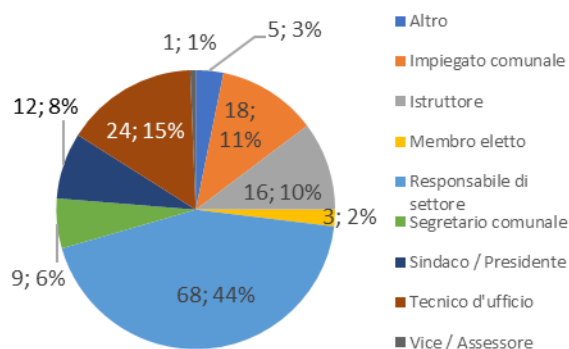


Figura 6 - Ruolo all'interno del Comune. Campione omogeneo

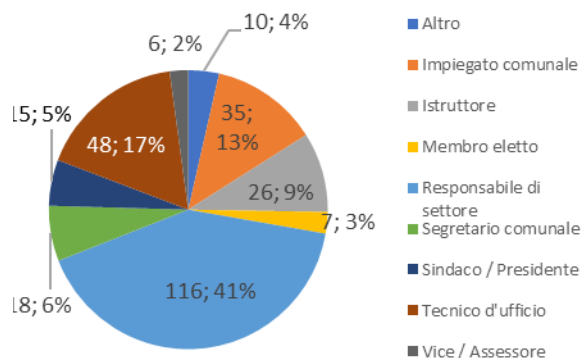


Figura 5 - Ruolo all'interno del Comune. Totale comuni risposte

esistente, oppure sono state create nuove categorie. I grafici mostrati e le rielaborazioni tengono conto di questa riassegnazione.

In questo paragrafo sono state confrontate le risposte del campione omogeneo (158 risposte) con le risposte totali, dai grafici presenti nell'Allegato 2, si può notare che la differenza è minima. Per quanto riguarda il totale, circa il 44% delle risposte è stata data da un responsabile di settore, la seconda categoria più rappresentata è quella dei tecnici d'ufficio e degli impiegati comunali, che insieme compongono il 30%, quindi gli istruttori amministrativi e i segretari comunali, 9 e 6%. La parte politica rappresenta l'10% delle risposte totali, Sindaci 5%, assessori o vice il 2% e il 3% da membri eletti. Confrontando le risposte totali con quelle del campione le percentuali variano poco.

4.1.4. Settore o ufficio d'appartenenza

La seconda domanda riguardava il ruolo dell'intervistato all'interno del comune. Anche in questo caso le risposte sono state omogeneizzate per ridurle nelle principali categorie, da oltre 50 sono state ridotte a 20 principali. Il settore più rappresentato risulta quello della segreteria generale, da cui provengono più di un quarto delle risposte. Altri settori importanti riguardano Informatica / Servizi ICT / Agenda digitale (14%) e il settore finanziario / tributi (9%). Le altre categorie sono minoritarie. In particolare, vorrei sottolineare la presenza dei settori di urbanistica e affini (edilizia privata e lavori pubblici) che messe insieme raggiungono solo il 9%. Il settore della mobilità e dei trasporti in cui generalmente vengono applicati i big data ha ricevuto una sola risposta. Probabilmente questo riguarda solo le metropoli, e meno i comuni. Una categoria che è stata inserita nella fase di riduzione e riassegnazione delle risposte è "più settori". Molti utenti, in particolare quelli dei piccoli comuni, specificavano il fatto di dover sovrintendere a tutte le attività. Anche i sindaci in genere non si concentrano su un settore, ma lavorano in modo trasversale su molti settori del comune, per questo, in base alla risposta originale, sono stati assegnati alla categoria "più settori".

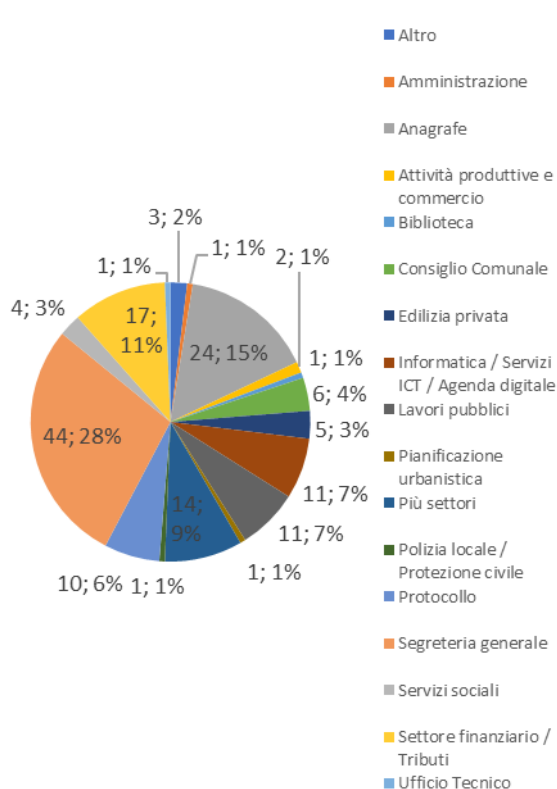


Figura 7 - Settore comunale. Campione omogeneo.

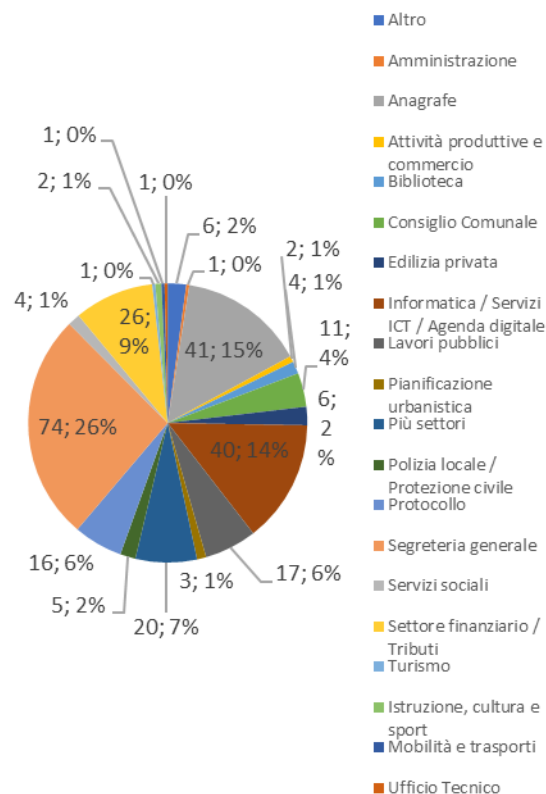


Figura 8 - Settore comunale. Totale risposte

4.1.5. Definizione di big data.

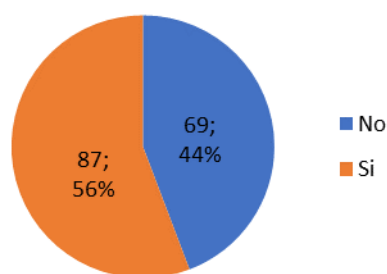


Figura 9 - Conosci la definizione di big data?

Le risposte binarie come questa si sono rivelate le più interessanti da valutare. Si richiedeva la conoscenza del termine big data con un secco sì o no. Le successive domande poi richiedevano una specificazione, anche da parte di chi nella prima risposta aveva negato la conoscenza del termine. Questo permette una valutazione del grado di conoscenza effettivo. La maggioranza degli intervistati conosce la definizione di big data.

Del campione omogeneo il 56% degli intervistati risponde affermativamente alla domanda. Lo scarto con chi risponde “no” è solo di 18 voti. Se si considerano tutte le risposte la percentuale di sì sale al 62%.

Come accennato in precedenza, i grafici di ogni sezione si possono dividere in tre categorie. Una relativa alle risposte del campione omogeneo (158) a cui si potranno legare ragionamenti che riguardano l'intera penisola e che quindi rappresentano la totalità dei comuni. Ragionamenti generali per quanto riguarda la data literacy italiana. La seconda relativa a tutte le risposte ricevute (283), che compara tra loro le macroregioni, ricercando eventuali differenze nella data literacy su

base geografica di macroregione o regione. La terza sempre relativa a tutte le risposte ricevute, che compara questa volta i gruppi di comuni suddivisi per fasce di popolazione, ricercando eventuali differenze nella data literacy tra comuni piccoli, medi o grandi in base alla popolazione. Tra le macroregioni la conoscenza più alta è quella del centro Italia, oltre il 70%, la più bassa il nord ovest, 57%. In questo caso, mentre se si ragiona comparando le risposte in base alla provenienza geografica non si trovano particolari differenze, sulla base dei gruppi le differenze sono più mirate. Si potrebbe dire che la conoscenza del termine cresce al crescere degli abitanti del comune. Infatti si parte da percentuali di “sì” del 52% nei gruppi 1 e 2, fino all’89% del gruppo 4 e 84% del gruppo 5. Una differenza importante che mostra come i dipendenti o i membri eletti di un comune con oltre 15.000 abitanti siano più avvezzi a big data, rispetto ai colleghi di piccole realtà. Inizia qui a delinearsi un tema, quello della competenza e delle risorse. Questa chiara divisione potrebbe essere dovuta a una diversa possibilità di accesso a risorse e competenze. Per capire meglio questi aspetti è utile entrare nel vivo della definizione attraverso l’analisi dettagliata dei parametri. Mantenendo lo schema utilizzato fino ad ora, cominciamo l’analisi dal campione omogeneo, ovvero dal risultato “italiano”. Ecco quali caratteristiche hanno i big data secondo i partecipanti:

Volume - grande quantità di dati	4,2
Velocità - rapidità di acquisizione / raccolta di dati	4,0
Varietà - diverse tipologie di dati	3,9
Procedure - capacità di gestione e analisi	4,2
Pubblicità - condivisione dei dati	3,6
Fonti - da dove vengono acquisiti i dati	3,9

La possibile risposta aveva un range tra 1 e 5 in cui 1 significava superfluo, mentre 5 significava molto importante. Nella tabella viene riportata la media pesata delle risposte. È chiaro che a tutte le caratteristiche viene attribuita una certa importanza. Sicuramente al volume, ma anche la parte di gestione ha ricevuto lo stesso riscontro, così come anche la velocità. Solamente la pubblicità intesa come condivisione dei dati è più bassa 3,6/5, ma comunque propende verso una certa importanza. Tra le voci messe a disposizione non ce ne sono alcune considerate superflue. Questo è indice di una certa consapevolezza. Nonostante molti abbiano risposto di non conoscere la definizione di big data, esiste una consapevolezza di quello che il termine big data rappresenta e la tabella di fianco lo dimostra.

Come per la statistica sulla definizione, anche per quanto riguarda le caratteristiche dei big data confrontate su base macroregionale il risultato è omogeneo e propende per l’importanza di tutte le caratteristiche. Considerando il criterio della popolazione questa volta i risultati cambiano, mettendo parzialmente in discussione il risultato precedente, che mostrava una maggiore conoscenza da parte dei medio-grandi comuni. Il verdetto non viene ribaltato, ma da questa seconda parte della statistica non emergono particolari differenze. Il campione confrontato è omogeneo, perciò la definizione può essere sconosciuta, ma nelle risposte i partecipanti dimostrano una consapevolezza su cosa siano i big data. O almeno le differenze importanti segnalate in precedenza in questa fase vengono limare.

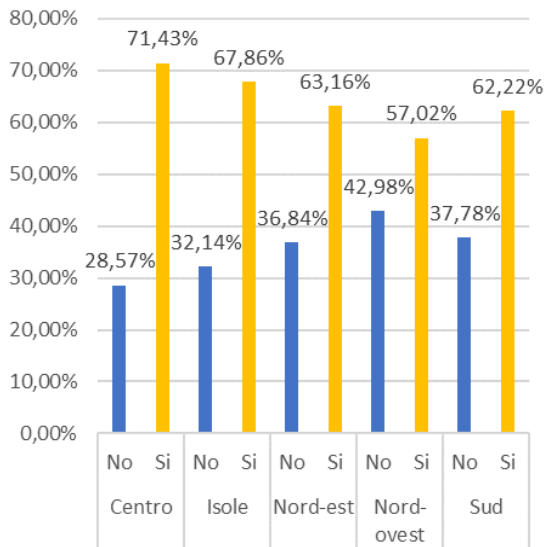


Figura 10 - Conosci la definizione di "big data"?

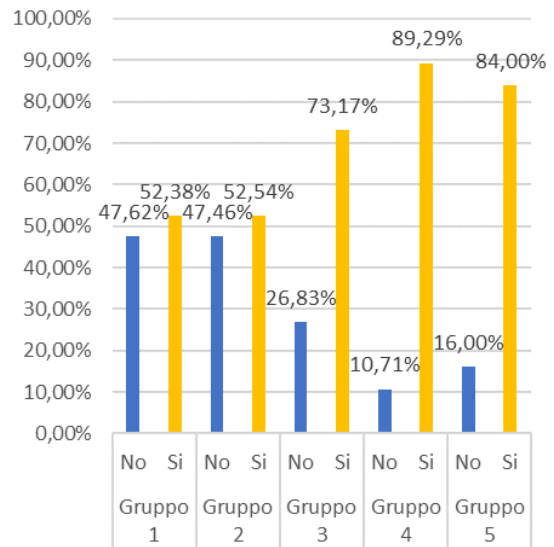


Figura 11 - Conosci la definizione di "big data"?

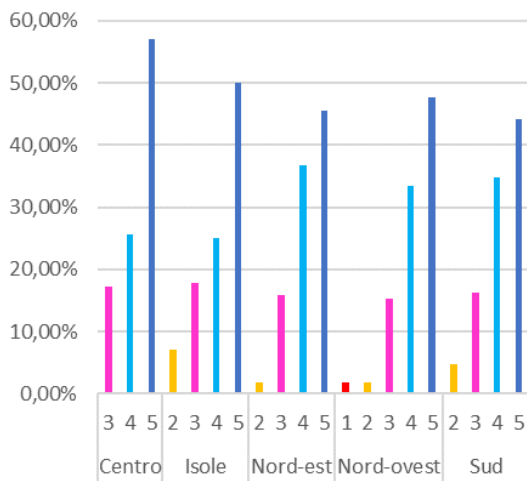


Figura 12 - Volume - grande quantità di dati

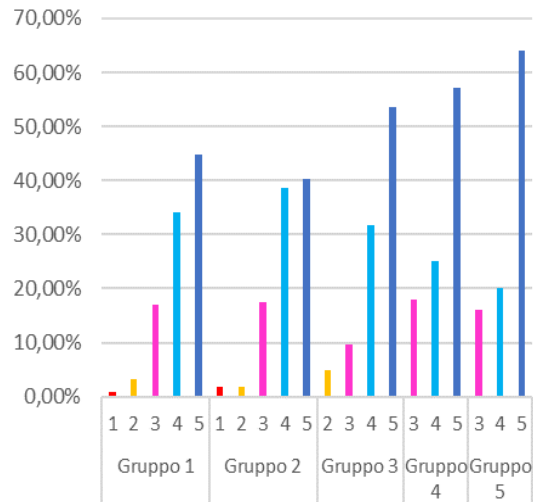


Figura 13 - Volume - grande quantità di dati

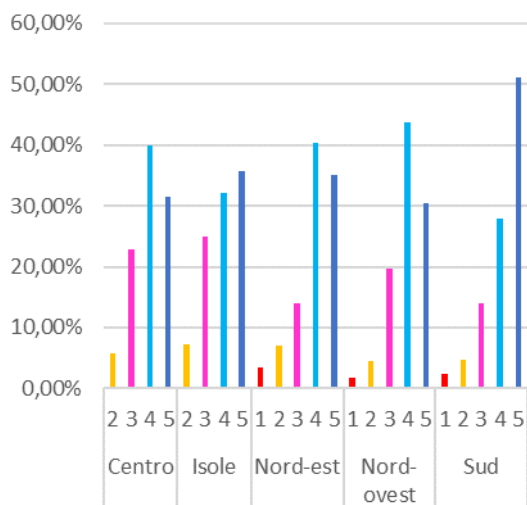


Figura 14 - Velocità - rapidità di acquisizione / raccolta di dati

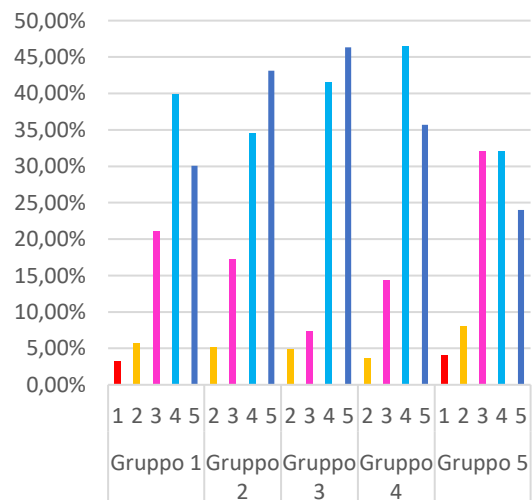


Figura 15 - Velocità - rapidità di acquisizione / raccolta di dati

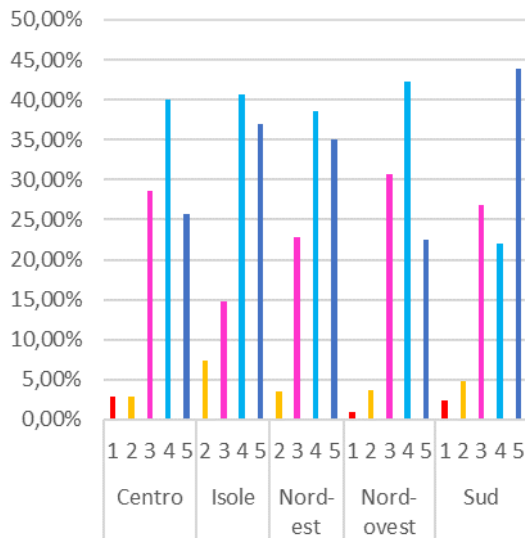


Figura 16 - Varietà - diverse tipologie di dati

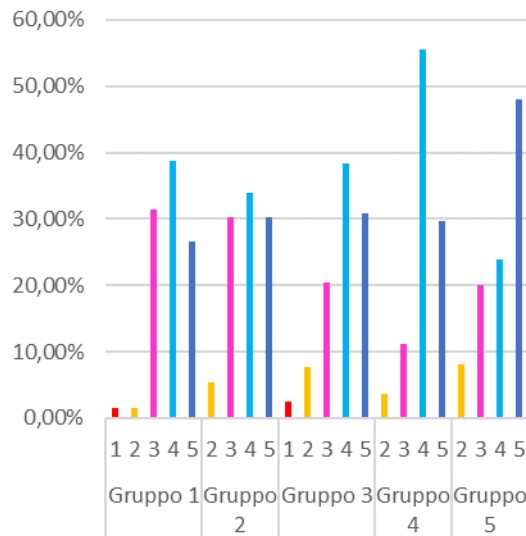


Figura 17 - Varietà - diverse tipologie di dati

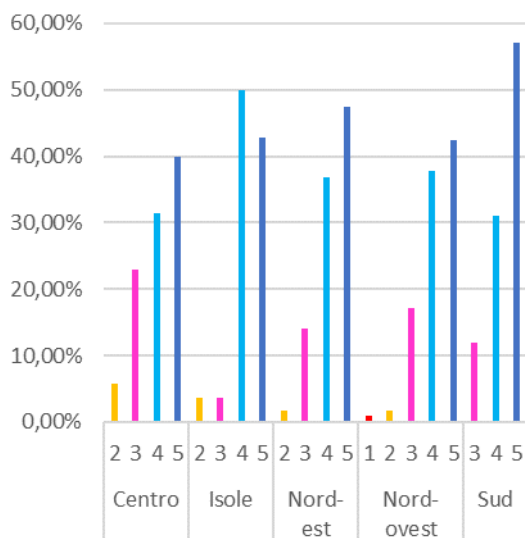


Figura 18 - Procedure - capacità di gestione e analisi

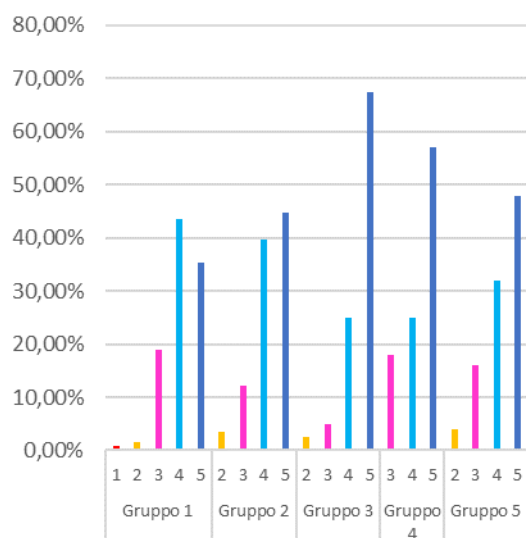


Figura 19 - Procedure - capacità di gestione e analisi

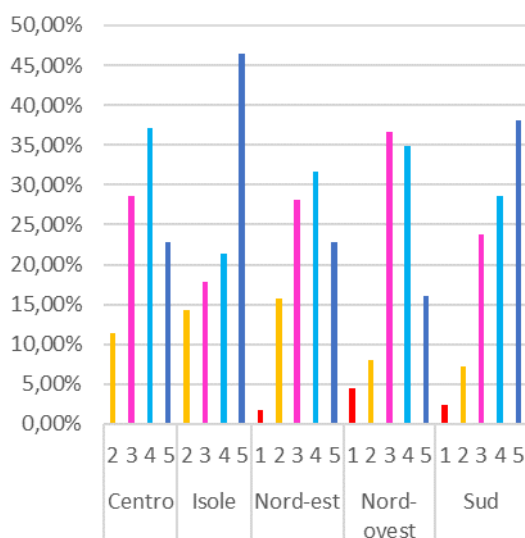


Figura 20 - Pubblicità - condivisione dei dati

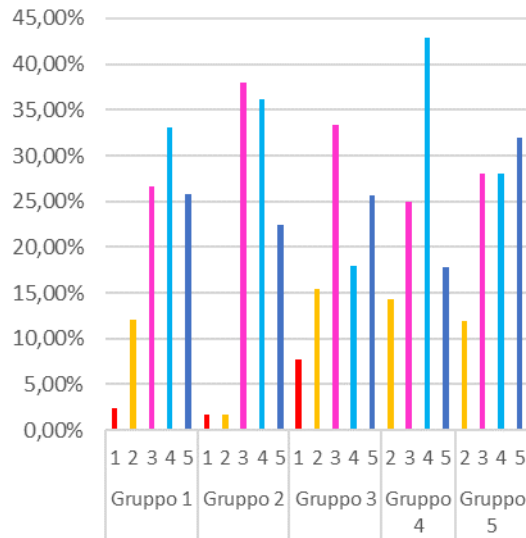


Figura 21 - Pubblicità - condivisione dei dati

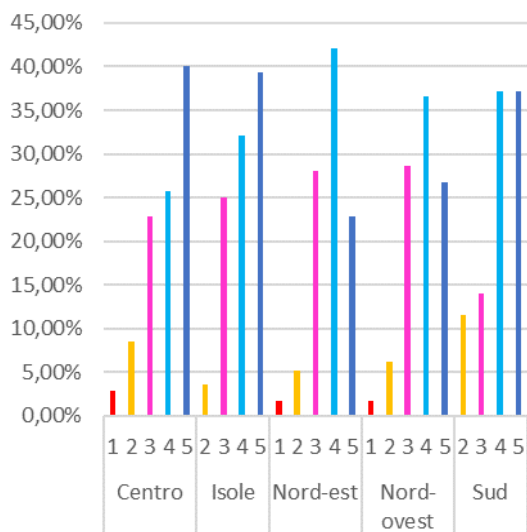


Figura 22 - Fonti - da dove vengono acquisiti i dati

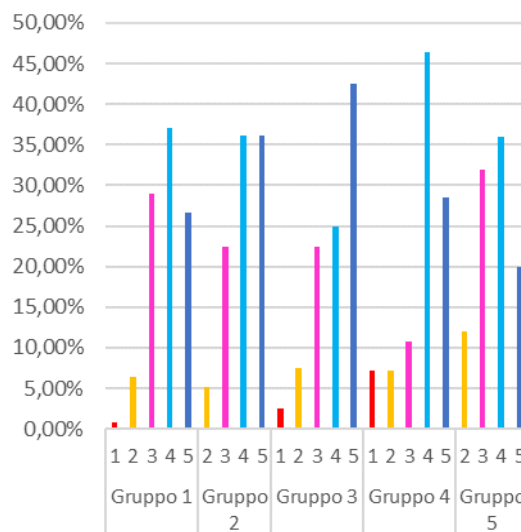


Figura 23 - Fonti - da dove vengono acquisiti i dati

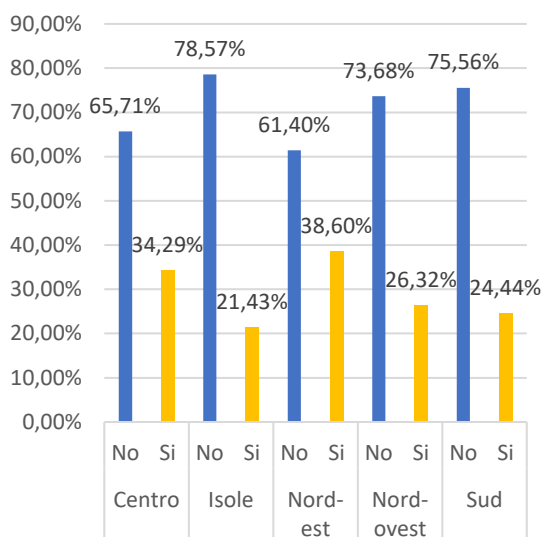


Figura 24 - Utilizzo dei big data

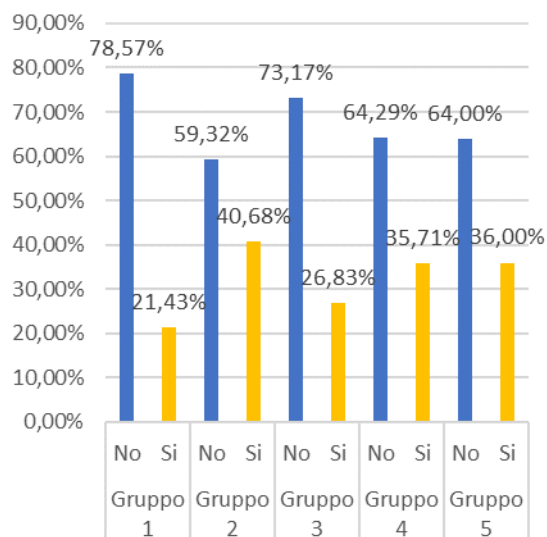


Figura 25 - Utilizzo dei big data

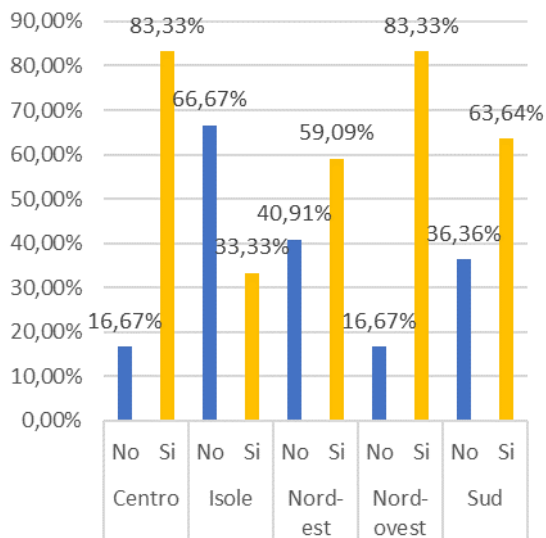


Figura 26 - Produzione interna di big data

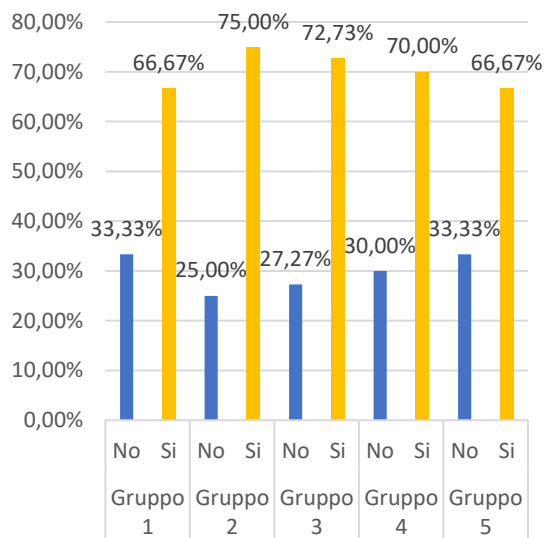


Figura 27 - Produzione interna di big data

4.1.6. Utilizzo dei big data.

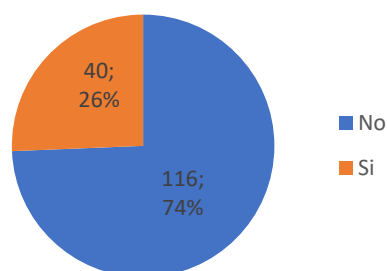


Figura 28 - Dove lavori vengono UTILIZZATI dei "big data"?

Altra domanda fondamentale del questionario, che riguarda l'utilizzo o il non utilizzo dei big data all'interno del comune. In questo caso la maggioranza è più netta e riguarda i comuni che non utilizzano big data, ben 74% del campione.

Soltanto il 26% afferma di utilizzare big data nel comune in cui lavora, solo un quarto. Questa percentuale ribalta il risultato della ricerca PoliVisu in cui il 72% aveva risposto affermativamente. Le possibili ragioni sono state ricercate comparando tra loro macroregioni e gruppi. Come prima questione va evidenziato il fatto che nei comuni del gruppo 1 (i più numerosi nel campione

omogeneo) la percentuale di sì è ferma al 21%, la più bassa. Tutte le altre sono oltre il 25%, intorno al 35% nei gruppi 4 e 5 e addirittura del 40% per quanto riguarda il gruppo 2. Su base geografica invece la macroregione in cui i big data sono utilizzati meno sono le Isole, 21% delle risposte affermative. Rispetto al gruppo 1 quest'ultima percentuale è meno indicativa, infatti, il numero di comuni delle isole è minoritario. La macroregione del Nord-Est si presenta come la più virtuosa, il 38,6% dei comuni utilizza big data. Un utile confronto, che potrebbe anche spiegare le ragioni di un tale risultato nella ricerca PoliVisu, potrebbe essere quello dei comuni capoluogo di regione rispetto a tutti gli altri. Alla data dell'aggiornamento di questo testo non sono ancora pervenute risposte al questionario da parte di capoluoghi regionali, ma soltanto da 8 capoluoghi provinciali.

Per ora, come si può notare dalle prime risposte del questionario, i piccoli comuni (la gran parte in Italia) hanno una maggiore difficoltà nell'utilizzo dei big data, che è stata una delle ragioni per cui il campione omogeneo si basava su questo criterio. Dagli abitanti deriva il numero di dipendenti, pochi dipendenti e scarse risorse, senza una regia pubblica di livello sovralocale rendono molto ardua la possibilità che nel comune si faccia uso di big data. Questa ipotesi inizia a emergere dalle risposte. Per una maggiore legittimazione di questa ipotesi, oltre che proseguire nell'analisi, sarebbe necessario un approfondimento che tenga conto delle differenze regionali e non solo macroregionali. La politica dei dati / IT a livello regionale potrebbe determinare ripercussioni importanti sui propri comuni.

La risposta positiva o negativa determina due strade diverse all'interno del questionario. Chi ha risposto negativamente (116) passa ad una sezione legata alle barriere che impediscono l'utilizzo dei big data, mentre chi risponde affermativamente (40), prosegue dettagliando utilizzo, gestione e tipologie di dati.

4.1.7. Produzione di big data.

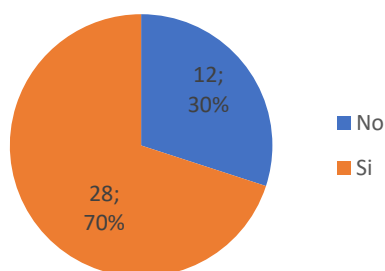


Figura 29 - Dove lavori vengono PRODOTTI INTERNAMENTE "big data"?

La produzione di big data è variegata come dimostrato nel primo capitolo. Sono molti gli strumenti interni ad un comune che permettono questa operazione, ovviamente non tutti ne sono a conoscenza e ancora di meno sono coloro che utilizzano i dati prodotti internamente. Vediamo il dettaglio nel grafico.

Dei 40 comuni in cui si utilizzano big data, solo 12, il 30% non producono dati al loro interno. Considerando la comparazione tra i gruppi di popolazione la differenza è quasi nulla, le percentuali di "no" oscillano tra il 25% e il 33%, la percentuale maggiore di "produttori" appartiene al gruppo due che aveva avuto anche la

percentuale più alta tra gli utilizzatori di big data. Ma ancora più interessante è la comparazione tra macroregioni dove le differenze si fanno più marcate: Centro e Nord-Ovest i produttori sono l'83%, la percentuale scende drasticamente a Sud e a Nord-Est, intorno al 60%, mentre nelle Isole i non produttori diventano maggioritari, il 67% contro il 32% dei produttori.

Le domande successive vanno dettagliare ulteriormente questa questione, per i 28 comuni produttori di big data viene chiesto il settore in cui vengono prodotti, la tipologia e l'eventuale pubblicizzazione. I grafici che seguono mostrano il nuovo dettaglio.

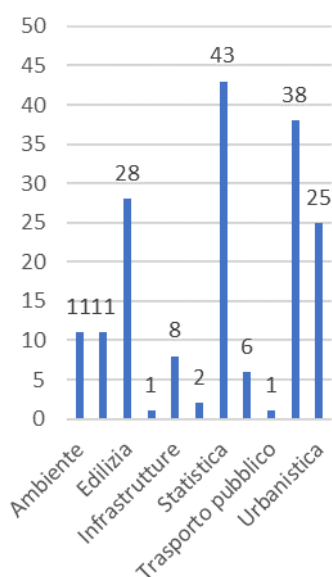


Figura 30 – Settore di provenienza

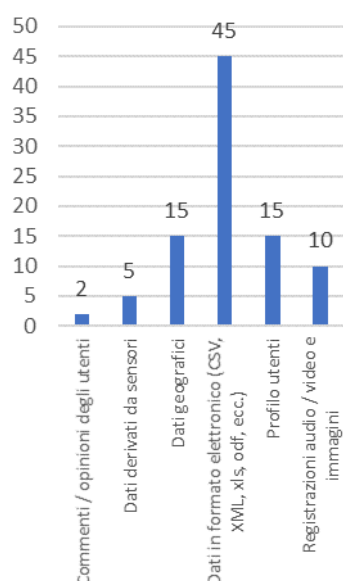


Figura 31 - Formato



Figura 32 - Fornitura dati a terzi

Le risposte sono chiaramente orientate e i grafici lo dimostrano, la maggior parte dei dati riguardano la statistica generale, quindi i tributi. Interessante il piazzamento di edilizia e urbanistica che sottolineano l'interesse che dovrebbe avere la nostra disciplina per i big data. Rimangono di poco conto i dati che riguardano la mobilità e i trasporti, contro ogni pronostico iniziale, che li vedeva i più indicati e che nella teoria sono anche i più utilizzati nella disciplina pubblica di utilizzo e produzione dei big data. Se si considera la tipologia di dati prodotti, l'osservazione sulla scarsa produzione nel settore della mobilità è segnalato dal fatto che la produzione di dati derivati da sensori (che generalmente riguardano traffico) è minoritaria.

Interessante anche in questo caso il piazzamento dei dati geografici. La risposta più cliccata riguarda la tipologia più comune, i dati in formato elettronico. L'ultimo grafico riguarda i dati che vengono messi a disposizione di altre organizzazioni o che vengono resi pubblici come open data. Secondo il campione selezionato la metà dei comuni (53%) rende disponibile solo una parte minoritaria di questi dati. Il 29% invece dichiara che nel suo comune vengono messi a disposizione di terzi la maggior parte dei dati. Solo il 18% non rende disponibile nessun dato raccolto. Insomma, nei comuni dove vengono prodotti big data si tende a pubblicizzarli e non tenerli chiusi.

Se ci addentriamo nella statistica analizzando le organizzazioni a cui questi dati vengono forniti, scopriamo che molti indicano Stato e Regione, ma anche altre pubbliche amministrazioni. A quanto pare i dati entrano in circolo soprattutto all'interno della stessa pubblica amministrazione. Mentre l'opzione open data (pubblici) viene selezionata la metà delle volte rispetto alle due opzioni precedenti. Nel caso del formato in cui vengono messi a disposizione, la statistica è allineata con il grafico sul formato dei dati raccolti. La maggior parte delle scelte è rappresentata da dati strutturati o formato aperto. Molti sono anche coloro che ammettono di non sapere il formato in cui vengono pubblicizzati o resi disponibili a terzi.

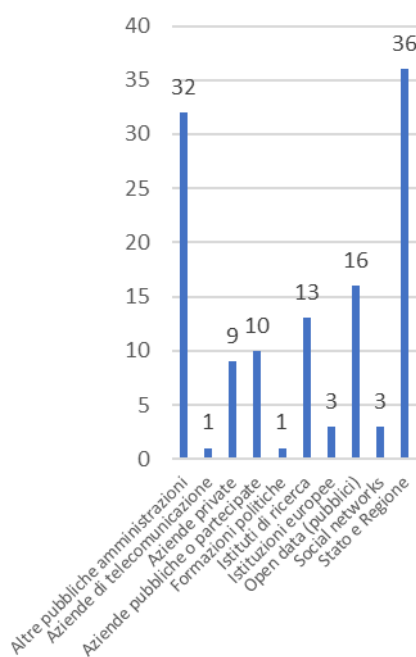


Figura 33 - Organizzazioni a cui vengono forniti i dati

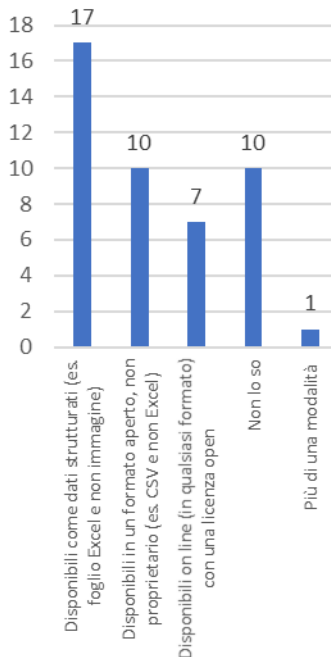


Figura 34 - Formato dei dati forniti a organizzazioni terze

4.1.8. Utilizzo di big data prodotti da terzi.

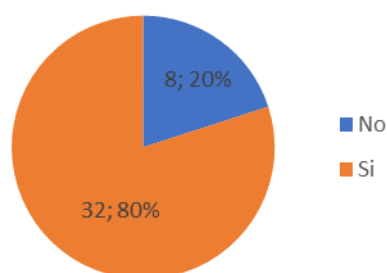


Figura 35 - Utilizzo big data di terze organizzazioni

Innanzitutto, notiamo che l'80% dei comuni che utilizza big data utilizza anche dati che provengono da organizzazioni terze, quindi non direttamente prodotti internamente. Confrontando questa statistica in base ai gruppi di popolazione notiamo che tutti i gruppi hanno una percentuale che si avvicina all'80% oppure superiore, mentre nel gruppo due si concentrano le risposte negative cioè di chi afferma di non utilizzare dati di terzi (quasi il 40%). La concentrazione di risposte negative in una sola categoria si ripete anche confrontando le macroregioni, nel sud Italia le risposte negative sono maggioritarie (oltre il 60%), mentre nelle restanti macroregioni sono tutte inferiori al 20%. Questa concentrazione è significativa e andrà approfondita nella parte riguardante le barriere rispetto l'utilizzo di dati terzi.

Così come i dati vengono prodotti internamente dal comune, altre organizzazioni hanno la possibilità di produrre dati utili alla pubblica amministrazione. Nel capitolo primo sono state presentate le modalità e i soggetti che tipicamente svolgono questa operazione, ovviamente le possibilità sono ampie, in funzione del settore in cui ci si trova ad operare e del tipo di dati che si intende utilizzare. Proviamo a esaminare quali informazioni utili e dettagli emergono dal questionario.

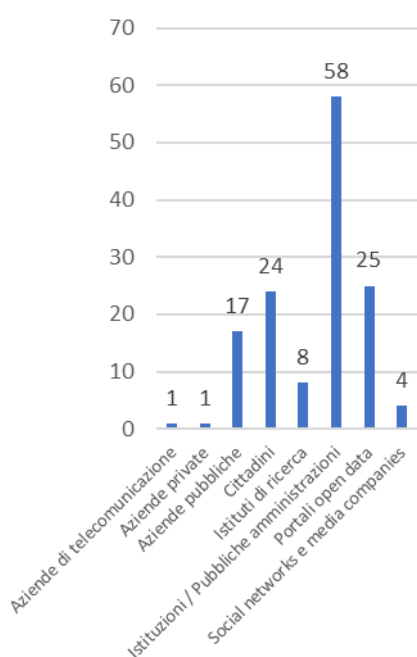


Figura 36 - Organizzazione da cui derivano i dati

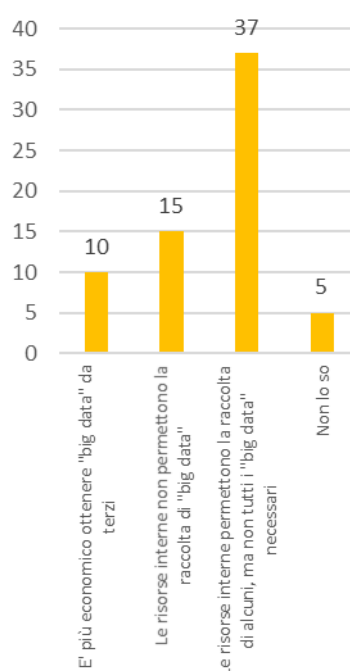


Figura 37 - Ragioni utilizzo dati terze organizzazioni

Nel grafico che rappresenta la provenienza dei dati terzi utilizzati nel comune viene riconfermata una dinamica che si è presentata precedentemente (ma in maniera inversa) nella fornitura di dati a terzi. Anche in questo caso la provenienza maggioritaria dei dati è rappresentata da altre pubbliche amministrazioni. Altro aspetto che potrebbe essere approfondito in fasi successive è proprio la ricerca di un'eventuale presenza di un fenomeno "circuito chiuso" dei dati che vengono

prodotti e riutilizzati all'interno della pubblica amministrazione senza trovare la possibilità di uscire attraverso l'apertura in open data. La voce open data infatti non raggiunge nemmeno la metà dei voti della precedente opzione, come anche l'opzione cittadini. Il grafico sulle ragioni di utilizzo di big data di terze organizzazioni ha un valore fondamentale per stabilire alcune dinamiche interne ai comuni. Questo grafico si concentra sulle risorse a disposizione dall'amministrazione, quello che emerge è molti comuni utilizzano dati terzi perché le risorse non permettono la raccolta di alcuni ma non tutti i big data necessari.

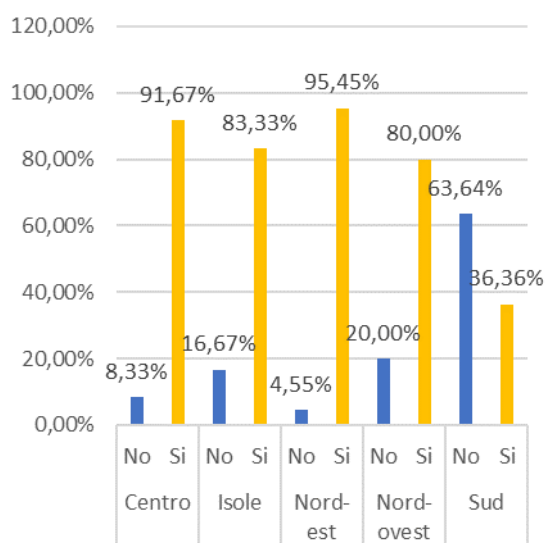


Figura 38 - Nella pubblica amministrazione in cui lavori vengono utilizzati dati prodotti da altre organizzazioni?

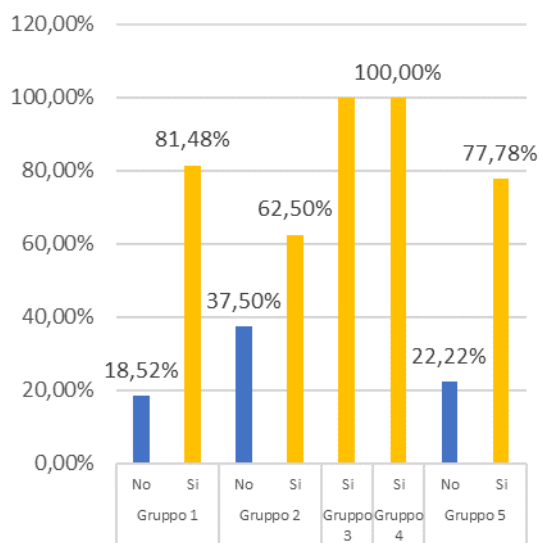


Figura 39 - Nella pubblica amministrazione in cui lavori vengono utilizzati dati prodotti da altre organizzazioni?

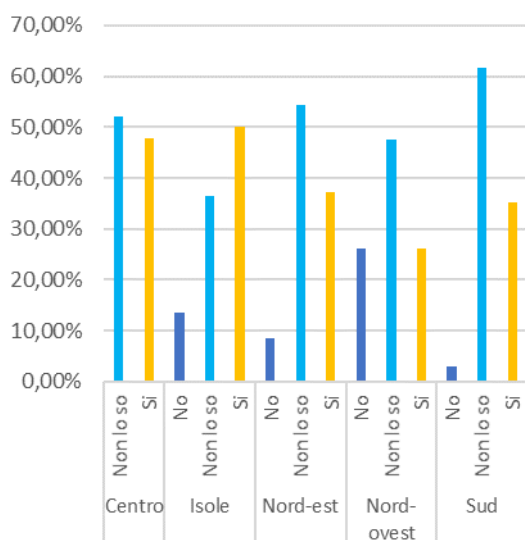


Figura 40- Secondo te i "big data" potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?

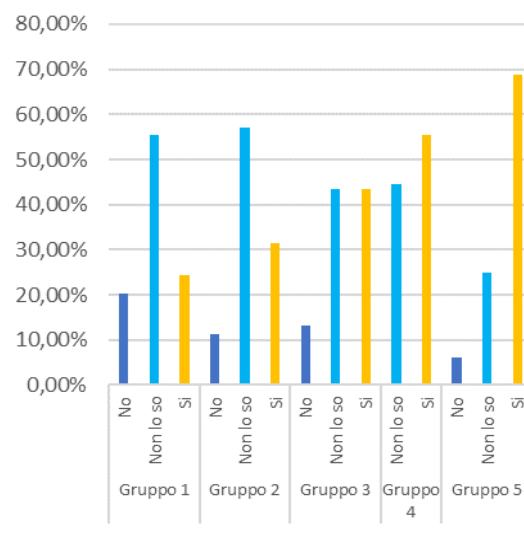


Figura 41 - Secondo te i "big data" potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?

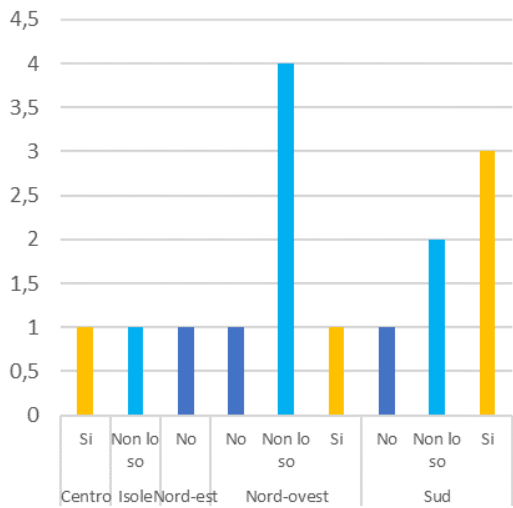


Figura 42 - Secondo te "big data" di altre organizzazioni potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?

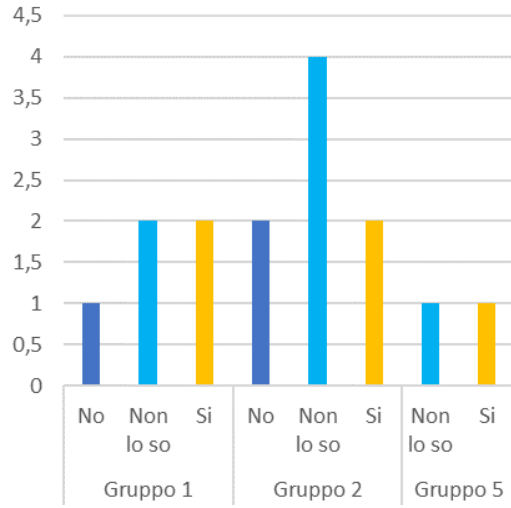


Figura 43 - Secondo te "big data" di altre organizzazioni potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?

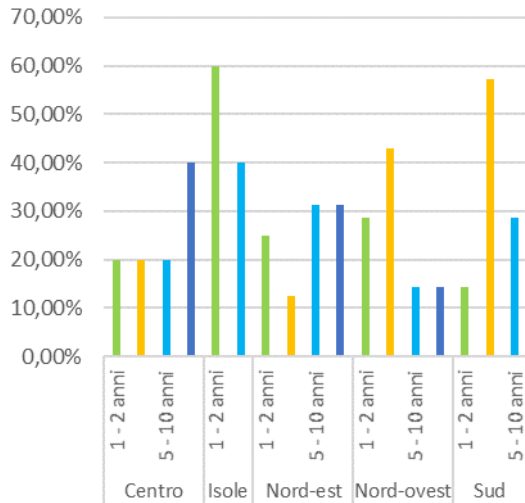


Figura 44 - Da quanto tempo la pubblica amministrazione in cui lavori utilizza "big data"?

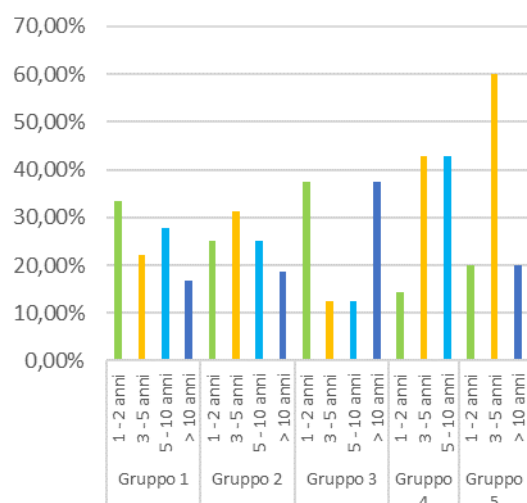


Figura 45 - Da quanto tempo la pubblica amministrazione in cui lavori utilizza "big data"?

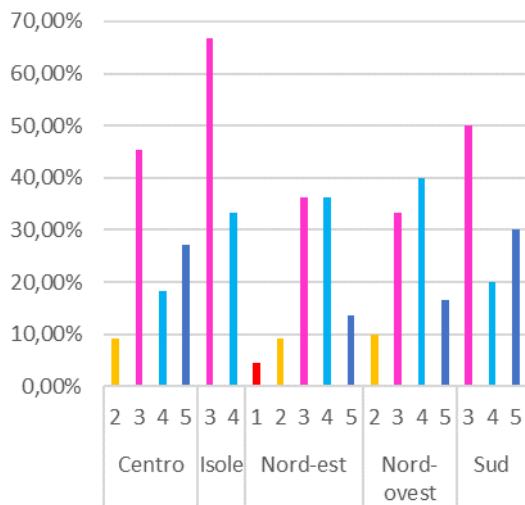


Figura 46 - I "big data" hanno prodotto maggiore efficienza

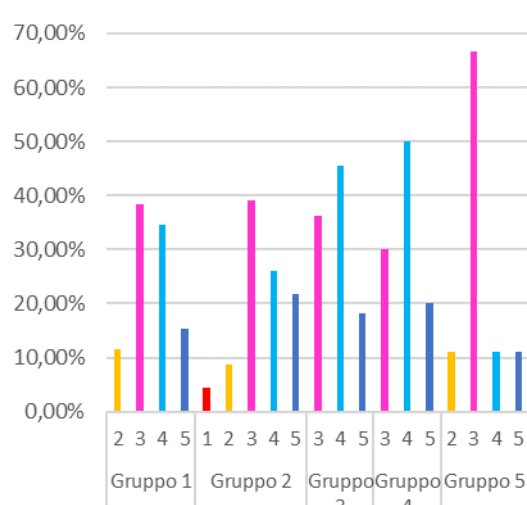


Figura 47 - I "big data" hanno prodotto maggiore efficienza

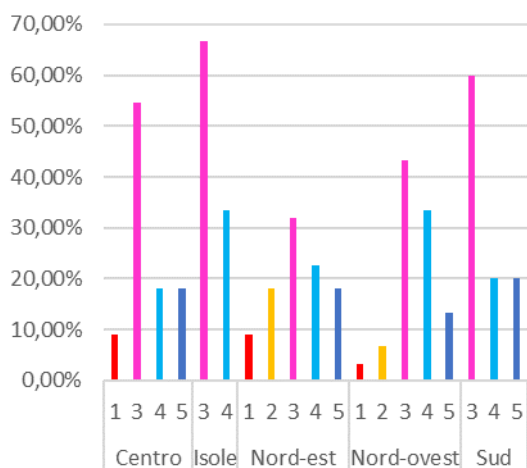


Figura 48 - I "big data" hanno migliorato i processi decisionali

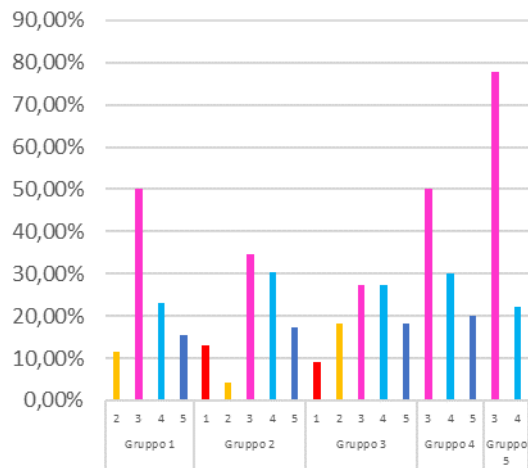


Figura 49 - I "big data" hanno migliorato i processi decisionali

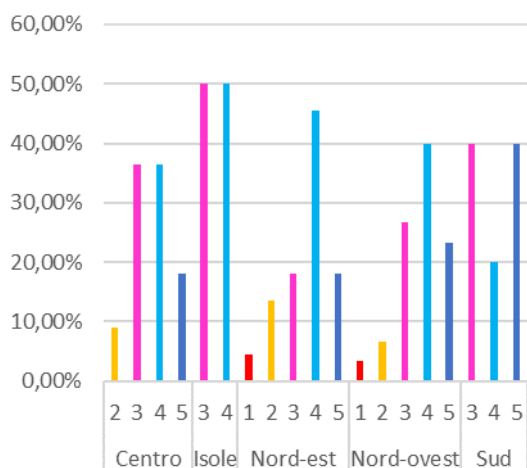


Figura 50 - I "big data" hanno aiutato a migliorare i servizi al cittadino

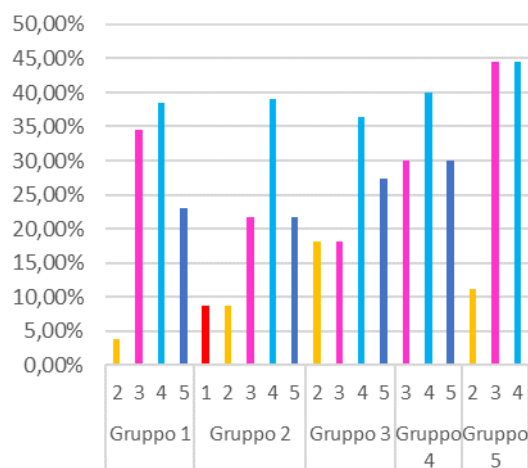


Figura 51 - I "big data" hanno aiutato a migliorare i servizi al cittadino

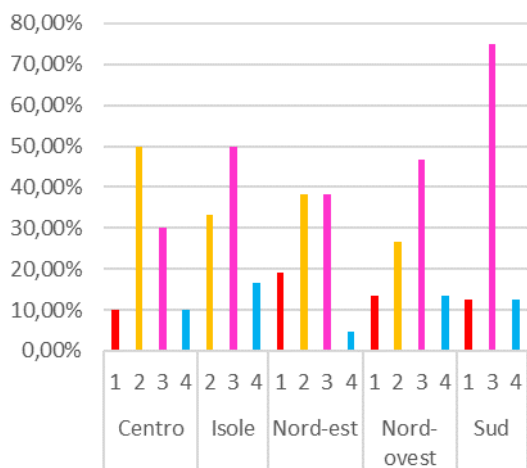


Figura 52 - Credi che l'amministrazione pubblica per cui lavori utilizzi a pieno le potenzialità dei "big data"?

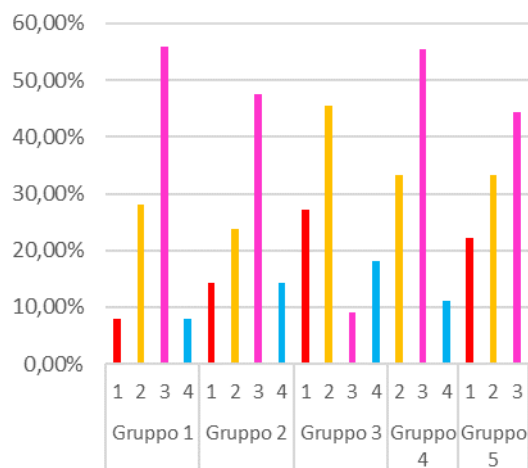


Figura 53 - Credi che l'amministrazione pubblica per cui lavori utilizzi a pieno le potenzialità dei "big data"?

4.1.9. Utilizzo dei big data.

Questa sezione risponde a due quesiti sull'utilizzo dei big data, in particolare la prima domanda riguarda gli obiettivi perseguiti attraverso l'utilizzo di big data. Secondo il campione omogeneo il principale risultato sperato riguarda un migliore rendimento dei servizi. Questa voce appare generica, capace di abbracciare diverse realtà comunali (i servizi offerti spaziano in tutti i settori), ma allo stesso tempo è indicativa di una dinamica positiva, quella di una ricerca del miglioramento del servizio offerto attraverso i dati. Le voci che vengono subito dopo, nettamente staccate dalla prima, riguardano invece l'orientamento degli indirizzi programmatici e risultati nel processo decisionale. Una voce quindi afferma di utilizzare dati nella fase di definizione degli obiettivi e degli indirizzi, rientrando nella sfera politica che precede la presa di decisione, l'altra si pone invece nella fase successiva, quella di valutazione delle decisioni politiche prese in base agli indirizzi definiti dagli stessi dati. Seconda e terza scelta, con percentuali di risposta molto simili, si pongono al principio e al termine del processo, dall'indirizzo alla valutazione finale della strada intrapresa. Segue al quarto posto, l'opzione "opinioni dei cittadini". Da notare che tutte le opzioni che riguardano aspetti di indagine della realtà, come ad esempio l'analisi delle condizioni ambientali, oppure la prestazione delle infrastrutture, rappresentano la minoranza delle scelte. Quindi indirizzi, valutazioni e opinioni sono più importanti nell'utilizzo dei big data rispetto all'analisi di aspetti tecnici. Una voce non presente nel questionario originale, ma che è stata aggiunta a seguito della ridefinizione delle risposte altro è "adempimenti burocratici". Le risposte sono minoritarie, ma rimane il dubbio su quanti altri l'avrebbero scelta se questa voce fosse stata disponibile sin dall'inizio.

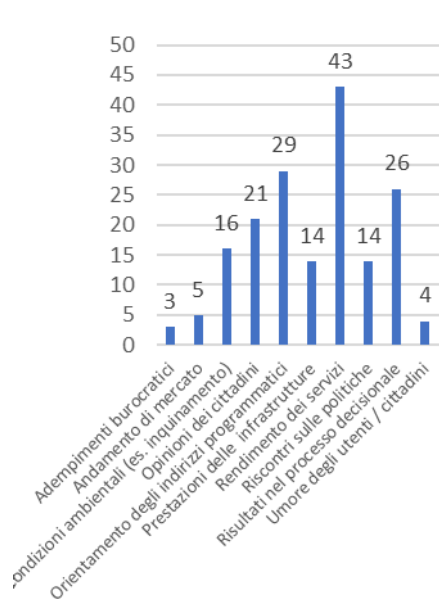


Figura 54 - Quali risultati si cercano dall'analisi dei "big data"?

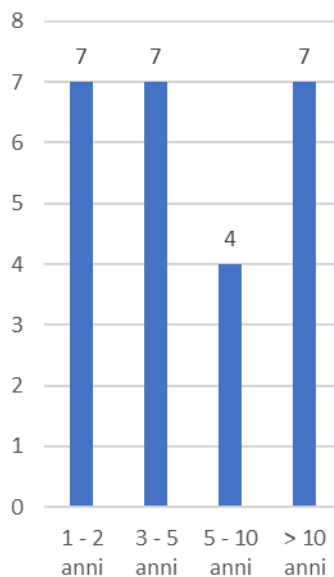


Figura 55 - Da quanto tempo la pubblica amministrazione in cui lavori utilizza "big data"?

Il grafico che mostra da quanti anni vengono utilizzati i big data nel comune, pare poco indicativo considerando il campione omogeneo e cercando di fornire una risposta a livello statale. Infatti, le risposte si distribuiscono abbastanza omogeneamente. Anche considerando le differenze su base abitanti non ci sono particolarità, se non un leggero vantaggio da parte dei comuni più popolosi, che tendono a utilizzare i big data da più tempo rispetto agli altri. La particolarità più evidente riguarda il centro Italia, che si differenzia dagli altri perché la maggior parte dei comuni afferma di utilizzare i big data da oltre 10 anni. Al contrario nelle Isole e nel Nord-Ovest la maggior parte

dei comuni utilizza big data da meno di cinque anni. Nelle isole il 60% di chi risponde afferma di farlo da meno di due anni.

4.1.10. Rendimento dei big data.

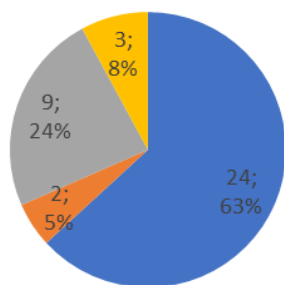
Questa sezione fornisce uno sguardo ancora poco analizzato dal questionario, la valutazione degli impatti per capire se l'utilizzo è stato utile per il comune. Secondo il campione omogeneo tutti i risultati sono stati sufficienti. Il punteggio più alto è stato ricevuto rispetto al miglioramento dei servizi al cittadino, il risultato medio è 3,7 su 5. Come in precedenza è stata utilizzata una scala di valori da uno a cinque dove uno significa per nulla d'accordo e cinque molto d'accordo. La voce che ha una media più bassa, ovvero in cui i comuni sono meno d'accordo, riguarda il pieno utilizzo delle potenzialità dei big data. Il risultato è appena sufficiente, nessun comune ha selezionato cinque (molto d'accordo) e 14 votanti ritengono che l'utilizzo della potenzialità sia ancora insufficiente. Con il senno di poi questi argomenti (in una forma interrogativa) sarebbe stato bene proporli a coloro che non utilizzano big data, questo avrebbe potuto determinare un interesse oppure uno scarso interesse sul tema big data, non solo da parte di chi già ne fa uso. Analizzando singolarmente le risposte, i comuni affermano di essere d'accordo sulla produzione di maggiore efficienza. Non ci sono particolari differenze considerando i criteri di popolazione e geografia separatamente. Lo stesso vale per il miglioramento dei processi decisionali, in questo caso però le risposte sono state leggermente più tiepide, con una leggera tendenza da parte dei piccoli comuni ad avere una considerazione più negativa. Sono i comuni del quarto e quinto gruppo ad essere più decisi nell'affermare che i processi decisionali hanno subito miglioramenti. Considerando le macroregioni, questa visione positiva l'hanno avuta in particolare le ragioni del Sud Italia e delle Isole. La stessa tendenza per le stesse categorie, quindi Sud Italia e Isole e per grandi comuni, si è registrata a proposito del miglioramento dei servizi al cittadino. Per quanto riguarda l'utilizzo pieno delle potenzialità dei dati le maggiori differenze hanno riguardato il gruppo 4, nettamente più d'accordo degli altri gruppi. Mentre tra le macroregioni il primato positivo appartiene alle Isole.

I "big data" hanno prodotto maggiore efficienza	3,6
I "big data" hanno migliorato i processi decisionali	3,2
I "big data" hanno aiutato a migliorare i servizi al cittadino	3,7
Credi che l'amministrazione pubblica per cui lavori utilizzi a pieno le potenzialità dei "big data"?	2,6

4.1.11. Gestione dei big data.

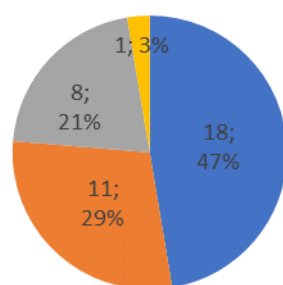
La dinamica chiara che emerge da questo paragrafo è certamente la mancanza di una struttura per la gestione dei big data a livello comunale. I quesiti posti non indagano la presenza di una maggiore organizzazione a livello provinciale, regionale o statale, cosa che parrebbe abbastanza ovvia, ma si limitano a verificare questo aspetto all'interno del comune. Alla prima domanda la maggioranza assoluta (63%) risponde che non esistono gruppi o nemmeno persone specifiche deputate ai big data. A proposito di questo sembra evidente che si generi un problema di competenza. Non tutti sono in grado di gestire questo tipo di dati, e poche persone sono in grado di seguire l'iter interamente, dalla raccolta alla restituzione grafica. Perciò questo grafico appare

critico, a meno che, chi afferma la mancanza di struttura organizzativa, non consideri la presenza di una struttura a livelli superiori di governo. Questo potrebbe spiegare almeno in parte, un “utilizzo senza personale dedicato”. Secondo il 37% una struttura esiste, e tra loro il 24% dei comuni la identifica in un ufficio o settore. Considerando ora le comparazioni tra i due criteri si nota che la strutturazione è molto presente nei grandi comuni del gruppo 5, mentre tra le macroregioni le più organizzate sembrano essere il centro e le Isole.



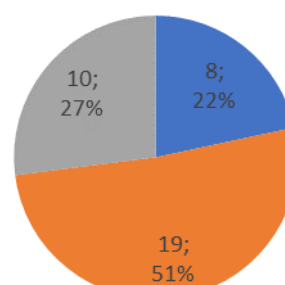
- Non sono definite persone o team
- Un solo tecnico
- Un ufficio / settore
- Una team di lavoro

Figura 56 - Chi si occupa dei "big data" dove lavori?



- No
- Non lo so
- Si (come parte di una strategia generica sui dati e l'informatica)
- Si (esiste una strategia specifica sugli open data)

Figura 57 - Esiste una strategia definita per l'utilizzo degli OD?



- No
- Si
- Si (come parte di una strategia generica sui dati e l'informatica)

Figura 58 - Esiste una struttura per il trattamento dei dati confidenziali?

Oltre una struttura, quello che viene indagato è la presenza di una strategia di apertura dei dati, ovvero di creazione, pulizia e rilascio di open data. Solo un quarto dei comuni prevede open data, e anche questa volta i più preparati in questo senso sono i comuni del gruppo 5. Solo i grandi comuni posseggono portali open data e una struttura in grado di gestire questo processo, ma vale anche in questo caso l'ipotesi secondo cui per i piccoli comuni esiste una struttura a livello quantomeno regionale che permetta le stesse operazioni. Molti sono anche i comuni che dichiarano di non sapere dell'esistenza di questa strategia, nonostante abbiano affermato di utilizzare i big data nel proprio comune. Chiaramente l'ipotesi che esista una struttura regionale che supporti i piccoli e medi comuni nell'uso dei big data va approfondita. Allo stesso modo la differenza, se esiste, tra le diverse regioni italiane.

Anche riguardo la presenza di una struttura di valutazione delle fonti la tipologia di risposte varia poco. La maggior parte risponde in maniera negativa, si assottiglia di poco il fronte dei no, mentre aumenta chi afferma la presenza di una struttura, ma soltanto generica interna a una più generale sull'IT.

Giunti a questo punto sembra quasi sorprendente il grafico che riguarda il trattamento dei dati confidenziali e della privacy. La maggioranza assoluta risponde che esiste una struttura in grado di gestire informazioni personali e privacy, e per un altro 27% sono tenuti in considerazione all'interno di una strategia più ampia sull'informatizzazione della pubblica amministrazione. Solo il 22% risponde negativamente. A proposito di privacy, le macroregioni più avanzate da questo punto di vista risultano il Sud e le Isole, senza nemmeno una risposta negativa. In tutte le macroregioni le risposte affermative sono maggioritarie, ma la percentuale più alta di no è quella del Nord-Est. Nello stesso modo rispetto ai tre grafici precedenti si presentano le differenze rispetto ai gruppi. Anche in questo caso i comuni del quarto e del quinto gruppo possiedono una struttura, nemmeno una risposta negativa da questi due gruppi.

4.1.12. Livello di data Literacy.

Le domande che hanno fatto parte di questa sezione sono servite come una sorta di auto valutazione del livello di data literacy nel proprio comune. I risultati non sono esaltanti, una sola voce ha raggiunto la sufficienza piena, tutte le altre no. Il voto finale infatti è un sei meno. Le voci più critiche sono rappresentate dalla capacità di utilizzare i dati nella risoluzione di problemi correnti e nell'utilizzo di strumenti tecnici avanzati. Conoscenza dei dati e interpretazione delle informazioni sfiorano la sufficienza, ma ancora una volta, come dimostrato nel paragrafo precedente, la privacy e la riservatezza nel trattamento dei dati personali ad alzare la media.

Capacità di valutare l'affidabilità e l'imparzialità delle fonti	2,8
Conoscenza di cosa è un dato, come si raccoglie, analizza, visualizza e condivide	2,9
Capacità di interpretare le informazioni derivate dai dati	2,9
Capacità di usare i dati per prendere decisioni, risolvere problemi correnti	2,7
Capacità di usare strumenti tecnici avanzati nelle diverse fasi del processo	2,7
Conoscenza di come trattare la privacy e la sicurezza dei dati	3,2

Per quanto riguarda conoscenza, capacità di interpretazione e capacità di utilizzo esiste una marcata differenza tra i comuni sotto i 15 mila abitanti rispetto ai più popolosi. Questi ultimi registrano una media più alta di data literacy. Non vale lo stesso quando si considera utilizzo di strumenti avanzati, che secondo i comuni del gruppo cinque non è affatto sufficiente.

Il dato certamente più evidente è che si può considerare valido per entrambe le categorie esaminate, è che la maggior parte dei comuni vota tre, si posiziona esattamente al centro. La data literacy italiana è mediocre, lievemente insufficiente a detta degli stessi comuni intervistati. In aggiunta si consideri che il dato non considera quei comuni che non utilizzano big data (non classificati), ma che certamente abbasserebbe la media di data literacy complessiva dei comuni italiani.

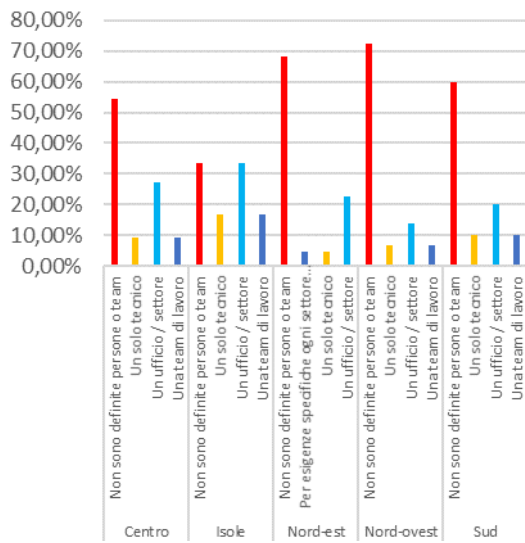


Figura 59 - Chi si occupa dei "big data" dove lavori?

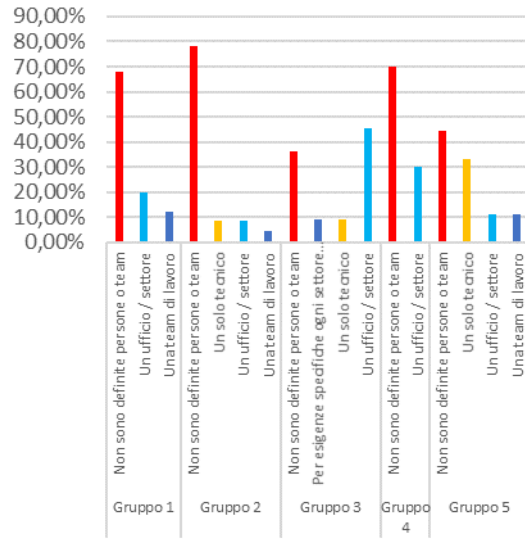


Figura 60 - Chi si occupa dei "big data" dove lavori?

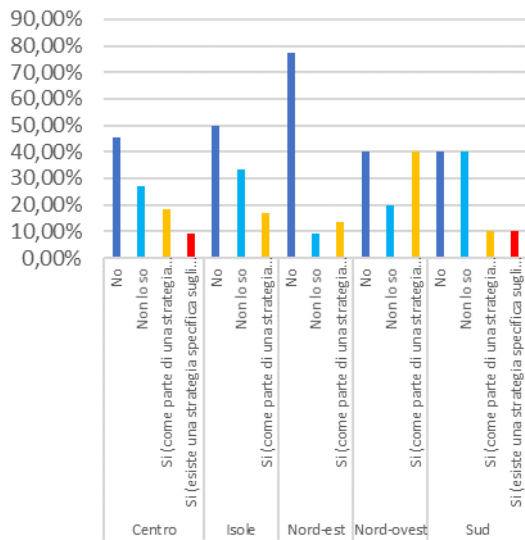


Figura 61 - Esiste una strategia definita per l'utilizzo degli OD?

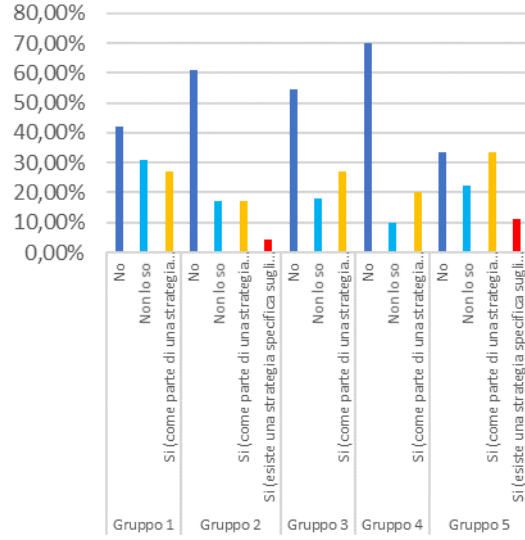


Figura 62 - Esiste una strategia definita per l'utilizzo degli OD?

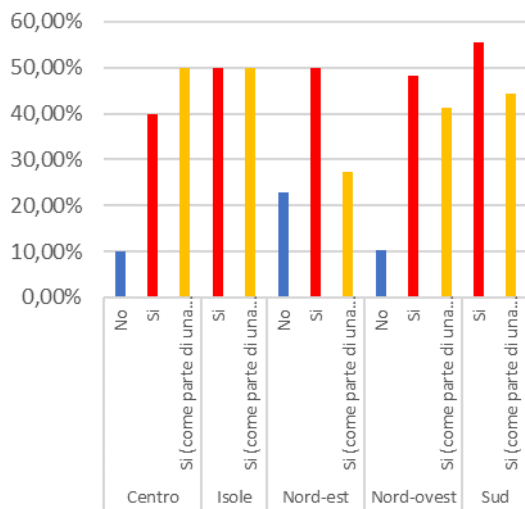


Figura 63 - Esiste una struttura per il trattamento dei dati confidenziali e della privacy?

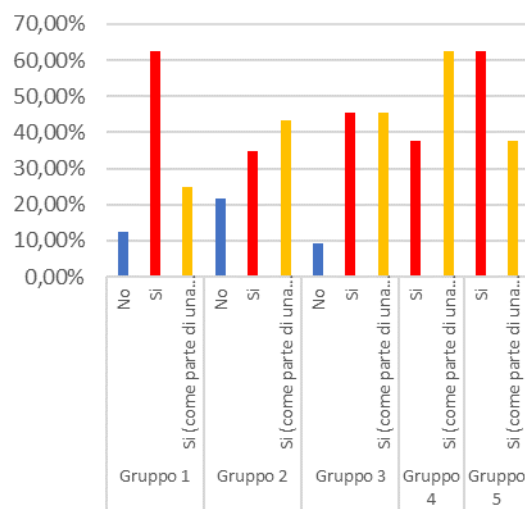


Figura 64 - Esiste una struttura per il trattamento dei dati confidenziali e della privacy?

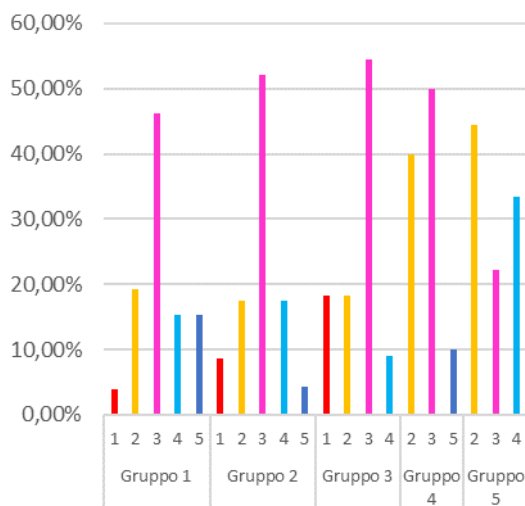


Figura 65 - Conoscenza di cosa è un dato, come si raccoglie, analizza, visualizza e condivide

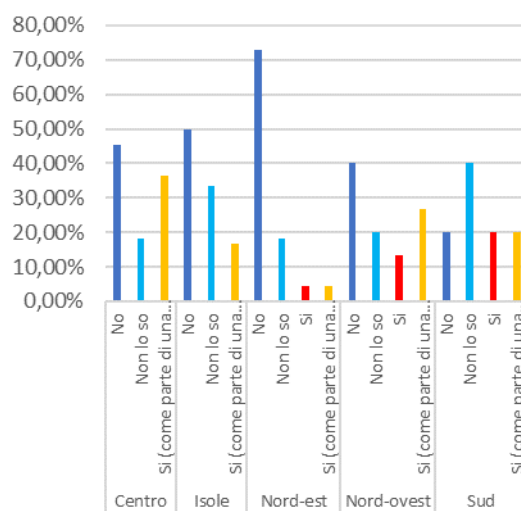


Figura 66 - Esiste una struttura di valutazione delle fonti, dei processi e dei risultati dei "big data"

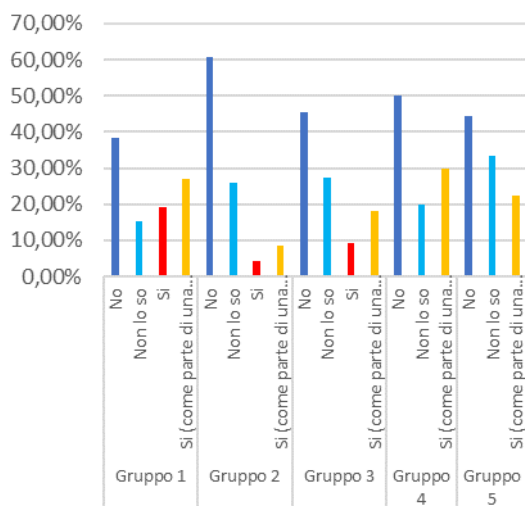


Figura 67 - Esiste una struttura di valutazione delle fonti, dei processi e dei risultati dei "big data"

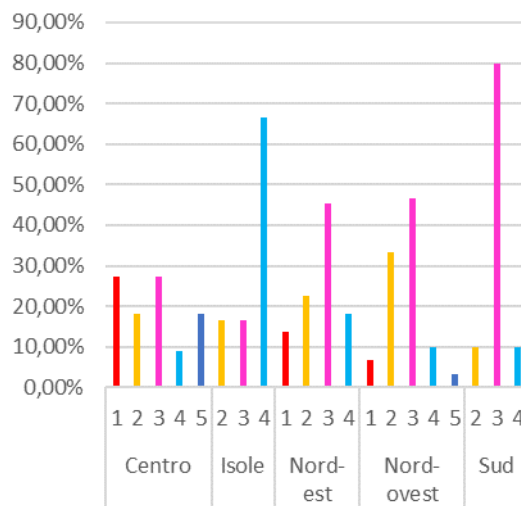


Figura 68 - Capacità di valutare l'affidabilità e l'imparzialità delle fonti

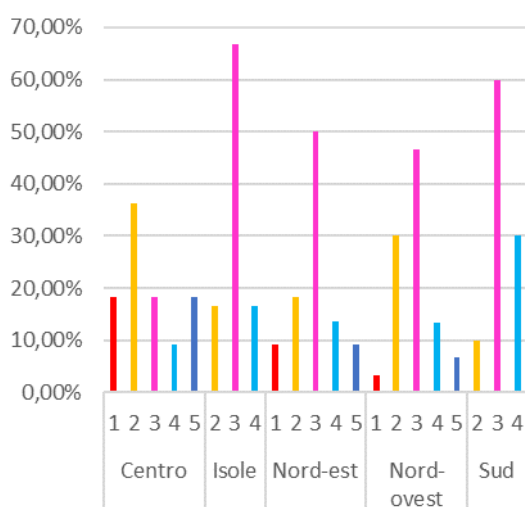


Figura 69 - Conoscenza di cosa è un dato, come si raccoglie, analizza, visualizza e condivide

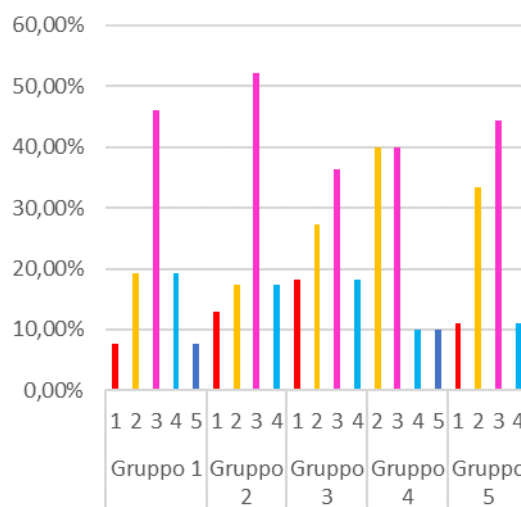


Figura 70 - Capacità di valutare l'affidabilità e l'imparzialità delle fonti

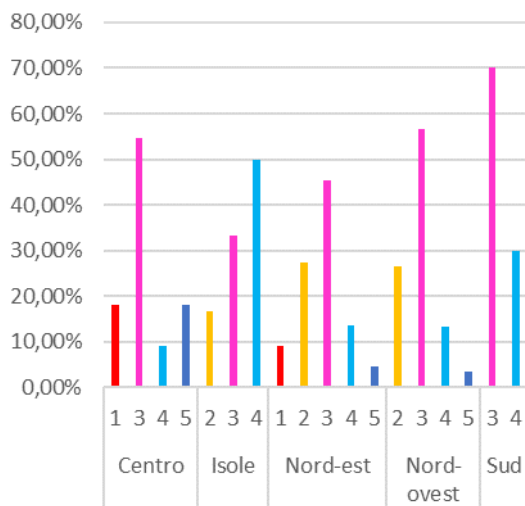


Figura 71 - Capacità di interpretare le informazioni derivate dai dati

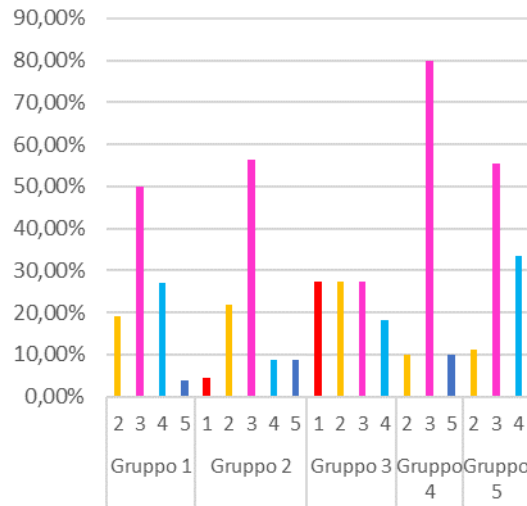


Figura 72 - Capacità di interpretare le informazioni derivate dai dati

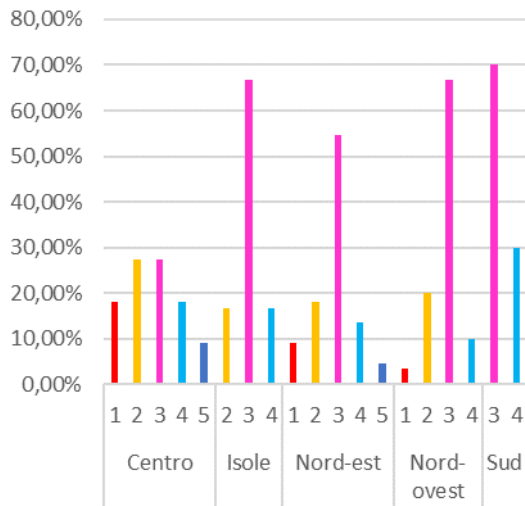


Figura 73 - Capacità di usare i dati per prendere decisioni, risolvere problemi correnti

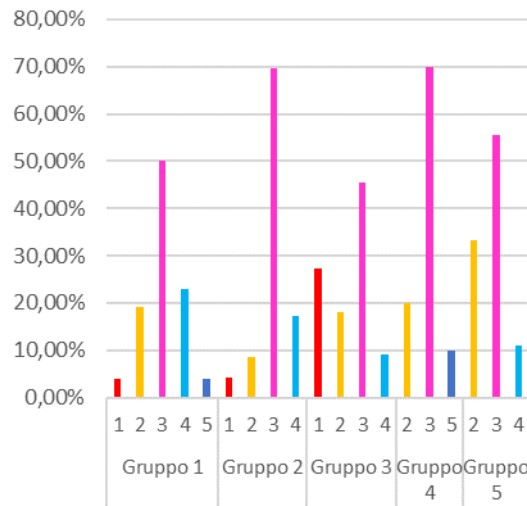


Figura 74 - Capacità di usare i dati per prendere decisioni, risolvere problemi correnti

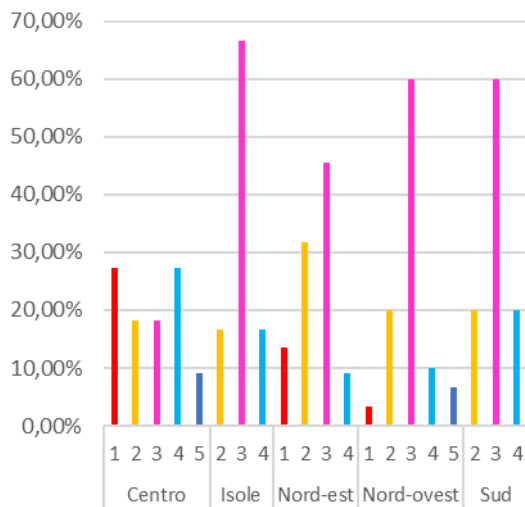


Figura 75 - Capacità di usare strumenti tecnici avanzati nelle diverse fasi del processo

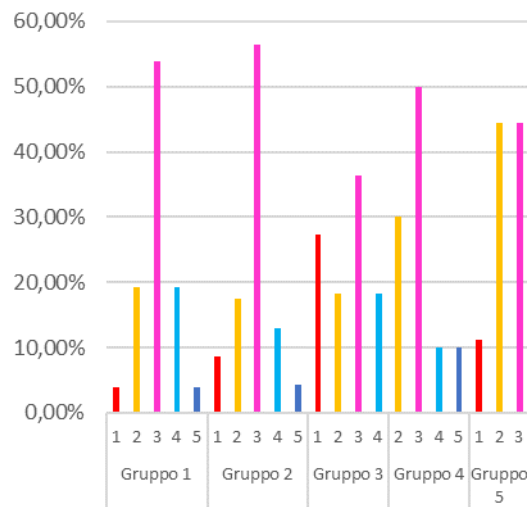


Figura 76 - Capacità di usare strumenti tecnici avanzati nelle diverse fasi del processo

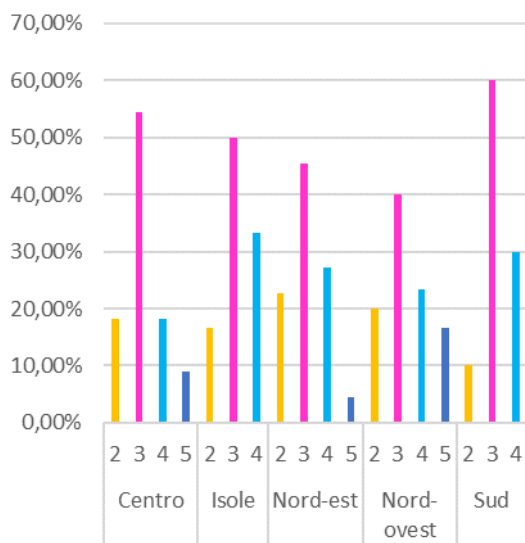


Figura 77 - Conoscenza di come trattare la privacy e la sicurezza dei dati

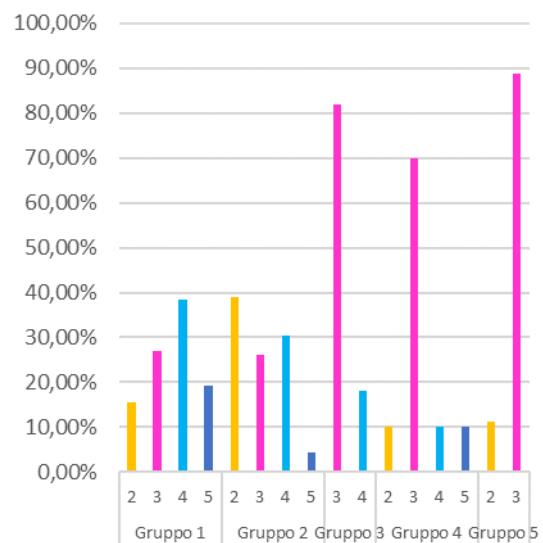


Figura 78 - Conoscenza di come trattare la privacy e la sicurezza dei dati

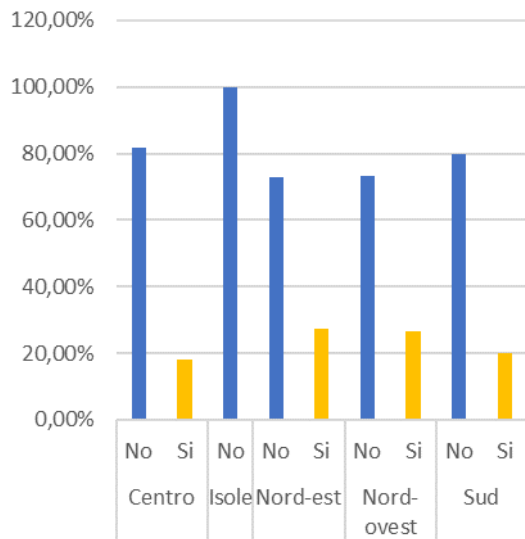


Figura 79 - La pubblica amministrazione dove lavori ha organizzato corsi di formazione per facilitare l'utilizzo dei "big data"?

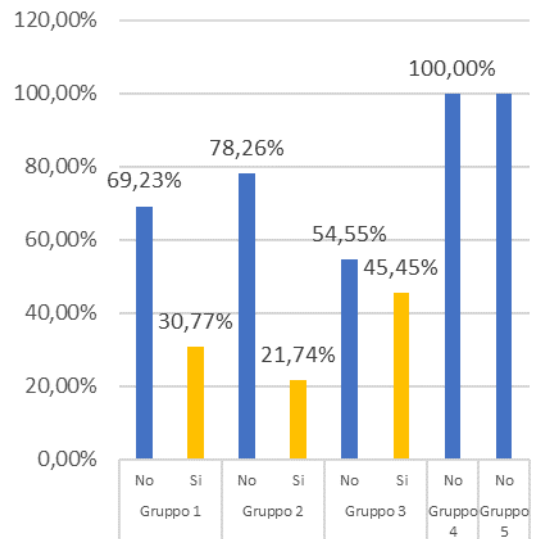


Figura 80 - La pubblica amministrazione dove lavori ha organizzato corsi di formazione per facilitare l'utilizzo dei "big data"?

4.1.13. Formazione sui big data.

Per incrementare il dato scoraggiante del precedente paragrafo, il metodo più adatto consiste in una adeguata formazione, utile anche per rimanere al passo con i costanti progressi del campo big data. Sono poco più di un quarto i comuni che mettono a disposizione corsi di formazione che facilitino l'uso dei big data. Se si considerano a questo punto gli argomenti più frequenti nei corsi, si spiega in parte la positiva conoscenza degli argomenti di privacy e sicurezza dei dati, la stragrande maggioranza dichiara di aver frequentato corsi proprio di questi argomenti.

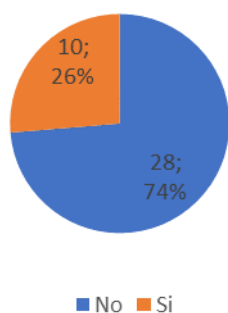


Figura 81 - La pubblica amministrazione dove lavori ha organizzato corsi di formazione per facilitare l'utilizzo dei "big data"?

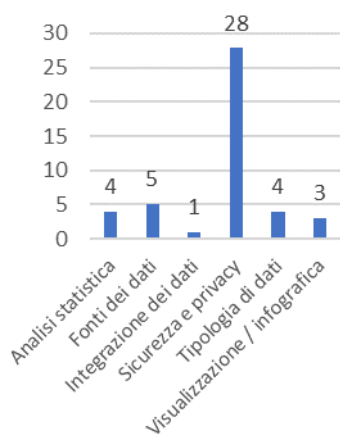


Figura 82 - Argomento del corso

In generale si scopre che la formazione dei dipendenti è assolutamente scarsa, i comuni dovrebbero dare maggiore accesso a corsi di formazione per qualificare i propri dipendenti. Non ci sono grosse differenze tra le macroregioni in tal senso, mentre si nota ancora una volta la differenza tra i gruppi 4 e 5 rispetto al resto dei comuni, questa volta il verdetto sembra ribaltarsi, i pochi corsi di formazione sono stati organizzati nei comuni con meno di 15 mila abitanti.

4.1.14. Barriere all'utilizzo di big data.

Ho posto questo paragrafo al termine, per non interrompere il processo del questionario. Fin qui le domande hanno proseguito in maniera regolare, con argomenti simili, spesso consecutivi, in sequenza. La parte relativa alle barriere dei big data esula da questo processo, si tratta di una serie di domande che ha come intento quello di verificare i motivi principali per cui i big data non vengono valorizzati internamente al comune. Il grado della barriera può essere totale, nel caso di quei comuni che non utilizzano assolutamente big data, o può essere parziale, nel caso di quei comuni che utilizzano big data, ma non utilizzano dati provenienti da altre organizzazioni. Se teniamo presente la definizione di big data, diventa riduttivo pensare che l'utilizzo venga limitato ai dati prodotti internamente, l'integrazione con diverse banche dati, che necessariamente provengono anche da fonti esterne, determinano un altro livello nell'utilizzo di big data.

Queste domande sono state poste solamente a quei comuni che non utilizzano big data nei propri processi interni. La prima domanda riguarda le motivazioni che adducono a quanto appena detto. La prima motivazione, certamente la più rappresentativa tra le opzioni fornita riguarda la mancanza di risorse interne competenti. Come riscontrato nei paragrafi precedenti, si pone una questione di competenza, che era anche stata l'ipotesi a inizio tesi e che in questo grafico trova una certa dimostrazione. Questa voce è stata selezionata oltre il doppio delle volte rispetto alle voci che seguono. La gara per la seconda posizione si gioca tra tre opzioni, costi, obiettivi e small

data, che risultano più o meno allo stesso livello. La spunta di una manciata di voti la motivazione per cui piccole quantità di dati sono sufficienti per i bisogni attuali del comune. In questo caso si determina una negazione completa del fatto che i big data possano risultare efficaci durante i processi comunali. Non vengono utilizzati perché non sono utili, perché pochi dati sono più che sufficienti per i servizi offerti. Questa risposta che rappresenta la seconda scelta tra coloro che hanno risposto nel campione omogeneo, è la più negativa, perché rispetto alle altre, indica un uso superfluo e quindi inutile. Senza la consapevolezza dell'importanza dei big data, non succederà mai che essi vengano utilizzati, nemmeno presi in considerazione, e quando imposti da un organo superiore, percepiti come puro adempimento burocratico.

Infatti, collegandoci alla domanda successiva il risultato riprende la dinamica precedente. La maggioranza assoluta, con il 55% dei voti, dichiara di non sapere se i big data siano realmente utili per la pubblica amministrazione. Mentre solo il 31% risponde positivamente. Infine, secondo il 14% dei comuni che non li utilizzano, i big data non sono utili. Molto importante per chi si occupa di agenda digitale, è lavorare su quella percentuale che ignora le potenzialità dei big data. Per questo grafico è stata effettuata la comparazione da cui emerge una crescita progressiva della percentuale di risposte positive al crescere della dimensione del comune (in termini di popolazione). I grandi comuni sono molto più consapevoli dell'utilità dei big data che i piccoli comuni. Nel Nord-Ovest invece si concentrano le risposte negative, che superano il 25%.

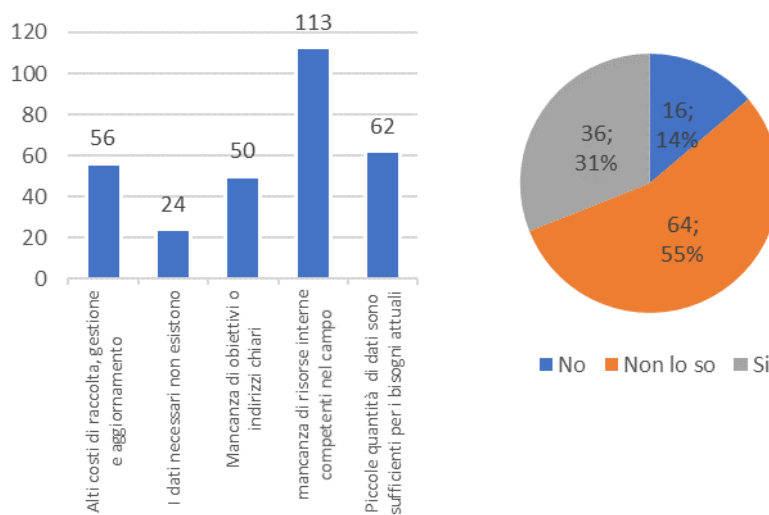


Figura 83 - Secondo te perchè la pubblica amministrazione in cui lavori non utilizza "big data"?

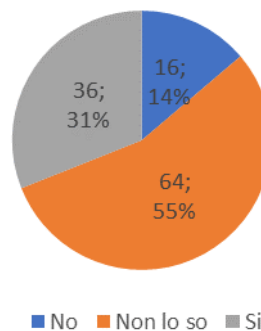


Figura 84 - Secondo te i "big data" potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?

4.1.15. Barriere all'utilizzo di dati terzi.

Le domande di questa sezione interessano i comuni che utilizzano big data internamente, ma che dichiarano di utilizzare solo dati prodotti in casa, e non dati di terze organizzazioni. Le domande tentano di definire il perché questo accade. Come mostra il paragrafo la maggior parte specifica di non sapere perché l'utilizzo è limitato ai dati prodotti dal comune stesso. Al secondo posto con il 29% dei voti i comuni affermano di non usare dati terzi perché non disponibili in formato open. In questo caso la questione si pone in termini di risorse, cioè i dati non vengono utilizzati perché non sono allocabili risorse per acquistarli. Solo un voto riguarda la mancanza di competenze per poter gestire questi dati. Infine, alla domanda sull'utilità dell'utilizzo di dati terzi, il 57% del campione risponde positivamente.

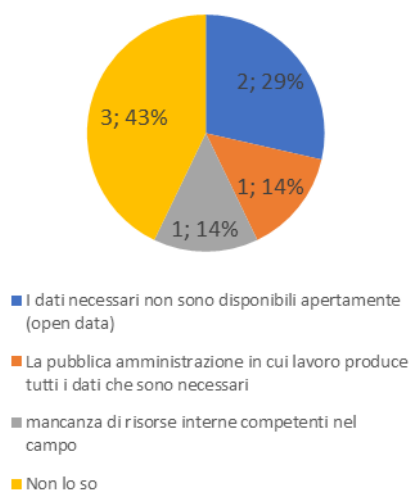


Figura 85 - Secondo te perchè la pubblica amministrazione in cui lavori non utilizza dati di altre organizzazioni?

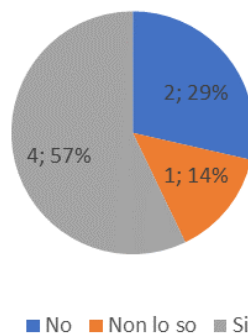


Figura 86 - Secondo te "big data" di altre organizzazioni potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?

5. Conclusioni

Dall'ultimo aggiornamento svolto il 05 Aprile 2019 sono state ricevute 322 risposte da 297 comuni italiani. Dopo aver raggiunto i comuni tramite pec le partecipazioni sono incrementate sensibilmente, dei 1.878 comuni italiani contattati hanno risposto quasi il 16%. La possibilità di partecipare non è chiusa, si spera di ottenere altre risposte dai comuni che ancora non hanno svolto il questionario. Inoltre, è possibile inviare mail ai comuni non ancora contattati. Il numero di questionari raccolti può essere considerato un campione numericamente valido secondo i parametri statistici definiti nel capitolo 3.

L'elaborazione dei risultati può essere migliorata ed integrata considerando altri aspetti per ora trascurati. Potrà essere valutato l'emergere di differenze tra le regioni oltre che tra macroregioni, migliorando lo studio del grado di digitalizzazione - IT regionale e della normativa ad essa associata. Importante sarebbe considerare l'utilizzo di big data da parte dei capoluoghi regionali e delle città metropolitane, di cui si iniziano ad ottenere alcune risposte. Parallelamente andrebbe studiato nel dettaglio l'universo dei piccoli comuni, ponendo attenzione alle pratiche più innovative e facilmente replicabili.

Un ulteriore sviluppo riguarda i comuni che rispondendo si sono messi a disposizione per essere ricontattati per eventuali interviste che chiariscano il rapporto del comune con i big data e il livello di data literacy.

Conclusioni che riguardano la data literacy.

Il criterio più importante per definire la data literacy si è rivelato l'abilità di comprendere i dati per come sono, mantenendo uno sguardo d'insieme. L'organizzazione presente nel comune insieme alle competenze tecniche e politiche presenti nel comune si sono dimostrate, come ipotizzato, rilevanti.

Competenze specifiche nel campo informatico sono risultate utili ma non indispensabili, come esemplificano le risposte relative alla privacy, che risulta il settore in cui si mostra più competenza. La necessità di adempiere a una normativa comporta la ricerca di competenze, anche esterne alla PA, insieme a una formazione adeguata.

I dati o i big data utilizzati maggiormente risultano quelli prodotti dallo stesso comune, certamente l'utilizzo di dati provenienti da organizzazioni terze offre un incremento delle potenzialità, ma è ristretto a causa dei costi, quindi dalla mancanza di risorse.

Conclusioni che riguardano l'utilizzo dei big data.

Le differenze tra macroregioni esistono e riguardano alcuni temi in particolare. Per una valutazione più efficace di questo fenomeno si attendono ulteriori risposte che vadano a rinforzare il campione soprattutto nel Sud Italia e nelle Isole. Per lo stesso motivo è ancora difficile studiare fenomeni nelle singole regioni.

Le differenze tra gruppi di comuni, ovvero considerando i comuni suddivisi per fasce di popolazione ha restituito i risultati più interessanti. Come ipotizzato è stato possibile dimostrare che i comuni del gruppo 5 (oltre 30.000 abitanti) possiedono un livello di data literacy più alto degli altri. Questo determina una criticità rilevante se si pensa che i comuni con una popolazione inferiore ai 15.000 abitanti sono oltre i tre quarti del totale.

I due criteri utilizzati nella definizione del campione e nella fase preliminare di ipotesi si sono rilevati utili mostrando importanti differenze, dando evidenza statistica nell'utilizzo di big data.

Un utilizzo chiaro e consapevole dei big data emerge soltanto da un numero ristretto di comuni. Ci sono ancora troppi comuni in particolare i piccoli (e quindi la maggior parte) che non utilizza big data.

Bibliography

- A. Longo, R. Natale. (2018, May 26). *GDPR, tutto ciò che c'è da sapere per essere in regola*. Tratto da Agenda Digitale: <https://www.agendadigitale.eu/cittadinanza-digitale/gdpr-tutto-cio-che-ce-da-sapere-per-essere-preparati/>
- AGID. (2019, 03 31). *Informazioni sul catalogo*. Tratto da Basi di dati della pubblica amministrazione: <http://basidati.agid.gov.it/catalogo/informazioni.html>
- Alissa Black, Rachel Burstain. (2013). Local Scale and Local Data. In L. D. Brett Goldstein, *Beyond Transparency: Open Data and the Future of Civic Innovation* (p. 173-183). San Francisco: Code for America Press.
- Batty, M. (2013). Big data, smart cities and city planning. In *Dialogues in Human Geography* (p. 274-279). SAGE journals.
- Big Data*. (2019, 03 30). Tratto da IT Glossary: <https://www.gartner.com/it-glossary/big-data>
- Boyd D., Crawford C. (2012). Critical questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 662-679.
- Bruun E., Givoni M. (2015). Sustainable mobility: Six research routes to steer transport policy. *Nature*, 29-31.
- Calzada Prado J, Marzal M. A. (2013). *Incorporating data literacy into information literacy programs: Core competencies and contents*.
- Carlson J, Fosmire M, Miller CC, et al. (2011). *Determining data information literacy needs: A study of students and research faculty*. Libraries and the Academy.
- Cecconi Gianfranco, Radu Cosmina. (2018). *Open Data Maturity in Europe*. European Data Portal.
- Chaoming Song, Zehui Qu, Nicholas Blumm, Albert-László Barabási. (2010, February 18). Limits of Predictability in Human Mobility. *Science*, p. 1018-1021.
- Crusoe, D. (2016). Data Literacy defined pro populo. *The Journal of Community Informatics*, p. 27-46.
- Data & Technical innovation deliverables*. (2019, March 30). Tratto da PoliVisu EU Project: <https://www.polivisu.eu/public-deliverables>
- Dennedy, M. (2018). Data Cocktail. *Wired*, 65.
- Eaton C., Deroos D., Deutsch T., Lapis G., Zikopoulos P. (2012). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. New York: McGraw Hill.
- FluidSurveys, T. (2014, Jul 8). *Calculating the Right Survey Sample Size*. Tratto da Fluid Surveys University: <http://fluidsurveys.com/university/calculating-right-survey-sample-size/>
- Giovanni Vecchio, Pavel Kogut. (2018). *Data literacy survey report*. Milan: PoliVisu.
- Kitchin, R. (2014, February). The real-time city? Big data and smart urbanism. *GeoJournal*, p. 1-14.

- Koltay, T. (2017). *Data literacy for researchers and data librarians*. Journal of Librarianship and Information Science.
- Mandinach EB, Gummer ES. (2013). *A systemic view of implementing data literacy in educator preparation*. Educational Researcher 42.
- Mayer-Schonberger, V. (2018). Big data. *Wired*, 45-48.
- Michael Flowers, Beth Blauer. (2013). Driving Decision with data. In L. D. Brett Goldstein, *Beyond Transparency. Open Data and the future of civic innovation* (p. 183-211). San Francisco: Code for America Press.
- OECD. (2017). *Skills for a High Performing Civil Service*. Paris: OECD Public Governance Reviews.
- Open Definition 2.1*. (2019, 03 30). Tratto da Open Definition: <http://opendefinition.org/od/2.1/en/>
- Paola Pucci, Giovanni Vecchio, Grazia Concilio. (2018). *Experimental dimension of policy making*. Milan: PoliVisu.
- Poorthuis A., Zook M. (2017). Making Big Data Small: Strategies to Expand Urban and Geographical Research Using Social Media. *Journal of Urban Technology*, 115-135.
- Qin J, D'Ignazio J. (2010, June). *Lessons learned from a twoyear experience in science data literacy education*. Tratto da Proceedings of the 31st annual IATUL conference: <http://docs.lib.purdue.edu/iatul2010/conf/day2/5>
- Rabary C., Stoper M. (2014). The digital skin of cities: urban theory and research in the age of the sensed and metered city, ubiquitous computing and big data. *Cambridge Journal of Regions, Economy and Society, Volume 8*, 27-42.
- S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives. (2007). DBpedia: A Nucleus for a Web of Open Data. In C. B. Sören Auer, *The Semantic Web* (p. 722-735). Springer.
- Shneider, R. (2013). Research data literacy. *Proceedings of the 1st European Conference on Information Literacy (ECIL)* (p. 134-150). Istanbul: Springer Communication.
- Sopra Steria. (2016). *Government Digital Trends Survey*. Hemel Hempstead: Sopra Steria.
- Task Team on Skills, Training and Capacity Building. (2015). *Analysis of Big Data Survey 2015 on Skills, Training and Capacity Building*. UN Global Working Group on Big Data for Official Statistics.
- W. Tortorella, G. Marinuzzi, F. Chiovoloni, T. Ulivieri. (2015). *Il personale dei comuni italiani*. Roma: IFEL fondazione ANCI.
- Wylie, C. (2018). Colonialismo. *Wired*, 71-73.

Allegato 1 - Questionario

Benvenuto o benvenuta!

Grazie del contributo che stai per dare a questa ricerca. Il questionario cerca di investigare il grado di "alfabetizzazione dei dati" nelle Pubbliche Amministrazioni italiane.

Per maggiori informazioni sul progetto o per problemi tecnici, puoi contattare il seguente indirizzo mail:

francesco.fagiani@mail.polimi.it

***Campo obbligatorio**

Premessa

Il seguente questionario ha una durata variabile tra i 5 e i 12 minuti.

Il numero e il tipo di domande varia in base alle risposte dell'utente, con un minimo di 14 fino ad un massimo di 38 domande.

Delle 38 domande soltanto 29 hanno obbligo di risposta.

Introduzione

Il termine "data literacy" o alfabetizzazione dei dati, riguarda la capacità di leggere, comprendere, analizzare e comunicare i dati.

All'interno del questionario saranno presenti altri due termini legati al primo:

"big data" - che descrive le tecnologie e i metodi per l'analisi di una grande mole di dati - e "open data" - dati aperti, con cui si indicano i dati liberamente accessibili a tutti, riutilizzabili e ridistribuibili liberamente.

Informazioni utente

1. Nome della pubblica amministrazione o dell'ente in cui lavori *
Ad esempio "Comune di Milano" o ente "Parco del Ticino".
2. Il tuo ruolo all'interno della pubblica amministrazione *
Contrassegna solo un ovale.
 - Sindaco / Presidente
 - Vice / Assessore
 - Membro eletto
 - Segretario comunale
 - Direttore
 - Responsabile di settore
 - Tecnico d'ufficio
 - Altro:
3. Settore o ufficio in cui operi all'interno della pubblica amministrazione *
Contrassegna solo un ovale.
 - Anagrafe
 - Attività produttive e commercio
 - Avvocatura comunale
 - Biblioteca
 - Consiglio Comunale
 - Ecologia e Ambiente
 - Edilizia privata
 - Informatica / Servizi ICT / Agenda digitale
 - Istruzione, cultura e sport
 - Lavori pubblici

- Mobilità e trasporti
- Pianificazione urbanistica
- Polizia locale / Protezione civile
- Protocollo
- Segreteria generale
- Servizi sociali
- Settore finanziario / Tributi
- Turismo
- Altro:

La tua opinione sui BIG DATA

4. Conosci la definizione di "big data"? *
- Contrassegna solo un ovale.*
- Si
 - No

La tua opinione sui BIG DATA

Anche se non conosci la definizione di "big data", prova a rispondere alle seguenti domande

BIG DATA

Con il termine "big data" vengono definite le tecnologie e i metodi per l'analisi di una grande mole di dati.

Secondo te, quali dei seguenti elementi caratterizzano maggiormente questo tipo di dati? Definisci il livello *da uno a cinque (1- Superfluo; 5 – Importante)*

5. Volume - grande quantità di dati
6. Velocità - rapidità di acquisizione / raccolta di dati
7. Varietà - diverse tipologie di dati
8. Procedure - capacità di gestione e analisi
9. Pubblicità - condivisione dei dati
10. Fonti - da dove vengono acquisiti i dati *Contrassegna solo un ovale.*
11. Se vuoi puoi aggiungere altre caratteristiche che secondo te sono importanti

I BIG DATA nella pubblica amministrazione in cui lavori

Nelle sezioni successive i "big data" vengono considerati come ampi dataset ovvero grandi serie di dati caratterizzate dal volume e della velocità di raccolta, e che richiedono archiviazione, gestione e analisi efficienti.

12. Dove lavori vengono UTILIZZATI dei "big data"? *
- Contrassegna solo un ovale.*
- Si
 - No *Passa alla domanda 23.*

I BIG DATA nella pubblica amministrazione in cui lavori

Nelle sezioni successive i "big data" vengono considerati come ampi dataset ovvero grandi serie di dati caratterizzate dal volume e della velocità di raccolta, e che richiedono archiviazione, gestione e analisi efficienti.

13. Dove lavori vengono PRODOTTI INTERNAMENTE "big data"? *
- Si
 - No *Passa alla domanda 19.*

I BIG DATA nella pubblica amministrazione in cui lavori

Nelle sezioni successive i "big data" vengono considerati come ampi dataset ovvero grandi serie di dati caratterizzate dal volume e della velocità di raccolta, e che richiedono archiviazione, gestione e analisi efficienti.

14. In quali settori vengono prodotti questi dati? *

Seleziona tutte le voci applicabili.

- Infrastrutture
- Trasporto pubblico
- Traffico stradale
- Parcheggi pubblici
- Statistica
- Ambiente
- Edilizia
- Urbanistica
- Tributi
- Altro:

15. Quali tipi di dati vengono prodotti?

Seleziona tutte le voci applicabili.

- RegISTRAZIONI audio / video e immagini
- Dati derivati da sensori
- Dati geografici
- Dati in formato elettronico (CSV, XML, xls, odf, ecc.)
- Profilo utenti
- Commenti / opinioni degli utenti
- Altro:

16. Quali di questi "big data" vengono forniti ad altre organizzazioni o resi pubblici? *

Contrassegna solo un ovale.

- Nessuno Passa alla domanda 19.
- Alcuni dati sono resi disponibili (meno del 50%)
- Molti dati sono disponibili (oltre il 50%)
- Tutti i dati sono resi disponibili

BIG DATA nella pubblica amministrazione - Fornitura dei dati a terzi

17. A quali organizzazioni vengono forniti i "big data" raccolti? *

Seleziona tutte le voci applicabili.

- Istituti di ricerca
- Open data (pubblici)
- Stato e Regione
- Altre pubbliche amministrazioni
- Aziende pubbliche o partecipate
- Social networks
- Aziende di telecomunicazione
- Aziende private
- Istituzioni europee
- Altro:

18. In quale formato vengono forniti questi "big data"?

- Disponibili on line (in qualsiasi formato) con una licenza open
- Disponibili come dati strutturati (es. foglio Excel e non immagine)
- Disponibili in un formato aperto, non proprietario (es. CSV e non Excel)

- Non lo so
- Altro:

BIG DATA nella pubblica amministrazione - Utilizzo di dati da terzi

19. Nella pubblica amministrazione in cui lavori vengono utilizzati dati prodotti da altre organizzazioni? * *Contrassegna solo un ovale.*
- Sì
 - No *Passa alla domanda 25.*

BIG DATA nella pubblica amministrazione - Utilizzo di dati da terzi

20. Da che tipo di organizzazione derivano i dati terzi? *

Seleziona tutte le voci applicabili.

- Portali open data
- Istituti di ricerca
- Cittadini
- Istituzioni / Pubbliche amministrazioni
- Aziende pubbliche
- Social networks e media companies
- Aziende di telecomunicazione
- Altro:

21. Qual è la ragione per cui si utilizzano dati di altre organizzazioni? *

Seleziona tutte le voci applicabili.

- Le risorse interne non permettono la raccolta di "big data"
- E' più economico ottenere "big data" da terzi
- Le risorse interne permettono la raccolta di alcuni, ma non tutti i "big data" necessari
- Non lo so
- Altro:

22. Se vuoi puoi fornire maggiori informazioni, esempi specifici o URL dei dati utilizzati
Passa alla domanda 27.

Le barriere ai BIG DATA

23. Secondo te perchè la pubblica amministrazione in cui lavori non utilizza big data? *

Seleziona tutte le voci applicabili.

- I dati necessari non esistono
- Alti costi di raccolta, gestione e aggiornamento mancanza di risorse interne competenti nel campo
- Mancanza di obiettivi o indirizzi chiari
- Piccole quantità di dati sono sufficienti per i bisogni attuali
- La PA per cui lavoro utilizza BIG DATA (in questo caso tornare indietro e modifica la risposta precedente)
- Altro:

24. Secondo te i "big data" potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori? *

Contrassegna solo un ovale.

- Sì
- No
- Non lo so

Interrompi la compilazione del modulo.

Le barriere ai BIG DATA

25. Secondo te perchè la pubblica amministrazione in cui lavori non utilizza dati di altre organizzazioni?*

Seleziona tutte le voci applicabili.

- I dati necessari non esistono
- I dati necessari non sono disponibili apertamente (open data)
- La pubblica amministrazione in cui lavoro produce tutti i dati che sono necessari
- Mancanza di risorse interne competenti nel campo
- Mancanza di obiettivi o indirizzi chiari
- Non lo so
- Altro:

26. Secondo te "big data" di altre organizzazioni potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori? *

Contrassegna solo un ovale.

- Sì
- No
- Non lo so

L'utilizzo dei BIG DATA

27. Quali risultati si cercano dall'analisi dei "big data"? *

Seleziona tutte le voci applicabili.

- Risultati nel processo decisionale
- Orientamento degli indirizzi programmatici
- Prestazioni delle infrastrutture
- Condizioni ambientali (es. inquinamento)
- Andamento di mercato
- Riscontri sulle politiche
- Opinioni dei cittadini
- Rendimento dei servizi
- Umore degli utenti / cittadini
- Altro:

28. Da quanto tempo la pubblica amministrazione in cui lavori utilizza "big data"?

Contrassegna solo un ovale.

- 1-2 anni
- 3-5 anni
- 5- 10 anni
- > 10 anni

Rendimento dei BIG DATA

Seleziona il grado con cui i "big data" hanno migliorato il lavoro all'interno della pubblica amministrazione in cui lavori.

29. "big data" hanno prodotto maggiore efficienza *

30. I "big data" hanno migliorato i processi decisionali *

31. I "big data" hanno aiutato a migliorare i servizi al cittadino *

32. Se vuoi puoi specificare le scelte precedenti

33. Credi che l'amministrazione pubblica per cui lavori utilizzi a pieno le potenzialità dei "big data"? *

34. Se vuoi puoi dare una breve spiegazione alla tua scelta

Gestione dei BIG DATA

35. Chi si occupa dei "big data" dove lavori? *
- Contrassegna solo un ovale.*
- Un solo tecnico
 - Un team di lavoro
 - Un ufficio / settore
 - Non sono definite persone o team
 - Altro:
36. Esiste una strategia definita per l'utilizzo degli OPEN DATA? *
- Contrassegna solo un ovale.*
- Sì (esiste una strategia specifica sugli open data)
 - Sì (come parte di una strategia generica sui dati e l'informatica)
 - No
 - Non lo so
37. Esiste una struttura per il trattamento dei dati confidenziali e della privacy? *
- Contrassegna solo un ovale.*
- Sì
 - Sì (come parte di una strategia generica sui dati e l'informatica)
 - No
 - Non lo so
38. Esiste una struttura di valutazione delle fonti, dei processi e dei risultati dei "big data" *
- Contrassegna solo un ovale.*
- Sì
 - Sì (come parte di una strategia generica sui dati e l'informatica)
 - No
 - Non lo so

Livello di DATA LITERACY

Come valuti complessivamente il livello della "data literacy" (l'alfabetizzazione e conoscenza dei big data) nella pubblica amministrazione dove lavori?

da uno a cinque (1- Insufficiente; 5 – Alto)

39. Capacità di valutare l'affidabilità e l'imparzialità delle fonti *
40. Conoscenza di cosa è un dato, come si raccoglie, analizza, visualizza e condivide*
41. Capacità di interpretare le informazioni derivate dai dati *
42. Capacità di usare i dati per prendere decisioni, risolvere problemi correnti *
43. Capacità di usare strumenti tecnici avanzati nelle diverse fasi del processo *
44. Conoscenza di come trattare la privacy e la sicurezza dei dati *

Formazione sui BIG DATA

45. La pubblica amministrazione dove lavori ha organizzato corsi di formazione per facilitare l'utilizzo dei "big data"? *

Contrassegna solo un ovale.

- Sì
- No *Dopo l'ultima domanda in questa sezione, interrompi la compilazione del modulo.*

46. Cosa riguardava il corso? *

Seleziona tutte le voci applicabili.

- Fonti dei dati
- Tipologia di dati
- Integrazione dei dati
- Analisi statistica Visualizzazione / infografica
- Sicurezza e privacy
- Altro:

Info e approfondimenti

il questionario è terminato. Ti ringraziamo per il contributo e la pazienza 47. Vuoi ricevere notizie e risultati al termine della ricerca?

Contrassegna solo un ovale.

- Sì
- No

48. Vuoi raccontarci di pratiche innovative sviluppate dalla tua pubblica amministrazione? Vuoi essere contattato per partecipare ad una intervista sul tema?

Lascia in descrizione il tuo indirizzo email o il numero di telefono per essere ricontattato

Allegato 2 - Percorso delle domande

Nome della pubblica amministrazione o dell'ente in cui lavori		
Il tuo ruolo all'interno della pubblica amministrazione		
Settore o ufficio in cui operi all'interno della pubblica amministrazione		
Conosci la definizione di "big data"?		
Volume - grande quantità di dati		
Velocità - rapidità di acquisizione / raccolta di dati		
Varietà - diverse tipologie di dati		
Procedure - capacità di gestione e analisi		
Pubblicità - condivisione dei dati		
Fonti - da dove vengono acquisiti i dati		
<i>Se vuoi puoi aggiungere altre caratteristiche che secondo te sono importanti</i>		
Dove lavori vengono UTILIZZATI dei "big data"?		
NO	SI	
Secondo te perchè la pubblica amministrazione in cui lavori non utilizza "big data"?	Dove lavori vengono PRODOTTI INTERNAMENTE "big data"?	
	NO	SI
Secondo te i "big data" potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?	Secondo te perchè la pubblica amministrazione in cui lavori non utilizza dati di altre organizzazioni?	In quali settori vengono prodotti questi dati?
	Secondo te "big data" di altre organizzazioni potrebbero aiutare in qualche modo la pubblica amministrazione in cui lavori?	Quali tipi di dati vengono prodotti?
<i>invio</i>		Quali di questi "big data" vengono forniti ad altre organizzazioni o resi pubblici?
		NESSUNO
		ALTRE RISPOSTE
		A quali organizzazioni vengono forniti i "big data" raccolti?
		In quale formato vengono forniti questi "big data"?
		Nella pubblica amministrazione in cui lavori vengono utilizzati dati prodotti da altre
		NO
		SI
		Da che tipo di organizzazione derivano i dati terzi?
		Qual è la ragione per cui si utilizzano dati di altre organizzazioni?
		<i>Se vuoi puoi fornire maggiori informazioni, esempi specifici o URL dei dati utilizzati</i>
		Quali risultati si cercano dall'analisi dei "big data"?
		Da quanto tempo la pubblica amministrazione in cui lavori utilizza "big data"?
		I "big data" hanno prodotto maggiore efficienza
		I "big data" hanno migliorato i processi decisionali
		I "big data" hanno aiutato a migliorare i servizi al cittadino
		<i>Se vuoi puoi specificare le scelte precedenti</i>
		Credi che l'amministrazione pubblica per cui lavori utilizzi a pieno le potenzialità dei "big data"?
		<i>Se vuoi puoi dare una breve spiegazione alla tua scelta</i>
		Chi si occupa dei "big data" dove lavori?
		Esiste una strategia definita per l'utilizzo degli OPEN DATA?
		Esiste una struttura per il trattamento dei dati confidenziali e della privacy?
		Esiste una struttura di valutazione delle fonti, dei processi e dei risultati dei "big data"?
		Capacità di valutare l'affidabilità e l'imparzialità delle fonti
		Conoscenza di cosa è un dato, come si raccoglie, analizza, visualizza e condivide
		Capacità di interpretare le informazioni derivate dai dati
		Capacità di usare i dati per prendere decisioni, risolvere problemi correnti
		Capacità di usare strumenti tecnici avanzati nelle diverse fasi del processo
		Conoscenza di come trattare la privacy e la sicurezza dei dati
		La pubblica amministrazione dove lavori ha organizzato corsi di formazione per facilitare l'utilizzo dei "big data"?
NO	SI	
<i>invio</i>		Cosa riguardava il corso?
		Vuoi ricevere notizie e risultati al termine della ricerca?
		Vuoi raccontarci di pratiche innovative sviluppate
		<i>invio</i>

Allegato 3 - Dettaglio comuni per invio questionario

Le seguenti tabelle riportano la matrice composta dal criterio geografico (regioni) e il criterio abitanti per comune (gruppi). Il numero di questionari si riferisce al campione omogeneo pari al 2% dei comuni totali (158), che corrisponde a un livello di confidenza del 95% e un errore del 7,5%.

Regione	GRUPPO 1			GROUP 2			GRUPPO 3		
	0 - 3.000		Questionari	3.001 - 7.000		Questionari	7.000 - 15.000		Questionari
	Popolazione	Comune		Popolazione	Comune		Popolazione	Comune	
Abruzzo	229.643	220	4	226.936	49	1	183.026	18	-
Basilicata	116.049	82	2	154.748	33	1	115.533	11	-
Calabria	370.335	256	5	437.993	96	2	324.548	33	1
Campania	406.303	264	5	504.234	112	2	903.162	89	2
Emilia-Romagna	140.470	82	2	498.353	102	2	898.634	89	2
Friuli-Venezia Giulia	192.864	131	3	226.241	45	1	279.880	28	1
Lazio	261.202	205	4	317.849	70	1	506.387	49	1
Liguria	173.941	164	3	158.223	34	1	248.213	25	1
Lombardia	1.048.864	787	16	1.889.636	405	8	2.070.000	211	4
Marche	170.736	125	3	224.226	50	1	304.459	31	1
Molise	113.621	116	2	62.900	14	-	27.461	3	-
Piemonte	886.717	957	19	618.407	142	3	483.357	50	1
Puglia	103.442	57	1	333.644	66	1	691.186	63	1
Sardegna	346.551	272	5	263.230	59	1	298.974	30	1
Sicilia	221.128	136	3	467.910	103	2	861.238	84	2
Toscana	123.118	78	2	313.035	67	1	764.614	74	1
Trentino-Alto Adige/Südtirol	274.245	207	4	274.634	64	1	100.130	11	-
Umbria	70.879	46	1	104.984	23	-	72.183	7	-
Valle d'Aosta	60.080	65	1	32.040	8	-	0	0	-
Veneto	320.081	187	4	854.657	182	4	1.430.883	140	3
Totale	5.630.269	4437	89	7.963.880	1.724	33	10.563.868	1.046	22

Regione	GRUPPO 4			Gruppo 5		
	15.000 - 30.000		Questionari	> 30.000		Questionari
	Popolazione	Comune		Popolazione	Comune	
Abruzzo	208.706	10	-	466.885	8	-
Basilicata	53.174	3	-	127.614	2	-
Calabria	204.187	11	-	619.624	8	-
Campania	962.383	44	1	3.050.778	41	1
Emilia-Romagna	703.913	36	1	2.211.259	22	-
Friuli-Venezia Giulia	128.474	7	-	389.394	4	-
Lazio	551.353	27	1	4.259.902	27	1
Liguria	145.717	6	-	830.887	5	-
Lombardia	1.487.296	74	1	3.540.462	39	1
Marche	207.275	11	-	625.057	12	-
Molise	21.666	1	-	82.845	2	-
Piemonte	592.797	29	1	1.794.587	19	-
Puglia	909.096	44	1	2.010.874	28	1
Sardegna	215.177	9	-	524.244	7	-
Sicilia	766.519	34	1	2.710.194	33	1
Toscana	689.981	33	1	1.846.220	22	-
Trentino-Alto Adige/Südtirol	113.015	6	-	305.624	4	-
Umbria	193.391	10	-	443.203	6	-
Valle d'Aosta	0	0	-	34.082	1	-
Veneto	940.265	46	1	1.357.836	16	-
Totale	9.094.385	441	9	27.231.571	306	5

Elenco dei comuni

Elenco dei comuni 296 che hanno risposto al questionario alla data del 5/04/2019.

Un ringraziamento, dovuto, alle persone che hanno partecipato per conto del comune di:

Acquaformosa; Adrara San Martino; Albiano d'Ivrea; Albissola Marina; Alcamo; Alfedena; Alseno; Altissimo; Annicco; Ariano Irpino; Arluno; Arsiè; Assago; Ateleta; Attimis; Avellino; Aversa; Badia; Bagnaria Arsa; Baia e Latina; Battipaglia; Battuda; Belvedere Ostrense; Benevento; Binasco; Bisignano; Blera; Bodio Lomnago; Bordighera; Borghetto Santo Spirito; Borgo Val di Taro; Borgomanero; Borgoricco; Bovolone; Brindisi Montagna; Bussoleno; Caines; Cairo Montenotte; Calamonaci; Calolziocorte; Camagna Monferrato; Camino; Campana; Campo nell'Elba; Campobello di Mazara; Canicattini Bagni; Canna; Cantello; Caraffa di Catanzaro; Carbognano; Carbonera; Carugate; Casale sul Sile; Casalgrande; Casatisma; Caselle Landi; Cassaro; Castel Ivano; Casteldelci; Castelfranco in Miscano; Castellana Sicula; Castelrotto; Castiglione Cosentino; Castiglione d'Adda; Castiglione di Garfagnana; Castions di Strada; Catania; Catanzaro; Cattolica; Celle Ligure; Cenadi; Cervia; Cesena; Chiusanico; Cinto Euganeo; Cirò Marina; Cogorno; Colico; Colle di Val d'Elsa; Colletterto Giacosa; Colorina; Concamarise; Contà; Corleone; Cortona; Cossato; Cotronei; Diano Marina; Eboli; Falcone; Fara Filiorum Petri; Fiesse; Fiorenzuola d'Arda; Firenze; Foglizzo; Fonte Nuova; Frascineto; Frosinone; Furore; Gadesco-Pieve Delmona; Gaglianico; Gaiole in Chianti; Gemonio; Geraci Siculo; Germagno; Giaveno; Gombito; Graffignano; Grandola ed Uniti; Griante; Grosotto; Grottazzolina; Guastalla; Guidonia Montecelio; Imer; Lagonegro; Lampedusa e Linosa; L'Aquila; Leggiuno; Lentini; Letino; Livigno; Locorotondo; Lonate Pozzolo; Maiolo; Marano di Valpolicella; Marsico Nuovo; Massafra; Mazara del Vallo; Mazzarrone; Milano; Modica; Mondragone; Monforte d'Alba; Monsummano Terme; Montagna in Valtellina; Montaldo Roero; Montanaso Lombardo; Monte Castello di Vibio; Monte Cerignone; Monte Porzio Catone; Monte San Pietro; Montecchio Precalcino; Montevarchi; Montorio al Vomano; Monvalle; Morazzone; Mornago; Morro Reatino; Morrovalle; Mulazzano; Napoli; Noale; Noicattaro; Nosate; Noto; Novara; Nule; Nulvi; Occimiano; Onano; Ortona; Orio Litta; Ornavasso; Ozzano dell'Emilia; Palazzago; Palermo; Palombara Sabina; Palosco; Pasiano di Pordenone; Pauli Arbarei; Paulilatino; Paupisi; Pavia; Pennadomo; Pescantina; Petilia Policastro; Pietrasanta; Pieve del Cairo; Pieve di Cento; Pieve Porto Morone; Pignataro Maggiore; Piscinas; Polino; Pollena Trocchia; Pomarico; Porte di Rendena; Porto Cesareo; Porto Recanati; Portogruaro; Pozzo d'Adda; Pozzomaggiore; Prali; Pramollo; Prascorsano; Premana; Primaluna; Propata; Pula; Quarrata; Quartucciu; Quero Vas; Rapone; Reggello; Riccione; Riccò del Golfo di Spezia; Rivello; Rocchetta di Vara; Romeno; Romentino; Rosolina; Rosora; Roverè della Luna; Sala Bolognese; Salgareda; San Cesareo; San Fior; San Germano Vercellese; San Giovanni a Piro; San Giuliano Terme; San Marco dei Cavoti; San Martino Buon Albergo; San Martino dall'Argine; San Martino Valle Caudina; San Vittore Olona; Sansepolcro; Sant'Ambrogio sul Garigliano; Sant'Angelo del Pesco; Sant'Angelo Muxaro; Santena; Santo Stefano di Cadore; Sauze di Cesana; Sauze d'Oulx; Savigliano; Savignone; Scandolara Ravara; Scilla; Segrate; Selva dei Molini; Selva di Cadore; Sergnano; Seriate; Sessano del Molise; Sesto ed Uniti; Sesto San Giovanni; Siena; Sondalo; Spino d'Adda; Stresa; Surano; Taglio di Po; Terre del Reno; Tonengo; Torgiano; Torrevecchia Teatina; Tovo San Giacomo; Tramonti di Sopra; Trapani; Travesio; Treccate; Treviglio; Uggiano la Chiesa; Ustica; Vaglia; Vaglio Serra; Vaiano Cremasco; Vairano Patenora; Valdengo; Valfabbrica; Valle San Nicolao; Vallermosta; Vallo della Lucania; Valtourneche; Varsi; Venaria Reale; Vercelli; Verolavecchia; Vesime; Vezzano Ligure; Viale; Vicchio; Villa di Chiavenna; Villa Lagarina; Villalfonsina; Villamar; Villanova Marchesana; Villanova Monferrato; Villaricca; Villasimius; Villeneuve; Visone; Voltido; Zerbolò; Zermeghedo; Zoppola; Zovencedo.