

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

DIPARTIMENTO DI INFORMATICA – SCIENZA E INGEGNERIA
Corso di Laurea Magistrale in Ingegneria e Scienze Informatiche

**COVID-19,
Open data e data visualization:
interazione con dati epidemiologici**

Elaborato in:
Tecnologie Web

Relatore:
Dott.ssa Silvia Mirri
Correlatore:
Dott.ssa Catia Prandi
Dott.ssa Chiara Ceccarini

Presentata da:
Stefania Riminucci

**Sessione II
Anno Accademico 2019-2020**

*Dedicato ai miei genitori
e a tutti coloro che
mi sono stati di aiuto e supporto
durante la mia carriera universitaria.*

Introduzione

Il lavoro presentato in questo documento è frutto del particolare momento che siamo vivendo attualmente.

L'argomento principale intorno a cui tutto ruota è la pandemia di COVID-19 che si è diffusa in Italia ad inizio 2020, portando con sé numerosi risvolti: sociali, economici e sanitari. Tutti questi aspetti sono guidati in molti casi dalla popolazione, essendo questa il fulcro centrale della diffusione del virus. A livello sociale è necessario gestire l'opinione pubblica relativa alla diffusione della malattia, alle misure di sicurezza e allo stato delle strutture sanitarie. A livello economico è necessario gestire gli effetti del lockdown, attuato per ridurre il numero di contagi.

A livello sanitario è necessario gestire l'emergenza dei ricoveri, raccogliere e valutare i dati dei pazienti che sono stati curati dei possibili contagiati.

Questo argomento ha invaso le pagine dei quotidiani, delle piattaforme online della maggior parte dei programmi televisivi.

Per tenere costantemente aggiornata la popolazione sulla diffusione del virus, sul numero di contagiati e di decessi, sullo stato degli ospedali e sulle condizioni degli altri stati sia europei che mondiali.

A seguito del primo caso di COVID-19, è iniziata la raccolta di dati: sia a livello sanitario con statistiche sui ricoveri, sui sintomi dei pazienti sia a livello di ricerca per tentare di individuare tutte le caratteristiche e comportamenti del virus. La condivisione dei dati tra diversi soggetti aiuta a velocizzare ed aumentare le possibilità di successo nella lotta contro la diffusione del virus e soprattutto nell'aiutare i malati a riprendersi.

L'enorme mole di dati raccolti relativi alla ricerca dei medici può dare un'enorme aiuto, ma è necessario renderli comprensibili e valutabili facilmente.

I dati, consultati dagli esperti, hanno fornito delle informazioni importanti ai Governi per decidere quale fosse la strada migliore da percorrere.

Così come anche per i cittadini è stato importante, con i dati aperti forniti dalle strutture sanitarie, avere informazioni sulle zone “rosse” (a più alto rischio di contagio), capire quali fossero le categorie più a rischio, l'andamento in generale della pandemia e così via.

Per rendere di facile e veloce comprensione tutti i dati raccolti dalle diverse strutture, è stato necessario avvalersi della *Data Visualization*. Questa non è utile solamente per chi non è del settore ed ha bisogno di capire informazioni altrimenti difficilmente leggibili, ma anche agli esperti di settore per l'analisi approfondita dei dati in modo rapido ed efficace.

È per tutto questo che negli ultimi mesi si sono viste tantissime infografiche, di ogni genere, colore e tipo, sia online che in TV che su riviste e quotidiani.

Per la rappresentazione di ogni tipologia di dato, ci sono grafici ed infografiche che possono risultare più o meno chiare a chi le consulta.

L'obiettivo di questo progetto è quello capire se ci siano dei grafici o infografiche che risultino più comprensibili al cittadino comune in base all'informazione che questo ha necessità di trovare. Ci sono grafici che risultano rispecchiare in modo più dettagliato e con aggiunta di informazioni i dati relativi ad un certo contesto, ma che risultano magari di difficile lettura a chi non è avvezzo alla consultazione di infografiche più complesse (che magari risultano essere più utili ed esplicative per gli esperti).

Per raggiungere l'obiettivo, come strumento può essere utilizzato il questionario per avere delle informazioni relative a quali siano le infografiche “preferite” dagli utenti per la presentazione dei dati.

Il seguito della tesi è così organizzato:

Capitolo 1 – Open Data e Data Visualization Questo capitolo parte introducendo gli Open Data, spiegando cosa sono, a cosa servono, i formati e alcuni casi di esempio.

A seguire viene presentata la Data Visualization, spiegando cosa sia, come ed il perché viene utilizzata ed alcuni casi di esempio.

Capitolo 2 – COVID-19: open data e data visualization Questo capitolo inizia con un'introduzione sulla pandemia del 2020 che ha colpito tutto il mondo. Viene proposta un'associazione agli open data, a come questi siano stati utilizzati per la gestione dell'emergenza a livello mondiale e come la data visualization sia uno strumento molto importante ed utile per la rappresentazione dell'enorme mole di dati prodotta in riferimento al virus negli ultimi mesi. In ultimo nel capitolo, vengono presentati alcuni esempi di utilizzo di open data e data visualization in riferimento al COVID-19.

Capitolo 3 – Questionario Online Questo capitolo introduce il progetto creato per sondare il livello di comprensione delle visualizzazioni dati proposte dalle diverse risorse online sul tema del COVID-19. Il capitolo presenta la struttura del questionario realizzato, riportando anche frammenti di codice creati per la gestione del questionario e la presentazione della struttura del database utilizzato per il salvataggio dei dati.

Capitolo 4 – Risultati del questionario In questo capitolo vengono presentati i dati estratti dal database. Con i risultati ottenuti sono state create delle infografiche per una comprensione più immediata dei dati raccolti.

Indice

Introduzione	i
1 Open Data e Data Visualization	1
1.1 Open Data	1
1.1.1 Cenni Storici	8
1.1.2 Formato e Distribuzione degli Open Data	9
1.1.3 Casi Studio	12
1.2 Data Visualization	15
1.2.1 Cenni Storici	20
1.2.2 Casi Studio	22
2 COVID-19: open data e data visualization	25
2.1 COVID-19 e Open data	25
2.2 COVID-19 e Data visualization	27
2.3 COVID-19 e Storytelling	32
3 Questionario Online	43
3.1 Contesto del problema	43
3.2 Questionari	44
3.3 Il progetto	48
3.3.1 Requisiti	49
3.3.2 Analisi e modello del dominio	50
3.3.3 Architettura e Design	55
3.3.4 Sviluppo	57

4 Risultati del questionario	65
4.1 I dati	65
4.1.1 Dati demografici e personali	67
4.1.2 Fonti di informazioni utilizzate	71
4.1.3 Comprensione delle infografiche proposte	74
4.1.4 Valutazione sull'utilizzo di piattaforme esterne	78
 Conclusioni	 89
 Bibliografia	 93
 Ringraziamenti	 100

Elenco delle figure

1.1	Diagramma concettuale della roadmap equa per i “dark data”	3
1.2	I benefici degli Open Data	7
1.4	Infografica dati S3 - Emilia Romagna	23
1.5	Comune di Bologna: infografica sulla produzione di energia dai rifiuti	24
1.6	Eni – Gender Equality, hashtag più utilizzati	24
2.1	WHO - Cartogramma interattivo del numero di casi	29
2.2	WHO - Grafici a barre, interattivi	30
2.3	Protezione Civile – Dashboard COVID-19 Situazione Italia	31
2.4	Proporzione casi per stato clinico ed età in Italia	31
2.5	OWD - Andamento dei tassi di crescita dei casi	33
2.6	OWD - Andamento del numero di decessi in alcuni paesi	35
2.7	OWD - Differenza percentuale settimanale morti in eccesso rispetto al lustro precedente	37
2.8	Numero deceduti per fascia d’età	38
2.9	Percentuale decessi con patologie preesistenti	39
2.10	Percentuale decessi con patologie preesistenti	40
3.1	Wireframe della pagina principale del questionario	51
3.2	Esempio domanda risposta multipla	52
3.3	Esempio domanda risposta singola	53
3.4	Esempio domanda risposta singola con sottodomanda	56
3.5	Schema del DB per il salvataggio dei dati	57

4.1	Risultati questionario: distribuzione temporale compilazioni	66
4.2	Risultati questionario: fasce d'età	68
4.3	Risultati questionario: livello d'istruzione	69
4.4	Risultati questionario: genere	70
4.5	Risultati questionario: dispositivo utilizzato per la compilazione del questionario	71
4.6	Risultati questionario: livello di comprensione della lingua scritta inglese	72
4.7	Risultati questionario: fonti di informazione sul COVID-19	73
4.8	Risultati questionario: fonti di consultazione visualizzazione dati	74
4.9	Risultati questionario: dispositivi utilizzati per la ricerca di informazioni	75
4.10	Risultati questionario: individuare il maggior numero di casi	76
4.11	Risultati questionario: individuare il paese con il maggior numero di decessi	77
4.12	Risultati questionario: maggior ricezione di materiali nelle regioni italiane	78
4.13	Risultati questionario: continente con il maggior numero di test effettuati	79
4.14	Risultati questionario: paese con il maggior numero di test effettuati ogni 1000 persone	80
4.15	Risultati questionario: ricerca della provincia italiana con il maggior numero di casi	82
4.16	Risultati questionario: ricerca dei casi totali e decessi per l'Italia	83
4.17	Risultati questionario: numero decessi per fascia d'età	85
4.18	Risultati questionario: ricerca della tipologia di mobilità con la percentuale più alta di modifica delle richieste di instradamento	87

Elenco delle tabelle

3.1	Lista domande questionario	53
-----	--------------------------------------	----

Listings

1.1	TPER – Estratto xml rivenditori biglietti	13
1.2	Comune di Bologna – Estratto geojson mappature colonnine .	14
3.1	Codice JS per il posizionamento random delle risposte contenenti immagini	59
3.2	Codice per la gestione della comparsa/scomparsa e scelta delle sottodomande	60
3.3	Codice per la validazione dei campi obbligatori	61
3.4	Codice per la gestione della gallery degli ingrandimenti e layer colorato sopra l'immagine corrispondente alla risposta scelta .	63

Capitolo 1

Open Data e Data Visualization

1.1 Open Data

Open data (in italiano *Dati aperti*) è un termine utilizzato a livello internazionale ed indica una tipologia di dato che rispecchia le seguenti caratteristiche:

I dati aperti sono dati che possono essere liberamente utilizzati, riutilizzati e ri-distribuiti da chiunque, soggetti eventualmente alla necessità di citarne la fonte e di condividerli con lo stesso tipo di licenza con cui sono stati originariamente rilasciati.[1]

La definizione sugli open data su Open Definition[2] spiega nel dettaglio cosa implichi tutto ciò:

- **disponibilità e accesso:** i dati devono essere disponibili nella loro totalità, per un prezzo non superiore a quello che potrebbe essere considerato un ragionevole costo di riproduzione, preferibilmente mediante download da piattaforme online. I dati devono essere disponibili in un formato utile e modificabile;
- **riutilizzo e redistribuzione:** i dati devono essere forniti a condizioni tali da permetterne il riutilizzo e la redistribuzione. Ciò comprende la possibilità di poterli combinarli con altre basi di dati;

- **partecipazione universale:** tutti devono essere in grado di usare, riutilizzare e ridistribuire i dati. Non ci devono essere discriminazioni né in ambito di iniziativa né contro soggetti o gruppi. Ad esempio, la clausola ‘non commerciale’, che vieta l’uso a fini commerciali o restringe l’utilizzo solo per determinati scopi (es. quello educativo) non è ammessa.

L’importanza della definizione di “aperto” e del perché utilizzare proprio questo termine può essere definita con il seguente termine: l’*interoperabilità* [1].

Il punto fondamentale di un set di dati (o linee di codice) accessibili ed utilizzabili in modo condiviso è dato dal fatto che questi set di dati possano essere “mescolati” liberamente tra loro, anche se provenienti da fonti differenti.

L’interoperabilità consente di sfruttare il principale vantaggio pratico dell’apertura dei dati: aumentare in modo esponenziale la possibilità di combinare tra loro diverse basi di dati e quindi di sviluppare sia nuovi che migliori prodotti e servizi.

La definizione di apertura assicura quindi di evitare di avere a disposizione molti set di dati ma senza la possibilità di combinarli in sistemi più ampi, perdendo quindi il vero valore dell’interoperabilità (figura 1.1).

Perché utilizzare gli open data?

Gli Open Data, ed in particolare gli open government data, sono un’immensa risorsa per la maggior parte ancora poco utilizzata. Molte organizzazioni così come molte persone raccolgono una vasta gamma di dati diversi (per svolgere i propri compiti).

Quello che fa il Governo da questo punto di vista è molto importante, non solo per l’enorme mole di dati che viene raccolta e la centralità che questi hanno, ma anche perché la maggior parte dei dati governativi sono pubblici

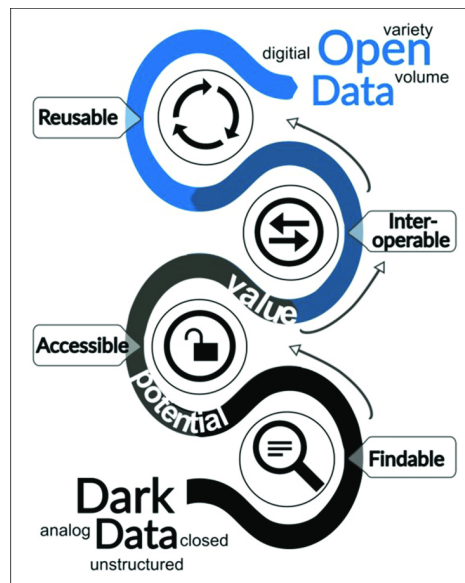


Figura 1.1: Diagramma concettuale della roadmap equa per i *dark data*: ogni curva rappresenta un passo verso l'aumento del valore e del potenziale dei “dati oscuri” in ambito scientifico [3].

per legge e dovrebbero essere quindi resi aperti e pubblici per consentirne l'utilizzo a chiunque.

Questo ci interessa perché ci sono tantissime circostanze in cui i dati aperti hanno un valore rilevante ed abbiamo già diversi esempi in cui questo accade.

Ci sono numerose categorie di persone ed organizzazioni che possono trarre un enorme beneficio dalla disponibilità dei dati aperti, soprattutto la pubblica amministrazione. Allo stesso tempo non è possibile predire come e dove l'utilizzo di questi dati creerà un valore.

È già possibile individuare un elevato numero di aree dove i dati pubblici stanno già creando o aggiungendo valore. Alcuni esempi sono [4]:

- controllo trasparente e democratico;
- partecipazione;
- auto-potenziamento;

- prodotti e servizi migliorati o privati nuovi;
- miglioramento dell'efficacia dei servizi governativi;
- rilevamento dell'impatto delle politiche;
- estrazione di nuova conoscenza dalla combinazione di diverse fonti di dati e dall'identificazione di regolarità che emergono dall'analisi di grandi masse di dati.

Esistono già diversi esempi per la maggior parte di queste aree.

Alcuni esempi relativi alla trasparenza li abbiamo con i progetti *Tax tree* [5] (Finlandese) e *Where does my money go* [6] (Britannico) che permettono di monitorare l'impiego, da parte del governo, dei soldi delle tasse dei cittadini. In Canada addirittura i dati aperti hanno consentito un risparmio di 3.2 miliardi di dollari in un caso di frode fiscale legato alla beneficenza [7]. Molti siti, tra cui *folketsting.dk* [8] (Danese) e l'italiano *parlamento18.openpolis.it* [9], tracciano le attività dei parlamentari ed il processo di formazione delle leggi, in modo da poter essere sempre aggiornati su cosa stia accadendo e quali parlamentari partecipino alle varie attività.

I dati resi pubblici dai governi possono essere di supporto anche nella nostra vita privata aiutandoci a prendere decisioni migliori o a renderci più attivi all'interno della società.

Alcuni esempi:

- **Danimarca:** una sviluppatrice ha creato *findtoilet.dk* [10] che consente l'accesso alla lista di tutti i bagni pubblici del paese (con questo servizio anche chi soffre di incontinenza si sente più rassicurato quando esce di casa);
- **New York:** un servizio che ti permette di sapere dove si può portare a spasso il proprio cane (*NYC Parks Dog Runs* [11]) così come trovare altre persone che utilizzano il tuo stesso parco;

- **Regno Unito e Germania:** esistono due servizi, rispettivamente *mapumental.com* [12] e *mapnificent.net* [13] che permettono di facilitare la scelta del luogo in cui andare ad abitare impostando i tempi massimi di percorrenza casa/ufficio, i prezzi delle case e la bellezza del quartiere.

Tutti questi esempi utilizzano dati aperti rilasciati dai governi.

I dati aperti offrono molti vantaggi anche dal punto di vista economico. Alcuni studi hanno stimato il valore economico dei dati aperti e questo si aggirerebbe su diverse decine di miliardi di euro, nella sola Europa [14]. Così come i governi ed i cittadini stessi, anche l'economia può trarre vantaggio da un più facile accesso a informazioni, contenuti e conoscenze, che a loro volta contribuiscono allo sviluppo di servizi innovativi e alla creazione di nuovi modelli di lavoro.

Ad esempio il sito danese *modstroem.dk* [15] aiuta gli utenti a risparmiare energia elettrica nelle proprie abitazioni ed aiuta anche nella pianificazione finanziaria e fornisce la possibilità di contattare gli artigiani che potranno eseguire il lavoro. Questa piattaforma funziona grazie all'utilizzo dei dati catastali, a informazioni sugli incentivi governativi e al registro delle imprese locali.

Altro esempio è fornito dal più celebre Google Translate, che utilizza l'enorme mole di documenti dell'Unione Europea, disponibili in tutte le lingue europee, per allenare gli algoritmi di traduzione automatica, aumentando così la precisione del servizio offerto agli utenti.

Anche per gli stessi governi, i dati aperti rappresentano una grande fonte per poter migliorare l'efficienza, non solo la propria, riducendo anche i costi. Una migliore efficienza dei processi e della fornitura dei servizi pubblici può essere ottenuta grazie alla condivisione dei dati tra un settore e l'altro, che può permettere per esempio di avere una visione d'insieme delle spese non necessarie [16].

Il Ministero Olandese dell'Istruzione ha reso pubblici on-line tutti i dati relativi al sistema educativo consentendone il ri-uso. Da quel momento il numero delle domande ricevute è sceso, riducendo così il carico di lavoro ed i costi ed è anche più facile per i dipendenti pubblici riuscire a rispondere alle domande residue in quanto sanno dove possono trovare tutti i dati necessari per fornire le risposte alle domande ricevute.

Un altro esempio sul miglioramento dell'efficienza delle strutture pubbliche con relativa riduzione dei costi, arriva sempre dal governo olandese, più precisamente dal dipartimento per il patrimonio culturale. Quest'ultimo sta attivamente rilasciando i propri dati e, collaborando con società amatoriali di storici e con gruppi come Wikimedia Foundation, sta gestendo i propri lavori in modo più efficace. Questo si traduce, non solamente in una migliore qualità di dati, ma anche in una riduzione delle dimensioni del dipartimento [4].

Oltre ai precedenti aspetti, tra i vantaggi degli open data ritroviamo anche il vantaggio a livello sociale con il miglioramento del benessere sociale in quanto la società ha a disposizione informazioni più trasparenti e accessibili (figura 1.2).

Gli open data sono un tassello fondamentale per la collaborazione e la partecipazione tra diverse entità, che possono portare anche al miglioramento dell'innovazione sociale [16].

Nonostante siano già presenti numerosi esempi in cui i dati aperti stanno già portando vantaggi economici e sociali, non siamo ancora in grado di prevedere quali saranno gli utilizzi futuri.

La combinazione dell'enorme mole di dati già presente (e quella che verrà man mano generata il futuro) renderà possibili infinite combinazioni che possono portare alla generazione di nuove conoscenze e studi che condurranno verso strade non ancora conosciute [4].



Figura 1.2: I benefici degli Open Data sono diversi e spaziano dall'aumento dell'efficienza delle pubbliche amministrazioni, alla crescita economica nel settore privato al benessere sociale in generale [16]

Nella Londra dell'800, ad esempio, il dottor Snow scoprì la correlazione tra l'inquinamento dell'acqua potabile e il colera combinando i dati sui morti per colera con quelli sull'ubicazione dei pozzi. A seguito di quanto era stato scoperto “incrociando” i due set di dati, si arrivò alla costruzione del sistema fognario, migliorando notevolmente le condizioni generali di salute della popolazione.

Questo potenziale non ancora completamente sfruttato può essere utilizzato se facciamo in modo di rendere aperti i dati delle amministrazioni pubbliche. Tuttavia, questo accade solo se l'apertura è completa, cioè se non ci sono limitazioni (giuridiche, finanziarie o tecnologiche) al riutilizzo da parte di altri. Ogni restrizione impedirà di ri-utilizzare i dati pubblici, e renderà più difficile il trovare altri modi per poter arrivare allo stesso risultato.

Sulla piattaforma italiana *dati.gov.it* [17] sono presenti le risorse utili per mettere in pratica l'apertura dei dati: riferimenti normativi, documenti di indirizzo, linee guida e specifiche tecniche.

1.1.1 Cenni Storici

Il termine open data è apparso per la prima volta nel 1995, in un documento di un'agenzia scientifica americana. Questa si è occupata della divulgazione di dati geofisici e ambientali. Diedero inizio ad uno scambio completo e aperto di informazioni scientifiche tra paesi diversi, un prerequisito per l'analisi e la comprensione di questi fenomeni globali.

L'idea del bene comune applicata alla conoscenza era già stata teorizzata, ben prima dell'invenzione di Internet. Robert King Merton era uno dei padri della sociologia della scienza e la teoria che porta il suo nome mostra i vantaggi di dati scientifici aperti. Già nel 1942, Merton spiegò l'importanza che i risultati della ricerca dovessero essere liberamente accessibili a tutti: ogni ricercatore deve contribuire e rinunciare ai diritti di proprietà intellettuale per consentire alle conoscenze di progredire [18].

Le moderne tecnologie dell'informazione hanno dato un nuovo contributo alla filosofia dei beni comuni. Nella sua ricerca, il premio Nobel per l'economia 2009 Elinor Ostrom ha mostrato le particolarità dei beni comuni di informazione. Sono molto simili ai beni pubblici, perché il loro uso da parte di una persona non ne impedisce l'uso da parte di altri. Tuttavia, a differenza dei beni pubblici, il loro uso non esaurisce la riserva comune, ma l'arricchisce.

Agli inizi, i dati aperti erano radicati nella prassi della comunità scientifica. I ricercatori sono stati i primi a percepire il vantaggio dell'apertura e della condivisione dei dati. Ma è l'incontro tra questa idea scientifica e gli ideali del software libero e dell'open source che ha dato forma ai dati aperti così come li conosciamo oggi.

Nel dicembre 2007, pensatori e attivisti di Internet hanno tenuto un incontro con l'obiettivo di definire il concetto di dati pubblici aperti e di farlo adottare dai candidati presidenziali statunitensi.

Tra questi, c'erano due personaggi famosi: Tim O'Reilly e Lawrence Lessig.

Il primo non è nuovo al settore tecnologico: questo autore ed editore americano è il creatore di molti movimenti informatici e web all'avanguardia; ha definito e diffuso espressioni come “*open source*” e “*Web 2.0*”.

Lawrence Lessig, professore di diritto alla Stanford University (California), è il fondatore delle licenze Creative Commons, basate sull'idea di copyleft e libera diffusione delle conoscenze.

Il contributo di Tim O'Reilly sull'*open government* ha gettato una nuova luce sulla relazione tra il movimento open source e i principi emergenti dei dati aperti, con le sue stesse parole: «*dobbiamo applicare i principi dell'open source e i suoi metodi di lavoro agli affari pubblici*» [18].

Nel 2007, sembrava un sogno, ma il risultato ha superato di gran lunga le aspettative iniziali. Poco più di un anno dopo, il presidente Barack Obama entrò in carica alla Casa Bianca e firmò tre memorandum presidenziali. Due di questi riguardavano il governo aperto, di cui i dati aperti rappresentano uno dei pilastri. Questi promemoria presidenziali pongono esplicitamente la questione dell'open source al centro dell'azione pubblica acquisendone i principi fondamentali: trasparenza, partecipazione e collaborazione.

1.1.2 Formato e Distribuzione degli Open Data

Gli open data devono essere disponibili a chiunque voglia accedervi, usarli e condividerli. In questa sezione vedremo quali formati massimizzano la fruibilità dei dati, la facilità di accesso per gli utenti e la semplicità di gestione per chi li pubblica.

Quando si vogliono rendere fruibili ed accessibili dei dati, è necessario tenere a mente questi 3 aspetti fondamentali:

- formato dei dati;
- struttura dei dati;
- modalità di distribuzione dei dati.

Il **formato dei dati** fa riferimento al modo in cui questi sono resi disponibili e comprensibili sia dalle persone che dalle macchine. Il formato più utilizzabile per i dati è probabilmente quello in cui i dati sono già stati gestiti ed utilizzati. Questo formato sarà quello maggiormente legato all'ambito in cui i dati vengono utilizzati attualmente, ma potrebbe non essere quello maggiormente disponibile o comprensibile, quindi potrebbe essere necessario pubblicare in un formato aggiuntivo (oltre a quello originale) come ad esempio in .csv. Si dovrebbe considerare quindi la possibilità di dover pubblicare in formati diversi ed essere pronti al fatto che alcuni di questi potrebbero diventare obsoleti.

In riferimento alla **struttura dei dati**, ce ne sono di diverse con cui questi vengono organizzati all'interno di un file. Le principali che bisogna conoscere sono tre [19]:

- tabulare;
- gerarchica;
- network.

La struttura più comune è ancora il formato **tabulare**, dove i dati sono organizzati in righe e colonne (appunto tabelle) che elencano valori sequenziali (ad esempio le tabelle relative alle spese). Se i dati sono basati su voci separate che non sono collegate tra loro, una struttura tabulare in un formato come CSV è l'ideale.

Un esempio di dati con un'ideale struttura tabellare è quella dei dati del portale italiano relativi alle presenze nei musei [20].

Nella struttura **gerarchica**, i dati mostrano le relazioni tra dati puntuali, come ad esempio un albero genealogico o i comuni di ciascun paese. Se il set di dati dipende dalle relazioni tra dati puntuali e segue una struttura in cui i punti sono collegati verticalmente come in un albero. Il formato JSON è l'ideale per questi tipi di strutture gerarchiche.

Con i dati strutturati in **network** possono sussistere relazioni tra qualsiasi combinazione di elementi in qualsiasi direzione. Questo tipo di dati si riescono ad immaginare difficilmente in termini di formati di file. Un buon esempio di una struttura di dati a network è una rete sociale. Pensiamo alla nostra rete di amici e ai loro amici su Facebook oppure consideriamo i contatti di primo, secondo e terzo grado su LinkedIn. I siti web sono un altro esempio di una struttura di dati a network, in cui le pagine sono collegate ad un qualunque numero di altre pagine in qualsiasi direzione.

Relativamente alla **modalità della distribuzione**, possiamo avere le seguenti opzioni:

- il download di file;
- l'utilizzo di flussi per i dati con aggiornamenti in real time.

Non tutti i dati sono adatti per il download o perché di dimensioni troppo grandi, o perché aggiornati troppo frequentemente, o perché troppo complessi per produrli in file statici. Alcuni dati possono essere suddivisi in parti più piccole. Potrebbe anche essere necessario pubblicare dei dati come flussi real time per assicurarsi che siano costantemente aggiornati.

I dati tabulari sono i più adatti per il download. Questo è il motivo per cui gli open data della maggior parte dei portali delle pubbliche amministrazioni sono prevalentemente tabulari. Tuttavia, la gestione di milioni di righe di dati potrebbe richiedere altre considerazioni:

- potrebbe essere opportuno dividere i dati in dataset più piccoli?
- quanto spesso devono essere aggiornati i dati?
- se si cambia il modo in cui vengono pubblicati i dati, che conseguenze avrà sui dati già pubblicati in precedenza?

Come già accennato sopra, alcuni dati non sono adatti per essere messi a disposizione come file scaricabili. Molti di questi dati vengono aggiornati

così spesso che i file da scaricare risulterebbero troppo pesanti per la maggior parte degli utenti.

Questo tipo di struttura può essere resa attraverso un'interfaccia per macchine nota come interfaccia di programmazione delle applicazioni (*API*).

Nel web ci sono molti servizi che rendono disponibili delle API. Questi servizi possono essere integrati direttamente in altre applicazioni. Ad esempio, l'italiano Istituto Nazionale di Geofisica e Vulcanologia [21] (INGV) mette a disposizione dei webservice per l'utilizzo dei propri dati (generati seguendo le specifiche internazionali [22]).

Potremmo dire che quindi è sempre consigliato cercare di fornire la più vasta selezione possibile di formati, strutture e modalità distribuzione dei dati. Ma il “cuore” di tutti questi formati è il formato .csv, che è attualmente quello più diffuso online per gli open data.

1.1.3 Casi Studio

TPER a Bologna

Tper (Trasporto Passeggeri Emilia Romagna) è un'azienda per il trasporto pubblico .

Le aree di attività della società coprono diversi segmenti del settore del trasporto, da quello automobilistico, a quello filoviario, a quello ferroviario di passeggeri. TPER gestisce il TPL su gomma nei bacini provinciali di Bologna e Ferrara ed il trasporto passeggeri in ambito ferroviario regionale, in partnership con Trenitalia.

Tper è una delle aziende italiane che mette a disposizione gli open data [23] relativamente ai propri mezzi di trasporto. I dati sono soggetti alla licenza Creative Commons Attribuzione 3.0 Italia e all'accettazione delle condizioni riportate nelle note legali sull'utilizzo degli open data di Tper.

Tper fornisce anche servizi real-time interattivi, accessibili anche tramite webservice, sulla disponibilità dei propri mezzi, attraverso i servizi *Chiama Treno* [24] e *Hello Bus* [25].

Di base, i dati forniti sono:

- csv;
- xml;
- formato tabellare consultabile online;
- definizioni del file per la consultazione dei dati tramite web service;
- gtfs (General Transit Feed Specification): un formato comune per orari ed informazioni geografiche per i trasporti pubblici.

La piattaforma fornisce i dati per le seguenti sezioni:

- rivenditori biglietti;
- zone (comuni) con i relativi codici di riferimento;
- linee e stazioni ferroviarie;
- archi, linee, fermate, percorsi bus.

Riportiamo un estratto dalla lista dei rivenditori di biglietti per i trasporti (codice xml 1.1), dal formato .xml [26]

```
<Table>
  <codice>3910</codice>
  <ragione_sociale>
    TABACCHERIA DESIDERATI DAVIDE
  </ragione_sociale>
  <giorno_chiusura>DOM</giorno_chiusura>
  <indirizzo_ubicazione>
    VIA BENTIVOGLIO , 180
  </indirizzo_ubicazione>
  <cap_indirizzo_ubicazione>44121</
    cap_indirizzo_ubicazione>
```

```
<comune_indirizzo_ubicazione>
  PONTELAGOSCURO -FE-
</comune_indirizzo_ubicazione>
<provincia_indirizzo_ubicazione>
  FE
</provincia_indirizzo_ubicazione>
<coordinate>44.8582,11.6083</coordinate>
<codice_zona_stimer>600</codice_zona_stimer>
<rivenditore_titoli_sosta>0</rivenditore_titoli_sosta
  >
</Table>
```

Listing 1.1: TPER – Estratto xml rivenditori biglietti

Comune di Bologna – colonnine per la ricarica di veicoli elettrici

Il comune di Bologna mette a disposizione diversi data set, relativi a differenti settori (Ambiente, Salute, Economia e Finanza, ecc...). Uno di questi riguarda la mappatura delle colonnine per la ricarica di veicoli elettrici.

I formati forniti sono il csv e il geojson. Il formato geojson è un formato aperto utilizzato per archiviare una collezione di geometrie spaziali i cui attributi sono descritti attraverso JavaScript Object Notation. Le geometrie spaziali possono indicare una serie di punti (indirizzi o toponimi), linee (strade, percorsi, confini) o poligoni (comuni, laghi, ecc...) e collezioni multiple di queste tipologie.

Nel codice 1.2 viene riportato un estratto del formato geojson relativo al posizionamento di queste colonnine nel comune di Bologna, con i dati descrittivi relativi alle colonnine e le coordinate esatte [27]:

```
{
  "type": "Feature",
  "properties": {
    "name": "Emilianauto Group, Bologna ",
```

```
        "description": "Via De' Carracci 8/6 - Accesso  
        alla colonnina da via de' Carracci, di fianco  
        al salone Mitsubishi e di fronte al salone  
        Suzuki. Lun-Ven 08:30-12:30/14:30-19 e Sab  
        9-12:30/15-18:30 Dom chiuso. colonnine: 1**"  
    },  
    "geometry": {  
        "type": "Point",  
        "coordinates": [ 11.331561, 44.508826 ]  
    }  
}
```

Listing 1.2: Comune di Bologna – Estratto geojson mappature colonnine

1.2 Data Visualization

La **Data visualization** è, la rappresentazione grafica dei dati. Questo implica la produzione di immagini che comunichino le relazioni tra i dati rappresentati a chi consulta queste immagini.

Questa rappresentazione stabilisce come i valori dei dati saranno mostrati visivamente, determinando come ed in che misura una proprietà di un punto sul grafico (ad esempio dimensione e colore) cambierà in modo da riflettere le variazioni del valore di un dato [28].

Una visualizzazione efficace aiuta gli utenti ad analizzare i dati e le prove. Rendono i dati complessi molto più accessibili, comprensibili ed usabili.

La rappresentazione grafica dei dati utilizza vari elementi visivi come grafici statistici, diagrammi, tabelle e mappe. Per questo, è spesso sia una disciplina creativa, sia una pratica tecnica [28].

Il professor Edward Tufte ha spiegato che gli utenti che consultano una presentazione di informazioni stanno eseguendo particolari compiti analitici

come fare confronti. Il principio di progettazione del grafico informativo dovrebbe supportare l'attività analitica.

I diversi elementi grafici aiutano nell'attività analitica in modi più o meno efficaci. Ad esempio, i grafici a punti e i grafici a barre superano i grafici a torta [29].

Nel suo libro *Visual Display of Quantitative Information* [30] del 1983, Edward Tufte definisce le “*presentazioni grafiche*” ed i principi per un'efficace visualizzazione grafica.

Secondo Tufte, il modo migliore per comunicare con chiarezza, precisione ed efficienza idee complesse è quello di far sì che l'attenzione dello spettatore sia concentrata sui dati che vengono mostrati e quindi alla sostanza delle informazioni rappresentate. L'occhio dello spettatore deve essere in grado di confrontare agevolmente diversi dati ed avere una presentazione dei dati a vari livelli di dettaglio (da una panoramica più ampia ad una rappresentazione più dettagliata).

Sarà molto importante far sì che la visualizzazione non presenti una visione distorta di ciò che devono comunicare i dati. Sarà necessario rendere coerenti i set di dati di grandi dimensioni e presentare l'elevato numero di dati in un piccolo spazio.

La rappresentazione dovrà avere uno scopo ben chiaro e definito che sia descrittivo, esplorativo, decorativo o tabulare. Inoltre dovrà essere corredata da descrizioni scritte o verbali relative al set di dati utilizzato. Molto spesso la grafica dà voce ai dati e può essere più precisa e illuminante rispetto ai classici calcoli statistici [30].

La mancata applicazione di questi principi può portare alla creazione di grafici fuorvianti, che distorcono il messaggio o che portano ad una conclusione errata. Secondo Tufte, il termine “*chartjunk*” fa riferimento a tutti gli elementi visuali nei grafici che non sono necessari alla comprensione dell'informazione che si vuole rappresentare o che distraggono dall'acquisizione di quell'informazione [30].

L'Ufficio del Congresso del Bilancio americano ha sintetizzato diverse best practice per la rappresentazione grafica in una presentazione di giugno del 2014.

Queste includevano [31]:

- conoscere il proprio pubblico;
- progettazione di elementi grafici che possono essere autonomi al di fuori del contesto del rapporto;
- progettazione di grafici che comunichino i messaggi chiave nel rapporto.

Nella presentazione qui sopra riportata, si fa riferimento al documento *An Economist's Guide to Visualizing Data* (di Jonathan A. Schwabish) [32] dove sono presenti molte considerazioni relative alla data visualization.

Schwabish scrive che siccome i nostri occhi rilevano una serie limitata di caratteristiche visuali, come forme o contrasti, possiamo facilmente combinare queste caratteristiche e inconsciamente percepirle come un'immagine. In contrasto ad una elaborazione "*attentiva*" - la parte cosciente della percezione che ci permette di percepire le cose in serie - l'elaborazione "*preattentiva*" viene eseguita in parallelo ed è molto più veloce.

Quest'ultimo tipo di elaborazione permette al lettore di percepire simultaneamente elementi multipli visuali di base. Nel documento viene riportato un semplice esempio, dove il lettore deve contare il numero di 3 nella seguente lista di numeri [32]:

1269548523612356987458245

0124036985702069568312781

2439862012478136982173256

ed a cui viene chiesto di ripetere lo stesso test con la seguente lista di numeri [32]:

126954852**3**612**3**56987458245

01240**3**6985702069568**3**12781

24**3**98620124781**3**698217**3**256

Le occorrenze nel secondo set si trovano molto più facilmente perché sono evidenziati utilizzando un attributo “*preattentivo*” (ovvero una caratteristica in grado di allertare in maniera inconsapevole il cervello umano) — in questo caso l’intensità del tipo grassetto.

L’autore scrive la creazione di immagini di alta qualità, che mostrino i dati accuratamente ed efficacemente, è un lavoro di immaginazione: la rappresentazione creata dovrà essere compresa ed apprezzata da parte di chi la consulerà e ne dovrà acquisire le informazioni [32].

Esistono migliaia di approcci alla presentazione dei dati, ma, secondo Schwabish, tre sembrano essere i principi base utili alla creazione di visualizzazioni grafiche.

1. Mostrare i dati

Le persone leggono i grafici nei report delle ricerche, articoli o blog per capire la “storia che viene raccontata” [32]. I dati sono la parte più importante di un grafico e dovrebbe essere presentata nella maniera più chiara possibile. Questo implica che non è necessario dover mostrare tutti i dati.

2. Ridurre la confusione e il disordine

L’utilizzo di elementi visuali che distraggono o non necessari porteranno ad una riduzione dell’efficacia della trasmissione dell’informazione. La confusione deriva da:

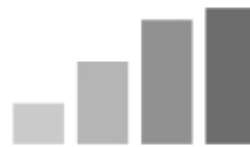
- linee di griglia scure o spesse, segni di spunta, etichette o testi non necessari;
- icone o immagini non necessarie;
- ombre e gradienti ornamentali;

- dimensioni inutili.

Ad esempio, troppo spesso dei grafici vengono texturizzati o riempiti con gradienti (figura 1.3a) quando semplicemente diverse tonalità di un colore svolgerebbero lo stesso compito (figura 1.3b).



(a) Grafico con patterns differenti
[32]



(b) Grafico con variazione tonalità
[32]

3. Integrazione del testo e del grafico

I report di ricerca standard soffrono spesso dell' “*effetto slideshow*” [32] dove lo scrittore descrive gli elementi testuali che appaiono nel grafico. Un modello migliore sarebbe quello in cui le visualizzazioni sono costruite per completare il testo ed allo stesso tempo che contengano abbastanza informazioni da essere comprensibili anche al di fuori del contesto del report.

Come semplice esempio: le legende che definiscono o spiegano una linea, una barra o un punto sono spesso posizionati lontano dal contenuto del grafico (ad esempio strettamente sulla destra o sotto il grafico). Le legende integrate appena subito sotto il titolo, direttamente sul grafico o alla fine di una linea sono molto più accessibili.

I tre principi descritti da Schwabish racchiudono l'idea che l'autore delle rappresentazioni dei dati dovrebbe fare in modo che il lettore acquisisca le informazioni in modo facile e veloce.

Quasi tutte le visualizzazioni di dati sono create per l'utilizzo umano. La conoscenza della percezione e cognizione umana è necessaria quando si progettano visualizzazioni intuitive [33]. La cognizione si riferisce a processi

umani come la percezione, attenzione, apprendimento, memoria, pensiero, formazione di concetti, lettura e problem solving [34]. L'elaborazione visiva umana è efficiente nel rilevare modifiche e nel fare confronti tra quantità, grandezze, forme e variazioni di spessore. Quando le caratteristiche di dati simbolici sono mappati in proprietà visive, gli umani possono navigare tra un grande numero di dati in modo efficiente. È stato stimato che i $\frac{2}{3}$ dei neuroni possano essere coinvolti nell'elaborazione visiva. Una corretta visualizzazione fornisce un approccio differente per mostrare nuovi collegamenti, relazioni, ecc.. che non sono evidenti in set di dati non rappresentati graficamente [35].

1.2.1 Cenni Storici

Non esiste una “storia completa” relativa alla visualizzazione dati. Non ci sono resoconti che racchiudano l'intero sviluppo del pensiero visivo e della rappresentazione visiva dei dati e che raccolgano i contributi di varie discipline [36].

Michael Friendly e Daniel J. Denis della York University sono impegnati in un progetto che tenta di fornire una storia completa della visualizzazione. Contrariamente alla credenza generale, la visualizzazione dei dati non è nata nei tempi moderni. Dati stellari o informazioni come la posizione delle stelle sono stati riportati sulle pareti delle caverne (come quelle trovate nella grotta di Lascaux nel sud della Francia) sin dall'era del Pleistocene [37].

La prima visualizzazione dati documentata può essere fatta risalire al 1160 a.C. con la mappa del papiro che si trova a Torino che illustra accuratamente la distribuzione delle risorse geologiche e fornisce informazioni sull'estrazione di tali risorse [36].

Tali mappe possono essere classificate come cartografia tematica, che è un tipo di visualizzazione dati che presenta e comunica dati e informazioni specifiche attraverso un'illustrazione geografica progettata per mostrare un tema particolare connesso con una specifica area geografica. Le prime forme documentate di visualizzazione dei dati erano varie mappe tematiche di culture

diverse, ideogrammi e geroglifici che fornivano e consentivano l'interpretazione delle informazioni illustrate.

L'idea delle coordinate è stata utilizzata dagli antichi geometri egiziani nella disposizione delle città, le posizioni terrene e celesti erano individuate da qualcosa di simile alla latitudine e alla longitudine almeno nel 200 a.C., e la proiezione cartografica di una terra sferica in latitudine e longitudine da Claudio Tolomeo ad Alessandria serviranno come standard di riferimento fino al XIV secolo [36].

Il filosofo e matematico francese René Descartes insieme a Pierre de Fermat hanno sviluppato la geometria analitica e il sistema di coordinate bidimensionale che ha fortemente influenzato i metodi pratici di visualizzazione e calcolo dei valori. Il lavoro di Fermat e Blaise Pascal sulla statistica e la teoria della probabilità ha gettato le basi per ciò che ora concettualizziamo come dati [36].

Secondo la Interaction Design Foundation, questi sviluppi hanno permesso e aiutato William Playfair, che vedeva il potenziale per la comunicazione grafica di dati quantitativi, a generare e sviluppare metodi grafici di statistica [33].

John Tukey e Edward Tufte, ai loro tempi, hanno superato i limiti della visualizzazione dei dati; Tukey con il suo nuovo approccio statistico all'analisi esplorativa dei dati e Tufte con il suo libro *The Visual Display of Quantitative Information* hanno aperto la strada al perfezionamento delle tecniche di visualizzazione dati non strettamente correlati all'ambiente statistico.

L'avanzamento della tecnologia ha portato alla progressione della visualizzazione dei dati: iniziando con visualizzazioni disegnate a mano ed evolvendo in applicazioni più tecniche, inclusi progetti interattivi che portano alla visualizzazione dati tramite appositi software [36].

1.2.2 Casi Studio

Regione Emilia Romagna – S3 Monitoraggio [38]

La Strategia di specializzazione intelligente (S3 - Smart Specialisation Strategy) è uno strumento utilizzato in tutta l'Unione europea per migliorare l'efficacia delle politiche pubbliche per la ricerca e l'innovazione.

La S3 individua gli ambiti prioritari di ricerca e innovazione su cui intervenire, con l'obiettivo di garantire un maggiore orientamento al risultato degli interventi, in particolare di quelli rivolti alla ricerca e all'innovazione.

La S3 dell'Emilia-Romagna definisce gli obiettivi da raggiungere per il sistema economico regionale nel suo complesso e, al tempo stesso, declina le sinergie con il mondo della ricerca e con quello della formazione, così come – ad esempio – con i temi dell'ambiente e dello sviluppo sostenibile, delle nuove tecnologie e dell'Ict, della salute e dell'attrattività turistica.

La strategia individua cinque grandi ambiti produttivi su cui concentrare l'azione delle politiche regionali di innovazione: tre di essi - **agroalimentare, mecatronica e motoristica, costruzioni** - rappresentano gli attuali pilastri dell'economia regionale, gli altri due - **salute e benessere, cultura e creatività** - costituiscono invece ambiti produttivi con alto potenziale di espansione e di cambiamento anche per altre componenti del sistema produttivo.

Nell'immagine 1.4 presentiamo un'infografica dei dati relativi allo strumento S3 della regione Emilia Romagna.

Comune di Bologna - Bilancio Ambientale [39]

Il comune di Bologna mette a disposizione diverse infografiche basate sui dati raccolti negli ultimi anni per il bilancio ambientale. Nella figura 1.5 è presente l'esempio di visualizzazione di dati relativa alla generazione di energia dai rifiuti.

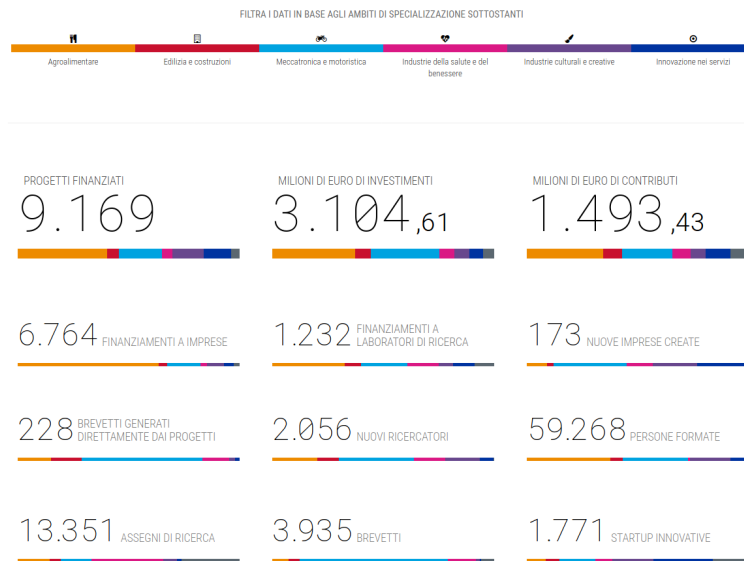


Figura 1.4: Infografica dati S3 - Emilia Romagna [38]

Eni - La gender equality vista attraverso gli occhi della Rete [40]

Eni [40] ha pubblicato un articolo relativo alla parità di genere nel corso degli ultimi anni con considerazione anche degli effetti della pandemia su un argomento delicato come questo. Nell'immagine 1.6 viene mostrata un'infografica di esempio dell'articolo che ben rappresenta i dati relativi agli hashtag maggiormente utilizzati in relazione all'argomento e che mostra quanto possa essere *artistica* la creazione di una rappresentazione grafica di un set di dati.

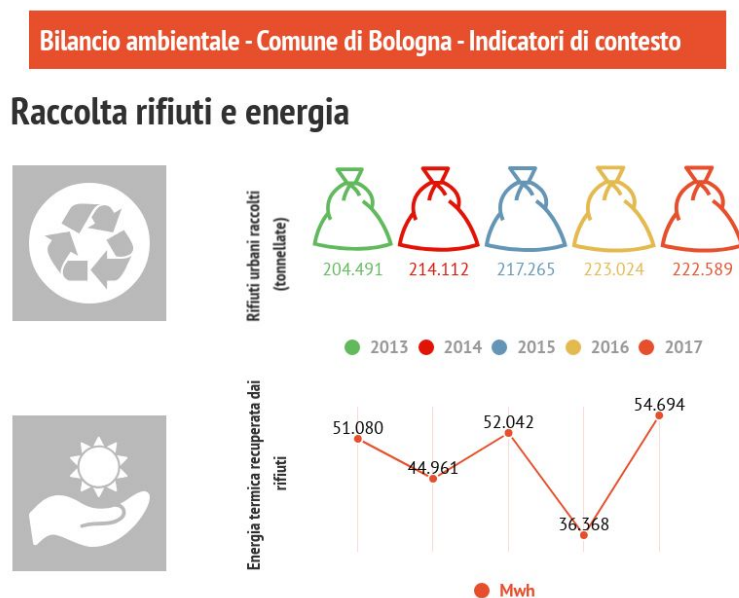


Figura 1.5: Comune di Bologna: produzione di energia dai rifiuti, dal 2013 al 2017 [39]

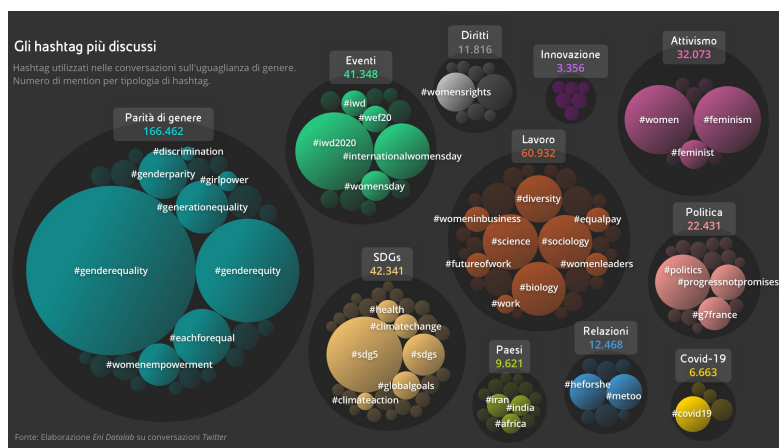


Figura 1.6: Gender Equality - Hastag più usati relativi all'argomento

Capitolo 2

COVID-19: open data e data visualization

2.1 COVID-19 e Open data

L'attuale pandemia di COVID-19 solleva importanti domande sull'apertura, la condivisione e l'utilizzo dei dati ed evidenzia le sfide associate all'uso di questi.

Questa pandemia ha portato alla luce la necessità di affrontare due obiettivi gemelli: aprire dati per aiutare ad affrontare insieme la diffusione del virus COVID-19 ed assicurarsi che la privacy e i diritti di individui e comunità siano protetti.

Come possiamo quindi garantire che questi due punti fondamentali vengano rispettati facendo sì che i dati siano liberamente consultabili ed utilizzabili ma al tempo stesso vengano rispettati i principi etici fondamentali? [41].

Uno dei modi che fornisce la risposta alla precedente domanda è rendere open i dati di cui le persone hanno bisogno per sostenere la collaborazione, consentire azioni, facilitare il controllo e creare fiducia per mitigare l'impatto di questa malattia sull'economia e società mondiale.

L'OCSE[42] (Organisation for Economic Co-operation and Development) e ODC[43] (Open Data Charter) hanno cooperato per avviare una collabo-

razione internazionale per identificare gli open data di alto interesse in una pandemia. Lo scopo di questa collaborazione è quello di:

- identificare i dati necessari alle diverse società;
- trovare esempi di come questi dati possono essere utilizzati;
- definire gli standard e le garanzie esistenti su come tali dati dovrebbero essere raccolti e strutturati;
- dare indicazioni su come comportarsi all'insorgenza di eventuali problematiche.

Il lavoro che stanno portando avanti si basa sulle linee guida di apertura dati dell'ODC per la pubblicazione di questi ultimi [44].

Le risorse disponibili online sono molteplici e molto diversificate tra loro: dai dati relativi ai decessi, ai casi, ai test effettuati paragonati ai test risultati positivi, dati sul cambiamento della mobilità fino ad arrivare addirittura a dati aperti relativi a risorse di AI (Artificial Intelligence) e immagini.

Un esempio dell'ultimo caso è la piattaforma della Stanford University che ha collezionato una lista di dati immagine e AI supportati da tutti i colleghi in giro per il mondo. L'obiettivo che si è posta la Stanford University è quello di dare una mano a livello di ricerca e formazione, sforzi che sono fondamentali per una migliore comprensione ed un più rapida diagnosi del COVID-19 [45].

Altro grande ed importante lavoro di collaborazione lo ritroviamo in un archivio di oltre 29.000 articoli accademici sulla famiglia dei corona-virus, provenienti da tutto il mondo. L'archivio COVID-19 (acronimo per *COVID-19 Open Research Dataset*) è stato rilasciato gratuitamente il 16 Marzo 2020 ed è in continuo aggiornamento da allora [46].

Il rilascio di questo dataset è stato il risultato del lavoro di Microsoft Research, dell'Allen Institute for AI, della National Library of Medicine presso il National Institutes of Health (NIH), del White House Office of Science and

Technology (OSTP) e altri.

Comprende ricerche leggibili dalle macchine per oltre 13.000 articoli accademici. L'obiettivo è consentire alle comunità di ricerca medica e alle macchine per l'apprendimento automatico di applicare i meccanismi di text mining sui dati che possano aiutare a combattere il COVID-19. Il text mining può consentire ai ricercatori di valutare ipotesi, formulare piani di ricerca, aiutare nella comprensione di lavori importanti ed una serie di altri aiuti come come creare bot che rispondano alle domande.

Un ulteriore esempio della condivisione di dati e della collaborazione tra diversi soggetti è data da oltre 500 scienziati di tutto il mondo che hanno raccolto e condiviso dati raccolti dall'uso di tecniche spettrometria di massa che esaminano il sangue delle persone ed altri biomarcatori [47]. La collaborazione è stata coordinata dalla Manchester University che, insieme ai principali esperti mondiali di spettrometria, ha gestito il lavoro di analisi dei modi questo virus si presenta nel sangue dei pazienti e com'è strutturato. L'obiettivo in questo caso è quello di affinare gli approcci di test, stratificare le opzioni di trattamento, determinare i requisiti di isolamento ed aumentare la velocità negli aspetti di misurazione di nuovi programmi di sviluppo terapeutico, sia per il COVID-19 che per necessità future.

Questi sono solo alcuni degli esempi di come la collaborazione tra diversi soggetti, la condivisione di dati, teoria e studi possa portare ad un miglioramento delle tecniche a disposizione per affrontare al meglio e con più preparazione le problematiche (come nel caso del COVID-19) che coinvolgono tutto il mondo (e non solo).

2.2 COVID-19 e Data visualization

Il modo migliore per tenere traccia dell'epidemia di COVID-19 è utilizzare la visualizzazione dei dati basandosi sulle fonti messe a disposizione, come gli open data. Gli schemi relativi al COVID-19 a volte sono contraddittori

ed è qui che entrano in gioco le visualizzazioni, dando enorme supporto ad una corretta e più immediata valutazione dei dati.

La visualizzazione dei dati consente la comprensione della pandemia grazie alle numerose visualizzazioni di dati offerte da diverse piattaforme online. Inoltre, le visualizzazioni “scompongono” le tendenze del COVID-19 in informazioni semplificate per la comprensione. I data scientist utilizzano le visualizzazioni per prevedere il prossimo breakpoint, prevedere i livelli di picco e misurare l’appiattimento della curva.

La maggior parte dei paesi affetti dall’epidemia utilizzano diversi sistemi di sorveglianza gestiti in collaborazione con dipartimenti sanitari statali, locali e territoriali, sanità pubblica, laboratori commerciali e clinici, uffici di statistica vitale, operatori sanitari, dipartimenti di emergenza e partner accademici per monitorare il COVID-19. Le visualizzazioni dati di supporto a questi sistemi di sorveglianza forniscono un riepilogo e un’interpretazione dei dati che verranno utilizzati per monitorare la progressione e la gravità della malattia durante il corso della pandemia.

Qui di seguito verranno proposti degli esempi di data visualization relativi al COVID-19.

WHO – World Health Organization

Qui vengono presentate alcune infografiche presenti nella pagina principale dell’Organizzazione Mondiale della Sanità. Nella prima immagine (figura 2.1) abbiamo un cartogramma. Sulla piattaforma dell’Organizzazione Mondiale della Sanità, la mappa è interattiva. L’utente passando con il puntatore del mouse sopra uno stato (o con un tocco nel caso di dispositivi mobili) può visionare il numero totali di casi e di decessi per quel paese. Sulla sinistra è presente una barra laterale dove è possibile modificare i parametri:

- scegliere la tipologia di mappa (coropletica o a “bolle”);
- scegliere il tipo di dato rappresentato (casi o decessi);

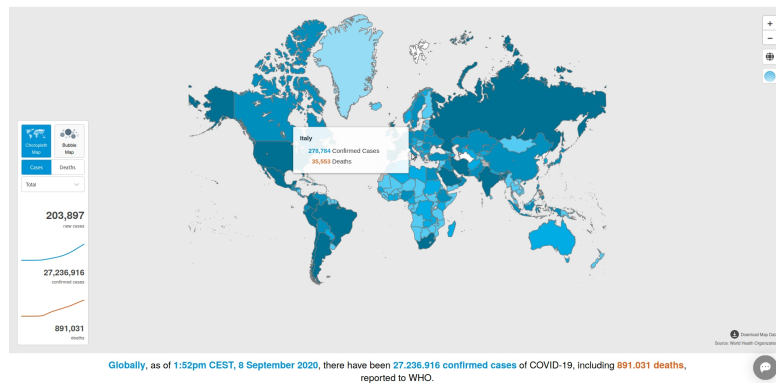


Figura 2.1: WHO - Cartogramma interattivo del numero di casi [48]

- scegliere la tipologia di “raggruppamenti” dei dati per la rappresentazione (conteggio totale, conteggio totale su un milione di persone, % di variazione negli ultimi sette giorni, ecc...).

Nella barra a sinistra è riportato anche un breve riepilogo dei totali mondiali per: nuovi casi, casi, decessi.

Sulla destra invece troviamo i comandi per lo zoom della mappa e la legenda dei colori (o dimensioni bolle) per poter interpretare al meglio i dati.

Le immagini successive mostrano una combinazione di due grafici della stessa tipologia: a barre orizzontale e a barre impilate. I due grafici sono interattivi ed il secondo dipendente in parte dal primo. infatti fermandosi nel primo grafico su una delle barre colorate, nel secondo grafico a destra rimarranno evidenziati i risultati solamente della regione associata.

Sul secondo grafico, fermandosi su una delle colonne, è possibile vedere i dati di riferimento a quella specifica sezione.

Protezione Civile

Nell'immagine 2.3, viene presentata la schermata della dashboard della protezione civile che fornisce diverse informazioni inerenti la diffusione

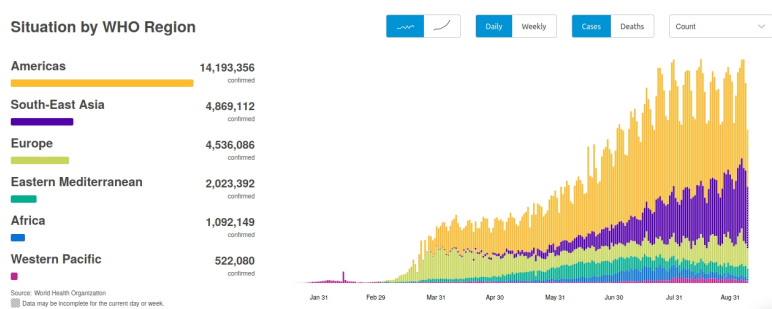


Figura 2.2: WHO - Grafico a barre orizzontali (sinistra) e a barre impilate (destra), interattivo, basato sul numero di casi [48]

del virus in Italia: numero dei casi, numero dei decessi, totale dei casi e numero dei guariti; l'andamento nazionale dei casi registrati e dei guariti ed un grafico (in basso a sinistra) che riporta l'andamento dei nuovi casi registrati. Nella mappa centrale, le bolle rappresentano il numero dei positivi localizzati nelle diverse aree italiane: più è grande è la bolla e maggiore sarà il numero dei casi accertati. Le diverse infografiche sono interattive in quanto, sostando con il puntatore sopra le linee (grafici laterali a sinistra) o sulle bolle della mappa centrale, vengono fornite informazioni aggiuntive relativamente ai dati "sottostanti" il puntatore.

ISS - Istituto Superiore della Sanità

Nel grafico a ciambella presente sul sito dell'Istituto Superiore della Sanità [50], vengono indicate, per ogni settore colorato, la proporzione percentuale dei casi di COVID-19 registrati in Italia negli ultimi trenta giorni per classe d'età. Ogni settore quindi riporta una colorazione relativa alla fascia d'età corrispondente ed ha un'ampiezza proporzionale alla percentuale dei casi. Con il posizionamento del puntatore sopra una delle sezioni del grafico a torta, verrà mostrata la percentuale con il relativo numero di individui che rientrano in quella classificazione. Il grafico a barre impilate (figura 2.4) mostra invece, in base alle fasce d'età, quali siano le percentuali dei diversi stati clinici dei positivi.



Figura 2.3: Protezione Civile – Dashboard COVID-19 Situazione Italia [49]

Questo grafico, a differenza del precedente, non è interattivo in questo caso.

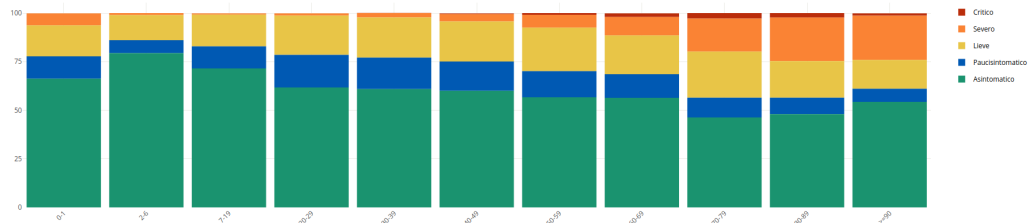


Figura 2.4: Proporzione (%) di casi di COVID-19 segnalati in Italia negli ultimi 30 giorni per stato clinico e classe di età (dato disponibile per 21.823 casi)[50] – Aggiornato al 08/09/2020

2.3 COVID-19 e Storytelling

Tanti governi, organizzazioni, istituti pubblici e di ricerca hanno contribuito alla generazione di molte banche dati relative al COVID-19 per cercare di aiutare e facilitare il lavoro di tutte le persone impegnate nella gestione della situazione economica, sociale e sanitaria sia a livello mondiale che locale. In tutto questo le visualizzazioni dati sono molto importanti per la comprensione del fenomeno, del suo andamento passato per intuirne lo sviluppo futuro.

Ma ad oggi, qual è la situazione a livello mondiale? Cosa ci raccontano tutti i dati raccolti in banche dati a disposizione di tutti per la consultazione? Cosa ci indicano le tante infografiche troviamo già pronte?

L'obiettivo è lo stesso per tutto il mondo: riuscire a portare i casi di COVID-19 a quota zero. Con i dati a nostra disposizione ed aggiornati quotidianamente, possiamo individuare “*facilmente*”, tramite le visualizzazioni dati, quali siano i paesi che stanno facendo progressi da questo punto di vista (ed ovviamente quali no).

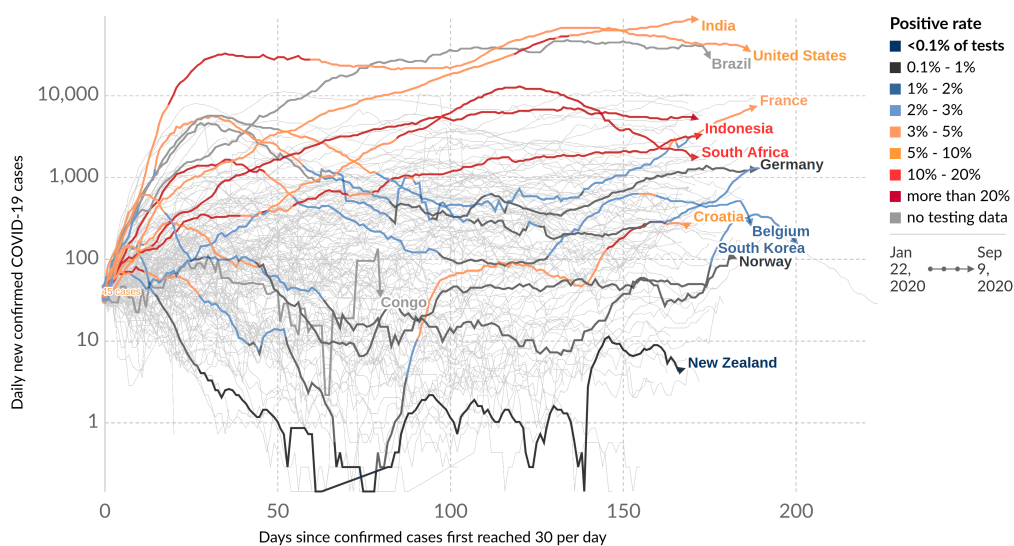
Nel grafico in figura 2.5, aggiornato al 09/09/2020, che ci mostra alcuni di questi paesi.

Le linee mostrano l'andamento del numero giornaliero di casi confermati. Ovviamente i dati a disposizione sul numero dei casi confermati diventano significativi solo quando possono essere interpretati alla luce di quanto un paese stia effettuando o meno i test sulla popolazione. La colorazione delle linee nel grafico mostrano se un paese sta eseguendo il numero adeguato di test.

Nello specifico un paese non sta effettuando un numero di test adeguati quando risulta esserci un caso positivo per un numero molto ridotto di test eseguiti. In questa situazione è probabile che il numero reale dei nuovi casi sia molto superiore rispetto a quelli confermati dai test. Quando la positività risulta alta, la linea assumerà una colorazione rossa.

Daily new confirmed COVID-19 cases

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Our World
in Data

Source: European CDC - Situation Update Worldwide - Last updated 9 September, 09:35 (London time), Official data collated by Our World in Data
CC BY

Figura 2.5: Our World in Data: Andamento di alcuni dei paesi con maggior tasso di crescita del numero di casi negli ultimi 7 giorni [51]

Al contrario, le linee sulle tonalità più scure del blu, indicano che un paese effettua molti test, perché il rapporto tra i test effettuati ed i casi confermati è molto basso. Significa quindi che i paesi con un tasso basso si stiano comportando adeguatamente all'emergenza.

Dal grafico nella figura 2.5 possiamo vedere come alcuni paesi siano stati in grado di abbassare la curva dei contagi in anticipo, applicando misure preventive adeguate a differenza di altri paesi che invece non hanno reagito prontamente alla diffusione del virus [52].

Essendo un problema attuale diffuso in tutto il mondo, per poterci considerare al sicuro, ogni regione del mondo deve compiere dei passi in avanti nella lotta alla diffusione del virus. Questo significa che tutte le linee dovranno avvicinarsi sempre più lo zero fino a raggiungerlo.

I miglioramenti sono possibili: alcuni paesi hanno abbassato la curva dei

nuovi casi e stanno monitorando bene l'epidemia. Ma a livello globale siamo ancora molto lontani dall'obiettivo e il numero globale di casi confermati purtroppo sta aumentando estremamente rapidamente.

La domanda principale è: possiamo fare progressi?

La domanda subito successiva è: come possiamo ottenere dei miglioramenti?

Alla prima domanda abbiamo risposto precedentemente visionando l'andamento dei dati di alcuni paesi, che si sono comportati egregiamente fornendo una risposta rapida ed efficace nella diffusione della pandemia.

Allora come può un paese far sì che ci siano dei progressi?

La risposta alla domanda precedente è: limitare l'impatto diretto e quello indiretto della pandemia.

I paesi che sono stati in grado di rispondere in modo migliore hanno evitato di dover trovare un compromesso tra un'alta mortalità ed un alto impatto socio-economico sul paese.

La Nuova Zelanda è stata in grado di ridurre al minimo le infezioni e poter quindi rendere liberi gli spostamenti internamente. Anche altre nazioni insulari sono state in grado di prevenire quasi completamente un'epidemia (come Taiwan, Australia e Islanda).

Ma non solamente le isole rientrano nella lista dei paesi con un "buon comportamento": Norvegia, Uruguay, Svizzera, Corea del Sud e Germania sono altri esempi. Questi paesi hanno subito un impatto diretto minore, ma hanno anche limitato gli impatti indiretti perché sono stati in grado di ridurre la durata delle contromisure atte a bloccare la diffusione del virus.

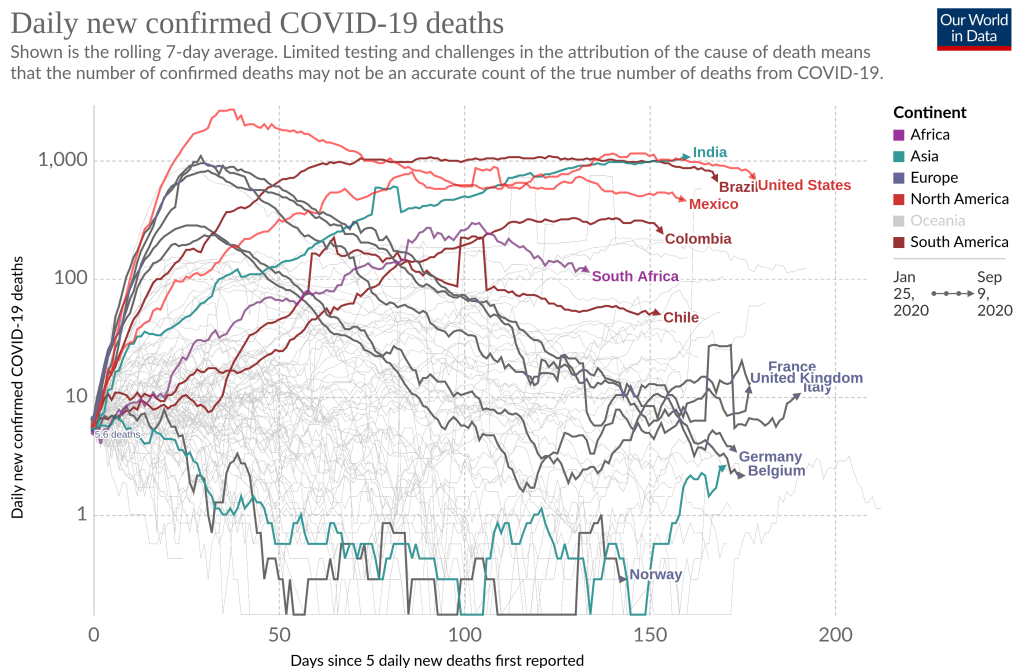
Viene proposto un altro grafico (figura 2.6) che riporta l'andamento delle morti causate dal COVID-19 in un certo lasso di tempo per alcuni paesi.

La pendenza delle linee mostra la rapidità con cui il numero di morti stava aumentando in un particolare momento dell'epidemia.

Come si può notare dal grafico, tra i paesi con il più alto numero di vittime ci sono alcuni di quelli più popolosi del mondo come Stati Uniti, Brasile e Messico.

Daily new confirmed COVID-19 deaths

Shown is the rolling 7-day average. Limited testing and challenges in the attribution of the cause of death means that the number of confirmed deaths may not be an accurate count of the true number of deaths from COVID-19.



Source: European CDC - Situation Update Worldwide - Last updated 9 September, 13:35 (London time), Our World In Data

CC BY

Figura 2.6: Our World in Data - Andamento del numero dei decessi in alcuni paesi a partire dal primo registrato (media settimanale) [52]

Osservando il grafico, possiamo individuare tre differenti modi in cui i paesi hanno reagito alla pandemia:

- alcuni paesi non sono stati in grado di far fronte alla pandemia ed il bilancio dei morti continua a salire rapidamente;
- alcuni paesi sono stati inizialmente colpiti duramente dalla pandemia ma sono poi riusciti a ridurre il numero dei morti. In questo tipo di andamento ritroviamo Italia, Germania ed in generale la maggior parte degli altri paesi europei;
- alcuni paesi sono stati in grado di far fronte fin da subito alla pandemia limitando notevolmente il numero dei decessi registrati. Tra questi troviamo la Corea del Sud e la Norvegia (presenti nel grafico). Questi

paesi hanno avuto dei focolai a cui sono state applicate adeguate misure limitando così la diffusione del virus ed il numero effettivo di decessi.

Sempre dal grafico possiamo notare come, mentre nella fase iniziale c'è stato un enorme incremento di numero di decessi per molti paesi, ad oggi molti paesi (che rientrano nella seconda e terza fascia descritte nell'elenco precedente), abbiano ridotto e, in un certo senso, “stabilizzato” in modo significativo il numero dei decessi.

Come già visto in precedenza, questa è un'ulteriore prova del fatto che sia possibile affrontare la pandemia e che è possibile combatterla con adeguate misure di sicurezza.

Considerando l'elevato numero di decessi registrati l'analisi della mortalità in eccesso è utile per comprendere l'impatto totale della pandemia sui decessi, sia diretti che indiretti.

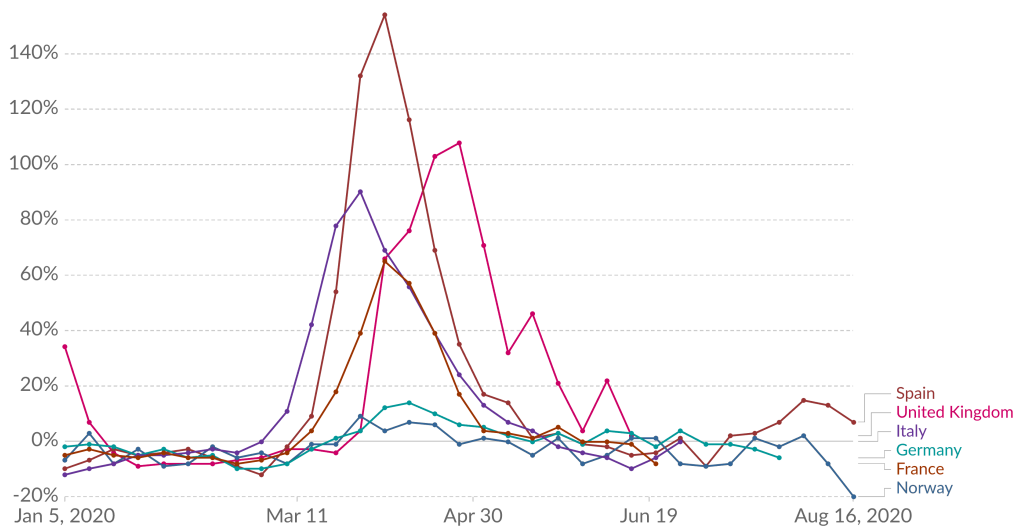
Ci aiuta a capire l'impatto diretto individuando i decessi causati dal COVID-19 che non sono stati correttamente diagnosticati e segnalati, ad esempio perché non è stato condotto alcun test per il virus.

Ci aiuta anche a comprendere l'impatto della mortalità indiretta individuando i molti modi in cui la pandemia ha colpito i sistemi sanitari e le condizioni di vita. Ad esempio, se la pandemia travolge i sistemi sanitari o devia le risorse utilizzate per altri problemi di salute, più persone di quante ci aspetteremmo potrebbero morire per cause non inerenti al COVID-19. La mortalità in eccesso è quindi una misura più completa dell'impatto della pandemia sui decessi rispetto al solo conteggio dei decessi confermato da COVID-19. Il grafico nella figura 2.7 sopra mostra la mortalità in eccesso durante la pandemia come differenza percentuale tra il numero di decessi settimanali nel 2020 e il numero medio di decessi nella stessa settimana nei cinque anni precedenti (2015-2019). Ad esempio, se un paese avesse un punteggio P del 100% in una determinata settimana nel 2020, ciò significherebbe che il conteggio dei decessi per quella settimana era del 100% superiore, cioè il doppio, del conteggio medio dei decessi nella stessa settimana negli ultimi cinque anni.

Excess mortality during COVID-19: The number of deaths from all causes compared to previous years



Shown is how the number of weekly deaths in 2020 differs (as a percentage) from the average number of deaths in the same week over the previous five years (2015–2019). This metric is called the P-score.



Source: HMD (2020), UK ONS (2020)

Note: Dates refer to the last day in each week for most but not all countries. Deaths in recent weeks might be undercounted due to reporting lags. More details can be found in the Sources tab.

OurWorldInData.org/coronavirus • CC BY

Figura 2.7: Our World in Data - Differenza percentuale settimanale delle morti in eccesso rispetto allo stesso periodo dei 5 anni precedenti [53]

Poiché il punteggio, utilizzato per il grafico, misura la differenza percentuale all'interno di un paese, consente confronti diretti tra paesi. Si può vedere dal grafico come alcuni paesi - come il Regno Unito, Spagna e Italia - abbiano affrontato livelli elevati di mortalità in eccesso, mentre altri - come la Germania, Francia e Norvegia - abbiano registrato aumenti molto più modesti.

Nel precedente grafico non erano presenti differenziazioni relative ai rischi di mortalità delle diverse fasce d'età.

Addentrando nei dettagli di un paese, l'Italia, vedremo le analisi effettuate relativamente ai gruppi più a rischio, valutando i grafici relativi ai deceduti risultati positivi per il COVID-19.

Il primo grafico (figura 2.8), in riferimento ai decessi in Italia su individui risultati positivi al COVID-19, mette a confronto il tasso di mortalità per

diverse fasce d'età. Come possiamo intuire dal grafico, la fascia d'età più



Figura 2.8: Numero di decessi per fascia d'età [54]

colpita è quella tra gli 80 e i 90 anni. L'età media risulta essere di 80 anni. Dal grafico possiamo anche osservare che il numero di deceduti di sesso femminile, sotto i 90 anni è sempre inferiore a quello degli uomini. Sotto gli 80 anni risulta quasi dimezzato il numero di decessi di donne rispetto agli uomini. Relativamente all'età, possiamo aggiungere che l'età mediana dei pazienti deceduti positivi a SARS-CoV-2 è più alta di oltre 20 anni rispetto a quella dei pazienti che hanno contratto l'infezione (età mediane: pazienti deceduti 82 anni – pazienti con infezione 58 anni). Le donne decedute dopo aver contratto infezione da SARS-CoV-2 hanno un'età più alta rispetto agli uomini (età mediane: donne 85 – uomini 79) [54].

Un altro parametro interessante da analizzare è la presenza di patologie croniche preesistenti (diagnosticate prima di contrarre il COVID-19) nei soggetti deceduti (figura 2.9).

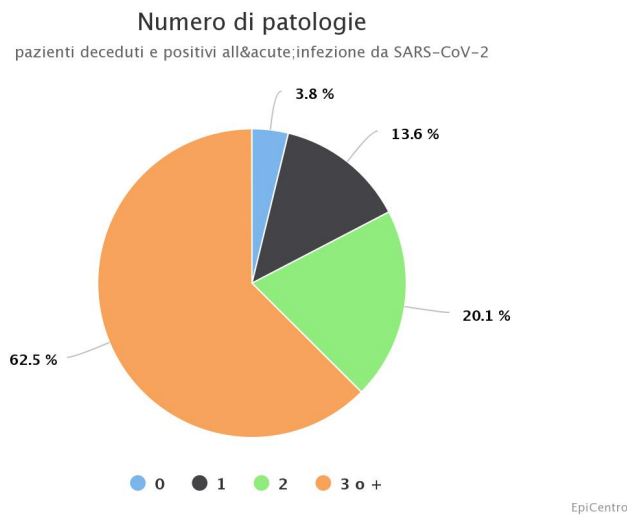


Figura 2.9: Percentuale dei decessi in base al numero di patologie preesistenti [54]

Da questo grafico a torta possiamo facilmente dedurre che il maggior numero di decessi si è verificato per individui con 3 o più patologie preesistenti. La percentuale dei decessi ed il numero di patologie preesistenti sono direttamente proporzionali: man mano che calano il numero di patologie, cala anche la percentuale dei decessi.

Il dato di riferimento per il grafico è stato ottenuto da 4190 deceduti per i quali è stato possibile analizzare le cartelle cliniche. Il numero medio di patologie osservate in questa popolazione è di 3,4 (figura 2.10).

Nel grafico qui sopra abbiamo un dettaglio delle patologie principali riscontrate nei pazienti deceduti. Le patologie registrate più frequentemente sono:

1. *ipertensione arteriosa*: 65.8%;
2. *diabete mellito - Tipo 2*: 29.5%;
3. *cardiopatía ischemica*: 28.0%;
4. *fibrillazione atriale*: 23.4%.

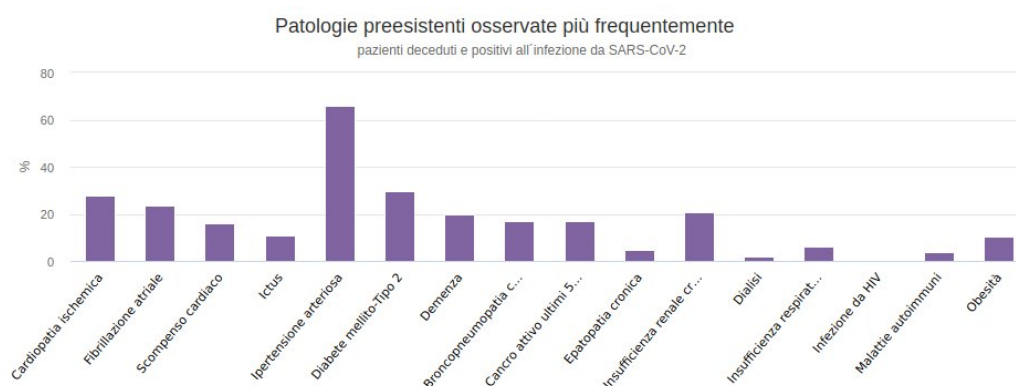


Figura 2.10: Percentuale dei decessi in base al numero di patologie preesistenti [54]

Relativamente alle complicanze presenti nei pazienti deceduti, l'ISS ci mette a disposizione anche un resoconto relativo alle diagnosi effettuate, dandoci delle percentuali relative ai decessi causati da sintomi compatibili con COVID-19 o altre diagnosi non correlate all'infezione.

“Nel 91,3% delle diagnosi di ricovero erano menzionate condizioni (per esempio polmonite, insufficienza respiratoria) o sintomi (per esempio, febbre, dispnea, tosse) compatibili con COVID-19. In 334 casi (8,7% dei casi) la diagnosi di ricovero non era correlata all'infezione. In 50 casi la diagnosi di ricovero riguardava esclusivamente patologie neoplastiche, in 113 casi patologie cardiovascolari (per esempio infarto miocardico acuto-IMA, scompenso cardiaco, ictus), in 46 casi patologie gastrointestinali (per esempio colecistite, perforazione intestinale, occlusione intestinale, cirrosi), in 125 casi altre patologie. [54]”

Facendo il punto per quanto riguarda la situazione in Italia, possiamo dire che l'Italia è uno di quei paesi che a livello mondiale si è comportata abbastanza bene. All'arrivo dell'infezione è stata colpa impreparata, avendo così un rapido aumento del numero di decessi nella fase iniziale. A seguito delle misure cautelative messe in atto dal Governo, il numero dei contagi e dei

morti ha iniziato la sua discesa, stabilizzandosi successivamente all'alleggerimento dei vincoli di spostamento e contatti. Nei primi mesi, come tutt'ora, la media dell'età dei deceduti si attesta tra gli 80 e i 90 anni, mentre negli ultimi mesi la fascia d'età dei contagiati è calata. Dai dati registrati, in confronto con tutto il resto del mondo, l'Italia risulta uno di quei paesi in cui attualmente è alto il numero dei test effettuati, il che consente di individuare tempestivamente eventuali focolai e rendere più facile ed immediata l'applicazione delle misure per limitare la diffusione del virus.

Capitolo 3

Questionario Online

3.1 Contesto del problema

Questo anno è iniziato con la pandemia causata dal virus SARS-Cov-2. Ciò ha portato ad un'esplosione di aggiornamenti continui sull'andamento della situazione, sulle previsioni, sulle statistiche delle settimane passate. Siamo stati quindi invasi, sia sui quotidiani, che sui social, che in TV da un'enormità di grafici ed infografiche in generale di qualsiasi genere colore e struttura.

C'è da chiedersi quindi: ma tutte queste infografiche possono essere facilmente e liberamente consultate da chiunque? Sono comprensibili anche isolate dal contesto scritto o verbale in cui erano collocate? Ci sono dei grafici o delle infografiche che riescono a rendere più comprensibile un'informazione rispetto ad un'altra?

In un periodo come questo, molto complicato e difficile da gestire, riuscire a mantenere aggiornato il pubblico generale con informazioni chiare è molto importante per riuscire a contenere il senso di paura e smarrimento iniziale. Avere qualche informazione in più dell'andamento nazionale rispetto a quello che possa essere il proprio quartiere o la propria città può renderci più consapevoli di quali possano essere i rischi (in questo caso) a cui potremmo andare incontro. Avere qualche nozione aggiuntiva ci consente anche di comprendere

meglio le scelte governative se supportate da studi e ricerche scientifiche che hanno messo a disposizione liberamente dati ed infografiche consultabili da tutti.

Ma quale sarebbe il modo migliore per riuscire ad avere un parere diffuso su quelle che siano le rappresentazioni grafiche migliori per il pubblico?

3.2 Questionari

Per quanto possa sembrare facile ed immediata la creazione e stesura di un questionario, è molto importante che questo abbia una presa considerevole su chi debba compilarlo. È per questo che numerosi studiosi e professionisti si tengono continuamente aggiornati su quali siano le migliori pratiche della metodologia di indagine nella progettazione di questionari, analisi dei dati e, più recentemente, visualizzazione dei dati.

I sondaggi solitamente sono il modo migliore, e spesso uno tra quelli più affidabili, per poter avere un'idea di cosa pensino e desiderino le persone. Quale modo migliore per sapere cosa pensa o desidera un individuo se non chiederglielo direttamente?. Le indagini vengono condotte per accrescere la nostra conoscenza sul pubblico di nostro interesse e risultano essere lo strumento più adatto [55].

Possiamo identificare quattro motivi principali per cui, in alcune occasioni, sia necessario utilizzare i questionari [55].

I sondaggi forniscono dati *importanti*

Nelle notizie i sondaggi vengono utilizzati per avvalorare quelle due o tre storie riportate in un articolo (se ad esempio facciamo riferimento ad un quotidiano cartaceo od online). Se per esempio in un articolo vengono raccontate le storie di due persone che sono state oggetto di una frode finanziaria, in quello stesso contesto potrebbe essere presentato il risultato di un sondaggio relativo agli individui che a loro volta si sono ritrovati nella stessa situazione.

Le storie hanno un impatto molto più forte sul pubblico se riflettono

le vicende di un considerevole numero di individui piuttosto che una ristretta minoranza.

Un altro esempio: i sondaggi sul servizio clienti sono importanti per capire se un cliente insoddisfatto stia esprimendo un sentimento proprio o un malcontento generalizzato. Allo stesso modo un solo cliente soddisfatto non implica che una piccola realtà agli inizi abbia imboccato la vera strada verso il successo.

Se creati correttamente, i sondaggi forniscono dati “critici” sulle opinioni e sui comportamenti delle persone, su cui ci si può basare per prendere decisioni importanti.

Le indagini forniscono parametri di riferimento fondamentali

Le indagini vengono abitualmente utilizzate per prendere decisioni individuali (ad esempio creare una particolare campagna pubblicitaria o un nuovo servizio da offrire) ma diventano ancora più efficaci ed utili se vengono ripetute nel tempo.

Il ripetere lo stesso sondaggio in diversi momenti può fornire preziose indicazioni su quale sia l'evoluzione dei vari aspetti che si stanno monitorando con il questionario.

Per un esempio su larga scala, possiamo prendere il censimento effettuato negli Stati Uniti, che possiamo considerare a tutti gli effetti un sondaggio. Questo sondaggio è estremamente utile per classificare i principali cambiamenti demografici del paese, ad esempio la composizione razziale, dell'età, del genere, ecc... Questo tipo di informazioni potrebbero essere molto utili al Governo per attuare delle politiche adeguate ai propri cambiamenti demografici [55].

Le indagini rivelano il “perché”

I “big data” rappresentano un argomento molto gettonato recentemente. Con il termine “big data” ci si riferisce, per la maggior parte, a dati

impliciti o dati derivanti dall'osservazione e dall'analisi del comportamento di individui, sia online che altrove.

La quantità di dati accumulati aumenta sempre di più. Come esempio possiamo pensare ad un qualsiasi motore di raccomandazione per l'acquisto di prodotti. Se un individuo aggiunge al carrello un certo tipo di prodotto, il motore non può sapere se l'attenzione mostrata derivi da un reale interesse dell'individuo stesso o se sia indotto, ad esempio, dalla necessità di fare un regalo ad un amico/a o parente. Questo va ad intaccare i risultati dei prodotti suggeriti. Per poter capire per quale ragione la persona abbia aggiunto quel tipo di prodotto al proprio carrello, è necessario integrare i dati impliciti con dei dati espliciti che vadano a completare la panoramica che gli algoritmi del motore di raccomandazione di prodotti andrà ad analizzare.

Possiamo definire come dati espliciti quelle informazioni che sono chiare e precise e non presentano ambiguità [55].

E questi dati espliciti sono presi direttamente dalla persona, utilizzando solitamente la metodologia dei sondaggi. E questi danno una maggiore affidabilità nella comprensione delle motivazioni che portano ad eseguire certe azioni.

Le indagini danno una voce

L'economista Albert Hirschman, con il suo libro "*Exit, Voice, and Loyalty*" [56], ha esaminato i principali modi in cui le persone reagiscono di fronte a un'organizzazione scadente.

I business di qualsiasi tipo hanno più successo quando incoraggiano e supportano il "dar voce" piuttosto che l' "uscita". Incentivare i propri clienti a dar voce alle proprie preoccupazioni serve ad aumentare il coinvolgimento, a ridurre la probabilità che si rivolgano ad un altro fornitore (con conseguente perdita di investimento per l'organizzazione che perde il cliente).

Ora che abbiamo definito il perché, possiamo passare al come.

Come deve essere strutturato un questionario? Quali sono le best-practice per avere un buon risultato e soprattutto un buon numero di compilazioni?

Possiamo individuare quattro passaggi fondamentali per la creazione di un sondaggio online [57].

1. Definizione degli obiettivi

Prima di tutto è necessario farsi un'idea mentalmente del risultato che si vuole ottenere.

Alcune indicazioni su quali domande porsi possono essere le seguenti:

- soggetto del feedback che si vuole ricevere (feedback su un prodotto, un servizio, un argomento, ecc...);
- specificità delle informazioni ricercate (servono informazioni generiche o specifiche?);
- specificità dei destinatari del sondaggio (deve essere un pubblico specializzato?).

2. Creazione di una lista di domande

Le tipologie di domande che possono essere inserite in un questionario sono diverse: domande aperte, domande a risposta singola, domande a risposta multipla, ecc..

La maggior parte delle persone che prende parte ai questionari preferisce domande brevi a scelta multipla.

Il linguaggio di scrittura delle domande deve essere molto semplice e sarebbe bene evitare ambiguità.

Uno dei vantaggi dei sondaggi online è che il questionario può essere “interattivo” nel senso che in base alle risposte già fornite dall'utente, è possibile variare le domande successive, migliorando i tassi di risposta.

3. Diffusione del sondaggio

Oggi giorno i canali per la diffusione di un questionario sono molteplici. Si possono inviare mail agli iscritte a liste, pubblicare il sondaggio sui

social, utilizzare le applicazioni per la messaggistica e chat sul proprio smartphone o creare banner pubblicitari su piattaforma esterne che riportino alla pagina del proprio sondaggio.

4. Raccolta delle risposte

è importante monitorare il tasso di risposta, dato che la dimensione del campione finale dipenderà dal numero di partecipanti. In molti casi, il partecipante può essere “invogliato” a completare il questionario a fronte, ad esempio, di un regalo, la possibilità di vincita ad una lotteria, donazioni per la beneficenza, un sistema di “loyalty” e così via. Un altro modo per aumentare il tasso delle risposte è quello di promettere di condividere i risultati del sondaggio con i partecipanti.

Successivamente all’ultimo punto sopra descritto, andranno analizzati i dati, facendosi supportare da visualizzazioni dati che ne rendano l’interpretazione più facile e gestibile. Successivamente alla creazione delle infografiche relative ai risultati, solitamente ci si aspetta un report completo relativo al questionario che spieghi i risultati ottenuti e se sono stati raggiunti gli obiettivi che ci si era prefissati inizialmente. Un sondaggio di successo fornirà risposte affidabili alle domande che sono state poste agli utenti.

3.3 Il progetto

Il questionario ha lo scopo di raccogliere dati relativamente alla comprensione delle infografiche utilizzate dalle diverse fonti nel contesto del COVID-19.

Mettendone a confronto diverse sullo stesso argomento, vogliamo cercare di capire quali siano le migliori dal punto di vista intuitivo e di comprensibilità per le varie categorie di utenti. Il questionario prevede domande di diverso tipo.

Le tipologie di domande possono essere suddivise in tre categorie:

1. domande di carattere personale;

2. domande a scelta singola con un'immagine;
3. domande a risposta mista dove è richiesta un' "azione" da parte dell'utente.

1. Domande di carattere personale Questo tipo di domande sono state inserite per avere una statistica dell'età, sesso e livello di istruzione degli utenti che hanno compilato il questionario

2. Domande a scelta singola con un'immagine Questo tipo di domanda presenta una serie di risposte a scelta singola in cui sono proposte delle immagini di grafici. L'utente deve scegliere quale tra le immagini proposte fornisca in modo più chiaro ed intuitivo l'informazione che viene richiesta nella domanda corrispondente.

3. Domande a risposta mista dove è richiesta un' "azione" In questa tipologia di domanda abbiamo un set di risposte principale che è a scelta singola. L'utente deve ricercare l'informazione richiesta in un link esterno (presente all'interno della stessa domanda) ed indicare poi, nel set principale di risposte, con qualche facilità/difficoltà è riuscito a reperire l'informazione.

In base alla risposta scelta l'utente dovrà scegliere, in un altro set di risposte multiple presentate successivamente alla scelta principale, quali siano le motivazioni per cui ha trovato facilmente/difficilmente l'informazione richiesta. Il set di risposte secondarie varia in base al tipo di risposta segnata nel set principale. La variazione dei set della risposta secondaria varia in base al livello di "difficoltà" indicato nella risposta primaria.

3.3.1 Requisiti

Qui di seguito è presente l'elenco dei requisiti del questionario:

- il questionario è anonimo

- come primo elemento nel documento dovrà essere presente (ovviamente) il titolo, seguito da un'introduzione in cui viene indicato lo scopo del questionario, da chi e come verranno raccolti i dati per fornire tutte le informazioni necessarie agli utenti
- a seguire il punto precedente dovrà essere richiesta l'accettazione al trattamento dei dati (casella da spuntare obbligatoriamente per inviare i dati)
- dopo la lista delle domande dovrà essere, ovviamente, presente un pulsante per l'invio dei dati
- in base all'esito dell'elaborazione dati, l'utente sarà reindirizzato sulla pagina di riferimento: pagina di ringraziamento in caso di successo, pagina con un messaggio di errore nel caso in cui il salvataggio dati non vada a buon fine.
- alto livello di esperienza utente per rendere il questionario di facile compilazione per tutte le tipologie di utenti

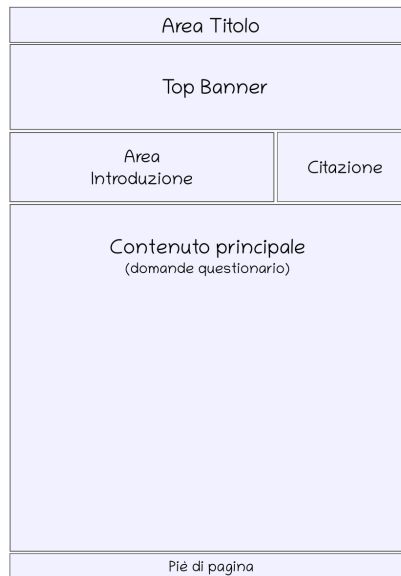
3.3.2 Analisi e modello del dominio

La pagina web che presenterà le domande del questionario agli utenti, può essere suddivisa in due parti principali (figura 3.1).

La prima parte, che chiameremo "*top content*", conterrà:

- il titolo;
- eventuale immagine rappresentativa (per catturare l'attenzione e "riempire/rendere più accattivante" la pagina;
- eventuale citazione relativa alla visualizzazione dati, per rendere la pagina un po' più ricca;
- una breve introduzione al questionario con riportato lo scopo ed indicazioni relativamente al trattamento dei dati inseriti e di eventuali suggerimenti sulla compilazione;

Figura 3.1: Wireframe della pagina principale del questionario



- primo campo di input per la conferma per l'accettazione al trattamento dei dati.

Nella seconda parte, che chiameremo “*content*”, sarà presente tutta la struttura html e javascript necessaria per la presentazione delle domande del questionario.

Le domande possono essere strutturate in tre maniere diverse:

1. domande a risposta multipla con una risposta libera per riportare indicazioni differenti da quelle proposte;
2. domande a risposta singola;
3. domande a risposta singola con sottodomanda a risposta multipla.

Tutte le domande sono obbligatorie.

1. Domande a risposta multipla

Le domande con risposta multipla, avranno, oltre alle opzioni proposte, un'ulteriore opzione “altro” associata ad un campo di testo libero dove l'utente

potrà inserire eventuali opzioni di risposta non presenti tra quelle già proposte.

L'utente deve scegliere *almeno* una risposta tra quelle presenti. Un esempio dal questionario nella figura 3.2.

Domanda 1

Durante il periodo di pandemia Covid-19 dove hai consultato informazioni online sulla diffusione del virus?
(puoi scegliere più di una risposta)

<input type="checkbox"/> Unità sanitaria locale ("AUSL")	<input type="checkbox"/> Ministero salute / Protezione civile	<input type="checkbox"/> OMS - Organizzazione Mondiale della Sanità <small>WHO - World Health Organization</small>
<input type="checkbox"/> Siti di quotidiani nazionali	<input type="checkbox"/> Siti di quotidiani internazionali	<input type="checkbox"/> ECDC - Centro europeo per la prevenzione e il controllo delle malattie <small>European Centre for Disease Prevention and Control</small>
<input type="checkbox"/> Johns Hopkins University	<input type="checkbox"/> Our World in Data	<input type="checkbox"/> Altro <small>Inserisci qui le altre fonti utilizzate</small>

Figura 3.2: Esempio domanda risposta multipla

2. Domande a risposta singola

Le domande a risposta singola riporteranno, ad ognuna di esse, un'immagine di un grafico di riferimento alla domanda.

L'utente dovrà scegliere quale, fra le immagini proposte, risulta essere un'infografica più chiara e comprensibile da cui recuperare l'informazione richiesta. Come indicato nell'introduzione ad inizio pagina, l'utente avrà la possibilità, cliccando sull'immagine di visualizzare un ingrandimento della stessa (gestito tramite libreria javascript) in modo da poter vedere più chiaramente i dati in essa rappresentati. L'utente in questo caso potrà scegliere solo una risposta. Un esempio dal questionario nella figura 3.3.

3. Domande a risposta singola con sottodomanda

In questa tipologia di domanda, all'utente viene richiesto di eseguire un'azione su un sito esterno e di indicare il livello di difficoltà riscontrato nel recuperare l'informazione richiesta. A seconda se la risposta a questa domanda

Domanda 4

Se volessi sapere quale sia il continente che ha registrato il maggior numero di casi, quale tra le seguenti immagini ti consente di reperire più facilmente l'informazione?

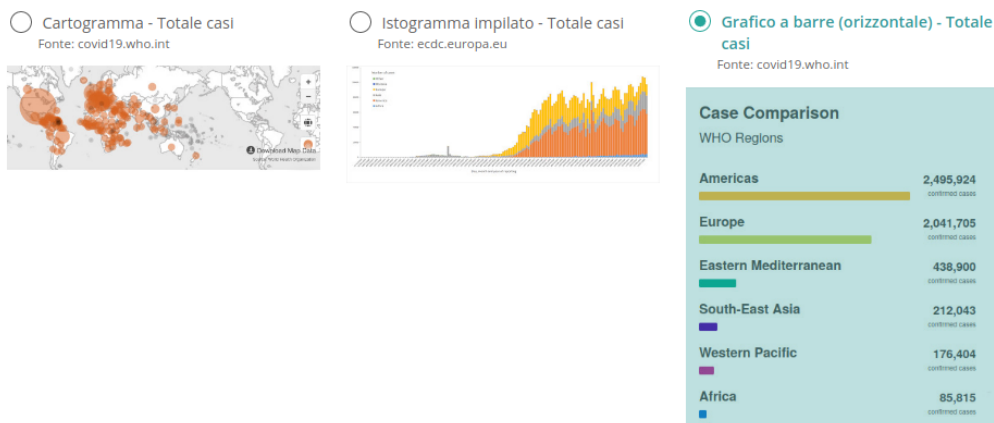


Figura 3.3: Esempio domanda risposta singola

possa considerarsi positiva o negativa, all'utente verrà successivamente richiesto (sottodomanda) di motivare la prima risposta indicando quali elementi abbiamo facilitato o complicato il reperimento dell'informazione (sarà presente anche un campo per la risposta libera se necessario per aggiungere ulteriori informazioni).

Un esempio dal questionario nella figura 3.4.

Il salvataggio dei dati inseriti all'interno della pagina principale verrà gestita lato server dall'applicativo. I dati in ingresso verranno letti, elaborati ed inseriti in un database per la memorizzazione. Nella tabella 3.1 è proposta la lista delle domande.

Tabella 3.1: Lista domande presenti nel questionario

#	Domanda
1	Durante il periodo di pandemia Covid-19 dove hai consultato informazioni online sulla diffusione del virus?

Continua nella pagina successiva

Tabella 3.1 – *Continua dalla pagina precedente*

#	Domanda
2	Considerando in particolare le visualizzazioni grafiche dei dati, quali risorse/siti/pagine web hai consultato?
3	Solitamente, che dispositivo utilizzavi per consultare tali informazioni?
4	Se volessi sapere quale sia il continente che ha registrato il maggior numero di casi, quale tra le seguenti immagini ti consente di reperire più facilmente l'informazione?
5	Naviga sul link sottostante e trova quale sia la provincia italiana con il maggior numero di casi registrati. – Protezione Civile Come ti sei trovato nella ricerca dell'informazione richiesta?
5 bis	Perchè?
6	Se volessi sapere quale sia il paese con il maggior numero di decessi a livello mondiale, quale tra queste immagini ti rende più facile l'acquisizione dell'informazione cercata?
7	Naviga sul link sottostante e trova il numero totale di casi e di decessi attuali per l'Italia – World Health Organization Qual è stato il livello d'impegno per recuperare l'informazione richiesta?
7 bis	Perchè?
8	Vuoi sapere quale sia la regione italiana in cui è stato distribuito il maggior numero di materiali per l'emergenza COVID-19. Quale tra queste immagini ti rende più facile l'acquisizione dell'informazione richiesta?
9	Naviga sul link sottostante e trova quale sia la fascia di età in cui si è avuto il maggior numero di decessi – Istituto Superiore della Sanità Qual è stato il livello d'impegno per recuperare l'informazione richiesta?
9 bis	Perchè?

Continua nella pagina successiva

Tabella 3.1 – *Continua dalla pagina precedente*

#	Domanda
10	Vuoi sapere quale sia il continente dove risulta essere maggiore il numero di test effettuati per il COVID-19 risultati positivi. Quale tra queste immagini ti rende più facile l'acquisizione dell'informazione richiesta?
11	Naviga sul link sottostante. Individua quale sia, ad oggi, la tipologia di mobilità con la percentuale più alta di modifica delle richieste di instradamento, nel tuo stato – Apple Qual è stato il livello d'impegno per recuperare l'informazione richiesta?
11 bis	Perchè?
12	Quale delle seguenti visualizzazioni ti rende più facile individuare la nazione dove sono stati effettuati il maggior numero di test ogni 1000 persone?
13	A quale fascia di età appartieni?
14	Il tuo titolo di studio?
15	Genere?
16	Che dispositivo hai utilizzato per compilare questo questionario?
17	Comprensione dell'inglese scritto?

3.3.3 Architettura e Design

Qui di seguito verrà illustrata l'architettura ed il design dell'applicativo in modo da renderne chiaro il funzionamento e tutti i passaggi relativi al salvataggio dei dati. L'applicativo è costituito da tre pagine: la pagina principale con le domande del questionario, la pagina di ringraziamento e la pagina di errore che verranno mostrate successivamente all'invio dei dati. I dati vengono inviati ad un endpoint lato server che si occuperà della verifica e della memorizzazione dei dati sul DB.

Domanda 5

Naviga sul link sottostante e trova quale sia la provincia italiana con il maggior numero di casi registrati.

> [Protezione Civile](#)

Come ti sei trovato nella ricerca dell'informazione richiesta?

- L'ho trovata molto facilmente
- L'ho trovata facilmente**
- Mi ci è voluto qualche minuto per capire dove poterla reperire
- L'ho trovata con molta difficoltà
- Non l'ho trovata

Perchè?

- Informazioni ben presentate, intuitive (facile reperibilità) Informazioni chiare (facile lettura) I colori usati sono piacevoli ("attrattività/eleganza")

Altro

Indicare qui le altre motivazioni

Figura 3.4: Esempio domanda risposta singola con sottodomanda

Il db è stato progettato come segue:

1. **Tabella con le domande:** in questa tabella è presente la lista delle domande, dove sono riportate anche le relative risposte disponibili per gli utenti. è stato pensato in questo modo per rendere più *dinamica* e *rapida* l'aggiunta di nuove domande e/o risposte, senza dover ogni volta mettere mano alla pagina php principale e dove aggiungere del codice html;
2. **Tabella con le risposte:** in questa tabella viene memorizzato ogni singolo invio ricevuto di dati. Ogni invio è identificato da un nuovo id associato al record della tabella, da un cookie che viene generato dal browser in fase di accesso alla pagina del questionario, il valore di accettazione della privacy. Il cookie viene registrato per ovviare ad eventuali compilazioni duplicate che potrebbero, (se in numero elevato)

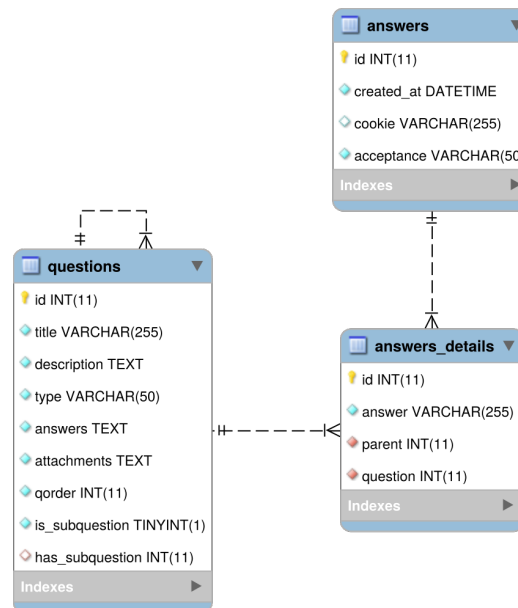


Figura 3.5: Schema del DB per il salvataggio dei dati

falsare i dati in ingresso. In fase di analisi dei dati, verranno dunque presi in considerazione le compilazioni con valori di cookie distinti;

3. **Tabella con il dettaglio delle risposte:** in questa tabella vengono memorizzate le risposte vere e proprie, con riferimenti esterni alla tabella delle domande e alla tabella delle risposte. In questo modo ogni record corrisponderà alla singola risposta di una domanda, associata ad una singola compilazione.

Nella figura 3.1 presentiamo il design della pagina principale e nella figura 3.5 la struttura del database.

3.3.4 Sviluppo

Per la realizzazione dell'applicativo prima di tutto ho creato la lista delle domande, ricercando il materiale e valutando quante e quali domande proporre. Una volta definite le domande con i relativi materiali (lista risposte ed immagini), ho provveduto alla creazione del template della pagina.

Il modello della pagina era un template preimpostato gratuito disponibile online [58]. Il template è stato successivamente personalizzato per rispondere meglio alla tipologia di progetto richiesta ed ottimizzato in base ai contenuti inseriti.

Il template è stato integrato con i seguenti elementi:

- **personalizzazioni css:**

- *jquery.fancybox.min.css* [59]: file css associato alla libreria javascript *jquery.fancybox.min.js* utilizzata per l'ingrandimento delle immagini presenti in alcune domande;
- *custom.css* [60]: file con css personalizzato per il miglioramento e l'ottimizzazione della grafica della pagina.

- **personalizzazioni javascript**

- *jquery.fancybox.min.js* [59]: libreria javascript per l'ingrandimento delle immagini e funzione di scorrimento delle stesse se incluse nello stesso gruppo (il gruppo viene definito in un attributo del tag relativo all'immagine);
- *jquery.validate.js* [61]: libreria per la validazione dei campi input presenti nel form;
- *www.google.com/recaptcha/api.js* [62]: libreria google per l'utilizzo delle api per l'aggiunta di un recaptcha invisibile all'interno della pagina per evitare compilazioni ed invio dati da parte di bot;
- *codice js personalizzato* [63]: codice inserito in fondo alla pagina per aggiungere funzionalità alla stessa ed esecuzione/utilizzo delle librerie incluse.

Per quanto riguarda la personalizzazione javascript presente direttamente nella pagina, riporto qui di seguito un dettaglio del codice ed esplicazione della funzionalità.

Posizionamento random delle risposte: il codice 3.1 ha lo scopo di scambiare la posizione delle risposte in cui è presente un'immagine. Questo per cercare di rendere più distribuito il risultato della risposta in quanto alcune immagini potrebbero essere considerate equivalenti dal punto di vista di presentazione dell'informazione e l'utente potrebbe essere portato a selezionare la prima risposta, portando quindi ad un risultato non troppo realistico. Scambiando la posizione delle immagini in modo randomico, nel caso in cui l'utente sia portato sempre a scegliere la prima opzione rispetto a quelle equivalenti in questo caso il risultato dovrebbe risultare più equilibrato.

```
$.fn.shuffle = function() {
    var allElems = this.get(),
        getRandom = function(max) {
            return Math.floor(Math.random() * max);
        },
        shuffled = $.map(allElems, function(){
            var random = getRandom(allElems.length),
                randEl = $(allElems[random]).clone(true)
                    [0];
            allElems.splice(random, 1);
            return randEl;
        });
    this.each(function(i){ $(this).replaceWith($(shuffled
        [i])); });
    return $(shuffled);
};
$('.tm-q-choice-container.shuffle').each(function(){ $(
    this).children('label').shuffle(); });
```

Listing 3.1: Codice JS per il posizionamento random delle risposte contenenti immagini

Comparsa della sottodomanda alla scelta della risposta alla domanda principale: il codice 3.2 seguente ha lo scopo di gestire la

comparsa (o aggiornamento nel caso in cui cambi una risposta già data) della sottodomanda associata alla propria domanda principale quando un utente seleziona una risposta alla domanda principale. Questo perché in base al tipo di risposta scelta (se positiva o negativa), verrà fatta comparire uno dei due set di risposte alla sottodomanda. Questo perché ovviamente non è possibile già sapere cosa risponderà l'utente e presentare come già visibili entrambi i set di risposte per la sottodomanda causerebbe confusione nell'utente.

```
$( 'div.subquestion, div.subquestion .tm-q-choice-  
container' ).hide();  
$( ".has-subquestion .tm-radio-group" ).on( 'click',  
function () {  
    var refSubquestion = $( this ).attr( 'data-ref' );  
    var subQContainer = $( this ).parents( '.q-container' )  
        .first()  
        .children( '.subquestion' )  
        ;  
    var subSectionVisible = $( subQContainer ).children( '.  
visible' );  
    if ( $( subSectionVisible ).length > 0 &&  
        $( subSectionVisible ).first().attr( 'id' ) !=  
        refSubquestion ) {  
        $( subSectionVisible ).first().slideUp( 400,  
            function () {  
                $( this ).removeClass( 'visible' );  
                $( subQContainer ).children( '#' +  
                    refSubquestion )  
                    .slideDown( 400 )  
                    .addClass( 'visible' );  
            } );  
    } else if ( $( subSectionVisible ).length < 1 &&  
        $( subSectionVisible ).first().attr( 'id' )  
        != refSubquestion ) {  
        $( subQContainer ).children( '#' +  
            refSubquestion )
```

```
                .show()
                .addClass('visible');
            $(subQContainer).slideDown(400);
        }
    });
```

Listing 3.2: Codice per la gestione della comparsa/scomparsa e scelta delle sottodomande

Validazione campi input e relativi messaggi di errore: il codice 3.3 proposto utilizza la libreria `jquery.validate` per andare a validare i campi input presenti nel form. Sono stati definiti dei messaggi personalizzati di errore in base alle domande ed il submit del form nel caso in cui sia stata fornita almeno una risposta per ogni domanda (in quanto erano tutte obbligatorie).

```
jQuery.validator.defaults({
    debug: false,
    errorClass: "alert alert-danger col-12",
    errorElement: 'div',
    errorPlacement: function(error, element) {
        if($(element).hasClass('error-after')){
            error.appendTo(
                element.parents('.tm-q-choice-container')
                    .first()
            );
        } else {
            error.prependTo(
                element.parents('.tm-q-choice-container')
                    .first()
            );
        }
    },
    submitHandler: function(form) {
        form.submit();
    },
});
```

```
});

var rulesValidation = {};
var numDom = 17;
for(var i = 1; i == numDom; i++){
    rulesValidation['q' + i] = { required: true };
}
rulesValidation['acceptance'] = { required: true };
rulesValidation['submitBtn'] = {
    depends: function(element) {
        grecaptcha.ready(function() {
            grecaptcha.execute(
                '6LdHzwAVAAAAAJukug7-nsx7Zs6V8PIkojhQaewE',
                {
                    action: 'submit'
                }).then(function(token) {
                    return true;
                });
        });
        return false;
    }
};
$("#qform").validate({
    rules: rulesValidation,
    messages: {
        'acceptance': 'Questo campo \\'e obbligatorio per poter inviare le risposte del questionario',
        'q1[]': 'Scegliere ALMENO una delle seguenti opzioni',
        'q2[]': 'Scegliere ALMENO una delle seguenti opzioni',
        'qsub5[]': 'Scegliere ALMENO una delle seguenti opzioni',
        'qsub7[]': 'Scegliere ALMENO una delle seguenti opzioni',
        'qsub9[]': 'Scegliere ALMENO una delle seguenti
```

```
        opzioni',
        'qsub11[]': 'Scegliere ALMENO una delle seguenti
                    opzioni'
    }
});
jQuery.extend(jQuery.validator.messages, {
    required: "Scegliere una delle opzioni disponibili
              per questa domanda"
});
```

Listing 3.3: Codice per la validazione dei campi obbligatori

Ingrandimento immagini e layer colorato per rispota selezionata:

la prima riga di codice proposta utilizza la libreria *fancybox* per generare la gallery di ingrandimenti immagini. Ogni domanda (che abbia immagini nelle risposte) avrà una propria gallery scorrevole che consente così all'utente di visionare tutti gli ingrandimenti senza dover ogni volta chiudere e riaprire l'ingrandimento.

Tutto il codice 3.4 successivo è stato inserito per far sì che, quando un utente ha selezionato una risposta (ad una domanda che presenta delle immagini nelle risposte), l'immagine corrispondente venga *colorata* in modo che risulti evidente all'utente l'immagine scelta. L'effetto viene applicato aggiungendo un layer colorato a cui viene applicata un valore di 0.3 al canale alpha (via css, un colore semitrasparente) che rende comunque visibile l'immagine sottostante ma la renda ben evidenziata rispetto alle altre.

```
$( '[data-fancybox|"gallery"]' ).fancybox({ });

// Color image when its radio button is selected
$( ".tm-q-choice-container:not(.has-subquestion) .tm-radio
    -group" ).on(
    'click',
```

```
function () {  
    var parent = $(this).parent().parent();  
    $(parent).parent().find("div.selected-layer").  
        remove();  
    $(parent).parent().find("img.selected").  
        removeClass("selected");  
    var selectedImg = $(parent).find("img");  
    $(selectedImg) .addClass("selected")  
        .after('<div class="selected-  
            layer"></div>');  
}  
);
```

Listing 3.4: Codice per la gestione della gallery degli ingrandimenti e layer colorato sopra l'immagine corrispondente alla risposta scelta

Capitolo 4

Risultati del questionario

In questo capitolo andremo a visionare ed analizzare i dati raccolti tramite il questionario online.

4.1 I dati

Partiamo con qualche numero.

Ultimo aggiornamento dei dati: 11/09/2020.

Il numero di **compilazioni totali inviate** sono **109**.

Il numero di **compilazioni totali distinte** (per cookie), sono **99**.

Nella fase di progettazione dell'inserimento dei dati nel database è stato pensato di creare e salvare un cookie alla prima visita della pagina.

In questo modo, se un utente ha compilato ed inviato più volte lo stesso questionario, il cookie assegnato sarebbe lo stesso.

Ovviamente in questo caso risulterebbero come duplicate le compilazioni effettuate dallo stesso browser (non in anonimo), anche se create da due persone differenti.

La raccolta dati del questionario parte **dal 28/07/2020** fino **al 11/09/2020**.

Nell'immagine 4.1 è possibile vedere il range di date con l'andamento delle compilazioni nel corso del mese di Agosto e Settembre.

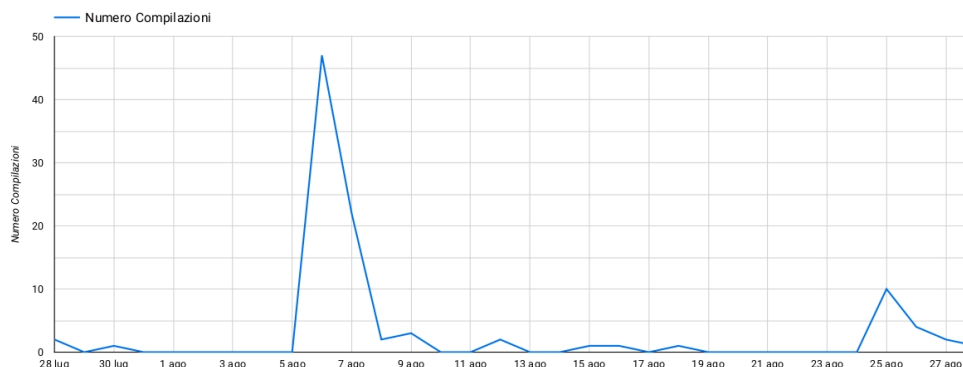
Distribuzione temporale delle compilazioni

Figura 4.1: Compilazioni questionario: distribuzione temporale compilazioni

Sono ben visibili 2 picchi nei giorni 06/08/2020 e 25/08/2020 che corrispondono rispettivamente alla diffusione su diversi canali del link relativo al questionario con richiesta di compilazione e condivisione del link stesso. La prima diffusione, quella che ha portato al maggior numero di compilazioni in un breve lasso di tempo, è avvenuta il 6 Agosto.

I canali sfruttati per la prima condivisione sono stati i social e di chat come Facebook, Telegram, Whatsapp. Subito dopo il 6 Agosto notiamo qualche compilazione sparsa. Potrebbero essere utenti che, ricevuto il link, non hanno avuto modo di compilarlo subito ed hanno rimandato la compilazione ad un secondo momento.

Altra possibilità è quella che il link sia stato condiviso e compilato successivamente alla prima diffusione.

Un altro picco delle compilazioni registrate, nel range di date, corrisponde al giorno 25 Agosto. In quella data è stato riproposto nuovamente il link sui canali precedentemente utilizzati, nel tentativo di raccogliere ulteriori compilazioni.

4.1.1 Dati demografici e personali

Per prima cosa analizziamo il pubblico attivo che ha portato a termine il questionario con qualche dato demografico e personale.

Le fasce d'età

Nella figura 4.2 possiamo vedere come i valori sulle fasce d'età siano abbastanza bilanciati:

- 18 – 25 anni: 11 compilazioni
- 26 – 30 anni: 30 compilazioni
- 36 – 45 anni: 20 compilazioni
- 46 – 55 anni: 18 compilazioni
- sopra i 55 anni: 20 compilazioni

L'aver un bilanciamento relativo alla fascia di età ci consente di valutare i risultati considerando che, nonostante le differenze di fascia d'età, se viene espressa una netta maggioranza su una certa tipologia di grafico o infografica, molto probabilmente è la più comprensibile su tutte le fasce. Se le compilazioni si fossero concentrate prevalentemente su un solo range di età, avremmo avuto la certezza che per quella fascia sarebbe stata ottimale una certa infografica piuttosto che un'altra ma non avremmo avuto informazioni rilevanti per considerare le altre fasce.

Livello di istruzione

Nella figura 4.3 (domanda 14) abbiamo una panoramica dei diversi livelli di istruzione degli utenti che hanno partecipato al questionario.

- Inferiore o equivalente al diploma: 52 compilazioni
- Laurea Triennale: 18 compilazioni
- Laurea Magistrale: 17 compilazioni
- Dottorato: 4 compilazioni

Domanda 13 - Fascia d'età

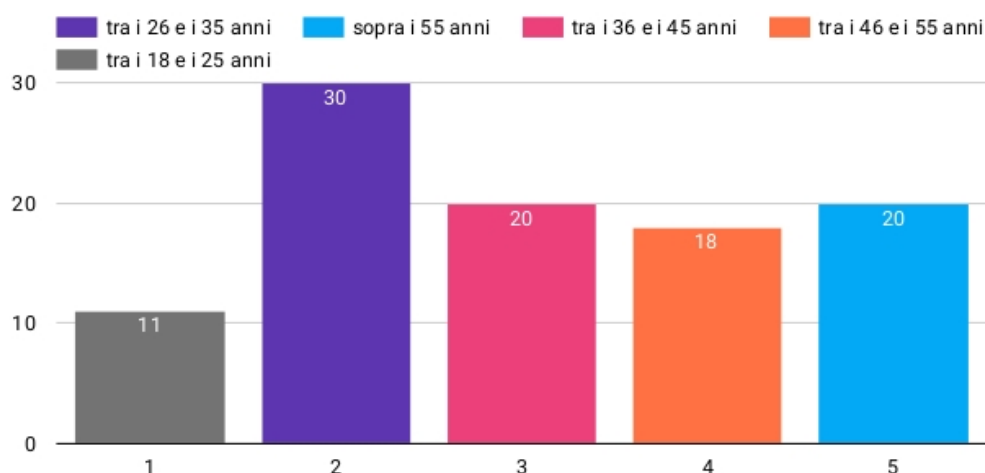


Figura 4.2: Compilazioni questionario: fasce d'età – Domanda 13

- Laurea a ciclo unico: 8 compilazioni

In questo caso, possiamo notare come ci sia una certa disuguaglianza nei risultati registrati. Abbiamo infatti una prevalenza di compilazioni di utenti che al massimo hanno il diploma di scuola superiore (52%), seguiti da quelli che hanno conseguito una laurea triennale o magistrale (rispettivamente 18% e 17%) ed in ultimo laurea a ciclo unico (8%) e dottorato (4%).

Il genere

Nella figura 4.4 (domanda 15) possiamo vedere quanti uomini o donne abbiano compilato il questionario. Era presente una terza opzione in cui l'utente poteva scegliere di non indicare il proprio sesso, ma non ci sono state risposte in riferimento a quell'opzione.

- Preferisco non rispondere: 0 compilazioni
- Uomo: 45 compilazioni

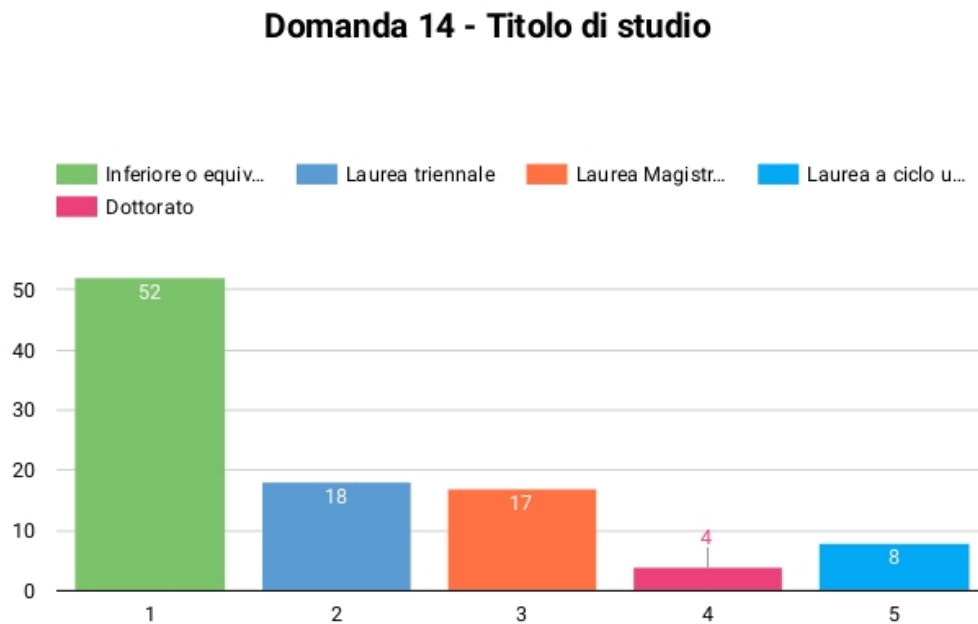


Figura 4.3: Compilazioni questionario: livello d'istruzione – Domanda 14

- Donna: 54 compilazioni

Possiamo notare che la percentuale dei 2 generi è abbastanza equilibrata, non c'è una prevalenza importante dell'una sull'altra. Anche in questo caso, come nelle fasce d'età, un equilibrio tra le risposte ricevute è importante per poter arrivare a dei risultati generici che ci possano portare ad identificare un eventuale standard per la creazione e generazione di infografiche relative alla diffusione di malattie.

Dispositivi utilizzati per la compilazione

Nella figura 4.5 (domanda 16) abbiamo una panoramica sul dispositivo utilizzato per la compilazione del questionario.

- Smartphone/Telefono cellulare: 90 compilazioni
- Computer: 8 compilazioni
- Tablet: 1 compilazione

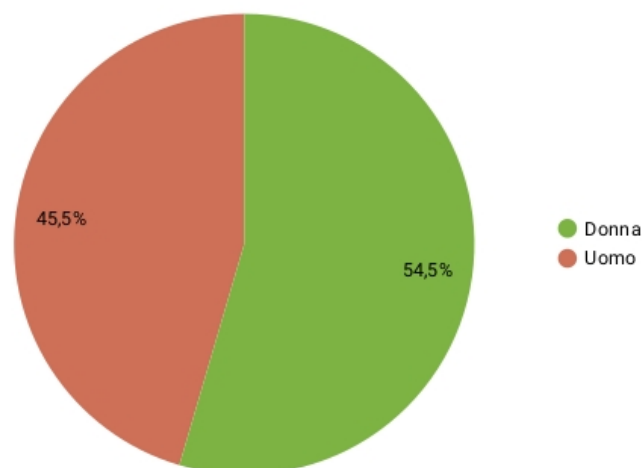
Domanda 15 - Genere

Figura 4.4: Percentuali compilazioni per genere – Domanda 15

Anche in questo caso, notiamo una notevole predominanza dell'utilizzo di un telefono cellulare/smartphone (90%) per la compilazione del questionario, con un 8% di compilazioni effettuate da computer ed 1% da tablet. Dati i canali utilizzati per la diffusione del link del questionario (prevalentemente Whatsapp con un paio di condivisioni su gruppi Facebook) e considerata la tendenza attuale che vuole un maggior utilizzo e sfruttamento dello smartphone per la navigazione web, la netta predominanza dell'utilizzo dello smartphone non è una sorpresa.

Livello di comprensione dell'inglese scritto

Nella figura 4.6 (domanda 17) abbiamo una indicazione di quale sia il livello di comprensione dell'inglese scritto. Questa domanda è stata inserita nel questionario in quanto alcune delle risorse presentate agli utenti facevano riferimento a siti esteri in lingua inglese.

- Buona/Ottima: 32 compilazioni
- Discreta: 47 compilazioni

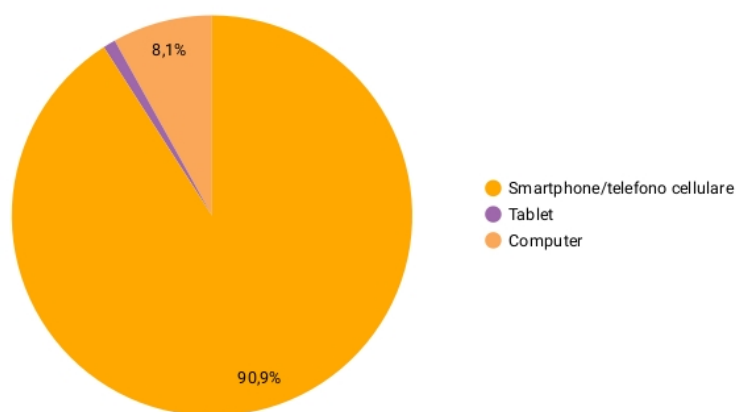
Domanda 16 - Dispositivo utilizzato per il questionario

Figura 4.5: Percentuali compilazioni per dispositivo utilizzato (al momento dell'invio del questionario) – Domanda 16

- Scarsa/Nulla: 20 compilazioni

4.1.2 Fonti di informazioni utilizzate

In questa sezione vedremo quali sono le fonti di informazione che gli utenti hanno utilizzato, prima di compilare il questionario, per ricercare dati relativi alla pandemia. Nella prima domanda (figura 4.7) veniva richiesto quale fonte è stata utilizzata per avere informazioni online sulla diffusione del COVID-19.

- Ministero della Salute/ Protezione Civile: 65 compilazioni
- Siti di quotidiani nazionali: 62 compilazioni
- Siti di quotidiani internazionali: 15 compilazioni
- Altro: 8 compilazioni
- ECDC - Centro Europeo per la prevenzione delle malattie: 4 compilazioni

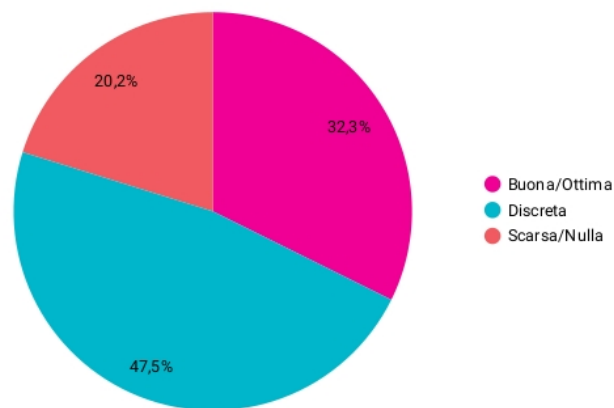
Domanda 17 - Conoscenza lingua scritta inglese

Figura 4.6: Percentuali compilazioni per livello di comprensione della lingua scritta inglese – Domanda 17

- John Hopkins University: 3 compilazioni
- Our World in Data: 3 compilazioni
- OMS - Organizzazione Mondiale della Sanità: 0 compilazioni
- Unità Sanitaria Locale (AUSL): 0 compilazioni

Possiamo vedere come ci siano solo 2 differenti fonti utilizzate dalla maggior parte degli individui: il sito del Ministero della Salute/Protezione Civile ed i siti dei quotidiani nazionali. Seguono poi i siti delle testate giornalistiche internazionali ed altre risorse consultate in minima parte (o per nulla nel caso dell'OMS e dell'AUSL).

Nella seconda domanda (figura 4.8) veniva richiesto, successivamente a quale fosse stata la fonte primaria di informazioni, quale fosse quella che era stata consultata per la visualizzazione dati sull'andamento della pandemia.

- Ministero della Salute/ Protezione Civile: 53 compilazioni

Domanda 1

Durante il periodo di pandemia Covid-19 dove hai consultato informazioni online sulla diffusione del virus?

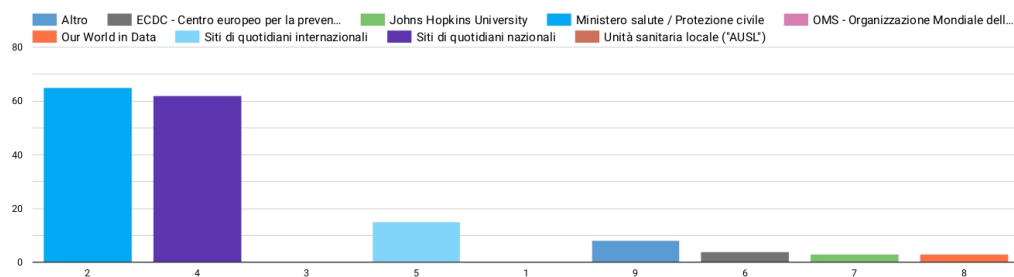


Figura 4.7: Distribuzione delle fonti consultate per la ricerca di informazioni sul COVID-19 – Domanda 1

- Siti di quotidiani nazionali: 51 compilazioni
- Siti di quotidiani internazionali: 16 compilazioni
- Altro: 5 compilazioni
- ECDC - Centro Europeo per la prevenzione delle malattie: 3 compilazioni
- John Hopkins University: 5 compilazioni
- Our World in Data: 3 compilazioni
- OMS - Organizzazione Mondiale della Sanità: 0 compilazioni
- Unità Sanitaria Locale (AUSL): 0 compilazioni

Anche in questo caso le fonti principalmente utilizzate sono 2: le stesse della precedente domanda ma in questo caso con una prevalenza dei siti dei quotidiani nazionali rispetto a quelli del Ministero della Salute o Protezione Civile. Per tutte le altre fonti proposte, i risultati sono pressoché simili ai precedenti (ritroviamo solamente qualche scelta in più sulla John Hopkins University a discapito di altre fonti di visualizzazione dati generiche).

Domanda 2

Considerando in particolare le visualizzazioni grafiche dei dati, quali risorse/siti/pagine web hai consultato?

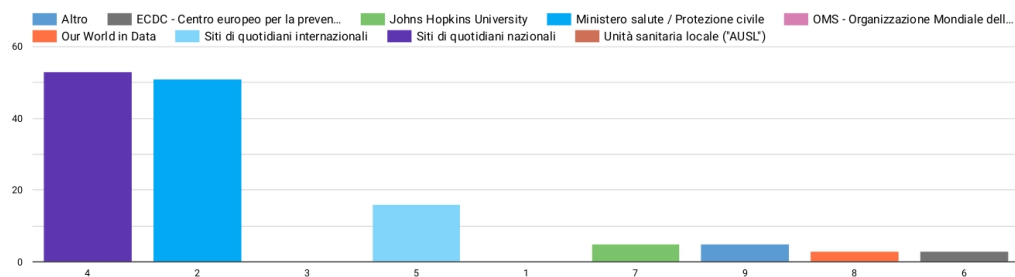


Figura 4.8: Distribuzione delle fonti consultate per la ricerca di visualizzazioni dati sul COVID-19 – Domanda 2

Nella terza domanda (figura 4.9) abbiamo chiesto agli utenti che hanno compilato il questionario quale sia stato il dispositivo maggiormente utilizzato per la consultazione delle risorse online relative al COVID-19.

- Smartphone/Telefono cellulare: 77 compilazioni
- Computer: 18 compilazioni
- Tablet: 4 compilazioni

Come forse ci si aspetterebbe ormai oggi, la maggior parte degli utenti ha utilizzato dispositivi mobili per la consultazione online delle risorse, con netta prevalenza di smartphone e telefoni cellulari (solo 4 hanno utilizzato il tablet).

Il risultato è molto simile a quello della domanda 16 (4.5), dove veniva richiesto con quale dispositivo avevano compilato il questionario. In questo caso però la percentuale delle consultazioni tramite computer è più elevata.

4.1.3 Comprensione delle infografiche proposte

In questa sezione analizzeremo i dati delle risposte registrate sulle domande in cui veniva proposto di scegliere una delle immagini proposte come più comprensibile.

Domanda 3

Solitamente, che dispositivo utilizzavi per consultare tali informazioni?

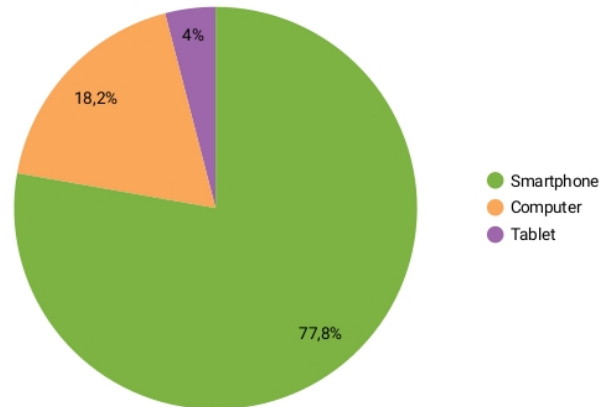


Figura 4.9: Percentuali di utilizzo dispositivi per la ricerca e consultazioni di informazioni sul COVID-19 – Domanda 3

Nella quarta domanda (figura 4.10) veniva richiesto di individuare il continente con il maggior numero di casi.

- Cartogramma: 29 compilazioni
- Istogramma impilato: 0 compilazioni
- Grafico a barre (orizzontale): 70 compilazioni

Possiamo notare come l'istogramma impilato non sia mai stato scelto come risposta, mentre abbiamo una notevole differenza tra il grafico a barre orizzontale ed il cartogramma.

Il grafico orizzontale, visivamente, rende immediatamente l'idea di quale sia il continente con più casi registrati, anche nel caso in cui i numeri tra due continenti possano essere molto simili.

Nel cartogramma abbiamo una diversa intensità del colore in base al numero dei casi. Questa intensità (assegnata ad ogni paese) rende probabilmente più complesso identificare quale sia il continente con più casi.

Domanda 4

Se volessi sapere quale sia il continente che ha registrato il maggior numero di casi, quale tra le seguenti immagini ti consente di reperire più facilmente l'informazione?

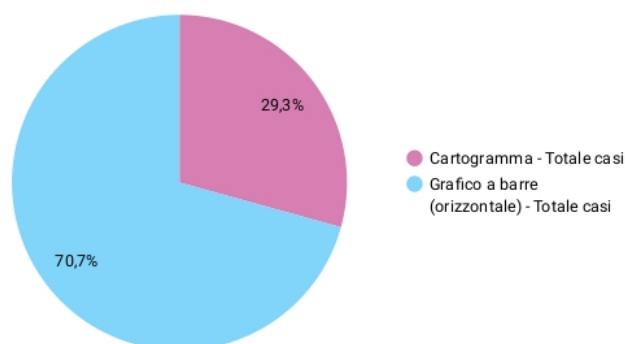


Figura 4.10: Percentuali di comprensione di ricerca del numero di casi totali – Domanda 4

Nella sesta domanda (figura 4.11) veniva richiesto di individuare nelle immagini proposte il paese che ha avuto il maggior numero di decessi.

- Grafico a linee: 2 compilazioni
- Cartogramma: 44 compilazioni
- Tabella dati: 53 compilazioni

In questo caso i risultati ci mostrano che gli utenti hanno preferito un dato tabellare (nell'esempio i paesi erano già ordinati per numero decrescente di decessi). La seconda scelta è ricaduta sul cartogramma che comunque ha ottenuto un bel punteggio. Il grafico a linee è quello che è stato scelto solamente 2 volte, quindi per gli utenti è risultato il più inefficace nella ricerca dell'informazione richiesta.

Nella domanda 8 (figura 4.12), veniva richiesto di individuare la regione italiana in cui era stato distribuito il maggior numero di materiali per l'emergenza.

Domanda 6

Se volessi sapere quale sia il paese con il maggior numero di decessi a livello mondiale, quale tra queste immagini ti rende più facile l'acquisizione dell'informazione cercata?

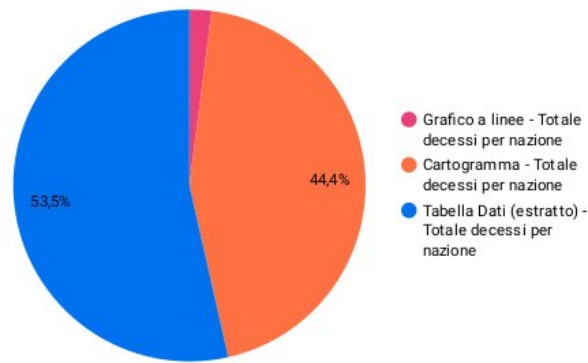


Figura 4.11: Percentuali di comprensione di ricerca del paese con il più alto numero di morti – Domanda 6

- Cartogramma: 20 compilazioni
- Grafico a barre (orizzontale): 79 compilazioni

In questo caso il grafico a barre orizzontale ha avuto la meglio sul cartogramma. Probabilmente perché un tipo di grafico a barre definisce, anche nel caso in cui la differenza non sia eccessiva, riesce a dare un valore “preciso”. Nel caso del cartogramma, l'intensità del colore, nel caso di differenze non eccessive, risulta essere molto simile, rendendo difficile l'effettiva ed immediata valutazione dei dati presentati nell'infografica.

Nella domanda 10 (figura 4.13), viene richiesto di individuare il continente con il maggior numero di test effettuati per il COVID-19.

- Cartogramma: 96 compilazioni
- Grafico a dispersione: 4 compilazioni

In questo caso possiamo vedere come sia stato scelto, in modo quasi assoluto, il cartogramma rispetto al grafico a dispersione.

Domanda 8

Vuoi sapere quale sia la regione italiana in cui è stato distribuito il maggior numero di materiali per l'emergenza COVID-19. Quale tra queste immagini ti rende più facile l'acquisizione dell'informazione richiesta?

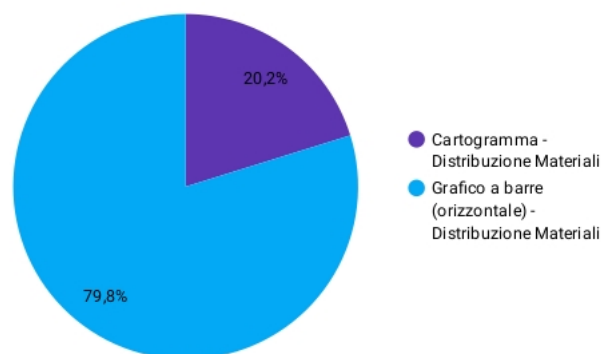


Figura 4.12: Percentuali di comprensione nella ricerca della regione italiana che ha ricevuto il maggior numero di materiali per l'emergenza – Domanda 8

Nella domanda 12 (figura 4.14), viene richiesto di individuare quale sia stato il paese in cui sono stati effettuati il maggior numero di test ogni 1000 persone.

- Grafico a linee: 1 compilazione
- Cartogramma: 17 compilazioni
- Grafico a barre (orizzontale): 81 compilazioni

In questo caso possiamo notare come il grafico a barre sia stata la scelta più gettonata per l'81%, seguita dal cartogramma.

4.1.4 Valutazione sull'utilizzo di piattaforme esterne

In questa sezione andremo ad analizzare i risultati relativi alle domande in cui era richiesto all'utente di cliccare sui link proposti e recuperare l'informazione presente nella domanda. Gli utenti poi dovevano scegliere di

Domanda 10

Vuoi sapere quale sia il continente dove risulta essere maggiore il numero di test effettuati per il COVID-19 risultati positivi.
Quale tra queste immagini ti rende più facile l'acquisizione dell'informazione richiesta?

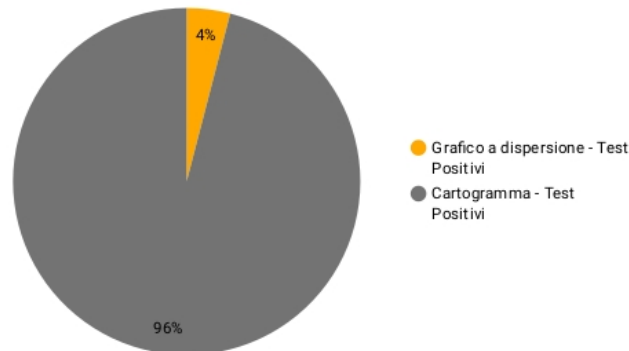


Figura 4.13: Percentuali di comprensione nella ricerca del continente con il maggior numero di test effettuati– Domanda 10

indicare un livello di facilità/difficoltà con cui erano riusciti a recuperare (o meno) l'informazione.

Una volta indicato il grado di difficoltà riscontrato, agli utenti veniva proposto un altro set di risposte (basato sulla risposta precedente) dove veniva richiesto di indicare quali erano gli elementi che avevano facilitato o complicato la ricerca dell'informazione.

Nella quinta domanda (figura 4.15) l'utente doveva navigare sul link che riporta alla dashboard delle infografiche della Protezione Civile.

L'utente doveva recuperare da questa pagina la provincia italiana con il maggior numero casi registrati.

Alla domanda principale hanno risposto come segue:

- L'ho trovata molto facilmente: 15 compilazioni
- L'ho trovata facilmente: 51 compilazioni
- Mi ci è voluto qualche minuto per capire dove poterla reperire: 22 compilazioni

Domanda 12

Quale delle seguenti visualizzazioni ti rende più facile individuare la nazione dove sono stati effettuati il maggior numero di test ogni 1000 persone?

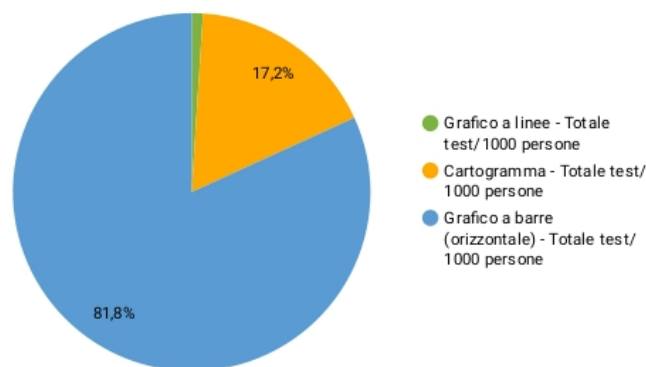


Figura 4.14: Percentuali di comprensione nella ricerca del paese con il maggior numero di test effettuati ogni 1000 persone – Domanda 12

- L’ho trovata con molta difficoltà: 5 compilazioni
- Non l’ho trovata: 6 compilazioni

Da queste prime risposte possiamo vedere come gli utenti abbiano risposto positivamente per un totale del 73%. Per il 15% la ricerca ha richiesto del tempo, mentre per l’11% la risposta è stata negativa (addirittura un 5% non è riuscito a recuperare l’informazione richiesta).

Vediamo qui le risposte relative agli elementi che hanno contribuito o hanno causato problemi nella ricerca dell’informazione.

Elementi positivi

- Informazioni ben presentate, intuitive (facile reperibilità): 32%
- Informazioni chiare (facile lettura): 55%
- I colori usati sono piacevoli (“attrattività/eleganza”): 10%

- Altro: 2%

I dati degli utenti, relativamente agli aspetti positivi, riportano in maggioranza (poco più della metà) che le informazioni erano presentate chiaramente, di facile lettura. Alcuni utenti hanno indicato che le informazioni erano ben presentate ed intuitive.

Elementi negativi

- Troppe informazioni insieme, difficile reperibilità: 61%
- Colori che confondono, “poco attraenti”: 4,9%
- Testi piccoli o non leggibili (“attrattività/eleganza”): 22%
- Altro: 7,3%

Tra gli utenti che hanno risposto negativamente alla quinta domanda, spicca in prevalenza un difetto nell’aver accorpato troppe informazioni e che queste fossero di difficile reperibilità. Inoltre è stato indicato anche che i testi erano piccoli e non leggibili.

Nella domanda 7 (figura 4.16), agli utenti era richiesto di consultare il sito dell’Organizzazione Mondiale della Sanità (WHO), in inglese, per trovare il numero di decessi e casi confermati totali per l’Italia.

- L’ho trovata molto facilmente: 25 compilazioni
- L’ho trovata facilmente: 46 compilazioni
- Mi ci è voluto qualche minuto per capire dove poterla reperire: 15 compilazioni
- L’ho trovata con molta difficoltà: 3 compilazioni
- Non l’ho trovata: 10 compilazioni

Dai risultati possiamo vedere come le risposte positive fornite ammontano quasi ad un 72%.

Domanda 5 - Protezione Civile

Trovare quale sia la provincia italiana con il maggior numero di casi registrati

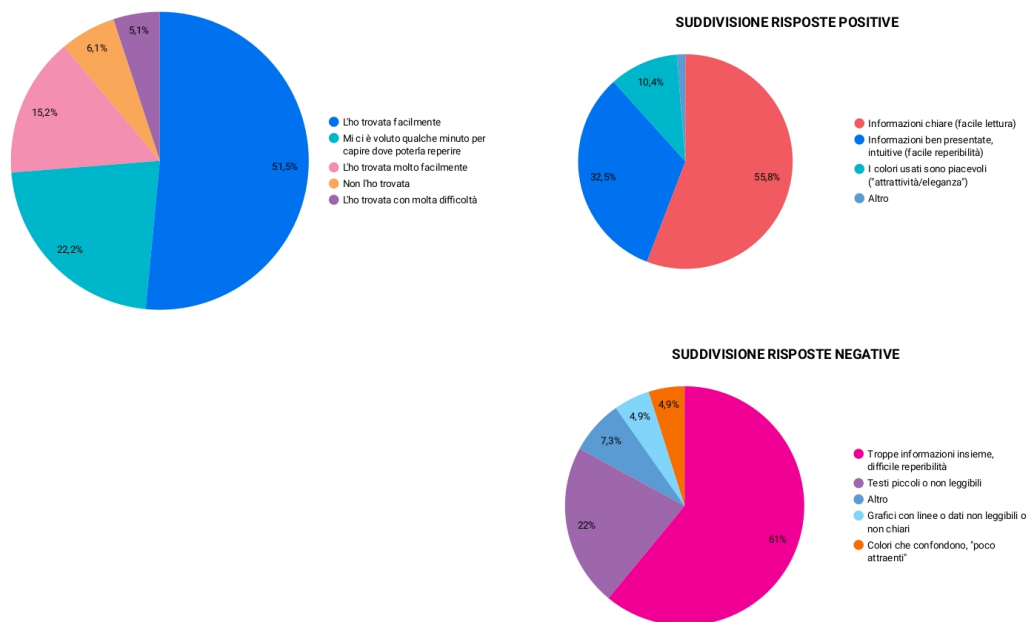


Figura 4.15: Percentuali di comprensione nella ricerca della provincia italiana con il maggior numero di casi registrati – Domanda 5

Elementi positivi

- Informazioni ben presentate, intuitive (facile reperibilità): 39,5%
- Informazioni chiare (facile lettura): 50%
- I colori usati sono piacevoli ("attrattività/eleganza"): 5,8%
- Altro: 4,7%

La maggioranza delle compilazioni positive riporta una presentazione delle informazioni soprattutto chiare, ma anche ben presentate ed intuitive.

Elementi negativi

- Troppe informazioni insieme, difficile reperibilità: 50%

- Colori che confondono, “poco attraenti”: 14,7%
- Testi piccoli o non leggibili (“attrattività/eleganza”): 2,9%
- Grafici con linee o dati non leggibili o non chiari: 8,8%
- Lingua Inglese: 8,8%
- Altro: 14,7%

Domanda 7 - World Health Organization

Trovare il numero totale di casi e di decessi attuali per l'Italia

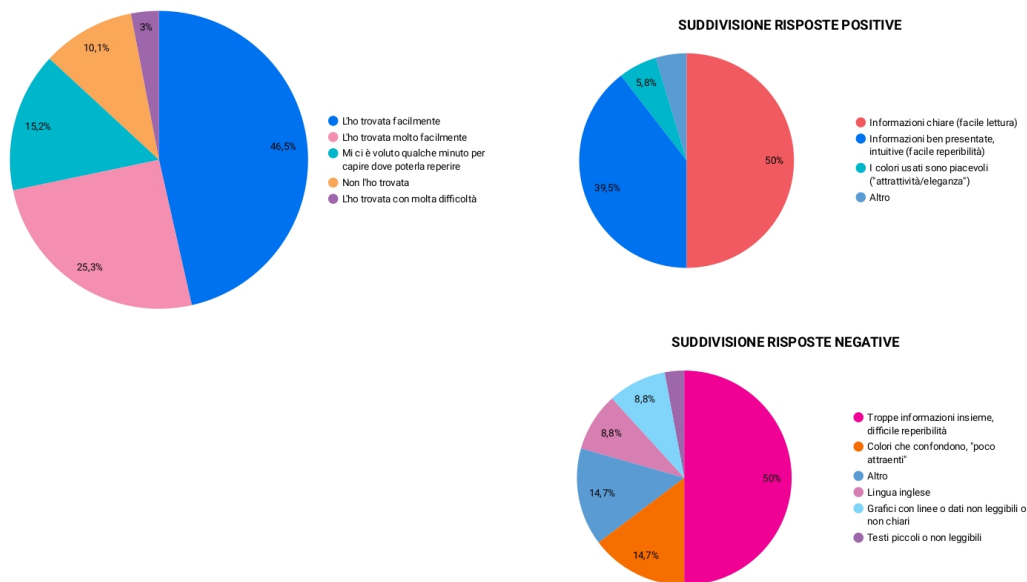


Figura 4.16: Percentuali di comprensione nella ricerca del numero totale di casi totali e decessi in Italia – Domanda 7

Nella domanda 9 (figura 4.17) è stato richiesto agli utenti di trovare sul sito dell'Istituto Superiore della Sanità, la fascia d'età per cui si sono avuti il maggior numero di decessi.

- L'ho trovata molto facilmente: 31 compilazioni
- L'ho trovata facilmente: 49 compilazioni

- Mi ci è voluto qualche minuto per capire dove poterla reperire: 13 compilazioni
- L'ho trovata con molta difficoltà: 6 compilazioni
- Non l'ho trovata: 0 compilazioni

In questo caso possiamo notare come, a differenza delle due domande precedenti, il numero di utenti che non sono riusciti a trovare l'informazione è pari a 0. Il conteggio totale delle risposte che possiamo considerare positive ammonta ad un 80% (più alto rispetto alle precedenti). Leggermente più basso (rispetto alle precedenti) il numero degli individui che ci hanno messo un pò di tempo a trovare l'informazione (13%).

Elementi positivi

- Informazioni ben presentate, intuitive (facile reperibilità): 40,5%
- Informazioni chiare (facile lettura): 57%
- I colori usati sono piacevoli (“attrattività/eleganza”): 1,2%
- Altro: 1,2%

In questo caso la quasi totalità di chi ha risposto positivamente alla domanda principale ha indicato che le informazioni erano molto chiare ma anche ben presentate ed intuitive.

Elementi negativi

- Troppe informazioni insieme, difficile reperibilità: 77,3%
- Colori che confondono, “poco attraenti”: 9,1%
- Testi piccoli o non leggibili (“attrattività/eleganza”): 13,6%
- Grafici con linee o dati non leggibili o non chiari: 0%

- Altro: 0%

Nel caso di chi ha risposto negativamente, al difficoltà maggiore è stata quella dell'aver troppe informazioni insieme, il che rendeva la ricerca specifica più complicata.

Domanda 9 - Istituto Superiore della Sanità

Trovare quale sia la fascia di età in cui si è avuto il maggior numero di decessi

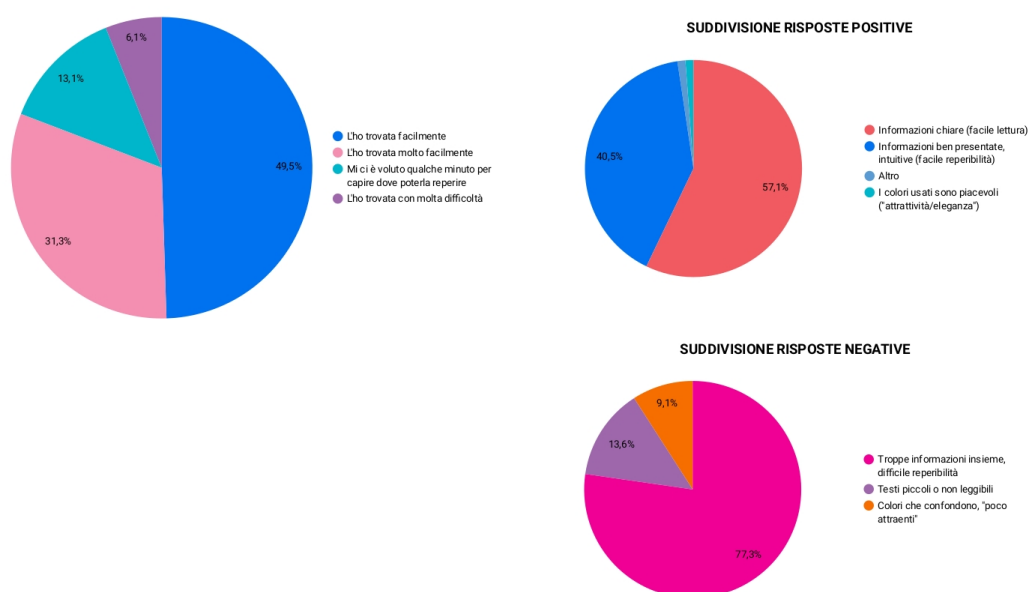


Figura 4.17: Percentuali di comprensione nella ricerca della fascia d'età con il maggior numero di decessi – Domanda 9

Per la domanda 11 (figura 4.18) è stato richiesto agli utenti di visitare il sito, gestito da Apple, relativo alle informazioni sulla mobilità durante il periodo della quarantena. Gli utenti dovevano trovare la tipologia di mobilità con la percentuale più alta di modifica delle richieste di instradamento nel proprio stato.

- L'ho trovata molto facilmente: 46 compilazioni
- L'ho trovata facilmente: 20 compilazioni

- Mi ci è voluto qualche minuto per capire dove poterla reperire: 18 compilazioni
- L'ho trovata con molta difficoltà: 7 compilazioni
- Non l'ho trovata: 8 compilazioni

Per le risposte positive, in questo caso raggiungiamo una quota del 66,7%, più bassa rispetto alle precedenti domande della stessa tipologia. Abbiamo un 18,2% di utenti che hanno impiegato un pò di tempo per trovare l'informazione cercata e quasi la stessa percentuale (7.1% e 8.1%) rispettivamente per chi ha trovato molta difficoltà e chi proprio non è riuscito a trovare nemmeno l'informazione richiesta.

Elementi positivi

- Informazioni ben presentate, intuitive (facile reperibilità): 35,1%
- Informazioni chiare (facile lettura): 51,4%
- I colori usati sono piacevoli (“attrattività/eleganza”): 12,2%
- Altro: 1,3%

Per chi ha risposto positivamente registriamo l'indicazione di informazioni chiare (metà circa degli utenti), mentre poco più di un terzo hanno indicato che le informazioni erano ben presentate ed intuitive.

Elementi negativi

- Troppe informazioni insieme, difficile reperibilità: 41,2%
- Colori che confondono, “poco attraenti”: 20,6%
- Testi piccoli o non leggibili (“attrattività/eleganza”): 11,8%
- Grafici con linee o dati non leggibili o non chiari: 14,7%

- Lingua inglese: 11,8%
- Altro: 0%

Nel caso degli utenti che hanno fornito una risposta negativa nella domanda principale, tra i motivi della loro precedente risposta rientra per la maggior parte il problema dell'aver troppe informazioni insieme, di difficile reperibilità. un quinto degli utenti ha riportato che i colori creavano confusione ed erano poco attraenti.

Domanda 11 - Apple

Individuare quale sia, ad oggi, la tipologia di mobilità con la percentuale più alta di modifica delle richieste di instradamento, nel tuo stato

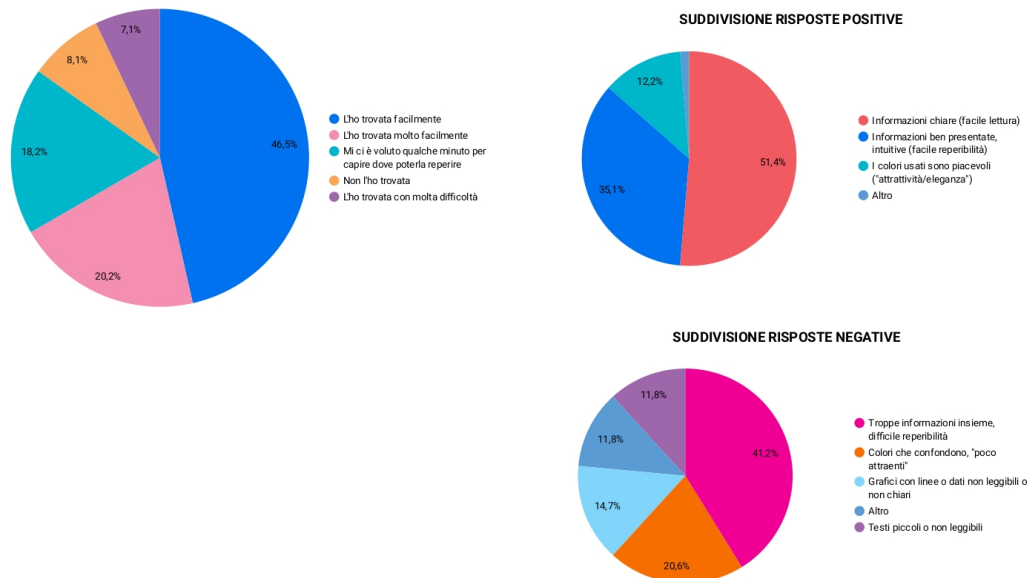


Figura 4.18: Percentuali di comprensione nella ricerca della tipologia di mobilità con la percentuale più alta di modifica delle richieste di instradamento – Domanda 11

Conclusioni

Questo progetto nasce con lo scopo di voler fornire qualche indicazione su quali possano essere i grafici e le presentazioni più efficaci che possono raggiungere gli individui comuni, che non sono avvezzi all'utilizzo di strutture complesse per l'analisi e valutazione dei dati.

Dal momento che è un aspetto soggettivo, l'unico strumento di cui possiamo avvalerci per avere un'indicazione media di quali siano le infografiche più adatte e comprensibili è il questionario.

Il questionario è uno strumento utile per sondare l'opinione di tanti e fare una statistica specifica rispetto ai dati che si stanno cercando di raccogliere.

Nel sondaggio di questo progetto abbiamo inserito domande personali (per poter valutare i risultati delle risposte in base a chi lo ha compilato); domande con proposte di immagini dove l'utente doveva scegliere quella più chiara e comprensibile; domande in cui era richiesto di visitare un link esterno e riportare poi nel questionario le indicazioni relative alla facilità/difficoltà riscontrate nell'esecuzione di un'operazione.

Il numero di compilazioni (circa un centinaio) ci fornisce una panoramica di quello che è il livello di comprensione di alcune delle infografiche proposte e quali siano quelle che meglio hanno soddisfatto l'utente nella ricerca dell'informazione richiesta nella domanda. Il questionario è stato realizzato ad hoc (ad esclusione del modello della pagina) con domande specifiche e salvataggio delle compilazioni su database.

I risultati del questionario mostrano che, per le immagini delle infografiche

proposte, la preferenza in media sia ricaduta principalmente sui cartogrammi o sui grafici a barre (orizzontali).

Nelle domande in cui era richiesta una scelta tra i grafici proposti emerge che: in 3 domande su 5 (figure 4.10, 4.12, 4.14), dove erano presenti sia il cartogramma che il grafico a barre orizzontali, il secondo ha ricevuto sempre un numero di votazioni sempre superiore al 70%, comunicando quindi una netta preferenza per gli istogrammi.

Nelle altre due domande della stessa tipologia, il cartogramma viene messo a confronto con un estratto di una tabella ed un grafico a linee (figura 4.11) in un caso e nell'altro con un grafico a dispersione (figura 4.13). In questi due confronti emerge come il grafico a linee e, soprattutto, il grafico a dispersione non siano efficaci nella comunicazione in quanto risultano scelti solo per un 1%-2%. Risulta invece quasi equa la scelta dell'estratto tabulare in confronto al cartogramma.

Per quanto riguarda invece le domande in cui era richiesta un'azione: il 10% - 15% ha riscontrato molta difficoltà (o non è riuscito a recuperare l'informazione richiesta), in media un 15% ci ha messo qualche minuto per trovarla, mentre in media un 70% - 75% è riuscito a recuperare l'informazione richiesta facilmente. Nelle quattro piattaforme su cui è stato richiesto un riscontro da parte degli utenti è stata quella dell'Istituto Superiore della Sanità che ha registrato una percentuale più bassa di risposte che indicavano un qualsiasi livello di difficoltà, dall'averci messo del tempo al non aver trovato l'informazione ricercata.

Tra le risposte positive, la maggioranza ha sempre riferito di aver trovato chiare le informazioni presentate (sempre sopra il 50%), mentre tra le risposte negative la maggioranza ha rilevato una difficoltà causata dalla presentazione di troppe informazioni che rendevano difficile reperirne una specifica (tra il 41% e il 77%).

Il questionario presentato in questo progetto è un primo punto di partenza per un nuovo studio che può andare più nel dettaglio e valutare tutte le combinazioni di dati realizzabili con i dati già raccolti per poi creare un

questionario più specifico ed indagare ancora più a fondo la tematica: l'utilizzo della visualizzazione dati nella comunicazione di informazioni relative ad epidemie e malattie.

I passaggi successivi possono sicuramente prevedere un'analisi approfondita dei dati per procedere con l'aggiornamento del questionario: si potrebbero strutturare, nel caso dei grafici proposti all'utente, delle domande dove vengono messe a confronto le tipologie di grafici indicati come più chiari e comprensibili, in modo da avere un dato più specifico relativo alle differenze tra cartogrammi, istogrammi e tabelle e verificare se effettivamente l'andamento dei nuovi dati rispecchia questi primi risultati.

Nel caso delle domande ad azione, si potrebbe anche pensare di ricreare delle *dashboard* simili a quelle già presentate nel sondaggio di questo progetto ed apportare delle modifiche in base ai risultati del questionario, chiedendo un nuovo parere agli utenti nel confronto tra quella precedentemente proposta e quella nuova modificata, per provare ad identificare eventuali elementi problematici.

Possiamo anche considerare di analizzare i dati raccolti verificando le risposte in base alle informazioni personali: verificare se ad esempio per un certo titolo di studio o fascia d'età ci sia stata una predominanza nella scelta di grafici rispetto ad altri e quale sia stato il livello di comprensione ed utilizzo delle piattaforme esterne.

In questo modo si potrebbe anche aggiornare il questionario introducendo delle domande alternative in base alle risposte personali fornite per approfondire e capire se effettivamente ci sono fattori demografici e/o culturali che influiscono maggiormente sul livello di lettura ed analisi nelle visualizzazioni di dati.

Dal momento che le informazioni raccolte non mostrano una direzione predominante, sicuramente una valutazione di questo tipo potrebbe fornire informazioni maggiori nel capire se questi possano essere fattori determinanti.

In questo modo si riusciranno ad avere dei risultati che potranno essere utilizzati per creare una sorta di guida relativa alle infografiche più utili

e comprensibili da poter utilizzare per la presentazione di dati al grande pubblico.

Bibliografia

- [1] Open Knowledge Foundation. Open Data Handbook - What is Open Data, 2020. URL <https://opendatahandbook.org/guide/en/what-is-open-data/>.
- [2] Open Definition. Open Definition, 2020. URL <https://opendefinition.org/okd/>.
- [3] Kelly Easterday, Tim Paulson, Proxima DasMohapatra, Peter Alagona, Shane Feirer, and Maggi Kelly. From the field to the cloud: A review of three approaches to sharing historical data from field stations using principles from data science. *Frontiers in Environmental Science*, 6, 10 2018. doi: 10.3389/fenvs.2018.00088. URL https://www.researchgate.net/figure/Conceptual-diagram-of-the-FAIR-roadmap-for-dark-data-each-curve-represents-a-step_fig1_328004345.
- [4] Open Knowledge Foundation. Why Open Data?, 2020. URL <https://opendatahandbook.org/guide/en/why-open-data/>.
- [5] Peter Tattersall. Tax Tree, 2017. URL https://hri.fi/data/en_GB/showcase/tax-tree.
- [6] Open Knowledge Foundation. Where does my money go, 2020. URL <https://app.wheredoesmymoneygo.org>.

-
- [7] David Eaves. Case study: How open data saved canada \$3.2 billion, 2010. URL <https://eaves.ca/2010/04/14/case-study-open-data-and-the-public-purse/>.
- [8] Folketsting. Folketsting, 2020. URL <https://www.folketsting.dk/>.
- [9] Fondazione Openpolis. Open Parlamento, 2020. URL <https://parlamento18.openpolis.it/>.
- [10] Tine Müller. Fint Toilet App, 2020. URL <http://www.findtoilet.dk/>.
- [11] The City of New York. NYC Dog Runs, 2020. URL <https://data.cityofnewyork.us/Recreation/NYC-Parks-Dog-Runs/8nac-uner>.
- [12] My Society. Mapumental, 2012. URL <https://www.mysociety.org/2012/11/10/mapumental-property-extra-insight-for-househunters/>.
- [13] Stefan Wehrmeyer. Mapnificent, 2020. URL <https://www.mapnificent.net/>.
- [14] Portale europeo dei dati. Impatto dei dati aperti, 2020. URL <https://www.europeandataportal.eu/it/impact-studies/open-data-impact>.
- [15] Modstrøm. Modstroem, 2020. URL <https://www.modstroem.dk/energiberegner/>.
- [16] Portale Europeo dei dati. Cosa sono i dati aperti? — Portale Europeo dei Dati, 2020. URL <https://www.europeandataportal.eu/it/training/what-open-data>.
- [17] Agenzia per l'Italia Digitale. Fare Open Data, 2020. URL <https://www.dati.gov.it/fare-opendata>.

- [18] Simon Chignard. A brief history of Open Data, 2013. URL www.paristechreview.com/2013/03/29/brief-history-open-data/.
- [19] European Data Portal. Choosing the right format for open data, 2020. URL <https://www.europeandataportal.eu/elearning/it/module9/>.
- [20] Istat. Visitatori musei pubblici e similari, per titolo d'accesso , 2013. URL http://www.datiopen.it/opendata/Visitatori_musei_publici_e_similari_titolo_d_accesso.
- [21] Istituto Nazionale di Geofisica e Vulcanologia. Istituto Nazionale di Geofisica e Vulcanologia, 2020. URL http://cnt.rm.ingv.it/webservices_and_software/.
- [22] International Federation of Digital Seismograph Networks. International Federation of Digital Seismograph Networks, 2020. URL <https://www.fdsn.org/webservices/>.
- [23] TPER. Open Data, 2020. URL <https://solweb.tper.it/web/tools/open-data/open-data.aspx>.
- [24] TPER. Chiama Treno, 2020. URL <http://www.tper.it/chiamata-treno>.
- [25] TPER. Hello Bus, 2020. URL <http://www.tper.it/hello-bus>.
- [26] TPER. Formato xml scaricabile, 2020. URL <https://solweb.tper.it/web/tools/open-data/open-data-download.aspx?source=solweb.tper.it&filename=rivendite&version=20200801&format=xml>.
- [27] Comune di Bologna. Colonnine di ricarica per veicoli elettrici, 2019. URL <http://catalogodati.comune.bologna.it/dataset/colonnine-di-ricarica-per-veicoli-elettrici>.

- [28] Wikipedia. Data Visualization, 2020. URL https://en.wikipedia.org/wiki/Data_visualization.
- [29] Wikipedia. Data Visualization - Characteristics of effective graphical displays, 2020. URL https://en.wikipedia.org/wiki/Data_visualization#Characteristics_of_effective_graphical_displays.
- [30] Edward R. Tufte. *The Visual Display of Quantitative Information (2nd Edition)*. 2009. ISBN 978-0961392147. URL <https://archive.org/details/visualdisplayofq00tufte>.
- [31] Robert Sunshine. Telling Visual Stories About Data, 2014. URL http://knesset.gov.il/mmm/oecd/Session_7_Robert_Sunshine.pdf.
- [32] Jonathan A. Schwabish. An economist's guide to visualizing data. In *Journal of Economic Perspective*, volume 28. 2014. URL <https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.1.209>.
- [33] Stephen Few. Data Visualization for Human Perception, 2020. URL <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception>.
- [34] SFU (Simon Fraiser University). Visualization - Lecture 11, 2016. URL https://web.archive.org/web/20160122203157/http://www.sfu.ca/gis/geog_x55/web355/icons/11_lec_vweb.pdf.
- [35] Wikipedia. Data Visualization - Human perception/cognition and data visualization, 2020. URL https://en.wikipedia.org/wiki/Data_visualization#Human_perception/cognition_and_data_visualization.
- [36] M. Friendly. A brief history of data visualization. In C. Chen, W. Härdle, and A Unwin, editors, *Handbook of Computational*

- Statistics: Data Visualization*, volume III. Springer-Verlag, Heidelberg, 2006. URL <http://www.datavis.ca/papers/hbook.pdf>. (In press).
- [37] Wikipedia. History of Data Visualization, 2020. URL https://en.wikipedia.org/wiki/Data_visualization#History_of_data_visualization.
- [38] Regione Emilia Romagna. S3 monitoraggio – smart specialisation strategy, 2020. URL <https://www.regione.emilia-romagna.it/s3-monitoraggio/>.
- [39] Comune di Bologna. Bilancio ambientale – laboratorio aria, 2020. URL <http://dati.comune.bologna.it/node/3852>.
- [40] Eni. Parità di genere, 2020. URL <https://www.eni.com/it-IT/trasformazione/parita-genere-coronavirus-analisi-rete.html>.
- [41] Open Data Charter. Open Covid-19 Data, 2020. URL <https://medium.com/opendatacharter/open-covid-19-data-461e1cbefbba>.
- [42] OCSE - Organisation for Economic Co-operation and Development. OCSE - Organisation for Economic Co-operation and Development, 2020. URL <http://www.oecd.org/>.
- [43] ODC - Open Data Charter. ODC - Open Data Charter, 2020. URL <https://opendatacharter.net/>.
- [44] Open Data Charter. Open Up Fields Guide, 2018. URL <https://drive.google.com/file/d/1itEjU0zSdn35K0o7VLoxrYzKHEwASYKV/view>.
- [45] Stanford University Center for Artificial Intelligence in Medicine & Imaging. COVID-19 + Imaging AI Resources, 2020. URL <https://aimi.stanford.edu/resources/covid19>.

- [46] Khari Johnson. Microsoft, White House, and Allen Institute release coronavirus data set for medical and NLP researchers, 2020. URL <https://venturebeat.com/2020/03/16/microsoft-white-house-and-allen-institute-release-coronavirus-data-set-for-medical-and-nlp-researchers/>.
- [47] The University of Manchester. More than 500 scientists in 18 countries to map COVID-19 effects and treatments in patients' blood, 2020. URL <https://www.manchester.ac.uk/discover/news/more-than-500-scientists-in-18-countries-to-map-covid-19-effects-and-treatments-in-patients-blood/>.
- [48] WHO World Health Organisation. WHO Coronavirus Disease (COVID-19) Dashboard, 2020. URL <https://covid19.who.int/>.
- [49] Protezione Civile. Protezione Civile – COVID-19 Situazione Italia, 2020. URL <https://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d4138b1>.
- [50] Epicentro – Istituto Superiore della Sanità. Epicentro – Istituto Superiore della Sanità, 2020. URL <https://www.epicentro.iss.it/coronavirus/sars-cov-2-dashboard>.
- [51] Our World in Data. Daily new confirmed COVID-19 cases (last 7 days), 2020. URL <https://ourworldindata.org/coronavirus-data-explorer?yScale=log&zoomToSelection=true&year=latest&time=2020-01-22..2020-09-09&country=USA~KOR~DEU~IND~BRA~ITA~IDN~ZAF~MEX~NZL~NOR~COG~BEL~HRV~FRA®ion=World&casesMetric=true&interval=smoothed&aligned=true&smoothing=7&pickerMetric=location&pickerSort=asc>.
- [52] Our World in Data. Coronavirus Pandemic (COVID-19), 2020. URL <https://ourworldindata.org/coronavirus>.

- [53] Our World in Data. Daily new confirmed COVID-19 deaths, 2020. URL <https://ourworldindata.org/coronavirus-data-explorer?yScale=log&zoomToSelection=true&year=latest&time=2020-01-25..2020-09-09&country=USA~GBR~BRA~IND~DEU~MEX~CHL~ZAF~COL~KOR~NOR~URY~ITA~FRA~BEL~ISL®ion=World&deathsMetric=true&interval=smoothed&aligned=true&smoothing=7&pickerMetric=location&pickerSort=asc>.
- [54] ISS - Istituto Superiore della Sanità. Report sulle caratteristiche dei pazienti deceduti positivi all'infezione da SARS-CoV-2 in Italia, 2020. URL <https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia>.
- [55] SurveyMonkey. Why? how survey research and methodology make a difference, 2020. URL <https://www.surveymonkey.com/mp/why-survey-understanding-survey-methodology/>.
- [56] Albert O. Hirschman. Exit, voice, loyalty revisited. 1970. URL https://en.wikipedia.org/wiki/Exit,_Voice,_and_Loyalty#References.
- [57] SmartSurvey. 6 steps to conducting an online survey, 2020. URL <https://www.smartsurvey.co.uk/articles/6-steps-to-conducting-an-online-survey>.
- [58] Template Monster. Templatemo - image survey, 2020. URL <https://templatemo.com/tm-523-image-survey>.
- [59] Fancyapps. Fancybox, 2020. URL <https://fancyapps.com/fancybox/3/>.
- [60] Stefania Riminucci. File di stile personalizzato, 2020. URL <https://www.stefaniariminucci.netsons.org/css/custom.css>.
- [61] Validate. Validate - jquery validation plugin, 2020. URL <https://jqueryvalidation.org/>.

- [62] Google Developers. Recaptcha – google, 2020. URL <https://developers.google.com/recaptcha/>.
- [63] Stefania Riminucci. Codice javascript personalizzato, 2020. URL <https://www.stefaniariminucci.netsons.org/>.
- [64] David Yakobovitch. Top 20 Visualization Dashboards for Mapping COVID-19, 2020. URL <https://www.analyticsvidhya.com/blog/2020/07/top-20-covid-19-dashboard/>.
- [65] Eni. Eni – Paura e speranza in Rete. Gli italiani nella Fase 1, 2020. URL <https://www.eni.com/it-IT/media/sfida-da-vincere-insieme/paura-e-speranza-in-rete-gli-italiani-nella-fase-1.html>.

Ringraziamenti

Ci tengo a ringraziare innanzi tutto la professoressa Silvia Mirri, la professoressa Catia Prandi e la dottoressa Chiara Cecchini per l'enorme supporto nella stesura di questo documento e per la creazione del progetto di tesi in generale, per la loro pazienza dimostrata in questi mesi.

Ringrazio Matteo Biagi per il supporto fornito negli ultimi anni nel completamento degli ultimi esami e della tesi.

Ringrazio Raffaele Mangiacasale, compagno di studi e lavoro per i primi anni, per il supporto agli studi e al superamento di molti esami.

Ringrazio tutti i miei amici, familiari e colleghi che mi sono sempre stati di supporto e mi hanno sopportato in questi anni di carriera universitaria che si è protratta più del dovuto.

Un ringraziamento generale va a tutte quelle persone che nel corso della storia dell'umanità si sono impegnate, a volte dando la loro stessa vita, per contribuire al progresso del settore scientifico, senza il quale oggi sicuramente non saremmo qui.