

# Sorgenti di dati del COVID-19 in Italia

Silvia Di Federico<sup>1</sup>, Tommaso Filippini<sup>1</sup>, Marco Vinceti<sup>1</sup>

## Introduzione e scopo

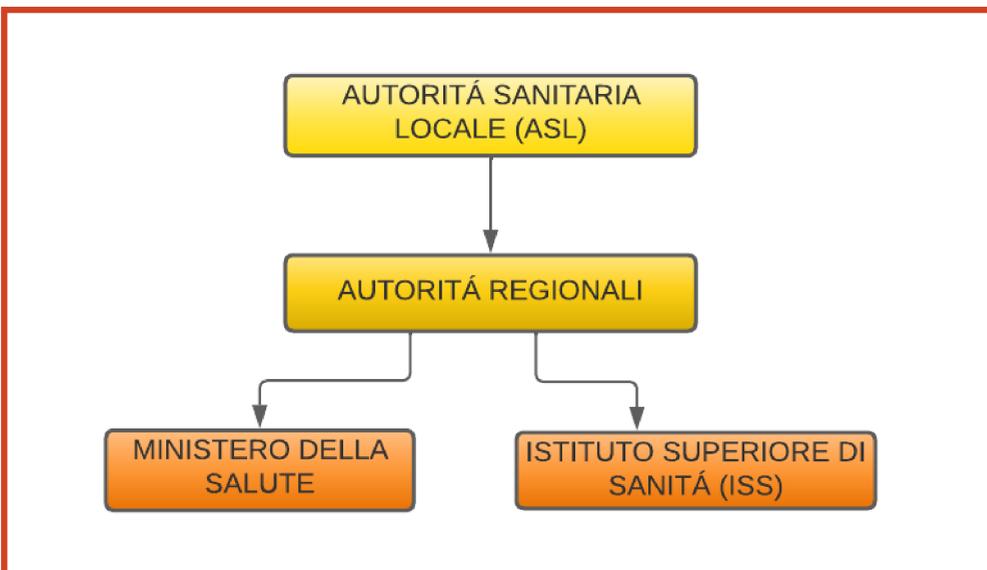
Dall'insorgenza della pandemia COVID-19, le autorità sanitarie di tutti i paesi hanno iniziato a pubblicare dati epidemiologici. Tali dati rappresentano una fonte importante di informazioni sia per i ricercatori sia per la popolazione. Lo scopo del nostro lavoro è indagare la disponibilità, la qualità, la completezza dei dati epidemiologici sul COVID-19 in Italia e se tali dati sono *open data* e *open format*.

## Metodi

Per prima cosa abbiamo indagato quali fossero le possibili sorgenti dei dati del COVID-19 in Italia. Abbiamo esaminato i dati riportati da tutte le Regioni Italiane (N=19) e dalle Province Autonome (PA) (N=2) con un particolare focus sui dati pubblicati dalle Autorità Sanitarie Locali (AUSL) della Regione Emilia-Romagna prese a titolo di esempio. Una prima analisi è stata condotta nel periodo novembre-dicembre 2020 e abbiamo eseguito un successivo aggiornamento a gennaio 2022. Abbiamo, inoltre, valutato i dati resi disponibili dal Ministero della Salute, dalla Protezione Civile e dall'Istituto Superiore di Sanità. Per ogni risorsa abbiamo esaminato: la disponibilità di *open data* e di *open format* per la creazione di un *database*, la disponibilità di un archivio contenente tutti i nuovi dati pubblicati ogni giorno, la capillarità territoriale dei dati epidemiologici e la frequenza di pubblicazione (giornaliera o settimanale). È stata condotta un'analisi dettagliata sia dei dati relativi ai casi sia dei dati relativi ai decessi.

## Risultati

Come mostrato dalla Figura 1, i dati sono raccolti inizialmente dalle ASL che trasmettono i dati alle Regioni. Queste trasmettono le informazioni all'Istituto Superiore di Sanità (ISS) e al Ministero della Salute che collabora con la Protezione Civile. La Tabella 1 ci mostra che le AUSL della Regione Emilia-Romagna né a fine 2020 né a gennaio 2022 pubblicavano dati in *open format*. In Tabella 2 possiamo vedere il confronto tra le modalità di pubblicazione dei dati da parte delle Regioni e PA a fine 2020 e a gennaio 2022. Le informazioni riportate in ogni sito erano eterogenee sia dal punto di vista contenutistico sia di formato. Il Ministero della Salute pubblica ogni giorno un aggiornamento in PDF e una *dashboard* interattiva, la Protezione Civile aggiorna quotidianamente i dati sul *repository* Github e l'ISS pubblica i dati in formato Excel.



La **Figura 1** mostra il flusso dei dati COVID-19 in Italia a partire dalla sorgente di dati locale a quelle nazionali.

Regione	Novembre-Dicembre 2020		Gennaio 2022	
	<i>open data</i>	<i>open format</i>	<i>open data</i>	<i>open format</i>
Veneto	Verde	Verde	Verde	Verde
Lombardia	Verde	Verde	Verde	Verde
Piemonte	Verde	Verde	Verde	Verde
Emilia Romagna	Verde	Verde	Verde	Verde
PA Trento	Verde	Verde	Verde	Verde
PA Bolzano	Verde	Verde	Verde	Verde
Friuli Venezia Giulia	Verde	Verde	Verde	Verde
Valle d'Aosta	Verde	Verde	Verde	Verde
Liguria	Verde	Verde	Verde	Verde
Toscana	Verde	Verde	Verde	Verde
Umbria	Verde	Verde	Verde	Verde
Marche	Verde	Verde	Verde	Verde
Lazio	Verde	Verde	Verde	Verde
Abruzzo	Verde	Verde	Verde	Verde
Molise	Verde	Verde	Verde	Verde
Campania	Verde	Verde	Verde	Verde
Puglia	Verde	Verde	Verde	Verde
Basilicata	Verde	Verde	Verde	Verde
Calabria	Verde	Verde	Verde	Verde
Sicilia	Verde	Verde	Verde	Verde
Sardegna	Verde	Verde	Verde	Verde

In **Tabella 2** si mostra la disponibilità di *open data* e *open format* (colore verde) delle Regioni e PA nel periodo novembre-dicembre 2020 e gennaio 2022.

AUSL	Novembre-Dicembre 2020		Gennaio 2022	
	<i>open data</i>	<i>open format</i>	<i>open data</i>	<i>open format</i>
Piacenza	Verde	Verde	Verde	Verde
Reggio Emilia	Verde	Verde	Verde	Verde
Bologna	Verde	Verde	Verde	Verde
Imola	Verde	Verde	Verde	Verde
Ferrara	Verde	Verde	Verde	Verde
Romagna	Verde	Verde	Verde	Verde
Modena	Verde	Verde	Verde	Verde
Parma	Verde	Verde	Verde	Verde

In **Tabella 1** si mostra la disponibilità di *open data* e *open format* (colore verde) delle AUSL nel periodo novembre-dicembre 2020 e gennaio 2022.

## Conclusioni

In conclusione possiamo dire che in Italia sono disponibili molteplici fonti di dati del COVID-19, ma la problematica maggiore è rappresentata dalla eterogeneità del formato. Pochissime Regioni o PA forniscono dati *open format* che permettono al ricercatore di analizzare facilmente l'andamento della pandemia. Per agevolare l'analisi dei dati sarebbe utile che tutte le Regioni pubblicassero le stesse informazioni in un formato *open* standardizzato. Tutti i dati dovrebbero essere di buona qualità e facilmente accessibili.

