

Report: Analisi e raccomandazioni per la fase 2

a cura del SGdL 7: “Big Data & AI for policy” della Task force data-driven anti Covid-19 del MID -- 14/04/2020

Il sottogruppo di lavoro 7 “**Big Data & AI for policy**” ha l’obiettivo di proporre metodi e strumenti per la progettazione e l’attuazione di politiche basate sui dati e sull’evidenza informativa, sfruttando tecnologie innovative di big data analytics e intelligenza artificiale. A tal fine il sottogruppo ha condotto dal 6 al 10 aprile 2020 un totale di **16 incontri con decisori ed esperti** impegnati nell’emergenza Covid-19 elencati in sez. 5 (epidemiologi, virologi, infettivologi, manager sanitari, manager di informatica sanitaria).

Le indicazioni del gruppo di lavoro, integrate con le informazioni raccolte, sono riportate in seguito. Esse comprendono le **raccomandazioni per affrontare le prossime fasi dell’epidemia**, e l’identificazione delle principali **esigenze di informazioni a supporto delle decisioni epidemiologiche e cliniche**. Si è riscontrato un forte **consenso tra gli esperti** intervistati su iniziative immediatamente cantierabili, supportate da una migliore comprensione dei fenomeni avvenuti nella fase 1.

-
- E’ necessario **potenziare con personale e tecnologie i presidi sanitari sul territorio** (servizi di igiene e prevenzione epidemiologica, medici di base, medicina del lavoro, servizi USCA di continuità assistenziale) mettendoli in grado di isolare e contenere tempestivamente catene di contagio e focolai.
 - E’ importante che il sistema di sorveglianza epidemiologica sia potenziato con la capacità di integrare **molteplici sorgenti di dati** (anche attivate ad hoc) in modelli di analisi e previsione per estrapolare le informazioni salienti: il **rischio e l’incidenza della malattia** nelle popolazioni e nei luoghi, gli elementi di **diagnosi precoce** a supporto della medicina del territorio.
 - E’ inoltre importante l’**apertura all’analisi dei dati clinici e radiologici** per rendere possibile l’ingaggio dei centri di ricerca su progetti di **big data analytics e intelligenza artificiale** per far avanzare la conoscenza sulla malattia mediante modelli predittivi e esplicativi per il decorso clinico dei pazienti Covid.
-

Si ritiene inoltre importante la disponibilità di piattaforme che consentano un **accesso tempestivo alle informazioni** da parte delle diverse figure sanitarie coinvolte nella gestione dell’epidemia (sia sul territorio che negli ospedali). Queste riguardano, da un lato, **materiali e infrastrutture** (disponibilità di dispositivi di protezione individuale; posti letto; test diagnostici e posti in terapia intensiva, etc.), e dall’altro **informazioni centrate sul paziente**.

Tutte le iniziative menzionate dovranno porre la dovuta attenzione agli aspetti di sicurezza informatica e di protezione della **privacy dei cittadini**.

1. Sorveglianza epidemiologica sul territorio

Per uscire progressivamente dal lockdown servono sistemi di **sorveglianza epidemiologica** molto efficaci (in termini di sensibilità e specificità) che accompagnino la riapertura, facendo tesoro delle lezioni imparate nella fase 1 (cosa è andato bene, cosa è andato male, cosa non è stato ancora compreso a fondo). Occorre identificare rapidamente i soggetti con sintomi (tosse, febbre, etc.) da sottoporre a tampone per **intercettare i positivi**; ricostruire le **catene di contagio** dei soggetti positivi; prevenire o spegnere sul nascere i **focolai**; affinare i **modelli epidemiologici** relativamente alla prima ondata -- corredando i dati epidemici con dati di **mobilità** dei cittadini ed altri *proxy* di **variabili demografiche, socio-economiche ed ambientali**, al fine di identificare i *determinanti* principali dell'epidemia. Un efficace sistema di sorveglianza ed una migliore comprensione dei determinanti dell'epidemia sono cruciali per pianificare la riapertura del paese, e in primo luogo del sistema produttivo.

1.1. Comprensione della *fase1* per affrontare la *fase2*

1. Affinare i **modelli epidemiologici** attraverso l'analisi retrospettiva della prima ondata dell'epidemia, che metta in correlazione i dati epidemici con dati sulla **mobilità** dei cittadini ed altre **variabili demografiche, socio-economiche ed ambientali**, a varie scale di risoluzione spaziale e temporale. Questa analisi consentirà di identificare i **determinanti principali per le stime dei parametri dell'andamento epidemico**, coadiuvando la pianificazione nella fase 2 (riapertura). In particolare:
 - a. Comprendere il ruolo della mobilità è cruciale per la ripartenza. A tal fine occorre analizzare le **matrici origine-destinazione dei movimenti** dei cittadini pre-lockdown per stimare il **rischio di diffusione** nei vari luoghi alla ripresa. Ad esempio, i dati da telefonia cellulare potrebbero consentire di monitorare **l'entropia e la varietà degli spostamenti** di ingresso/uscita per ciascun territorio. Questi dati possono poi essere combinati con altre variabili socio-economiche, quali ad esempio rapporto fra spazi e addetti in un sistema locale del lavoro, indici INAIL sulla sicurezza sul lavoro, etc.
 - b. Occorre quantificare e separare il contributo all'andamento dell'epidemia determinato dalla **mobilità** dei cittadini rispetto a quello dei **focolai ad alta intensità** (ospedali, RSA, centri commerciali, etc.). L'ipotesi da vagliare con questa analisi è che il contributo della mobilità sia stato relativamente limitato rispetto a quello dei focolai-- che potranno poi essere ulteriormente studiati in relazione a pratiche sanitarie non adeguate (per es., in pronto soccorso e RSA) e a carenze nella quantificazione dei contagi (esecuzione dei tamponi).

Azione collaterale: queste analisi richiedono il consolidamento e l'alimentazione continuativa di una piattaforma che rende disponibili i **dati provenienti dagli operatori telefonici nazionali, aggregati ed anonimizzati, relativi alle presenze e ai flussi origine-destinazione dei cittadini**.

2. Analizzare retrospettivamente i **positivi non-ospedalizzati** (i.e. seguiti a casa in quarantena) nella fase 1 in base a caratteristiche sanitarie e demografiche, al fine di:
 - a. estrapolare il **rischio di malattia** nell'intera popolazione;
 - b. Identificare elementi per la **diagnosi precoce** di malattia a supporto della medicina del territorio nella fase 2.

***Azione collaterale:** queste analisi richiedono la raccolta di dati attraverso un questionario rivolto ai positivi non-ospedalizzati. L'associazione nazionale degli infettivologi è a disposizione per disegnare il questionario, che andrà poi somministrato in collaborazione con le regioni interessate.*

3. Stimare accuratamente l'**incidenza del contagio da COVID** (velocità dell'infezione) nella popolazione, attraverso:
 - a. Una **indagine campionaria a livello nazionale**, su larga scala e appropriatamente stratificata, di test sierologici (in corso di organizzazione da parte dell'Istat e dell'ISS in collaborazione con la Croce Rossa);
 - b. In concomitanza con la riapertura, **indagini focalizzate** (stile Vo' Euganeo) da svolgersi in un certo numero di località appropriatamente selezionate, e su un campione di soggetti in ciascuna località. Questi soggetti verranno seguiti nel tempo mediante strumenti di telemedicina e con test periodici ripetuti (sia tamponi che test sierologici), che idealmente dovrebbero essere effettuati anche su individui asintomatici -- così da garantire un dettagliato follow-up dei positivi e una accurata caratterizzazione sia della evoluzione dei contagi nella popolazione, che della malattia nei contagiati.

1.2. Azioni e strumenti per la fase 2: *test & track*

1. **Screening precoce dei soggetti sintomatici** finalizzato a sottoporre a tampone soggetti a rischio e scoprire precocemente i positivi al Covid-19. Lo screening va implementato attraverso la rete territoriale dei medici di base, dei servizi di prevenzione delle ASL e medici del lavoro nelle imprese e nella PA, e può essere efficacemente supportato da strumenti di **telemedicina**. Potrebbe essere considerato anche l'impiego di volontari reclutati tra gli infermieri e gli studenti di medicina.
 - a. priorità: tenere i pazienti sintomatici, positivi e non, **lontani dall'ospedale** per quanto possibile attraverso il potenziamento della medicina del territorio. Quando non si riesca a garantire un efficace isolamento domiciliare, si devono considerare forme alternative di sorveglianza sanitaria, ad esempio presso alberghi.

2. **Tracciamento dei contatti dei nuovi casi confermati** e ricostruzione delle catene di contagio finalizzata all'isolamento dei positivi. Anche il tracciamento va implementato utilizzando e potenziando le strutture territoriali esistenti, e può essere efficacemente supportato da una tecnologia di **contact-tracing automatico** basata su dispositivi personali (per es., telefoni cellulari) su cui installare software specifici (con la **massima attenzione possibile al rispetto della privacy**)¹
3. Identificazione e contenimento tempestivi dei **focolai epidemici**, ossia di **cluster** di casi associati ad una comunità ristretta come un ospedale, una struttura sanitaria residenziale, un luogo di lavoro, una comunità cittadina, anche attraverso strumenti di supporto basati sulla **geo-localizzazione** dei positivi al test, sempre in una cornice di rispetto della privacy dei cittadini.

Si ritiene fondamentale per questa attività di sorveglianza *test & track* sul territorio un coordinamento degli aspetti logistici ed informativi, che permetta una ordinata registrazione e gestione dei casi uniforme sul territorio nazionale in modo da poter affrontare la complessità del processo in tempi rapidi.

2. “Open data” sanitari e intelligenza artificiale

1. **Aprire infrastrutture per l'accesso ai dati clinici** (in particolare le cartelle cliniche dei pazienti Covid) seguendo il modello “Data as a service” delle infrastrutture di ricerca (approccio FAIR - Findable, Accessible, Interoperable, Responsible) per rendere possibile l'**ingaggio dei centri di ricerca** italiani e internazionali su progetti di analisi di dati e intelligenza artificiale. L'apertura all'analisi dei dati clinici è ritenuta fondamentale per avanzare la conoscenza sulla malattia e lo sviluppo di **modelli predittivi e esplicativi per il decorso clinico di pazienti Covid**. Ad esempio, l'iniziativa della Regione Lombardia consentirà l'accesso ai dati clinici, rendendo possibili studi longitudinali anche di notevole profondità sugli attuali pazienti Covid. Il modello lombardo definisce il protocollo di selezione e conduzione delle ricerche proposte dai gruppi di ricerca, e i quadro delle responsabilità per il trattamento dei dati ed il rispetto delle norme del GDPR.
2. **Aprire le biobanche di immagini radiologiche** relative alla malattia Covid per rendere possibile l'**ingaggio dei centri di ricerca** Italiani e internazionali per le applicazioni dell'intelligenza artificiale in radiologia, finalizzate a migliorare la **capacità diagnostica**. Un esempio di obiettivo concreto è lo sviluppo di modelli predittivi che migliorino la specificità della TAC ai fini delle diagnosi Covid-19.
3. Pubblicazione tempestiva di tutte le **sequenze dei ceppi virali** identificati dai laboratori su tutto il territorio nazionale per consentire studi di epidemiologia molecolare che possano chiarire meglio le dinamiche di diffusione del virus.

¹ “La Corea del Sud non ha fatto solo la tracciabilità sui cellulari. Ha anche mobilitato 20 mila persone che hanno indagato e spezzato le catene di trasmissione. L'innovazione tecnologica deve essere accompagnata da uno sforzo umano”. Jean-François Delfraissy, direttore del CTS del governo francese su La Repubblica del 12/04/2020.
https://www.repubblica.it/cronaca/2020/04/12/news/coronavirus_infettivologo_macron-253823225/?ref=RHPPLF-BH-I253798132-C8-P1-S2.2-T1

3. Sistemi informativi sanitari centrati sul paziente e sulla comunità

1. Visione integrata del sistema informativo (regionale) centrato sul paziente, che **connetta i servizi della medicina sul territorio con l'assistenza ospedaliera**, gestendo tutto il **ciclo clinico ed extra-clinico del paziente Covid** (diagnostica, assistenza a domicilio, quarantena, assistenza ospedaliera) con vista differenziata per i diversi operatori (medico di base, USCA, epidemiologo e servizi di prevenzione, clinico, manager sanitario locale e regionale) e per le diverse sorgenti informative (test, cartelle cliniche, analitica). Un tale sistema deve poter offrire anche una visione aggregata a livello di comunità per mettere in evidenza trend, focolai. Carente o assente in diverse regioni.
2. Dashboard integrate e tempestive su risorse sanitarie/ospedaliere disponibili e sulle stime di occupazione futura a livello regionale o anche nazionale (posti letto, posti in terapie sub-intensive e intensive, ma anche disponibilità dei dispositivi di protezione individuale).

4. Azioni progettuali

Il gruppo di lavoro “Big Data & AI for policy” ha anche l'obiettivo di favorire azioni progettuali tempestive sui temi evidenziati: (i) agendo da catalizzatore e favorendo progettualità e hackathon (o datathon) nelle reti di ricercatori in epidemiologia, big data, AI e clinica medica; (ii) conducendo progetti collaborativi fra i membri del gruppo di lavoro stesso e altri soggetti. Questi obiettivi possono essere perseguiti sfruttando le relazioni fra i membri del gruppo di lavoro e le associazioni di ricercatori italiane e internazionali, anche in riferimento alla pluralità di proposte emerse nella call aperta del Ministero dell'Innovazione in collaborazione con il MUR. L'approccio è coerente con, anzi anticipa, l'*ERAvsCorona Action Plan* della Unione Europea pubblicato il 7 Aprile².

Casi di studio attivati e/o immediatamente cantierabili

A titolo di esempio, la seguente è una lista non esaustiva di attività in essere o in partenza di cui il gruppo di lavoro è a conoscenza allo stato attuale.

- Analisi delle cartelle cliniche dei pazienti Covid per studio del decorso clinico (Vineis Imperial College, IIT, CNR, UNIFI, vari ospedali Milano e Roma)
- Modelli Epidemiologici:
 - Vineis, Dorigatti (Imperial College), Merler (FBK), Crisanti (Uni. Padova)
 - Integrazione con dati di mobilità GSM (Lopalco Giannotti Pedreschi, CNR, UNIFI, WindTre)
 - Ferretti (Oxford Univ.) e collaboratori

²[10 priority actions for coordinated research and innovation actions: FIRST “ERAvsCORONA” ACTION PLAN short-term coordinated](#)

- Iniziativa **Open Data** Lombardia, potenzialmente anche in altre regione (ad esempio Regione Toscana)
- Analisi AI di immagini radiodiagnostiche (TAC): Grigioni (ISS), Cucchiara (Lab AIIS Cini), Grigetto (PoliTO, Osp. Molinette), Neri (UNIFI), Zobel e Quattrocchi (Campus BioMedico)
- Iniziativa Istat / ISS per indagine campionaria con test sierologici
- Drug repurposing: Barabasi Lab su scoperta di farmaci anti-Covid con network science A.-L. Barabasi (Northeastern Univ. USA), Andrea Cavalli (IIT)
- Modelli su osservazioni ambientali - (CNR, SIMA: Relazione circa l'effetto dell'inquinamento da particolato atmosferico e la diffusione di virus nella popolazione, Harvard Univ)
- AI & Devices: riconoscimento positività attraverso suono della tosse (ERC project Cecilia Mascolo), telemedicina (tantissime proposte di collegamento allo smartphone di dispositivi di rilevazione di parametri vitali, quali ossigenazione)
- Modelli per lo studio degli effetti di possibili strategie alternative per la fase 2: Marinari, Parisi, Ricci-Tersenghi (Sapienza), Bernaschi, Gori (Univ. Siena), Pedreschi

Altre necessità e azioni specifiche emerse dalle interviste

Rafforzamento delle pneumologie e delle strumentazioni per l'ossigenoterapia. Prevedere le esigenze **post-degenza dei reduci da malattia Covid** con riduzione del 30-40% delle capacità respiratorie da recuperare.

Counseling psicologico per la **gestione dell'ansia da parte di soggetti vulnerabili**. Focus su **donne incinte**, una indagine ha evidenziato livelli di ansia in forte aumento dopo l'inizio del lockdown, dal 15% all'80% (!), a fronte di una probabilità molto bassa delle donne incinte di contrarre la malattia e ancora più bassa di trasmetterla al feto.

5. Elenco degli stakeholder intervistati

1. Riccardo **Ammoscato**, Protezione Civile
2. Massimo **Andreoni**, Primario centro Covid Osp. U. Tor Vergata (infettivologo)
3. Andrea **Belardinelli**, Dir. Sanita' digitale e innovazione Reg. Toscana (manager IT)
4. Giuseppe **Costa**, Servizio Sovrazonale di Epidemiologia Piemonte (epidemiologo)
5. Andrea **Crisanti**, Dir. Lab. microbiologia e virologia Univ. Osp. Padova (virologo)
6. Rita **Cucchiara**, Prof. Univ. Modena Reggio E., Dir. Lab ImageLab, Dir. Lab Naz CINI su Intelligenza Artificiale (ing. informatica)
7. Gianpiero **D'Offizi**, Dir. Malattie Infettive Epatiche, INMI Spallanzani (infettivologo)
8. Iole **Fantozzi**, Commissario Straord. Osp. Reggio Calabria (manager sanitario)
9. Francesco **Frieri**, Dir. Gen. Reg. Emilia Romagna per la digitalizzazione (manager)
10. Pier Luigi **Lopalco**, Commissario Task force Covid Reg. Puglia (epidemiologo)
11. Paola **Pagliara**, Protezione Civile
12. Giuseppe **Rizzo**, Dir. Medicina Materno Fetale, Osp. Cristo Re, Roma (ostetricia)

13. Marcello **Savarese**, Chief Data Officer, WindTre (data scientist)
14. Salvatore **Scondotto**, Dir. Sorveglianza Epidemiologia Valutativa Reg. Sicilia (epidemiologo)
15. Roberto **Soj**, Dir. Lombardia Informatica, in-house IT Reg. Lombardia (manager IT)
16. Stefano **Vella**, Task force Reg. Sardegna (infettivologo)
17. Alessandro **Vespignani**, Prof. Northeastern Univ. Boston MA USA (epidemiologo computazionale)
18. Maurizio **Viecca**, Dir. Cardiologia Osp. Sacco Milano (cardiologo)

Il gruppo di lavoro ringrazia sentitamente tutte le persone elencate per aver dedicato, nonostante i loro impegni gravosi, un poco del loro tempo a condividere conoscenze ed esperienze preziose.

6. Gruppo di lavoro

Le interviste agli esperti e la stesura di questo rapporto sono stati curati da:

1. Dino **Pedreschi** Univ. Pisa, Dir. Phd Data Science, Dir. KDD Lab, coordinatore gruppo di lavoro 7 “Big Data & AI for policy” (Computer Science, Big data, AI)
2. Fosca **Giannotti** ISTI-CNR, Pisa, coordinatrice RI SoBigData.eu (Computer Science, Big data, AI)
3. Francesca **Chiaromonte** Scuola Superiore S.Anna, Pisa, coordinatrice scientifica Dipartimento di Eccellenza EMbeDS, Economics and Management in the era of Data Science (Statistica)
4. Paolo **Vineis** Imperial College, London, Environmental Epidemiology (Epidemiologia)
5. Massimo **Bernaschi** IAC-CNR, Roma, (Mathematics, Computer Science)
6. Serafino **Sorrenti** expert on Innovation (esperienze mondo bancario, telco e PA) - responsabile Agenda Digitale Regione Sicilia, Monitoraggio Migrazione per Ministero Salute (Economia)
7. Luca **Ferretti**, Univ. Oxford, Physics, Statistical Genetics and Pathogen Dynamics Epidemiology (Epidemiologia computazionale, Fisica)
8. Mauro **Grigioni**, Centro Nazionale Tecnologie Innovative in Sanità Pubblica, Istituto Superiore di Sanità
9. Paolo **De Rosa** Dipartimento Trasformazione Digitale, Responsabile Task force data-driven anti Covid-19, Min. Innovazione (Computer Science)