

INTEROPERABILITÀ DEI DATI AMBIENTALI E SANITARI

VERSO UN'INTEGRAZIONE SISTEMICA TRA AMBIENTE E SALUTE A VANTAGGIO DELLA COLLETTIVITÀ: LA PIATTAFORMA DIGITALE IDEAH E LA SFIDA PER IL FUTURO DELLA RETE DEL SISTEMA SANITARIO NAZIONALE E DEL SISTEMA NAZIONALE PER LA PROTEZIONE DELL'AMBIENTE, TRA COLLABORAZIONE DELLE RETI, SOSTENIBILITÀ, PRIVACY E SICUREZZA.

L'integrazione tra la protezione dell'ambiente e la prevenzione della salute rappresenta oggi una delle sfide più ambiziose per il Sistema sanitario nazionale e il Sistema nazionale per la protezione dell'ambiente (Snpa). In questo scenario di necessaria convergenza si inserisce il progetto Ideah (*Integrated database for environment and health*), pilastro tecnologico del Piano nazionale per gli investimenti complementari al Pnrr (Pnc, Investimento 1.5).

Il quadro normativo, delineato dal Dl 36/2022 e dal Dpcm 29 marzo 2023, ha sancito la necessità di una collaborazione strutturale tra il Sistema nazionale prevenzione salute dai rischi ambientali e climatici (Snps) e il Snpa. La piattaforma Ideah nasce proprio con l'obiettivo di superare la storica frammentazione dei dati, creando un'infrastruttura digitale nazionale capace di acquisire, integrare e analizzare flussi informativi eterogenei, con il coordinamento tecnico-scientifico dell'Istituto superiore di sanità (Iss). Il progetto si colloca inoltre nel più ampio quadro regolatorio europeo, che include

l'*European health data space* (Ehds), il regolamento Eu 2025/327 sull'uso secondario dei dati sanitari e la direttiva Inspire per l'infrastruttura di informazione geo-spaziale per scopi ambientali.

Gli obiettivi strategici e la visione del progetto

L'obiettivo primario di Ideah è la realizzazione di una piattaforma di rete digitale nazionale, basata sulle soluzioni di *cloud computing* del Polo strategico nazionale (Tim, Leonardo, Sogei, Cdp), che funga da hub centrale per la ricerca epidemiologica, biostatistica e ambientale. La visione alla base del progetto non è meramente tecnologica, ma strategica. Il primo obiettivo è connettere i sistemi ambientali e sanitari per identificare correlazioni dirette tra esposizione a inquinanti e insorgenza di patologie. A questo si affianca la creazione di un *data lake* nazionale, alimentato da procedure di interoperabilità avanzate con i flussi del nuovo sistema informativo sanitario

(Nsis) e le fonti *open data* ambientali. La piattaforma intende inoltre sviluppare interfacce computazionali che consentano ai ricercatori e agli analisti l'impiego di algoritmi di *machine learning* e analisi biostatistica su larga scala. Infine, si punta a mettere a disposizione di stakeholder, uffici di programmazione e prevenzione e di decisori politici *dashboard* interattive per monitorare l'andamento dei fenomeni e comunicare i rischi in modo tempestivo.

Architettura tecnica e data lake

La realizzazione della piattaforma si articola in quattro fasi logiche fondamentali: censimento e valutazione delle fonti dati, sviluppo delle interfacce di elaborazione, integrazione dei pacchetti applicativi statistici e realizzazione delle procedure di reportistica. Il cuore pulsante del sistema è il *data lake*, un'architettura flessibile alimentata da oltre 60 fonti censite (figura 1), selezionate per rappresentatività territoriale, profondità storica e uniformità delle misure.

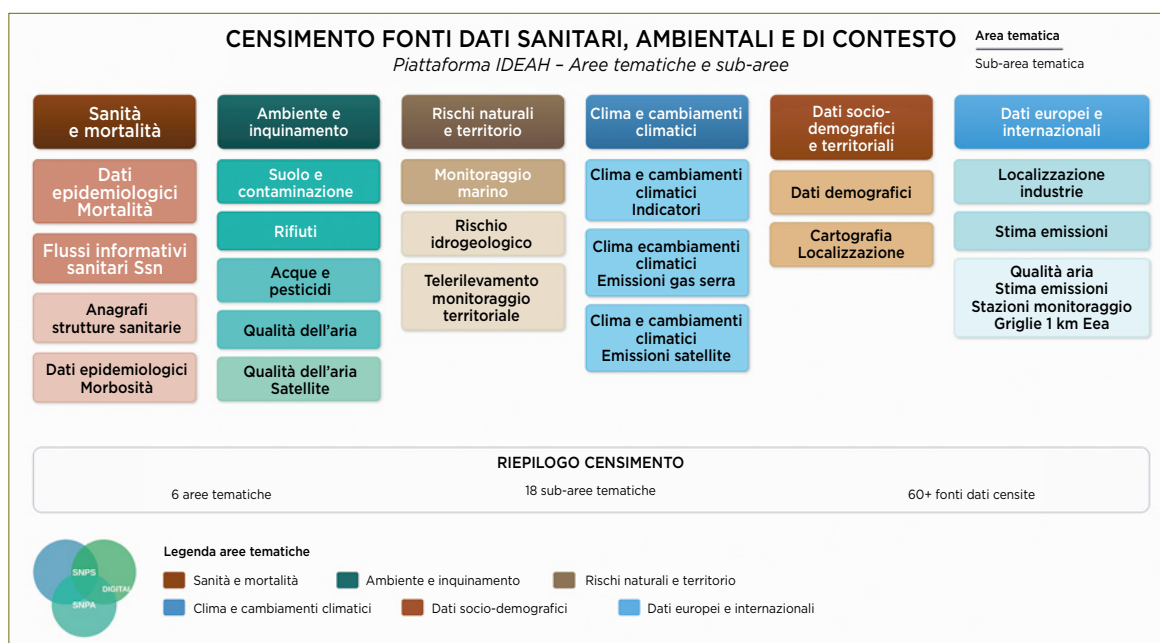


FIG. 1
PIATTAFORMA IDEAH

Censimento delle fonti di dati sanitari ambientali e di contesto.

Le fonti di alimentazione comprendono quattro comparti principali (figura 2). Il comparto sanitario attinge alle statistiche di mortalità Istat e ai flussi informativi del Nsis: dati Emur per gli accessi in pronto soccorso, schede di dimissione ospedaliera (Sdo), specialistica ambulatoriale, assistenza farmaceutica, registro tumori e certificati di assistenza

al parto (Cedap). Il comparto ambientale dialoga con Sinanet (Ispra/Arpa-Appa), acquisendo dati su inquinanti atmosferici, qualità delle acque e del suolo, rifiuti, siti contaminati e stime emissive. Il terzo comparto integra dati satellitari del programma europeo Copernicus (Cams e Land) che forniscono informazioni ad alta risoluzione su qualità dell'aria,

emissioni e uso del suolo, insieme a dati di contesto territoriale quali la localizzazione dei siti industriali, i dati sociodemografici e quelli di geo-localizzazione. Infine, il comparto *big data* include flussi informativi da sensori Iot e fonti media, utili per il monitoraggio in tempo reale e l'analisi della percezione del rischio.

FIG. 2
ARCHITETTURA
DELLE FONTI DATI

Integrazione tra flussi sanitari Nsis, dati ambientali Sinanet, dati satellitari e open data.

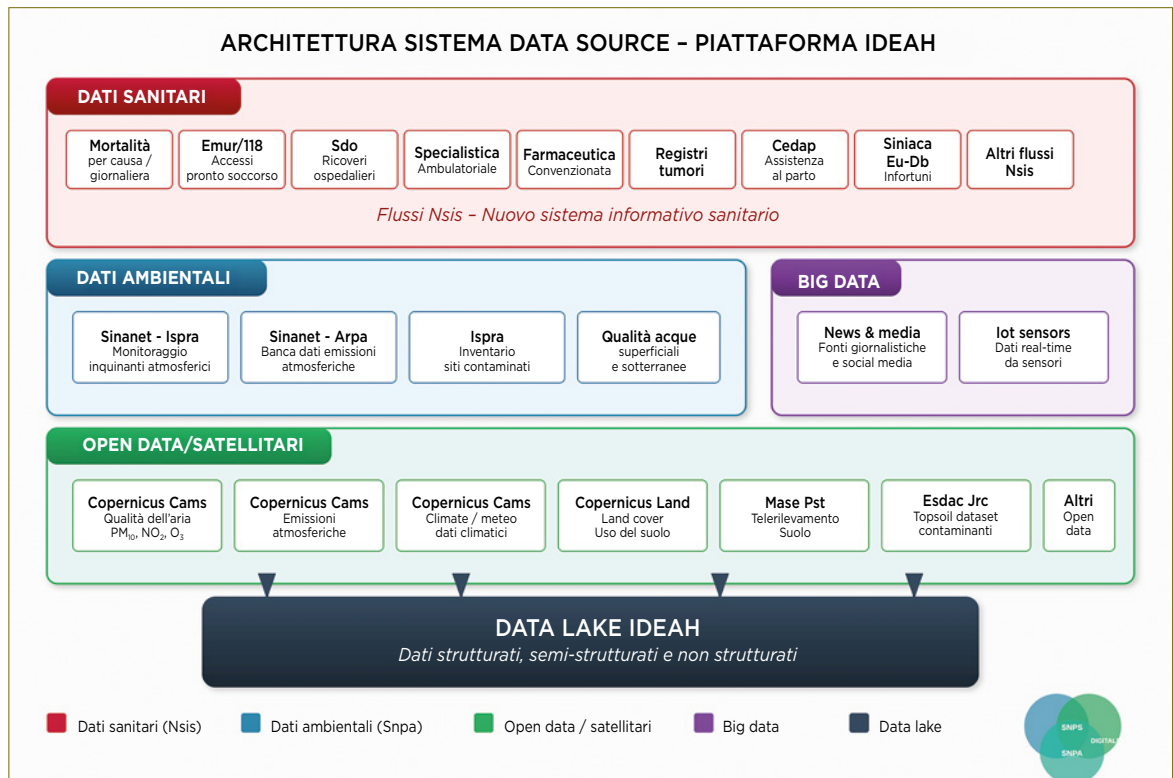
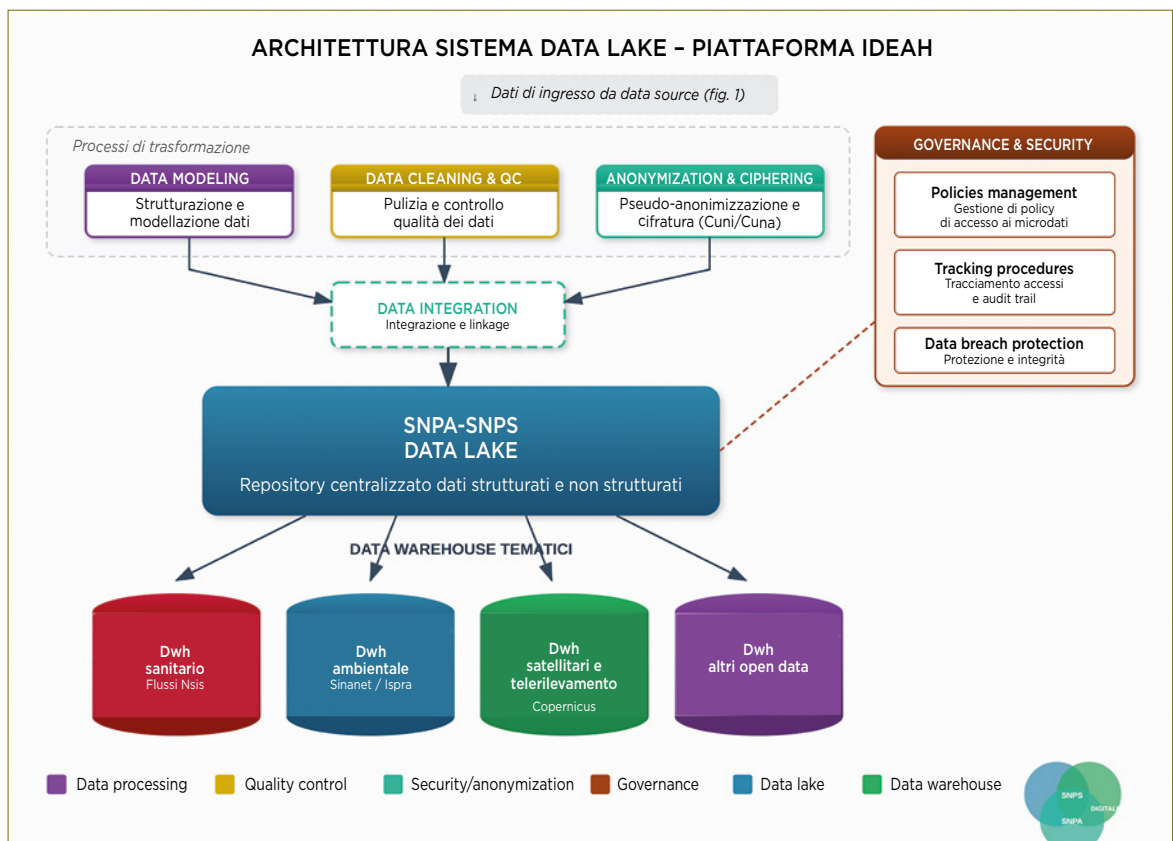


FIG. 3
ARCHITETTURA
DEL DATA LAKE

Processi di data modeling, quality control, anonimizzazione e gestione delle policy di accesso ai data warehouse tematici.



L'architettura tecnica si basa su microservizi specializzati: il *ms-Ispra-adapter* raccoglie dati sugli inquinanti atmosferici dalle Api di Ispra; il *ms-Copernicus-adapter* interroga i sistemi satellitari europei per rilevazioni orarie su PM₁₀, NO₂, O₃ ecc.; il *ms-Ideah-storage* riceve i dati da tutti gli adapter e li organizza nel *data lake* in modo coerente, applicando logiche di ordinamento per area geografica, sorgente, variabile e frequenza. I dati acquisiti vengono quindi sottoposti a processi di *data modeling*, *quality control* e anonimizzazione prima di essere distribuiti nei *data warehouse* tematici (figura 3).

Interoperabilità, sicurezza e privacy

La costruzione di un sistema così vasto comporta sfide significative. La necessità di armonizzare formati di dati eterogenei provenienti da diverse fonti regionali, nazionali e internazionali richiede procedure di *data cleaning*, *quality control* e protezione dei dati estremamente rigorose. È stato costituito un tavolo tecnico comprendente Iss, Polo strategico nazionale e consulenti specializzati (Università La Sapienza-Dipartimento di Ingegneria informatica automatica e gestionale, Scudomed, We-

Com) che ha definito i principi cardine delle procedure di trattamento dei dati e le *policy* d'accesso alle risorse della piattaforma.

Trattandosi di dati sanitari sensibili, la piattaforma adotta criteri di sicurezza allo stato dell'arte. I dati individuali non sono mai accessibili in chiaro: si utilizzano codici univoci non invertibili, come il Cuni (codice unico non invertibile) e il Cuna (codice univoco nazionale dell'assistito), per permettere l'interconnessione tra i flussi senza identificare il paziente. Gli uffici di prevenzione e gli analisti potranno accedere esclusivamente ai metadati

IDENTITY ACCESS MANAGEMENT - PIATTAFORMA IDEAH

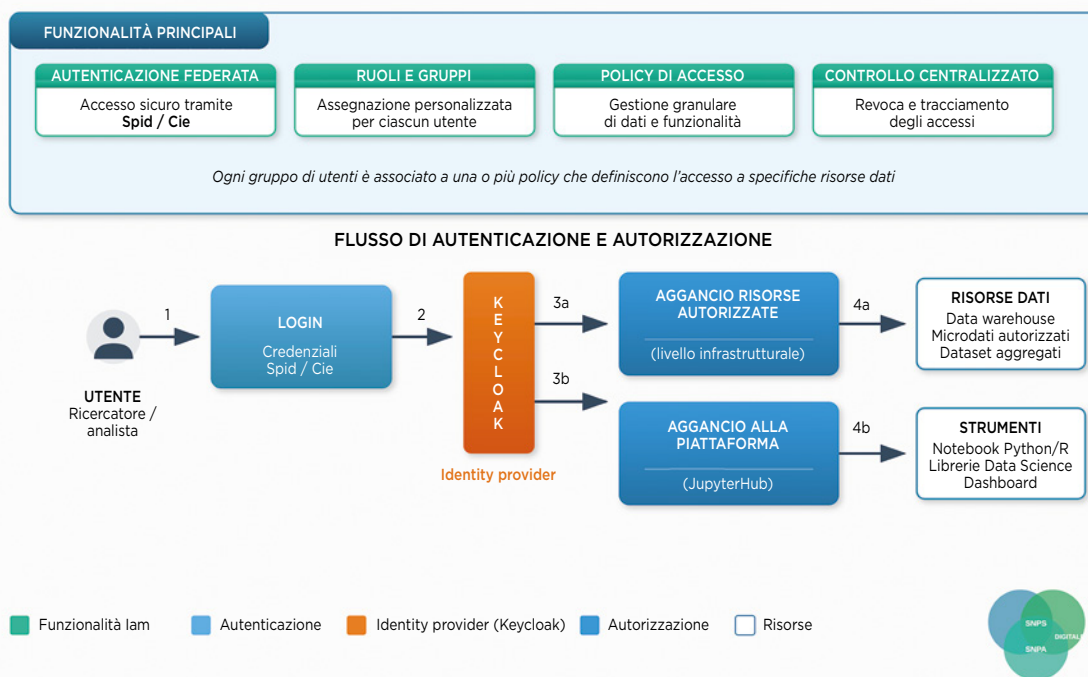


FIG. 4
SISTEMA DI
IDENTITY ACCESS
MANAGEMENT

Autenticazione federata Spid/Cie, assegnazione di ruoli e policy di accesso alle risorse dati.

CICLO DI VITA DEL DATO - PIATTAFORMA IDEAH

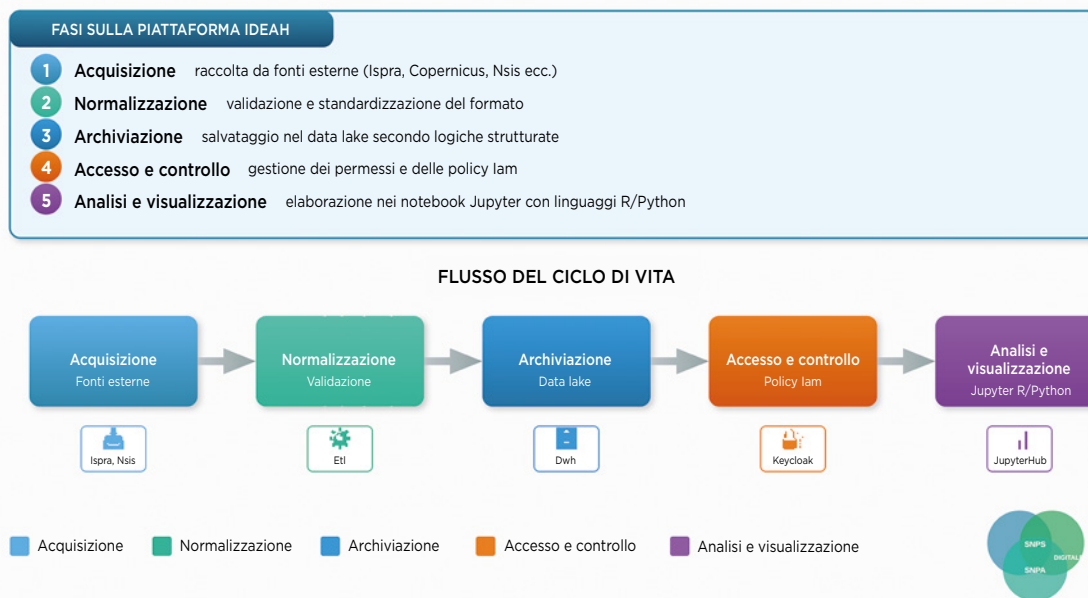


FIG. 5
CICLO DI VITA
DEL DATO SULLA
PIATTAFORMA IDEAH

Acquisizione, normalizzazione, archiviazione, controllo accessi, analisi e visualizzazione.

necessari per le finalità di analisi, con tracciamento rigoroso di ogni procedura di accesso e criterio di minimizzazione informativa (figura 4).

Un esempio concreto di questa complessità è l'analisi dell'impatto dell'inquinamento atmosferico sull'asma infantile. Sono state messe a punto delle *query* di test per incrociare i dati delle sezioni censuarie Istat (domicilio del paziente) con i flussi Emur, le Sdo e l'assistenza farmaceutica, filtrando per codici diagnostici Icd-9-Cm specifici e stagionalità mensile e oraria di assistenza, garantendo la massima riservatezza tramite cifratura e pseudo-anonimizzazione.

Verso lo spazio europeo dei dati sanitari

Il ciclo di vita del dato sulla piattaforma Ideah si completa con la fase di analisi e visualizzazione (figura 5). La piattaforma computazionale JupyterHub, già integrata nel sistema, consente la creazione di notebook interattivi combinando codice eseguibile in Python, R e Sql. Sono state testate librerie per *data science* (ad esempio Pandas, NumPy, Scikit-learn, TensorFlow), per l'epidemiologia (tra cui zEpid, EpiPy, surveillance, EnvStats) e per l'analisi geospaziale (come Geopandas, netCDF4, leaflet). Tra gli sviluppi previsti figura anche l'integrazione di pacchetti applicativi statistici proprietari, tra quelli a maggior diffusione, in configurazione

multiprocessore, che sono in fase di test. Inoltre, sono in fase d'implementazione strumenti di *dashboard* sia mediante l'uso di librerie di Python e R, sia mediante pacchetti applicativi *open source*, quale Apache Superset, o proprietari. Il progetto punta ad allinearsi con le direttive dello *European health data space* (Ehds), rendendo l'Italia un nodo centrale della rete HealthData@EU. Tra i possibili sviluppi futuri: l'implementazione di modelli predittivi di *epidemic intelligence* per l'allerta precoce su eventi sentinella legati a cambiamenti climatici; l'esplorazione dell'uso di *big data* da sensori per le misure d'esposizione e social media per monitorare la percezione del rischio; lo sviluppo di strumenti di *data storytelling* che permettano ai cittadini di consultare indicatori di salute e ambiente del proprio territorio in modo semplice e trasparente.

Conclusioni

La piattaforma Ideah rappresenta un passo fondamentale verso una gestione moderna e integrata della sanità pubblica. Superando la logica dei silos informativi, il progetto dota il sistema Snpa-Snps di uno strumento potente per comprendere i legami profondi tra la salute dell'ecosistema e quella dell'uomo. La sfida futura sarà garantire la sostenibilità di questa infrastruttura e la sua capacità di adattarsi alle rapide evoluzioni tecnologiche, mantenendo sempre al centro la sicurezza dei dati e il beneficio per la collettività.

Alessio Pitidis, Anna Carannante, Cecilia Fazio, Marco Giustini

Dipartimento Ambiente e salute,
Istituto superiore di sanità

RIFERIMENTI BIBLIOGRAFICI E SITOGRAFIA

Filippini T., Mecella M., Pitidis A., Giustini M., de Luzi F., Bianchini F., Macrì M., Marras M., Landi R., Vinceti M., "Ideah - Integrated database for environment and hHealth", *Eur J Public Health*, 2025 Oct 27, 35(Suppl 4), ckaf161.1349.doi: 10.1093/eurpub/ckaf161.1349. PMID: PMC12556613.

Harby A.A., Zulkernine F., 2025, "Data Lakehouse: A survey and experimental study", *Information Systems*, 127, 102460.

www.agendadigitale.eu/sanita/sistema-ideah-unire-salute-e-ambiente-per-prevenire-rischi/

<https://jupyter.org/hub>

www.snambiente.it/attivita/sistema-informativo-nazionale-ambientale/

www.copernicus.eu/en

<https://code.europa.eu/healthdataeu>

https://knowledge-base.inspire.ec.europa.eu/legislation/inspire-directive_en



IMMAGINE C. SCHNEIDER - UNSPLASH