PLANET: studio di correlazione Covid-19 e Inquinamento Atmosferico

Daniele Spiga, INFN-PG
Per il gruppo PLANET
09-10 Febbraio 2021

Outline

- Il progetto PLANET
  - Motivazioni e Obiettivi
- Le sfide che stiamo affrontando
- Esplorazione dei dati sulla regione Umbria
  - I primi passi
- Aggregazione e gestione dei dati
  - Sfide tecnologiche.. e non solo
- Sintesi e considerazioni

Breve premessa

Il progetto nasce a seguito di una proposta fatta da Medici (clinici, epidemiologi etc) dell'**Università degli Studi di Perugia** all'INFN

- Sviluppo del modello di analisi
- Integrazione e gestione dei dati

Contatto con **INFN-Perugia** (e successivamente **CNAF**) per valutare la fattibilità di una collaborazione finalizzata a

- **studio osservazionale** per la valutazione dell'incidenza dell'inquinamento atmosferico sull'infezione COVID-19.

Dopo alcune valutazioni preliminari abbiamo richiesto l'apertura di **sigla in CSN5** discussa a Settembre e finanziata per 2021/22

Covid-19 e inquinamento atmosferico: Ipotesi

La malattia Covid-19 è una sindrome respiratoria acuta grave causata dall'infezione Sars-Cov-2, che si **manifesta in modo molto variabile**:

- Forme completamente asintomatiche … fino a quadri clinici molto critici
- Comportamento molto eterogeneo sulla distribuzione geografica, con particolare attenzione alla prima fase pandemica

E' legittimo porsi domande sul possibile legame sia tra la velocità di trasmissione che la gravità del Covid-19 e l'inquinamento atmosferico, da cui l'ipotesi di due meccanismi:

- **Acuto**: le microparticelle derivate da combustibili fossili potrebbero agire come vettori aerei di virus;
- **Cronico**: l'esposizione a microparticelle e inquinanti chimici potrebbe causare lesioni polmonari croniche, esacerbando le conseguenze dell'infezione virale

Elevato livello di attenzione sul tema

Molteplici studi recenti hanno esaminato questo problema. **Tuttavia, le evidenze sulla relazione tra inquinamento atmosferico ed epidemia di Covid-19 non sono molte**

- È un fenomeno recente
- **Analisi non conclusive** per carenza di studi inclusivi di fattori potenzialmente confondenti e mediatori (vedi dopo)

Una possibile nesso è comunque suggerito dalla consolidata rapporto tra esposizione a inquinanti atmosferici e rischio di patologie e infezioni acute delle basse vie respiratorie

- particolarmente evidente in soggetti vulnerabili (e.g. anziani)

L’obiettivo (ambizione) del progetto PLANET

Effettuare uno studio di tipo osservazionale (ecologico) e valutare l’associazione tra Covid-19 e l’esposizione all’inquinamento atmosferico per contribuire alla comprensione del suo ruolo nella **suscettibilità al contagio** e nella **gravità dei sintomi**.

Come:

- **Includendo molteplici fonti di dati:**
  - si ipotizza che dati atmosferici, densità di popolazione, ambiente urbano o rurale, mobilità, condizioni socio-economiche etc influenzino sia i tassi di diffusione e infezione di SARS-COV-2 che la severità. Complementando anche con variabili cliniche (comorbidità, fragilità..)
- **Analizzando dati su scala spaziale comunale,** al livello Nazionale

Le sfide poste da questa attività

La realizzazione dello studio descritto pone molteplici sfide, di diversa natura e con diverso fattore di rischio

Raccolta dei dati

Molteplicità di interlocutori
- Non esiste referente unico
- **E tantomeno interfacce tecnologiche...**

Aspetti organizzativi
- Accordi formali tra enti (DPO), rispetto delle norme, classificazione dei dati, GDPR risk based approach

Analisi e interpretazione

Qualità del dato
- Dati eterogenei e raccolti con finalità diverse dall'uso fatto "Validazione", armonizzazione

Necessità di Competenze Multidisciplinari
- Tema ampio che richiede competenze di dominio, bisogna porsi le giuste domande

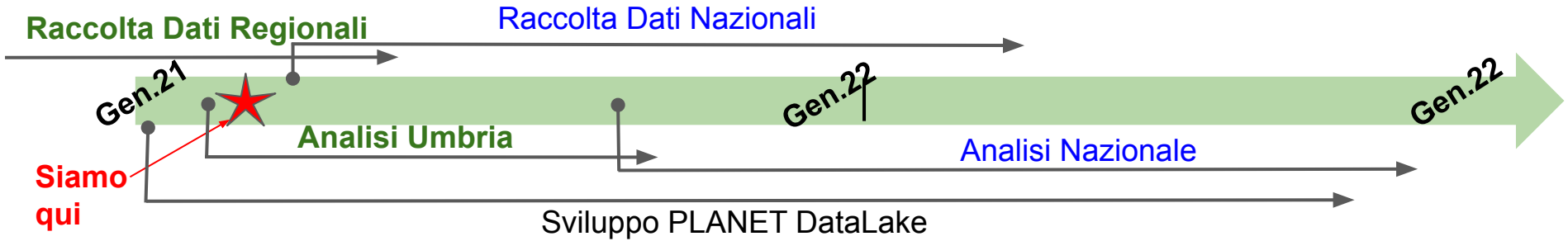
Gestione dei dati e delle pipelines

Infrastruttura (Datalake) scalabile, agile, collaborativa e integrata
- per la gestione di dati eterogenei

Procedure specifiche di gestione dei dati sensibili

Soluzione estendibile
- e riusabile in altri ambiti

Roadmap e strategia dello studio



Periodo: dal febbraio a dicembre 2020.

- Studio su finestre temporali

Area: nazionale

- Focus iniziale su regione Umbria
- Liguria, Lazio, Marche, Lombardia…

Esposizione: concentrazioni stimate di PM10, PM2,5, biossido di azoto e ozono.

- Modellistica ARPA/ISPRA/CAMS, valori giornalieri, griglia fino a 1kmx1km

Fattori di suscettibilità: demografici (dimensione della popolazione, percentuale per fascia età, reddito, percentuale laureati…), stato socio-economico, indici di fragilità, meteorologici (temperatura, umidità), clinici (tasso di mortalità, cartelle cliniche casi covid), politiche di lockdown.

I dati attualmente aggregati

ISTAT
ISTAT Censimento
ID (Dr. N. Caranci)
Dati covid Regione Umbria
ARPA Umbria

Partendo dai **dati RAW** raccolti e archiviati nel formato originale possiamo generare datasets integrati funzionali all'analisi

- primo dataset integrato per avviare la fase di "validazione" ed esplorazione

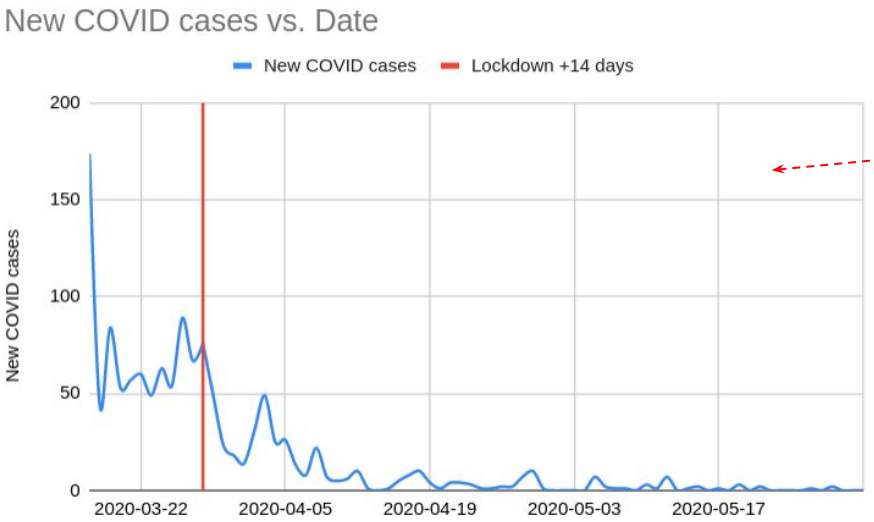
	City	Population	Density	Surface	Lattante	Primainfanzia	Secondainfanzia	Terzainfanzia	Adolescenza	Primaadulta	Secondaadulta	Terza
0	Acquasparta	4611	57.0	81.61	48.0	22.0	150.0	222.0	484.0	708.0	1368.0	915.0
1	Allerona	1722	21.0	82.61	20.0	13.0	53.0	63.0	140.0	326.0	503.0	370.0

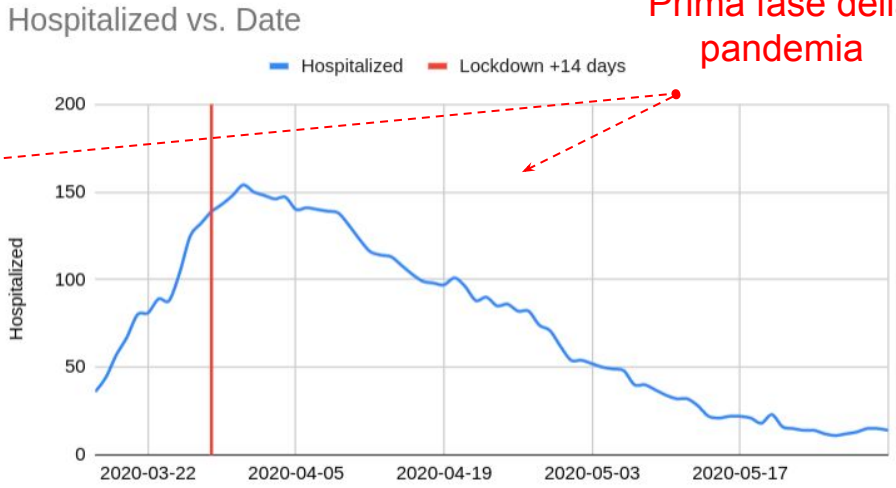
Quarta	Quinta	TotaleEta	MediaEta	Depriv_idx	Depriv_cat	CovidCases	MaxHospitalized	AvgHospitalized	Deceased	MaxIntensiveCare	AvgIntensiveCare
579.0	74.0	4570.0	47.441357	1.013507	3	146	3	0.380952	2	1	0.027211
207.0	29.0	1724.0	48.408353	-1.309272	1	41	2	0.262238	0	0	0.000000

mean_pm10_ug/m3_mean_2019	mean_pm10_ug/m3_std_2019	mean_pm10_ug/m3_median_2019	mean_pm10_ug/m3_mean_2020	mean_pm10_ug/m3_std_2020
16.183606	7.358666	14.376109	15.833302	6.970052
15.828832	7.167570	14.451391	14.905208	7.113094

Dati Covid: Controlli preliminari

- I dati epidemiologici disponibili per l'Umbria sono stati esaminati per controllare rispondessero con i trends attesi in conseguenza del lockdown
  - Dati disponibili per comune con frequenza giornaliera





Prima fase della pandemia

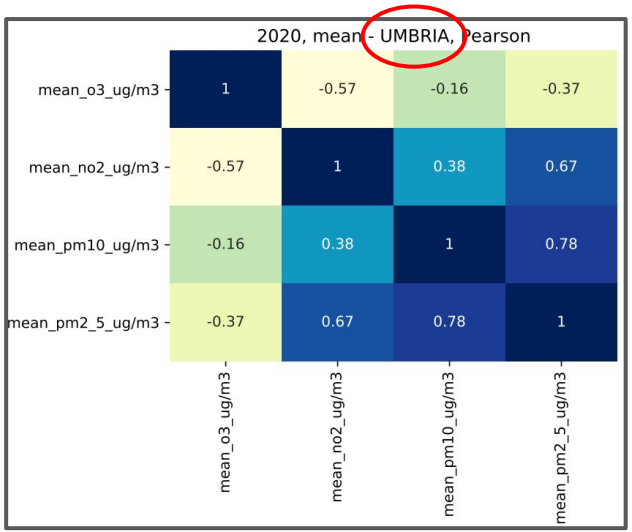
Dati Inquinamento: controlli preliminari

Istituto Nazionale di Fisica Nucleare

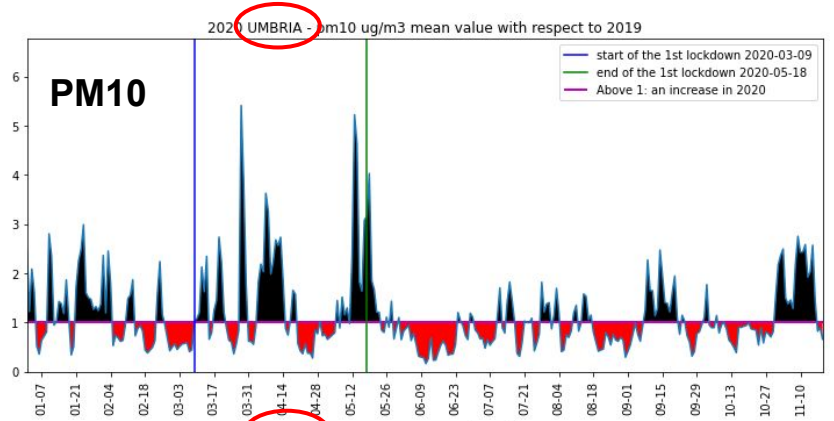
Dati della modellistica forniti da ARPA Umbria [M.Vecchiocattivi]

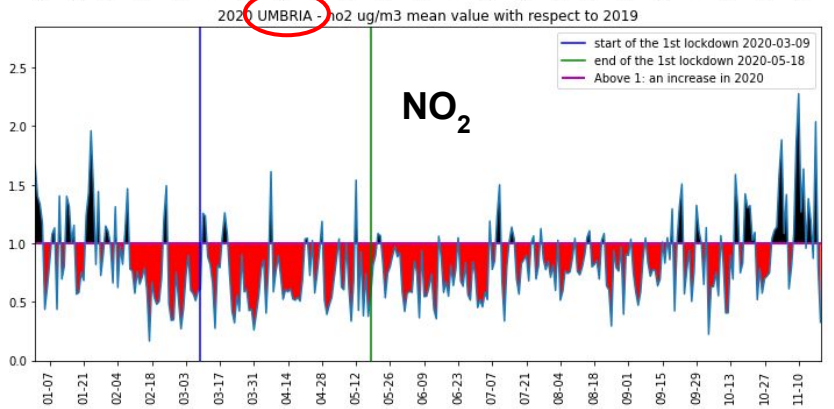
- Dati disponibili su griglia 1kmx1km, frequenza oraria, annualità 2019-2020 (in arrivo '17/'18)

Gao et al., 2017

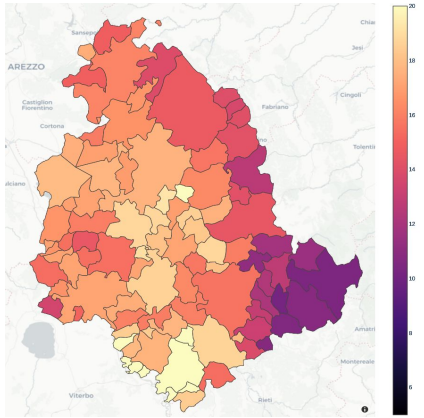


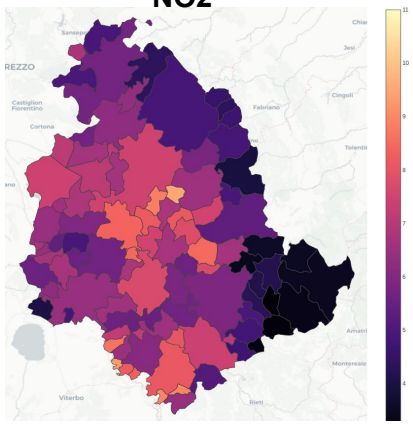
Report SNPA

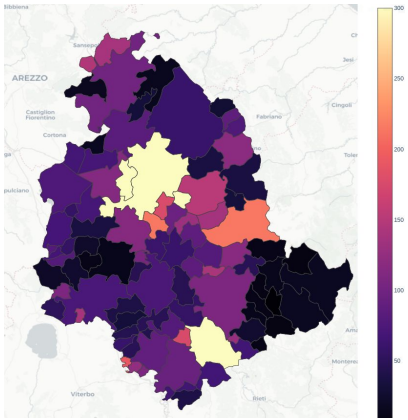


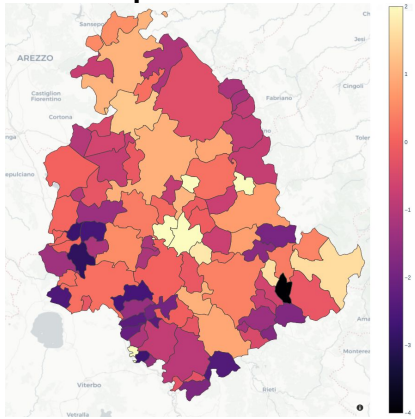


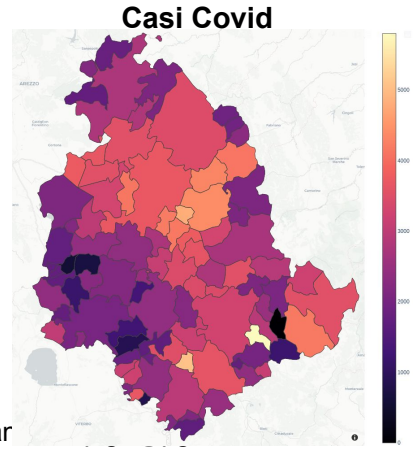
Scala spaziale e molteplicità di dati: **Esempi**

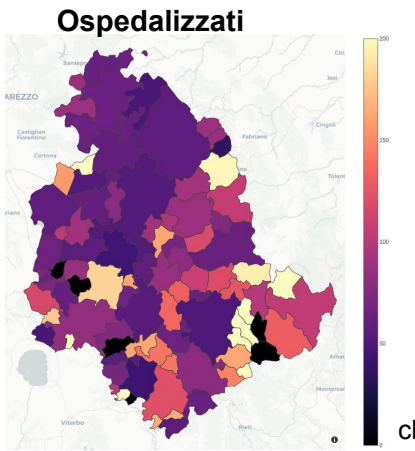


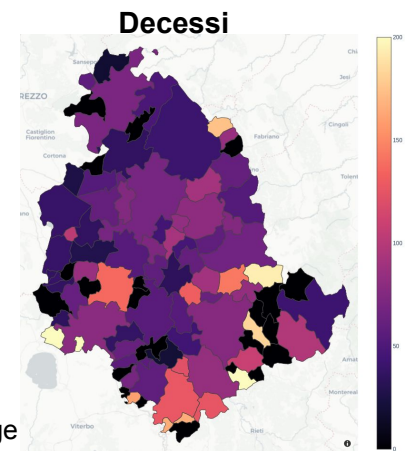


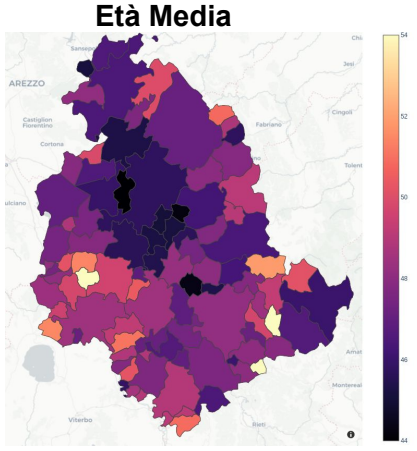












Dar challenge

Prossimi passi

Finalizzare la fase di esplorazione declinandola sui vari fattori di suscettibilità

- età, genere, comorbidità, fattori di contesto socio-economici e demografici, e di comunità (urbano-rurale, attività produttive...).
- Implementare clusterizzazioni per l'identificazione di set omogenei

Effettuare una analisi temporale considerando

- Politiche di lockdown
- Condizioni di variazione dei 4 inquinanti
  - studiando non solo i valori di soglia ma le condizioni di massimo/minimo

Cercare di determinare l'impatto di ciascuna variabile nella gravità dell'infezione

Gestione e accesso ai dati

Dal punto di vista tecnologico la sfida è **realizzare un ambiente di archiviazione dei dati nel loro formato nativo,** dove rimangono in tale condizione fin quando non è necessario definire una struttura per l'analisi

I Dati devono essere FAIR

- Facilmente recuperabili
- Accessibili
- Catalogati

Il sistema deve offrire:

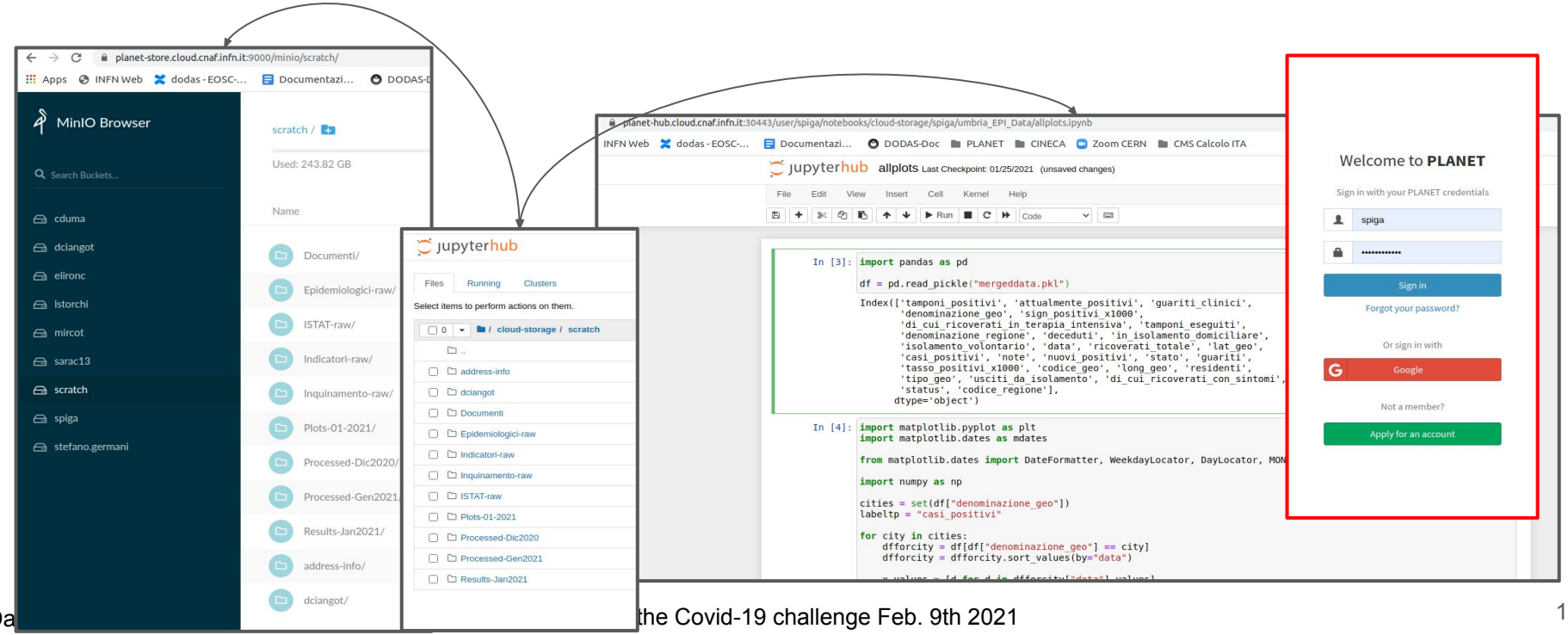
- Integrazione con soluzioni di data processing
- Supporto Multi-utente (autenticazione e autorizzazione)
- Federabile con sistemi distribuiti

L'ambizione è quella di realizzare un Hub di dati eterogenei ed eventualmente sensibili.

PLANET **DataLake** (prototipo)

Istituto Nazionale di Fisica Nucleare

Integrare e rendere disponibile una piattaforma “open” e generica (riutilizzabile) per
- Integrazione di dati da sorgenti informative multiple, Processamento, Federazione di risorse attraverso soluzioni di **INFN-Cloud**



I dati: Non solo tecnologia

Avviate interazioni con molteplici Interlocutori

- **ISRPA, ARPAE e ARPA Umbria, CMCC...**
- **Regione Umbria e Liguria, ASL di Viterbo…**
  - Dati epidemiologici
- **IRCCS Multimedica Milano, ASL di Viterbo, Fermo, Az. Ospedaliere di Perugia e di Terni**
  - Cartelle cliniche

Il trattamento di alcuni dataset clinico-sanitari richiede l'**attivazione di misure tecnico/organizzative che assicurino il rispetto del GDPR**:

- Redazione di template di convenzioni in **collaborazione con INFN-AC e DPO**
- Utilizzo della **certificazione ISO/IEC 27001 dell'INFN**
  - Policy e processi, *risk-based approach, accountability, privacy and security by-design and by-default*
  - Possibilità di utilizzare adeguate tecnologie di cifratura at-rest e in-transit, multi-factor authN, tamper-proof audit logs

Considerazioni e Sintesi

Il progetto è entrato nella fase operativa su tutti i fronti

- con buone prospettive a breve termine relativamente ai dati regione Umbria, Il focus ora è sul recupero dei dati a livello nazionale

Si è innescato un gioco di squadra virtuoso

- Tra la comunità medico-epidemiologica e quella INFN ma anche tra membri INFN con competenze complementari

Impatto e opportunità

- Collegamento con iniziative esistenti INFN (ML-INFN, ELIXIR, LifeWatch, Alleanza Contro il Cancro, Harmony / Harmony Plus...), Progetti Nazionali (PRIN), Europei… ma anche riutilizzo delle tecnologie di PLANET all'interno dell'ente.

Criticità

- La complessità del processo di recupero di dati epidemiologici (Covid19) su scala spaziale comunale, rappresenta a tutti gli effetti un criticità per il progetto.

Il Team

INFN-Perugia

- Diego Ciangottini
- Pasquale Lubrano
- Mirco Tracolli
- Loriano Storchi
- Sara Cutini
- Daniele Spiga

Medici/epidemiologi:

- Giuseppe Ambrosio
- Fabrizio Stracci
- Giampaolo Reboldi

INFN-CNAF

- Davide Salomoni
- Cristina Duma
- Elisabetta Ronchieri
- Barbara Martelli
- Alessandro Costantini