



Anitec - Assinform

White Paper

Una Data Strategy per la Sanità Italiana

a cura del gruppo di lavoro
Digital Transformation in Sanità di Anitec-Assinform

Maggio 2022

ANITEC-ASSINFORM

Associazione Italiana per l'Information and Communication Technology

Tel. 02 00632801 - Fax. 02 00632824

Sede e uffici di Milano:
Via San Maurilio 21, 20123 Milano

Uffici di Roma:
Via Barberini 11 00187 Roma

segreteria@anitec-assinform.it www.anitec-assinform.it Aderisce a



CONFINDUSTRIA



CONFINDUSTRIA DIGITALE

Realizzato da:
Anitec-Assinform

Gruppo di lavoro “Digital Transformation in Sanità”

Accenture	Ifm
Adecco Formazione	Indra Minsait
ADS Automated Data Systems	Inmatic
Amazon Italia Service	Insiel
Fondazione Asphi	Itinera
BTO Research	Kelyon
Cefriel	Leonardo
Cisco Systems	Livemote
Commvault Systems Italia	Lutech
Confindustria Bari-BAT	Maggioli
Consorzio Netcomm	MaticMind
Copying	Mediterraneo Lab 4.0
Dassault Systemes Italia	Microsoft
Data 4 Services Italy	Mychicjungle
DataCore Software Italy	Netconsulting
Dxc Technology Italy	NTT ITALIA
Edicom	Ocra
Eris	Oracle Italia
Eustema	Present
Exprivia	Red Hat
Fracarro Radioindustrie	Reply
GPI	Sesa
Hewlett Packard Enterprise	TIM
Hiperforming Research	Tj Point
Hp Italy	VMware Italy
IBM Italia	While True

Coordinatore:
Fulvio Sbroiavacca (Insiel)

Revisione editoriale:
Ettore Russo e Stefania Follador

Sommario

Nota accompagnatoria	5
Executive Summary	6
1. Capitolo primo – Data Strategy	10
1.1 Dai Data Warehouse al Data Mesh.....	10
1.2 Il mercato dei Big Data in sanità e l'utilizzo e la gestione dei dati nelle aziende sanitarie	14
1.3 I dati come fattore abilitante per i processi di Intelligenza Artificiale e Machine Learning	16
2. Capitolo secondo – I dati sanitari: evoluzione normativa e il problema della sicurezza	20
2.1 Evoluzione normativa	20
2.1.1 Introduzione: definizioni e cenni alla normativa pre GDPR	20
2.1.2 Il General Data Protection Regulation (GDPR), principi generali e dati sanitari	21
2.1.3 Cenni sul Fascicolo Sanitario Elettronico, dalla sua genesi alle modifiche introdotte dal “Decreto Rilancio”	23
2.2 Dati sanitari e sicurezza: confidenzialità, integrità e disponibilità nell'Information Security	25
2.3 Dati sanitari e sicurezza: Il rischio cibernetico della forza lavoro distribuita in ambito sanitario	29
2.3.1 L'importanza di un approccio “Zero – Trust”	30
3. Capitolo terzo - Contesto EU e Gaia-X	34
3.1 Strategia europea sui dati.....	34
3.1.1 Common European Data Spaces e European Health Data Space	34
3.1.2 Data Act	36
3.2 Gaia-X: Data driven Innovation.....	37
4. Contesto nazionale	40
4.1 Fonti e flussi di dati sanitari	40
4.1.1 Flussi sanitari e flussi clinici	40
4.2 Necessità di un'infrastruttura	41
4.3 Piattaforma Digitale Nazionale Dati (PDND).....	43

5. Capitolo quinto – PNRR	44
5.1 Azioni del PNRR.....	44
5.2 Investimento 1.3 (M1C1) – Dati e interoperabilità	45
5.3 Investimento 1.3 (M6C2) – Rafforzamento dell’infrastruttura tecnologica e degli strumenti per la raccolta, l’elaborazione, l’analisi dei dati e la simulazione	46
6. Capitolo sesto – Data Governance	49
6.1 Data Access Architecture	50
6.2 Data Classification e Code Systems	51
6.3 HL7 FHIR®: lo standard per l’interoperabilità dei dati in sanità	52
7. Capitolo settimo – Life Data	54
7.1 Dati genomici e dati esogeni.....	55
7.2 Prospettive per il futuro: un modello dati per il territorio	56
8. Conclusioni – Raccomandazioni e proposte	58
8.1 Competenze	60
8.2 Cybersicurezza.....	62
8.3 Data Governance	63
8.4 Infrastruttura	64
8.5 Interdisciplinarietà e Life Data	65
9. Appendice – Best Practices	66
9.1 Mission Community Hospital – una singola piattaforma software unifica tutte le risorse di storage.	66
9.2 Uk Biobank – Enabling scientific discoveries that improve human health	67
9.3 Mayo Clinic – Clinical Data Platform	69
9.4 National Health Service del Galles (NHS Wales) – National Data Resource.....	70
9.5 Regione Lombardia – Alfabeto (ALI FASTER BETTER TOGETHER).....	72
9.6 Agenzia Regionale per l’innovazione e gli acquisti (ARIA SpA) – Digital Information Hub (DIH).....	74
9.7 AI Score – Ospedale San Raffaele: una piattaforma di apprendimento autonomo per consentire interventi sanitari mirati e tempestivi e ridurre l’impatto sul sistema sanitario del COVID-19	76
9.8 Swarm Intelligence – Unanimous AI	78

NOTA ACCOMPAGNATORIA

Salute e benessere sono determinanti per la nostra vita felice, il futuro ci pone la sfida della sostenibilità e della capacità di prevenire, come ben ci ha insegnato il Covid. Solo atteggiamenti individuali responsabili ci potranno garantire il futuro.

È giusto chiedersi: pensiamo di poterlo fare?

Le aree prioritarie di intervento individuate e i fattori abilitanti descritti nel documento rappresentano i pilastri su cui deve poggiarsi una Data Strategy per la Sanità del futuro.

EXECUTIVE SUMMARY

La pandemia degli ultimi due anni ha reso evidente il ruolo centrale delle tecnologie digitali nella società dell'informazione: mai come ora siamo coscienti dell'importanza di prendere decisioni di qualità, precise e consapevoli, in tempi rapidi. Oggi l'elemento fondante di qualsiasi decisione informata non può che essere il dato. Il nuovo paradigma dell'interconnessione tra sistemi fisici e digitali, delle analisi complesse in real-time che sfruttano i Big Data, dei Digital Twin e dell'IA, pone il dato al centro, e, in ultima istanza, queste tecnologie servono a prendere decisioni migliori. La crescita della nostra capacità di produrre e scambiare dati (come dimostra lo sviluppo dell'IoT) ci pone di fronte alla sfida di migliorare la nostra capacità di sfruttare i dati. Qualsiasi riflessione sul futuro del nostro modo di vivere e sulla qualità della vita nei prossimi decenni passa attraverso il confronto con il mondo dei dati.

Un'analisi di NetConsulting Cube sulla Digital Transformation in Sanità ha fotografato come le aziende sanitarie italiane stiano approcciando la transizione verso una sanità *data-driven*. Il 59% del campione indica che i dati sono già fruibili per l'uso secondario e il 29% dichiara che lo saranno in futuro. A oggi, inoltre, poco più della metà del campione (52%) ha in corso progetti di revisione dell'architettura dati, la maggior parte dei quali riguardanti la revisione della data governance (82%) e il rinnovo della data platform (55%)¹. L'importanza dei dati è confermata anche dagli importanti valori registrati dal mercato dei Big Data. Secondo una stima di Mordor Intelligence, il Big Data Healthcare Market valeva 23,7 miliardi di dollari USA nel 2020, con una previsione di crescita sino a 58,4 miliardi entro il 2026, registrando così una CAGR (Compounded Average Growth Rate) del 16,24% . Per quanto riguarda l'Italia, NetConsulting Cube stima un valore di 135 milioni di euro per il mercato Big Data e Analytics in Sanità per il 2022 con una prospettiva di crescita a 200 milioni nel 2025². Il tutto in un quadro dove la dinamica di spesa del settore Sanità digitale ha registrato una forte accelerazione raggiungendo un tasso di crescita quasi a doppia cifra nel 2021.

Il presente documento vuole essere uno stimolo da parte dell'industria ICT italiana, rappresentata da Anitec-Assinform, per l'elaborazione di una Data Strategy per la sanità del nostro paese e per lo sviluppo di strategie e di iniziative che portino il dato al centro delle azioni delle organizzazioni che operano nel settore sanitario.

Nella redazione di questo documento abbiamo deciso di concentrarci più sugli strumenti, i requisiti, che dovrebbero abilitare lo sviluppo di una sanità *data-driven* in Italia piuttosto che sulle tecnologie che si dovrebbero servire dei dati. Per

¹ NetConsulting Cube. Indagine campionaria sanità digitale 2021

² NetConsulting Cube 2022.

questo la trattazione è articolata su sette capitoli cui si aggiungono un capitolo conclusivo con delle raccomandazioni sintetiche per i *policymaker* e per le organizzazioni che operano in ambito sanitario e un'appendice con alcune best practice relative alla *data-driven healthcare*.

Il White Paper si compone di otto capitoli e di un'appendice in cui vengono raccolte alcune *best practice*. Il documento fa il punto del quadro complessivo della sanità digitale *data-enabled*, sia in riferimento allo stato dell'arte, sia per quanto riguarda gli sviluppi futuri. La trattazione parte dai modelli per la gestione dei dati e, passando per i contesti nazionale ed europeo, arriva al concetto di *life data*. La sanità del futuro avrà bisogno della trasversalità dei dati per migliorare la qualità della vita delle persone.

Nel primo capitolo descriviamo alcuni modelli per la gestione dei dati nel settore della sanità, partendo dalla *data warehouse* fino al nascente paradigma del data mesh, Il capitolo è poi arricchito con dati di mercato e con cenni all'attività svolta nelle AO italiane sulla gestione dei dati.

Il secondo capitolo affronta il tema prismatico della sicurezza dei dati e i principali profili normativi affini al tema del White Paper.

Terzo e quarto capitolo fanno riferimento rispettivamente al contesto *macro* Europeo (con un riferimento speciale a Gaia-X) e nazionale.

Il quinto capitolo offre una disamina sintetica delle azioni del Piano Nazionale di Ripresa e Resilienza che toccano più da vicino il tema della *Data Strategy* (Dati e Interoperabilità e Rafforzamento della struttura tecnologica per l'analisi dei dati del SSN).

Il sesto capitolo fa riferimento ai temi della Data Governance come l'architettura d'accesso e il catalogo dei dati.

Il settimo capitolo, infine, approfondisce il tema emergente dei *life data* tradizionalmente i sistemi sanitari si sono concentrati sulla gestione degli episodi clinici relativi allo stato di malattia e sui dati da essi generati; nel futuro però si dovrà sempre di più tenere conto di fattori che esulano dagli episodi clinici come patrimonio genetico, condizioni economiche, ambientali, abitudini e stili di vita.

Il capitolo conclusivo sintetizza il messaggio del documento veicolando cinque raccomandazioni per i *policymaker*:

Le aree di policy individuate nel documento sono: Competenze, Cybersecurity, Data Governance, Infrastruttura, e Interdisciplinarietà e Life Data.

- **Competenze**

Il tema delle competenze rappresenta un nodo cruciale per il successo dei programmi di trasformazione digitale per il Sistema Paese, incluse Pubblica Amministrazione e Sanità. Per una vera *digital transformation* bisogna agire su due fronti: aumentare le competenze di base nella popolazione e rinforzare le competenze informatiche specialistiche nell'SSN.

- **Cybersecurity**

Il dato sanitario è per sua natura sensibile. La Cybersecurity è da considerare come componente essenziale di qualsiasi processo di trasformazione digitale in ambito sanitario. È fondamentale nel tradurre la Data Strategy in progetti concreti, la sicurezza dei dati rivesta un ruolo centrale, strutturando servizi e infrastrutture per contrastare potenziali attacchi informatici, soddisfacendo i requisiti normativi e garantendo la privacy di operatori e pazienti. Ogni progetto di digitalizzazione dei processi sanitari si deve sviluppare con una logica di “Security by Design” che contempli tutti gli elementi che compongono la catena di erogazione del servizio: in primis organizzazione, sicurezza delle apparecchiature e competenze dei dipendenti.

- **Data Governance**

A differenza dei dati amministrativi già organizzati e semanticamente standardizzati per verificare i costi del SSN, i dati clinici esistono ma sono molto frammentati e raccolti in modalità estemporanee. Oggi non è possibile fruire di tutto il valore dei dati clinici, sia per quanto riguarda la ricerca scientifica, sia per la pratica clinica. È urgente normalizzare le prassi di utilizzo dei dati partendo dalla standardizzazione semantica dei dati clinici per arrivare a un Common Data Model nazionale. Il modello HL7-FHIR è un ottimo punto di partenza.

Abbiamo bisogno di norme sulle possibilità di utilizzo e scambio dei dati per capitalizzare i buoni risultati ottenuti dai progetti abilitati dal Deep Learning, dall'Intelligenza Artificiale e dal Machine Learning, deve passare dall'accordo sulle possibilità normative e legittime di utilizzo e scambio di dati. Scalare a livello nazionale le esperienze già fatte di IA, accresce il loro utilizzo su larga scala e permette di agevolare il dialogo con il garante della privacy portando nel confronto il valore ottenuto di predittività in termini di efficienza ed efficacia nella pratica clinica.

- **Infrastruttura**

Una *Data Driven Healthcare* presuppone l'esistenza di un'infrastruttura capillare che consenta la raccolta, il trasferimento, l'elaborazione e la fruizione dei dati. L'infrastruttura deve abilitare l'interoperabilità dei dati attraverso una strutturazione logica degli stessi: un modello semantico che a partire dallo

standard FHIR consenta la comprensione del significato dei dati contenuti per tutti i possibili utilizzi.

È suggerita l'adozione di linee guida comuni, modelli di maturità, che aiutino le organizzazioni sanitarie a dotarsi di infrastrutture adeguate e affidabili che, una volta realizzate, dispongano di strumenti che permettano un attento monitoraggio funzionale a vantaggio della qualità d'erogazione dei servizi destinati all'utenza.

- **Interdisciplinarietà e Life Data**

Si deve passare da una logica di uso verticale dei dati finalizzata al singolo progetto, a una strategia che valorizzi standard, collaborazione e riuso dei dati in senso trasversale, multidisciplinare, verso il paradigma di Life Data, cioè sfruttare tutti i dati che possono influenzare il benessere, le nostre condizioni di vita.

I Life Data consentono di utilizzare algoritmi, Data Analytics, Intelligenza Artificiale (con particolare utilizzo di Machine Learning, Deep Learning, Chatbots) per valorizzare queste informazioni, garantire una visione complessiva ma nello stesso tempo altamente dettagliata e specialistica per ogni necessità di approfondimento di cui chi partecipa al processo di cura ha bisogno in ogni momento. In definitiva abilitando la medicina personalizzata o di precisione fino ad arrivare a tecnologie come i digital twin personali.

1. CAPITOLO PRIMO – DATA STRATEGY

Per delineare una "**Strategia Dati**" adeguata al compito di produrre valore, facendo leva sugli standard disponibili, sulla collaborazione e sul riuso dei dati in senso trasversale e multidisciplinare, occorre tenere presenti diversi elementi.

Bisogna, in primis, guardare al recente passato per non commettere negli stessi errori o adottare modelli già superati. Occorre poi considerare la situazione presente, partendo dalla strategia di interoperabilità dei dati definita a livello nazionale e che si basa sulla definizione di un'**Agenda Nazionale Dati (AND)**, sulle **Linee guida e del Modello di Interoperabilità (ModI)** e sull'attuazione della **Piattaforma Digitale Nazionale Dati (PDND)**, che renderà possibile l'interoperabilità dei sistemi informativi e delle basi di dati delle pubbliche amministrazioni e dei gestori di servizi pubblici.

Infine, come sarà illustrato nei successivi capitoli, non si può prescindere dalle **normative** e dai vincoli di riservatezza e protezione dei dati sanitari come quelli definiti dal **GDPR**, e in generale dalla cornice normativa europea

La Strategia Dati dovrà necessariamente coniugare modelli e piattaforme dati efficaci e di nuova generazione con modelli e piattaforme di interoperabilità per creare un ecosistema nazionale, e auspicabilmente europeo, per la "produzione e il consumo" dei dati di rilevanza sanitaria (dati clinici, economici, statistici, etc...).

1.1 Dai Data Warehouse al Data Mesh

Per quanto riguarda i modelli e le piattaforme dati, possiamo sicuramente affermare che molte realtà, in quasi tutti i settori industriali inclusa la sanità, hanno investito in **Data Warehouse** e stanno investendo in **Data Lake** con l'obiettivo di democratizzare, su larga scala, i dati in proprio possesso e renderli disponibili a potenziali "fruitori" interni o esterni alle organizzazioni.

L'Enterprise Data Warehouse, il monolite dei dati per antonomasia, ha rappresentato nel recente passato l'idea di un'unica fonte della verità per la creazione di "conoscenza" generalmente basata su un insieme comune di "aree tematiche" e domini di dati. Nella pratica, ciò che la maggior parte delle organizzazioni ha in genere attuato è una pletora di diversi Data Warehouse e Data Mart, spesso allineati alle esigenze di particolari Line of Business o gruppi di lavoro, che creavano in effetti dei "silos" piuttosto che aree di condivisione.

Gli indubbi successi del passato di questa prima generazione sono stati in parte oscurati da evidenti limiti come: la rigidità del modello dati, la staticità dei dati (spesso alimentati via batch), gli elevati costi di manutenzione (spesso superiori ai benefici ottenibili), l'incapacità di incamerare grandi volumi di dati, grezzi e

destrutturati da utilizzare, ad esempio, in combinazione con i dati strutturati e certificati, per addestrare modelli di analisi predittiva.

Questa **prima generazione** di Data Warehouse aziendali, spesso messi in campo su piattaforme proprietarie e soluzioni dagli elevati costi, hanno lasciato le aziende con una quantità altrettanto elevata di debito tecnologico. Questo è dovuto, nella pratica, a migliaia di job ETL, tabelle e report difficilmente manutenibili e comprensibili solo a un ristretto gruppo di persone specializzate.

Con la seconda generazione è stato introdotto il concetto di **Big Data** e di **Data Lake**. I Data Lake possono essere utilizzati principalmente per archiviare dati grezzi (raw data) e non elaborati che provengono direttamente dai vari sistemi dell'organizzazione e da fonti di dati esterne all'organizzazione.

Si tratta, come per il Data Warehouse, di un'architettura centralizzata caratterizzata da sistemi basati su un repository di dati archiviati nel loro formato naturale (non elaborato) con lo scopo di raccogliere, tipicamente in batch, tutte le informazioni disponibili presso l'organizzazione. Le informazioni possono provenire da diverse sorgenti, e avere strutture dati anche molto differenti fra loro. Un Data Lake può infatti includere dati strutturati provenienti da database relazionali, dati semi-strutturati (CSV, log, XML, JSON), dati non strutturati (e-mail, documenti, PDF) e dati binari (immagini, audio, video).

L'obiettivo è quello di aumentare la capacità di gestire dati eterogenei, in grandi volumi e in parte alimentati in **real-time**, superando i limiti dell'**Enterprise Data Warehouse**.

Con un approccio tipicamente "IT centrico", vengono individuati tutti le possibili sorgenti dati di interesse, si effettua un "caricamento" di questi dati che quindi vengono "puliti", "arricchiti" e "trasformati" in "data set" in grado, auspicabilmente, di soddisfare le esigenze di un insieme diversificato di "consumatori".

Spesso i potenziali "consumatori" dei dati non sanno che un "data set" esiste, o non sono in grado di interpretarne correttamente la semantica, o non lo usano a sufficienza, oppure l'IT, che non conosce le reali esigenze dei "data consumer", ha dedicato troppo tempo a cercare di renderlo disponibile per renderlo realmente utile.

Per ridurre il gap e accorciare la filiera tra "consumatore" e "produttore" si è affermato negli ultimi anni un nuovo approccio di ultimissima generazione denominata **Data Mesh**.

Il **Data Mesh** presuppone un cambio di paradigma sulla base del quale collezionare e rendere disponibili i dati. In letteratura esistono diverse descrizioni e declinazioni del concetto di **Data Mesh** (coniato e descritto per la prima volta nel 2019 da Zhamak Dehgani) e il presente documento non ha l'ambizione di

descriverlo dettagliatamente. In estrema sintesi possiamo dire che le strategie dati e le architetture fondate sul modello **Data Mesh** si basano generalmente sulla convergenza di alcuni principi base, alcuni dei quali imprescindibili che proveremo a descrivere brevemente:

- **Dati pensati come Prodotti:** il dato deve essere trattato come un "prodotto" che in quanto tale assolve uno specifico fabbisogno nel contesto di "domini" definiti. Diventa essenziale il ruolo di **domain data product owner** il quale deve permettere che i dati di uno specifico dominio vengano progettati e consegnati come "prodotto" secondo la teoria del "*job to be done*", adottando tutte le misure per garantire la qualità dei dati, la riduzione dei tempi di consegna ma più in generale, la piena soddisfazione delle necessità dei "data consumer". Il **domain data product owner** deve avere una profonda comprensione di chi sono gli utenti dei dati del dominio, come questi ultimi utilizzano i dati e quali sono i metodi nativi (interfacce) più congeniali per il loro utilizzo (consumo). Tale conoscenza approfondita di chi consuma i dati di un dominio si traduce nella progettazione di interfacce di "prodotti dati" (Data as a Product) che soddisfano a pieno le loro esigenze.
- **Dati Decentralizzati per Dominio:** seguendo un approccio di Domain Driven Design (DDD), al fine di decentralizzare la piattaforma dati affrancandosi da approcci monolitici, si deve invertire il modo in cui pensiamo ai dati, alla loro localizzazione e alla loro proprietà. Invece di trasferire i dati dai domini di origine a piattaforma centralizzata, i domini di origine devono ospitare e offrire i propri data set di dominio in modo facilmente consumabile.
- **Dati in movimento (Event Driven):** un approccio data driven presuppone di vedere i dati, generati o modificati dai diversi sistemi sorgente, come uno stream continuo di dati in movimento (stream di data event). Il valore che assumeranno una volta confluiti nei diversi data set di dominio sarà tanto superiore rispetto a quelli generati da batch periodici quanto è maggiormente rilevante la loro dimensione temporale (real time data vs batch data) nel contesto del caso d'uso per cui sono stati progettati. **Event-Driven Data Mesh** è quindi un'architettura dati event-centrica e in molti contesti è lo stesso stream di eventi a caratterizzare il "data product" da consumare.
- **Self-Service:** principio che serve ad abilitare l'autonomia dei domini attraverso una progettazione domain-oriented design che consente agli utenti di astrarre la complessità tecnica e concentrarsi sui singoli casi d'uso.

- **Data Infrastructure as a Platform:** per consentire a tutti i Domini di assumersi la piena responsabilità del trattamento dei dati fornendo gli strumenti tecnologici necessari.
- **Interoperabilità e standardizzazione delle comunicazioni:** Alla base di ogni dominio c'è un insieme universale di standard di dati che aiuta a facilitare la collaborazione tra i domini quando necessario, e spesso lo è. È inevitabile che alcuni dati (sia fonti grezze che set di dati puliti, trasformati e serviti) saranno preziosi per più di un dominio. Per consentire la collaborazione tra domini, una Data Mesh deve standardizzare i campi di formattazione, la governance, l'individuabilità e i metadati. Inoltre, ogni dominio di dati deve definire e concordare SLA e misure di qualità da "garantire" ai propri consumatori.

In sintesi, una strategia dati basata sul **Data Mesh** sposa il concetto di **dato federato e distribuito** in specifici domini che hanno la piena responsabilità di fornire tali dati come "prodotti", facilitando al contempo la comunicazione e l'interoperabilità tra dati distribuiti in diverse posizioni.

Mentre l'infrastruttura dati è responsabile di fornire a ciascun dominio le soluzioni con cui elaborarli, i domini hanno il compito di gestire l'inserimento, la pulizia, l'aggregazione, l'arricchimento dei propri dati di dominio per generare "asset" che possano essere utilizzati efficacemente dai "data consumer" (siano essi applicazioni, data science, sistemi analitici etc...).

Il Data Mesh è una strategia dati che abbraccia l'ubiquità dei dati sfruttando un design self-service e domain-oriented. Una Data Mesh supporta "data consumer" distribuiti e specifici del dominio fornendo "data-as-a-product", con ogni dominio che gestisce le proprie pipeline di dati in piena autonomia. Il tessuto che collega questi domini e le relative risorse di dati associate si deve, infine, basare su un **"layer" di interoperabilità** universale caratterizzato da standard e sintassi comuni.

Dal punto di vista macroscopico una **strategia dati** basata su **Data Mesh** può essere efficacemente rappresentata dalla seguente figura.

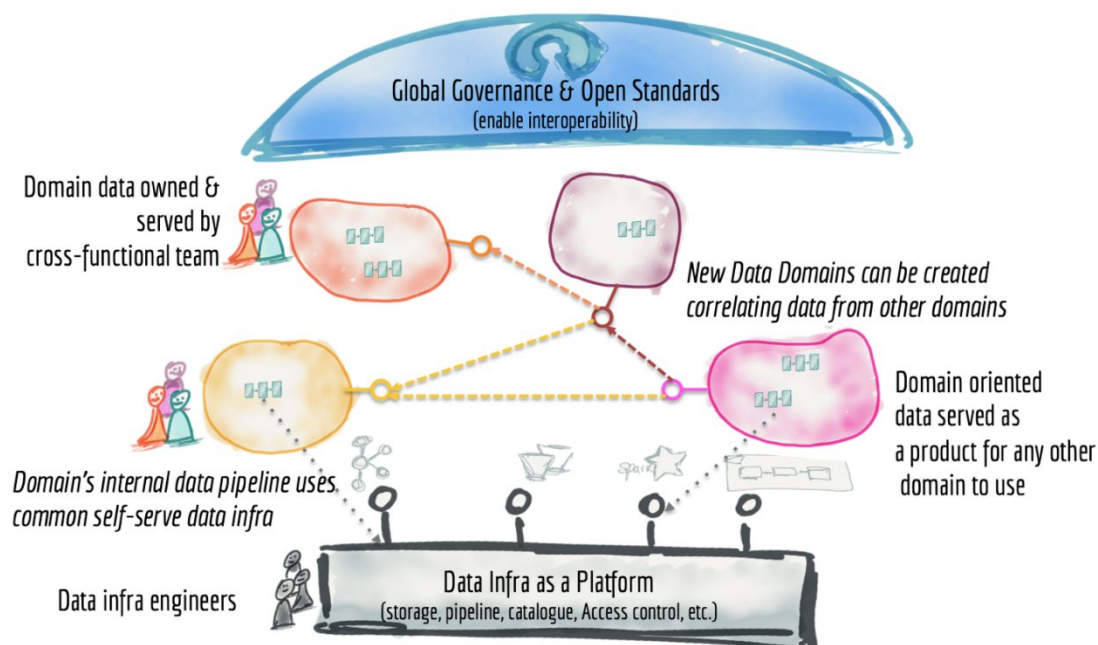


Figura 1. Schematizzazione di strategia dati basata su Data Mesh. Fonte: <https://martinfowler.com/articles/data-monolith-to-mesh.html>

In questa rappresentazione ad alto livello risulta facile trovare una evidente analogia e coerenza con quanto definito nelle **“Linee Guida sull’infrastruttura tecnologica della Piattaforma Digitale Nazionale Dati per l’interoperabilità dei sistemi informativi e delle basi di dati”** responsabili della definizione delle logiche di governance, degli standard da adottare e dell’infrastruttura a supporto (catalogo, access controll etc...), le quali dovranno probabilmente evolvere per svolgere anche le funzionalità di interoperabilità asincrona (storage, pipeline).

1.2 Il mercato dei Big Data in sanità e l’utilizzo e la gestione dei dati nelle aziende sanitarie

L’importanza dei Big Data nel mondo della sanità è confermato da i dati di mercato. Secondo una stima di Mordor Intelligence il *Big Data Healthcare Market* valeva 23,7 miliardi di dollari USA nel 2020 con una previsione di crescita sino a 58,4 miliardi entro il 2026, registrando così una CAGR (Compounded Average Growth Rate) del 16,24%³. Per quanto riguarda l’Italia, NetConsulting Cube stima un valore di 135 milioni di euro per il mercato Big Data e Analytics in Sanità per

³ Mordor Intelligence, Big Data Healthcare Market – Growth, Trends, COVID-19 impact and forecasts (2022-2027). Estratto disponibile online: <https://www.mordorintelligence.com/industry-reports/big-data-healthcare>

il 2022 con una prospettiva di crescita a 200 milioni nel 2025⁴. Il tutto in un quadro dove la dinamica di spesa del settore Sanità digitale ha registrato una forte accelerazione raggiungendo un tasso di crescita quasi a doppia cifra nel 2021. Mentre, complessivamente, secondo il rapporto annuale “Il Digitale in Italia”, il valore complessivo del mercato italiano Big Data è stimato a 1,5 miliardi di euro per il 2022 e dovrebbe raggiungere gli 2,1 miliardi nel 2025 (Tasso di crescita medio annuo 2021/25 del 13,1%)⁵.

È particolarmente interessante andare oltre le rilevazioni di mercato e focalizzarsi su come le aziende sanitarie italiane stanno approcciando la transizione verso una sanità *data-driven*. Un’analisi di NetConsulting Cube ha fatto luce sull’utilizzo e la gestione dei dati nelle aziende sanitarie italiane⁶. Per quanto riguarda la raccolta e la gestione dei dati, il 71% delle aziende dichiarano che i dati sono collocati all’interno dei sistemi che li generano e non sono condivisibili. Allo stesso tempo il 59% del campione dichiara che i dati sono già fruibili per l’uso secondario e il 29% che lo saranno in futuro. Risultano poco diffusi data lake/data platform di struttura o di gruppo: il 12% del campione dichiara di avvalersene e il 35% di progettarne l’adozione futura, il restante 53% del campione non utilizza tali risorse né ne pianifica l’utilizzo in futuro (Fig 3.)⁷.

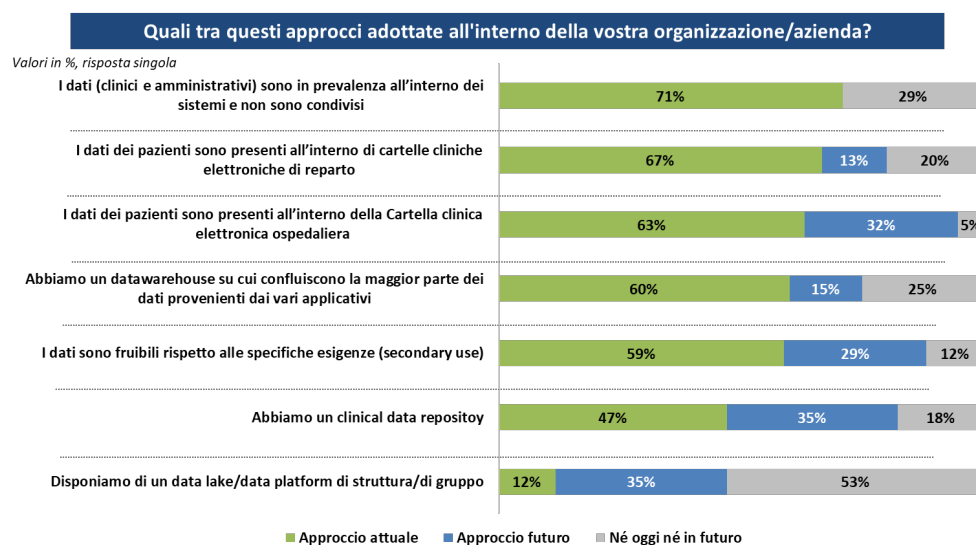


Figura 3. Approcci adottati per la raccolta e la gestione dei dati, fonte: NetConsulting cube, 2021

⁴ NetConsulting Cube 2022.

⁵ NetConsulting Cube 2022.

⁶NetConsulting Cube. Indagine campionaria sanità digitale 2021

⁷ Ibidem

Una seconda dimensione di analisi ha riguardato la revisione dell'architettura dati. Il 52% del campione ha in corso progetti di revisione: di questi, l'82% riguarda la revisione della data governance e il 55% il rinnovo dell'architettura della *data platform*; molto meno diffusi i progetti di introduzione del data catalogue (9% del campione) (Figura 4)⁸.

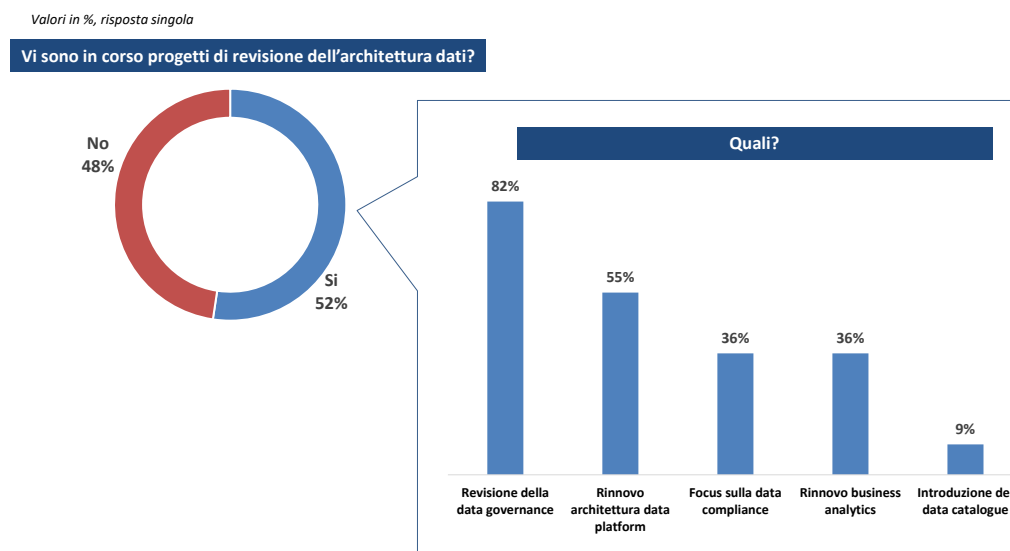


Figura 4. Revisione dell'architettura dati, fonte: NetConsulting cube, 2021

1.3 I dati come fattore abilitante per i processi di Intelligenza Artificiale e Machine Learning

Il comparto della sanità ha una base di dati enorme che si presta a essere trattata, e sfruttata, attraverso sistemi di intelligenza artificiale. Mai come nel recente periodo di pandemia si è osservato da un lato il valore dei dati e dall'altro la loro attuale frammentazione e difficoltà di utilizzo.

Secondo l'Annuario Statistico del Servizio Sanitario 2017, in Italia è stato prodotto quasi un miliardo di prestazioni di Laboratorio Analisi, quasi 54 milioni per la Radiologia Diagnostica, per citare due fra le fonti più dense⁹. È indubbio che l'informazione diagnostica, che naturalmente ha un valore in sé nella cura del singolo, contiene un potenziale significativo di rappresentazione della popolazione. In altri termini, può abilitare conoscenza aumentata della salute di gruppi estesi di persone che vivono in un territorio. I dati diagnostici, possono

⁸ Ibidem

⁹ [Annuario Statistico del Servizio Sanitario Nazionale - Assetto organizzativo, attività e fattori produttivi del SSN - Anno 2017 \(salute.gov.it\)](#). Pag 22.

abilitare, insieme a fonti già disponibili come i flussi NSIS, modelli di popolazione che caratterizzano la salute.

L'IA ha bisogno di dati, soprattutto se si finalizza ad applicazioni che supportino i processi di diagnosi e cura. Un esempio di applicazione dell'IA è quello dell'epidemiologia, specie considerando che questa inizia a virare verso il concetto di "Augmented Epidemiology", spostandosi, cioè, dal tradizionale approccio retrospettivo per concentrarsi invece sull'abilitazione di strumenti di sorveglianza e prevenzione. È opportuno sottolineare come importanti aree della ricerca clinica quali Precision Medicine, scienze omiche e trial clinici si servano di dati e possano venire supportati dall'aumento dell'informazione.

I dati quindi permettono lo sviluppo di applicazioni. In questo contesto il grande progresso del settore dell'IoT ha permesso di fare grandi passi in avanti alle applicazioni dell'IA in ambito sanitario. Difatti, stiamo assistendo alla nascita dell'Internet of Medical Things (IoMT), in cui i dispositivi medici e le applicazioni sono collegati ai sistemi IT sanitari tramite il Web e i dispositivi abilitati Wi-Fi possono facilitare la comunicazione machine-to-machine (M2M) e il collegamento a piattaforme cloud per l'archiviazione dei dati. L'IoMT include un panorama di dispositivi indossabili, sistemi di monitoraggio remoto dei pazienti, letti ospedalieri dotati di sensori e pompe per infusione, sistemi di monitoraggio dei farmaci, monitoraggio delle scorte mediche e delle apparecchiature e altro ancora. In questo scenario, i dispositivi indossabili connessi abilitati dall'IoMT possono migliorare l'affidabilità delle diagnosi e le prestazioni temporali, consentendo la raccolta dei dati per l'analisi e di conseguenza un'adattabilità più reattiva alle fluttuazioni delle condizioni. Il progresso dell'IoMT e delle tecnologie mobili è un vettore cruciale per l'evoluzione dei sistemi sanitari nel percorso di prevenzione, diagnosi precoce e cura. È poi fondamentale la trasparenza nelle decisioni prese. Una prospettiva di sviluppo è quella di associare una spiegazione semantica per arricchire le previsioni insieme ai dati sanitari mediante tecnologie del Explainable AI.

L'Intelligenza Artificiale può essere molto utile nella struttura sanitaria anche per migliorare le performance operative, in modo non dissimile dagli impieghi che si hanno in un'azienda normale per migliorare i processi. In questo senso, tecnologie di Natural Language Processing e Text Analytics permettono di analizzare automaticamente grandi quantità di dati non strutturati. Oggi molti fornitori tecnologici offrono servizi pre-allenati per il riconoscimento di entità nel campo sanitario, invocabili tramite APIs: in questo modo, sarà possibile utilizzare le preziose informazioni presenti in forma non strutturata per migliorare performances operative e cliniche.

In definitiva, l'evoluzione dettata dall'emergenza sanitaria impone un cambiamento, in cui l'Intelligenza artificiale può rappresentare il fulcro di un nuovo modello di sostenibilità per economia e società, raggiungendo una

consapevolezza di contesto aiutando gli *stakeholder* del settore sanitario a fornire preziosi suggerimenti nel processo decisionale.

Rispetto alle tre direttrici di evoluzione della Sanità, personalizzazione, proattività e delocalizzazione, l'Intelligenza Artificiale giocherà un ruolo crescente, facendo leva su data spaces (Gaia X), su apparecchiature consumer e su sistemi di intelligenza distribuita.

In particolare:

Personalizzazione: il singolo paziente avrà il suo “data space” in cui si aggregano sia i dati sanitari (esami, medicine, terapie/interventi, protesi, ...), sia il suo profilo (ad esempio genoma, e in futuro gli “omics”, statura, peso, fattori ereditari, alimentazione, attività fisica, s...) sia fattori ambientali (situazione abitativa, familiare, lavorativa, hobby...). Questi dati rappresentano la base di partenza e la “storia della persona”. Ciò permette a sistemi di intelligenza artificiale di analizzare i dati attuali di confrontarli con il pregresso, creando un profilo che non solo si aggiorna costantemente ma che diventa sempre più preciso nelle interpretazioni dei dati e nel suggerimento di azioni.

Entro fine decade ci si può attendere che l'insieme dei dati così come parte dell'intelligenza applicata ai dati faccia parte del Personal Digital Twin -PDT- della persona. Sarà il PDT che si farà carico di rendere comprensibili alla persona lo stato di salute, la necessità di approfondimenti e il livello di rischio conseguente a certi comportamenti. Sarà il PDT che assicurerà la condivisione dei dati sulla base di un contesto di riferimento in parte imposto dalle istituzioni e in parte definito dalla persona stessa (o da chi ne ha la responsabilità).

L'erogazione dei servizi sanitari si interfacerà con il PDT sia per la acquisizione dei dati sia per la personalizzazione dei servizi. Un ruolo fondamentale della intelligenza artificiale, in prospettiva, sarà la correlazione del genotipo al fenotipo, sia per arrivare a protocolli di cura sia per interpretare i sintomi nel contesto specifico della persona.

Proattività: anticipare significa, da un lato, diminuire i problemi alla persona, e, dall'altro, diminuire significativamente i costi. L'intelligenza artificiale, facendo leva sui dati della persona (storia e stato attuale) e analizzandoli nel contesto ambientale, può fornire sia informazioni sul rischio (awareness) sia anticipare l'esistenza di problemi mentre si è ancora in una fase asintomatica. Questo permette di iniziare protocolli di cura meno impattanti e meno costosi, spesso gestibili al di fuori di strutture sanitarie.

Da notare come le nuove tecnologie di IA come le GAN -Generative Adversarial Networks-, consentono di sviluppare IA a partire da basi di dati limitate, quali quelle relative a un singolo paziente, attraverso utilizzo di serie di dati storici,

generazione di dati sintetici e simulazione, aspetto ques'ultimo di notevole importanza per la decisione del protocollo di cura.

Delocalizzazione: l'accesso a strutture sanitarie comporta costi e in generale un impatto maggiore sulla qualità di vita del paziente. La tecnologia permette di erogare prestazioni sanitarie in maniera delocalizzata, avvalendosi di sistemi di monitoraggio, diagnosi e intervento a partire dalla persona, dalla abitazione, da centri distribuiti di primo intervento fino ad arrivare a strutture sanitarie (ospedali) che condividono risorse attraverso la rete. L'intelligenza artificiale rappresenta un importante abilitatore nella gestione delle risorse delocalizzate e un sistema efficace di supporto alla comunicazione. L'utilizzo crescente dei chatbots nella interazione con il paziente, già iniziato in alcune realtà e in forte espansione ne è un esempio. Il valore della "consulenza" attraverso il chatbot non è solo nella sua immediata disponibilità 24/7 ma anche nel fatto che questo, attraverso IA e tecniche ML diventa sempre più "legato" a quel particolare paziente, conoscendone la storia attraverso connessione con il suo PDT e attraverso le interazioni che ha con questo. In particolare, le interazioni vocali permettono di costruire la digital signature del paziente e di intercettare variazioni rispetto alla norma (il permette di rilevare potenziali emergenze di patologie, incluso infezione da Covid). Il chatbot, inoltre, fa parte della conoscenza del sistema complessivo, contribuendo ai dati e ricevendo informazioni di contesto che aiutano a meglio analizzare i dati del paziente.

2. CAPITOLO SECONDO – I DATI SANITARI: EVOLUZIONE NORMATIVA E IL PROBLEMA DELLA SICUREZZA

2.1 Evoluzione normativa

2.1.1 Introduzione: definizioni e cenni alla normativa pre GDPR

I dati sanitari sono considerati nella normativa italiana e in quella europea come una categoria particolare dei dati personali. La loro definizione più moderna e puntuale arriva dal Regolamento (UE) 679/2016 sulla protezione dei dati personali – il GDPR –, ivi i dati sanitari sono definiti come *dati personali attinenti alla salute fisica o mentale di una persona fisica, compresa la prestazione di servizi di assistenza sanitaria, che rivelano informazioni relative al suo stato di salute*¹⁰.

Grazie alle potenzialità offerte dall'innovazione tecnologica, i dati sanitari stanno acquisendo sempre maggiore importanza nel mondo contemporaneo. Ciò è dovuto alle loro molteplici applicazioni: i dati sanitari trovano impieghi che vanno dalla ricerca medica e farmaceutica e arrivano alla gestione dei servizi sanitari pubblici e privati. È chiaro che il tema dei dati sanitari intreccia più interessi costituzionalmente garantiti, fra tutti il diritto alla riservatezza e il diritto alla salute, e ne rende imperativo un bilanciamento. A un quadro di base già complesso si aggiunge il problema della sicurezza, i nuovi sistemi informativi e le banche dati digitalizzate permettono una gestione più efficiente dei dati, ma allo stesso tempo sono esposte sempre di più ad attacchi informatici che mettono a rischio sia la privacy degli individui sia la tutela della salute pubblica. Poiché in dialettica con una tecnica in continua evoluzione e con i molteplici interessi, di natura sia individuale che collettiva, la materia della protezione dei dati sanitari è stata definita un “diritto inquieto”¹¹. Per queste ragioni è utile per il presente documento di posizione accennare al contesto normativo che governa la materia di cui si tratta.

Già nei vent'anni precedenti all'emanazione del GDPR la materia dei dati sanitari aveva iniziato a emergere all'interno delle prime norme sul tema della riservatezza nell'orientamento italiano. La Legge n. 675 del 31 dicembre 1996, legge capostipite in materia privacy in Italia¹², offriva un primo inquadramento dei sanitari definendoli come una categoria di dati sensibili i quali, a loro volta,

¹⁰ Regolamento (UE) 679/2016, art. 4 n. 15.

¹¹ Rapporto ISS Covid-19, Protezione dei dati personali nell'emergenza COVID-19.

¹² Iusinitinere, I dati sanitari e l'evoluzione normativa nazionale e comunitaria fino all'introduzione del Regolamento UE 2016/679

considerati come un sottoinsieme bisognoso di tutele speciali della più ampia nozione di dati personali¹³. Successivamente, il D. Lgs n. 196 del 30 giugno 2003 – il cosiddetto Codice Privacy – arriva a dedicare un intero titolo – il quinto – ai dati sanitari. È interessante rilevare come l'elemento legittimante per il trattamento dei dati nella formulazione originaria del Codice fosse il consenso dell'interessato; la regola generale – della quale comunque sussistevano diverse deroghe – prevedeva che i dati personali potessero essere trattati solo con il consenso dell'interessato¹⁴. Questa impostazione verrà superata dal GDPR il quale annovera il consenso come una di diverse fonti legittimanti del trattamento dei dati (l'articolo 6 indica come **fondamenti di liceità**: il consenso, la salvaguardia di interessi vitali della persona interessata o di terzi; l'adempimento di obblighi di legge cui è soggetto il titolare; l'interesse pubblico e l'esercizio di pubblici poteri; il perseguimento del legittimo interesse prevalente del titolare o di terzi cui i dati vengono comunicati; l'esecuzione di un contratto di cui l'interessato è parte).

2.1.2 Il General Data Protection Regulation (GDPR), principi generali e dati sanitari

La cornice normativa nella quale si inserisce il tema dei dati personali fa riferimento in massima parte al GDPR. Un'analisi completa del Regolamento andrebbe ben al di là dei fini di questo documento. Qui ci si concentrerà su alcuni passaggi particolarmente rilevanti per la materia dei dati sanitari: i principi generali relativi al trattamento dei dati personali, il divieto di trattamento dei dati sanitari e le eccezioni a esso esistenti, la sicurezza dei dati attraverso crittografia e pseudoanonimizzazione.

L'art. 5 del Regolamento individua alcuni principi generali relativi al trattamento dei dati personali¹⁵:

- a. principi di **liceità, correttezza e trasparenza**;
- b. limitazione della **finalità**: i dati devono essere raccolti per finalità determinate, esplicite e legittime;
- c. principio di **minimizzazione**: i dati sono pertinenti e limitati rispetto alle finalità per le quali sono trattati;
- d. principio di **esattezza e aggiornamento**: i dati devono essere esatti e aggiornati, dati inesatti vanno rettificati o cancellati tempestivamente;
- e. limitazioni alla **conservazione**: i dati devono essere conservati in una forma che consenta l'identificazione degli interessati **per il tempo**

¹³ Ibidem.

¹⁴ [Trattamento dei dati sanitari, alla luce del GDPR: il quadro normativo - Cyber Security 360](#)

¹⁵ Rapporto ISS Covid-19, Protezione dei dati personali nell'emergenza COVID-19.

strettamente necessario al conseguimento delle finalità per le quali sono trattati

- f. principio di **integrità e riservatezza**: bisogna garantire un'adeguata sicurezza dei dati;
- g. principio di **responsabilizzazione**: il titolare del trattamento è responsabile dei principi appena esaminati ed è in grado di provarlo. Il titolare mette in atto le misure tecniche e organizzative adeguate a garantire, ed essere in grado di dimostrare, che il trattamento è effettuato conformemente al Regolamento (art. 24, paragrafo 1).

È importante evidenziare che il GDPR **vieta**, in via generale, il trattamento di alcune categorie di dati personali, tra le quali i dati relativi alla salute (art. 9). Sono però presenti delle eccezioni che è utile passare in rassegna:

- 1. Il divieto al trattamento dei dati sanitari non si applica per:
 - a. motivi di interesse pubblico rilevante;
 - b. motivi di interesse pubblico nel settore della sanità pubblica;
 - c. finalità di medicina preventiva, diagnosi, assistenza o terapia sanitaria o sociale e gestione dei sistemi e servizi sanitari o sociali
- 2. Non è richiesto il consenso per il trattamento dei dati sanitari per i trattamenti essenziali per finalità determinate e connesse alla cura della salute ed effettuati da un professionista sanitario soggetto al segreto professionale (o altra persona soggetta all'obbligo di segretezza).
- 3. Il trattamento dei dati sanitari è possibile **esclusivamente** con il consenso dell'interessato nei casi di:
 - a. consultazione del **Fascicolo sanitario elettronico**;
 - b. consegna del referto online;
 - c. utilizzo di app mediche;
 - d. fidelizzazione della clientela;
 - e. finalità promozionali o commerciali;
 - f. finalità elettorali.

Il GDPR prevede anche disposizioni relative alle tecniche volte a garantire la sicurezza dei dati. Principalmente si può fare riferimento a:

- 1. **la pseudonimizzazione** dei dati personali, questa è stata definita come "una metodologia che si pone l'obiettivo di "allontanare" il dato della persona, rendendo complessa o molto complessa la riferibilità del dato

- alla persona stessa. Diversamente dalle tecniche di anonimizzazione non viene “rotto” il legame che esiste fra dato e persona.¹⁶;
2. la **cifratura** dei dati, volta a rendere i dati personali incomprensibili a chiunque non sia autorizzato ad accedervi¹⁷.

2.1.3 Cenni sul Fascicolo Sanitario Elettronico, dalla sua genesi alle modifiche introdotte dal “Decreto Rilancio”

Una delle maggiori innovazioni nel Sistema Sanitario italiano degli ultimi anni è stata l'introduzione del Fascicolo Sanitario Elettronico (FSE). Questo strumento è stato previsto dal legislatore con il Decreto-Legge n. 179 del 18 ottobre 2012 convertito con modificazioni dalla L. n. 221 del 17 dicembre 2012, dove è definito come *l'insieme dei dati e documenti digitali di tipo sanitario e sociosanitario generati da eventi clinici presenti e trascorsi, riguardanti l'assistito*¹⁸. Tale definizione è stata poi estesa dal Decreto-Legge 34 del 19 convertito con modificazioni dalla L. n. 77 del 17 luglio 2020, a *tutti i documenti digitali sanitari e socio-sanitari, riferiti alle prestazioni sia a carico del Servizio Sanitario Nazionale (SSN) che fuori SSN*. Sin dalla definizione si evince come tale strumento sia legato in modo indissolubile al tema del dato sanitario (l'FSE è difatti un “*insieme di dati*”). L'FSE, registrando, la storia clinica degli utenti permette di migliorare i servizi di:

1. prevenzione, diagnosi, cura e riabilitazione;
2. studio e ricerca scientifica in campo medico e biomedico;
3. programmazione sanitaria¹⁹.

Il DL n. 179 del 2012 prevede che il Fascicolo Sanitario Elettronico sia istituito dalle Regioni e dalle Province Autonome **nel rispetto della normativa vigente in materia di protezione dei dati personali**. Un ulteriore passaggio nella definizione dell'architettura del FSE lo ha imposto il “Regolamento in materia di fascicolo sanitario elettronico” adottato con il DPCM n. 178 del 29 settembre 2015, il quale dispone che all'interno dei dati relativi alla storia clinica dei pazienti vadano distinti un “nucleo minimo” di dati e documenti da considerare obbligatori e una serie di elementi integrativi utili ad arricchire il FSE, ma che rimangono

¹⁶ D'Acquisto G., Naldi M. (2017). Big Data e Privacy by Design. Anonimizzazione, Pseudonimizzazione e Sicurezza. Torino: Giappichelli Editore.

¹⁷ Si fa riferimento per questo elenco al lavoro di A. Antonelli: Antonilli, Andrea. "Sicurezza informatica e trattamento dei dati in ambito sanitario." Opera a cui si rimanda anche per ulteriori approfondimenti sul tema.

¹⁸ Decreto-Legge n. 179 del 18 ottobre 2012 convertito con modificazioni dalla L. n. 221 del 17 dicembre 2012, art. 12.

¹⁹ HIMSS Italian Community, Il Fascicolo Sanitario Elettronico: Potenzialità ed Azioni Evolutive

facoltativi e dipendono dalle scelte delle singole regioni. Emerge un'organizzazione fortemente regionalizzata dello strumento che, se da un lato riflette la struttura della sanità in Italia, dall'altro ha portato alla creazione di disparità nell'adozione e nell'utilizzo dello strumento tra le varie regioni²⁰.

Un altro intervento normativo che ha inciso sull'FSE è stato il cd. Decreto Rilancio (DL n. 34 del 19 maggio 2020 convertito con modificazioni dalla L. n. 77 del 17 luglio 2020). Il decreto:

- a. abroga l'obbligo del consenso del paziente per alimentare il Fascicolo (obbligo che tuttavia permane per la consultazione del FSE e per l'inserimento dei dati relativi al "pregresso"). Si tratta di un'innovazione che indica una chiara intenzione di potenziare lo strumento facilitandone l'espansione.
- b. rafforza l'infrastruttura nazionale per interoperabilità (INI) con funzioni estese e a supporto delle regioni maggiormente in ritardo con l'adozione del FSE al fine di facilitare l'interoperabilità dei dati e promuovere l'attuazione dello strumento in tutto il territorio nazionale.

Parallelamente va ricordato l'avvio dei lavori al "Portale di accesso nazionale" del fascicolo che si dovrebbe affiancare ai portali regionali.

Infine, il DL Sostegni-ter (Decreto-Legge n. 4 del 27 gennaio 2022 convertito con modificazioni dalla L. n. 25 del 28 marzo 2022) ha introdotto ulteriori modifiche alla disciplina dell'FSE:

- a. sono state affidate nuove funzioni all'Agenas, tra cui: attuazione delle linee guida per la realizzazione del FSE, funzioni di garanzia per l'interoperabilità del Fascicolo con i dossier farmaceutici, per l'allineamento con l'Anagrafe Nazionale degli Assistiti e, in generale, per l'interconnessione dei soggetti e l'alimentazione del FSE.
- b. viene esplicitato che tutti i gli operatori della sanità (pubblici, privati e privati autorizzati) contribuiscono all'alimentazione del FSE inserendo le prestazioni nel Fascicolo entro 5 giorni.

Sommando a questi interventi la previsione di spesa di quasi 1.4 miliardi di euro si ha la misura di quanto il tema della gestione digitalizzata – e in sicurezza – dei dati sanitari dei cittadini stia assumendo centralità, evidenziando una tendenza di crescita parallela delle applicazioni e del perimetro dello strumento nonché di accentramento amministrativo.

²⁰ Si rimanda a HIMSS Italian Community, Il Fascicolo Sanitario Elettronico: Potenzialità ed Azioni Evolutive

In conclusione è utile fare riferimento alla direzione che sta prendendo il rinnovamento del FSE nella cornice del PNRR. L'impegno del Governo Draghi è quello di trasformare definitivamente il Fascicolo da semplice archivio di documenti a "ecosistema dei dati". Oggi cittadini e operatori sanitari utilizzano poco lo strumento e non esistono praticamente servizi per finalità di governo o ricerca (nonostante questi siano previsti dalla normativa). Con i fondi del PNRR si prevede di intervenire per migliorare lo strumento ampliando le funzionalità a disposizione di cittadini, operatori sanitari, regioni, governo e centri di ricerca. Per i cittadini l'FSE sarà un punto di accesso unico ai servizi sanitari digitali, raccoglierà tutti i dati clinici rilevanti (anche immagini o dati generati dai dispositivi) e avrà un'interfaccia omogenea su tutto il territorio nazionale. Per gli operatori sanitari sarà uno strumento di facile consultazione della storia clinica dei pazienti, anche grazie a servizi evoluti come le dashboard, abiliterà servizi di monitoraggio di aderenza alle cure e strumenti per prevenzione primaria e secondaria. Per regioni, governo e ricerca l'FSE permetterà l'utilizzo dei dati per migliorare gli outcome di cura e ridurre i costi. I dati, pseudonimizzati, saranno utilizzabili poi per la ricerca.

2.2 Dati sanitari e sicurezza: confidenzialità, integrità e disponibilità nell'Information Security

Il GDPR promuove l'uso delle migliori pratiche di sicurezza nel trattamento dei dati, richiedendo che i titolari del trattamento (controllers) e i responsabili del trattamento (processors) adottino misure di sicurezza appropriate per assicurare un livello di sicurezza adeguato al livello di rischio. Il Regolamento insiste sull'analisi del rischio e sull'attuazione delle misure di sicurezza più opportune per mitigare i rischi e garantire la confidenzialità, integrità e disponibilità dei sistemi e dei dati.

Confidenzialità, integrità e disponibilità sono i principi guida e gli elementi fondamentali di qualsiasi strategia di sicurezza e sono infatti anche i tre componenti principali della triade CIA (Confidentiality Integrity Availability) per l'Information Security, un modello di sicurezza delle informazioni ampiamente accettato e destinato a guidare le procedure e le politiche di sicurezza delle organizzazioni.



Figura 6. La triade CIA della sicurezza informatica.

Gli elementi della triade sono considerati le tre componenti più importanti per la cybersecurity e, in generale, per la sicurezza delle informazioni. La compromissione di un qualsiasi elemento della triade comporta la compromissione della sicurezza dei dati di un'organizzazione, il che, in ambito sanitario, può comportare anche conseguenze molto gravi.

Nel trattare il **dato sanitario** occorre coniugare e bilanciare l'attenzione da porre sui tre principi guida, in funzione della criticità e dei rischi correlati alla tipologia di dati e ai sistemi in esame.

Ad esempio, è chiaro che i livelli di confidenzialità, integrità e disponibilità di un sistema per la gestione dei trapianti debbano essere più elevati rispetto a un sistema che colleziona dati clinici a fini statistici o di prevenzione. Il compito degli specialisti di sicurezza delle informazioni sarà quello di analizzare e calibrare correttamente i tre aspetti, la relazione tra loro, come si sovrappongono e come potenzialmente si oppongono l'uno all'altro. L'obiettivo è determinare le priorità e i più appropriati processi di information security abbinati alle tecnologie e misure di sicurezza da mettere in campo.

Ai fini del presente documento possiamo così riassumere i tre principi guida del modello e le macro-raccomandazioni che ogni principio comporta:

- **Confidenzialità**

Il primo e migliore modo per proteggere i dati e le informazioni è limitare chi può accedervi. Meno persone possono accedere ai dati, minore è il rischio che possano essere effettuate delle violazioni alla sicurezza.

La confidenzialità (o riservatezza) ha a che fare quindi con il mantenere privati i dati di un'organizzazione. Ciò significa che solo gli utenti e i processi autorizzati dovrebbero essere in grado di accedere o modificare i dati. Per mantenere i dati riservati, si devono mettere in atto le protezioni appropriate, sia in termini di politiche di accesso, sia di tecnologie in grado di attuarle. Queste ultime dovrebbero essere presenti nativamente all'interno dei sistemi di gestione dati (DBMS) come, ad esempio, le regole per definire l'autenticazione e autorizzazione nell'accesso ai dati (tipicamente tramite la definizione e assegnazione di specifici privilegi a determini ruoli e profili).

Queste tecnologie devono incorporare dei meccanismi SoD (Separation of Duty) di controllo accessi e separazione delle responsabilità tra chi deve realmente accedere e modificare i dati e chi deve invece solo garantirne il pieno funzionamento (tipicamente i gestori dei sistemi o DBA). L'obiettivo è prevenire conflitti di interesse (reali o apparenti), atti illeciti, frodi, abusi e furti di informazioni. In tale modo utenti privilegiati (come, ad esempio, i DBA) devono poter eseguire le normali attività operative di gestione (ad esempio l'applicazione di *patch*, l'importazione, l'esportazione e il backup dei dati) senza mai poter accedere al contenuto dei dati trattati.

Infine, si garantisce una maggiore confidenzialità adottando meccanismi di crittografia (tipicamente basati su standard come l'AES – Advanced Encryption Standard) che proteggano i dati da accessi che tentano di aggirare i controlli applicati dai sistemi di gestione dati (DBMS) come, ad esempio, i tentativi di accedere direttamente ai dati presenti nei file system sottostanti ai DBMS o ai file di backup.

- **Integrità**

Integrità significa che ci si può fidare dei dati. L'integrità garantisce che i dati siano mantenuti in uno stato corretto, che non possano essere manomessi e quindi siano affidabili. Oltre a controllare l'accesso ai dati, le misure di sicurezza dovrebbero anche limitare ciò che le persone possono fare con tali dati. Ad esempio, la possibilità di copiare, spostare o alterare i dati dovrebbe essere limitata. Ciò protegge l'integrità dei dati, garantendo che gli utenti autorizzati accedano ai dati giusti in ogni momento.

Attuare le giuste misure di sicurezza in questa fase può fare un'enorme differenza per la quantità di danni che un utente malintenzionato potrebbe infliggere.

Anche in questo ambito entrano in gioco le corrette politiche di accesso e le tecnologie in grado di attuarle. Sono molto importanti anche le misure messe in atto dai sistemi di gestione dati (DBMS) per garantire la piena integrità dei dati anche in caso di accesso a sottocomponenti dell'infrastruttura tecnologica come i sistemi di storage (*data at rest*) o la rete di comunicazione (*data in transit*). Per

proteggere l'integrità dei dati dai possibili attacchi vengono utilizzati nativamente dai software di gestione dati (DBMS) algoritmi quali Message Digest Algorithm 5 (MD5) o Secure Hash Algorithm (SHA).

- **Disponibilità**

Il principio di disponibilità implica che i dati siano sempre disponibili per gli utenti autorizzati ogni volta che lo richiedano. Ciò significa mantenere sistemi, reti e dispositivi con cui trattare e accedere ai dati, sempre attivi e funzionanti.

La disponibilità consiste quindi nell'assicurarsi che i sistemi (e i dati gestiti) siano attivi e funzionanti in modo che l'organizzazione possa continuare a iperare anche di fronte a un attacco. Gli attacchi DDoS (Distributed Denial of Service) si basano, ad esempio, sul limitare o neutralizzare la disponibilità di un sistema. Gli attacchi di tipo ransomware hanno un diverso obiettivo (richiesta di riscatto e appropriazioni di dati) ma producano di nuovo una indisponibilità dei sistemi. Per questo motivo, la creazione di un piano di risposta DDoS, di un piano dei Backup/Recovery e la ridondanza nei sistemi è un modo per garantire la disponibilità. Tuttavia, quando non c'è alcun attacco, i sistemi possono ancora rendersi indisponibili a causa di guasti e malfunzionamenti, attività manutentive non programmate, errori umani, calamità naturali etc.

Per garantire la disponibilità, le organizzazioni devono utilizzare reti, server, storage e applicazioni ridondanti coadiuvati da sofisticati meccanismi di replica e protezione dei dati per garantirne l'integrità e la completezza in caso di evento ostile. Questi sistemi ridondati possono essere configurati per garantire la piena operatività – e quindi disponibilità – in caso di guasti su singoli componenti del sito primario che li ospita e possono essere configurati e collocati su un sito diverso dal primario (sito di *disaster recovery*) ed essere programmati per diventare disponibili quando l'operatività del sistema primario viene compromessa (attacco DDoS, attacco ransomware, catastrofe naturale, grave errore umano, o manutenzione/sostituzione di componenti chiave del sistema primario). È inoltre possibile migliorare la disponibilità tenendo sotto controllo gli aggiornamenti (upgrade) e le patch di sicurezza (security patch) che i produttori di software rendono disponibili per ovviare a problemi e vulnerabilità note dei software forniti. In questo modo, si rende meno probabile che un'applicazione non funzioni correttamente o che una minaccia che sfrutta una nuova vulnerabilità si infiltri nel sistema. Infine, opportuni piani di backup/recovery rappresentano l'ultima linea di difesa per ripristinare l'operatività e la disponibilità di un sistema nel caso tutte le precedenti misure non abbiano prodotto l'effetto desiderato.

2.3 Dati sanitari e sicurezza: Il rischio cibernetico della forza lavoro distribuita in ambito sanitario

Il dato sanitario è per sua natura sensibile. La trasformazione digitale aumenta il rischio di attacchi informatici nel settore della salute. È fondamentale strutturare servizi e infrastrutture per contrastare gli attacchi, soddisfacendo i requisiti normativi e garantendo la privacy di operatori e pazienti.

Molti report internazionali mostrano come, dallo scoppio della pandemia, gli attacchi informatici perpetrati nei confronti delle strutture operanti nel settore della salute siano aumentati. Ad esempio, una ricerca di Forrester Consulting ²¹ evidenzia come dal 2020 un numero senza precedenti di dipendenti coinvolti nell'assistenza sanitaria ha iniziato a lavorare da remoto. Le organizzazioni sanitarie, che per loro stessa natura richiedono una forza lavoro in loco, non erano preparate a gestire le sfide della sicurezza legate al lavoro a distanza.

Dalla stessa ricerca sono emerse le seguenti evidenze:

- Gli aggressori stanno sfruttando le alte vulnerabilità della forza lavoro distribuita nell'assistenza sanitaria, e i direttori sanitari ne stanno risentendo.
- Il mondo della Sanità era meno preparato per l'impatto sulla sicurezza dei lavoratori remoti rispetto ad altri settori
- La sanità dà la priorità alla protezione dei dati, ma divari di budget e la varia complessità degli attacchi, impedisce ai fornitori di migliorare le pratiche di sicurezza
- Le avversità condivise hanno migliorato il rapporto tra IT e gruppo di Sicurezza.
- Le organizzazioni sanitarie stanno cambiando per affrontare casi d'uso chiave e rendere le loro organizzazioni a prova di futuro.

A conferma di ciò, in una recente indagine²², il 62% degli intervistati del settore sanitario, ha riportato un aumento del volume degli attacchi negli ultimi 12 mesi. La frequenza delle violazioni dovute ad attacchi informatici subite dal settore sanitario negli ultimi 12 mesi è stata superiore alla media. La debolezza dei processi è stato un problema particolare nelle aziende sanitarie, causando più di un quinto (21%) delle violazioni. Il 46% delle aziende operanti nel settore sanitario ha detto di aver aggiornato la propria tecnologia di sicurezza per mitigare il rischio mentre il 58% degli intervistati nel settore sanitario sta già operando con una strategia di sicurezza incentrata sul cloud.

²¹ Forrester Consulting. "Hindsight Is 2020 – The Pandemic Provides A Wake-Up Call: Integrated Solutions Future-Proof Organizations." Febbraio 2021.

²² [Report Security Insights di VMware per l'Italia, 2021](#)

2.3.1 L'importanza di un approccio "Zero – Trust"

L'impatto di un attacco sul settore sanitario ha una rilevanza cruciale sulla vita delle persone. La perdita di informazioni o l'impossibilità di accedere tempestivamente alle cartelle cliniche, ai dati dei pazienti, alla programmazione degli appuntamenti o degli interventi, possono avere effetti disastrosi e, in alcuni casi, anche estremamente gravi.

Un altro grande problema nelle strutture sanitarie, che contribuisce a renderle ancor più vulnerabili agli attacchi informatici, riguarda l'utilizzo di sistemi e dispositivi spesso datati e non opportunamente aggiornati. Sebbene le attrezzature mediche e i dispositivi elettromedicali siano indispensabili per formulare diagnosi e cure, essi non sono concepiti a prova di cyber-security, aspetto che fa di questi elementi un facile punto d'accesso per i criminali informatici.

Non esiste una soluzione "One-Fits-All" bensì una serie di elementi interdipendenti che in una logica "Zero-Trust" devono essere considerati alla base di ogni progetto di digitalizzazione dei processi sanitari. Sarebbe corretto chiamarlo un approccio "Security by Design" che tocchi tutti gli ambiti di un servizio sanitario digitale, non da ultimo quello organizzativo.

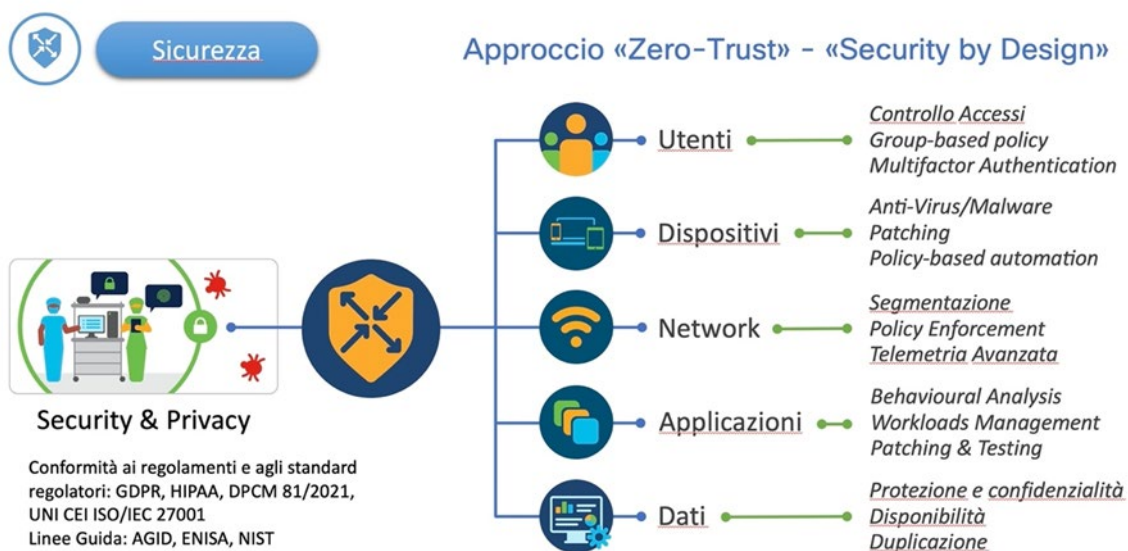


Figura 5. Approccio "Zero Trust" – "Security by Design".

La sicurezza è una responsabilità di tutti. È fondamentale creare una cultura diffusa della sicurezza che coinvolga l'intera organizzazione attraverso piani di formazione, figure dedicate (CISO – Chief Information Security Officer) e piani di resilienza informatica per gestire nel modo migliore eventuali incidenti.

È importante evidenziare che le attuali necessità di efficientamento stanno determinando cambiamenti significativi sia nella modalità con cui le applicazioni operano sia nelle infrastrutture necessarie alla loro erogazione. Le organizzazioni, private e pubbliche, non stanno solo spostando le applicazioni nel cloud dai data center locali, ma stanno sviluppando strategie multicloud che sfruttano la disponibilità, la resilienza e il risparmio sui costi rese possibili da questo nuovo approccio.

La gestione e il controllo della sicurezza per ambienti on-premise, cloud, ibridi e multicloud diventa rapidamente complessa, costosa e onerosa. Uno degli obiettivi principali diventa quindi come proteggere meglio applicazioni e dati senza compromettere l'agilità operativa.

In termini pratici, è necessario:

- Proteggere i WorkLoads, implementando soluzioni di sicurezza che siano il più "vicino" possibile alle applicazioni
- Analizzare costantemente le dinamiche applicative (Application's Behavior), definendo specifiche "Baselines" e segnalando prontamente eventuali scostamenti
- Applicare simultaneamente politiche di sicurezza e segmentazione dei flussi di traffico in logica Multi-Cloud

La protezione dei Workloads e dei Dati in contesti Multi-Cloud / Hybrid-Cloud deve necessariamente considerare la telemetria dei flussi di traffico, l'apprendimento automatico avanzato e le tecniche di analisi del comportamento applicativo. In particolare, per supportare i principali casi d'uso, si ritengono particolarmente rilevanti i seguenti aspetti:

- Definizione della Baseline comportamentale, analisi e identificazione delle anomalie relative ai workloads
- Adozione del modello "Zero Trust" che consente la microsegmentazione basata su caratteristiche funzionali
- Rilevamento di vulnerabilità e potenziali esposizioni associate alla tipologia e alla versione di software presente sui server
- Dislocamento in quarantena proattiva dei server quando vengono rilevate vulnerabilità e relativo blocco della comunicazione
- Analisi dell'interazione fra i diversi componenti applicativi e della modalità di Esposizione e Consumo delle API (API Security)
- Velocizzare la risposta alle minacce informatiche, migliorando la "posture" di sicurezza dell'organizzazione

Per approfondire i dettagli di quanto sopra espresso, anche nell'alveo delle direttive di Cyber Sicurezza nazionali e internazionali, si segnalano le seguenti linee guida:

- Decreto Presidente del Consiglio dei Ministri n.81 del 14.04.2021 ²³
- ENISA (European Union Agency for Cybersecurity) Regulation (EU) 2019/881 ²⁴
- ENISA Data Protection Engineering ²⁵
- European Commission NIS Directive (EU 2016/1148) ²⁶

In conclusione è necessario ripensare alla sicurezza come una parte intrinseca e distribuita dell'impresa (e della Sanità) moderna, incorporando continuamente tutti gli aspetti dello *stack* tecnologico per fornire una sicurezza più efficace attraverso un approccio Zero-Trust (o architettura Zero Trust).

Ciò significa mettere in pratica un approccio connesso, che unisce i punti critici di controllo tra utenti, dispositivi, applicazioni e reti. La sicurezza deve essere una parte integrante dei punti di controllo. Infine, le informazioni devono essere contestualizzate, combinando i dati provenienti da tutte le fonti in modo intelligente.

- **Mettere in sicurezza le cartelle cliniche elettroniche (Fascicolo Sanitario Elettronico (FSE)).**

La protezione deve concentrarsi sulla riduzione della superficie di attacco, e questo inizia con la microsegmentazione. Una forte strategia di microsegmentazione impedisce alle minacce di diffondersi lungo il vettore est-ovest e mantiene l'azienda il più sicura possibile, aggiungendo funzionalità automatizzate che semplificano l'applicazione delle politiche di sicurezza in futuro.

- **Segmentazione dei Dispositivi Sanitari**

²³ [Decreto Presidente del Consiglio dei Ministri n.81 del 14.04.2021](#)

²⁴ [ENISA \(the European Union Agency for Cybersecurity\) Regulation \(EU\) 2019/881](#)

²⁵ [ENISA Data Protection Engineering](#)

²⁶ [European Commission NIS Directive \(EU 2016/1148\)](#)

Proprio come la microsegmentazione tiene gli utenti fuori dalle aree a cui non possono accedere, così anche i dispositivi medici e/o accesso a Internet possono essere segmentati lontano dai flussi di rete, applicazioni e dati critici.

- **Il Valore di una strategia di micro-segmentazione**

Mentre una parte del ROI viene realizzata gradualmente, la mitigazione delle potenziali perdite associate con un attacco prende forma immediatamente. Un rapporto del Ponemon Institute mostra che in media un attacco informatico nel settore sanitario comporta in media costi pari a 6.45 milioni di dollari, oltre il 60% in più rispetto alla media intersettoriale²⁷.

- **Dispositivi Mobili all'interno di una rete sicura**

La mobilità sta presentando vaste nuove opportunità per l'efficienza e coinvolgimento del paziente, ma solo se i dispositivi fanno parte di una rete sicura che bilancia sicurezza di accesso con protezione. Questo tipo di mobilità è un aspetto importante della digitalizzazione dell'assistenza sanitaria.

- **Telemedicina**

Idealmente, i pazienti in presenza si aspettano di essere controllati da personale infermieristico incaricato e poi di avere una vera e propria conversazione dal vivo con un medico. Dal lato del fornitore di servizi sanitari, i professionisti richiedono l'accesso in tempo reale a raggi X, test, prestazioni dei dispositivi medici, di monitoraggio, video e registrazioni.

Queste aspettative sono altrettanto valide per la telemedicina. Con il giusto tipo di divulgazione digitale, le organizzazioni sanitarie possono fornire servizi e trattamenti che superano tali aspettative per pazienti ovunque si trovino. Per garantire la sicurezza le trasmissioni devono essere cifrate, così come i dati devono essere cifrati.

²⁷ <https://www.fiercehealthcare.com/tech/healthcare-data-breach-costs-average-6-45m-60-higher-than-other-industries-report>

3. CAPITOLO TERZO - CONTESTO EU E GAIA-X

L'Unione Europea, attraverso la spinta propulsiva di *policymaking* della Commissione, sta cercando di ritagliarsi un ruolo da protagonista globale nel governo della rivoluzione digitale. Uno degli obiettivi in questo processo è la creazione di un mercato unico dei dati nell'UE. Idealmente i dati potranno circolare all'interno dell'Unione in maniera trans-settoriale. Da un lato, nelle intenzioni della Commissione, vi è l'esigenza di proteggere la privacy e la concorrenza, dall'altro la necessità di costruire una *data economy* dinamica, attraverso regole chiare sull'accesso ai dati, investimento in strumenti e infrastrutture per archiviazione ed elaborazione di dati, costruzione di capacità cloud a livello europeo e creazione di spazi di dati interoperabili in alcuni settori strategici.

3.1 Strategia europea sui dati

A febbraio 2020 la Commissione Europea ha presentato la sua Strategia sui dati. Si tratta di un piano complesso che si compone di diversi elementi: regolamenti sullo scambio di dati (Data Governance Act, Data Act), misure per data spaces verticali (ad esempio lo European Health Data Space) e investimenti in infrastrutture per il trattamento e la condivisione dei dati. Alla base dell'iniziativa vi è la consapevolezza da parte della Commissione della crescente importanza dei dati nella società e nell'economia dell'UE. Per il 2025 si prevede un aumento del 530% del volume globale dei dati in Europa (da 33 a 175 zettabyte), l'economia dei dati raggiungerà il valore di 829 miliardi di euro e più di 10 milioni di professionisti lavoreranno nel settore dei dati.

3.1.1 Common European Data Spaces e European Health Data Space

Uno dei tasselli fondamentali della Strategia europea sui dati è la creazione di Data Space europei comuni in alcuni settori strategici per l'economia europea. I data space previsti dalla strategia sono nove: salute, industria e manifattura, agricoltura, finanza, mobilità, *green deal*, energia, pubblica amministrazione e skills (Figura 6); a questi si sono poi aggiunti i progetti relativi al data space per il settore media e quello per il settore dei beni culturali. L'obiettivo dei data space è mettere insieme infrastrutture e *governance frameworks* per facilitare la raccolta e la condivisione dei dati.

Si tratta di progetti complessi in quanto si compongono di interventi sia normativi che infrastrutturali. Vanno stabilite regole riguardanti la condivisione dei dati e allo stesso tempo assicurate efficienza e sicurezza dei sistemi.

La Commissione europea ha indicato diversi principi guida per il *design* dei Data space comuni: controllo sui dati, governance, rispetto dei valori e delle regole dell'UE, infrastruttura tecnica comune, interoperabilità, *openness* intesa come

apertura alla partecipazione di tutti gli attori (organizzazioni o individui) che rispettano i valori e le regole dell'Unione²⁸.

Dal punto di vista del finanziamento sono previsti diverse azioni, principalmente nella cornice del programma "Digital Europe" per quanto riguarda l'infrastruttura digitale, e nell'ambito di "Horizon Europe" per ciò che concerne la ricerca e l'innovazione.

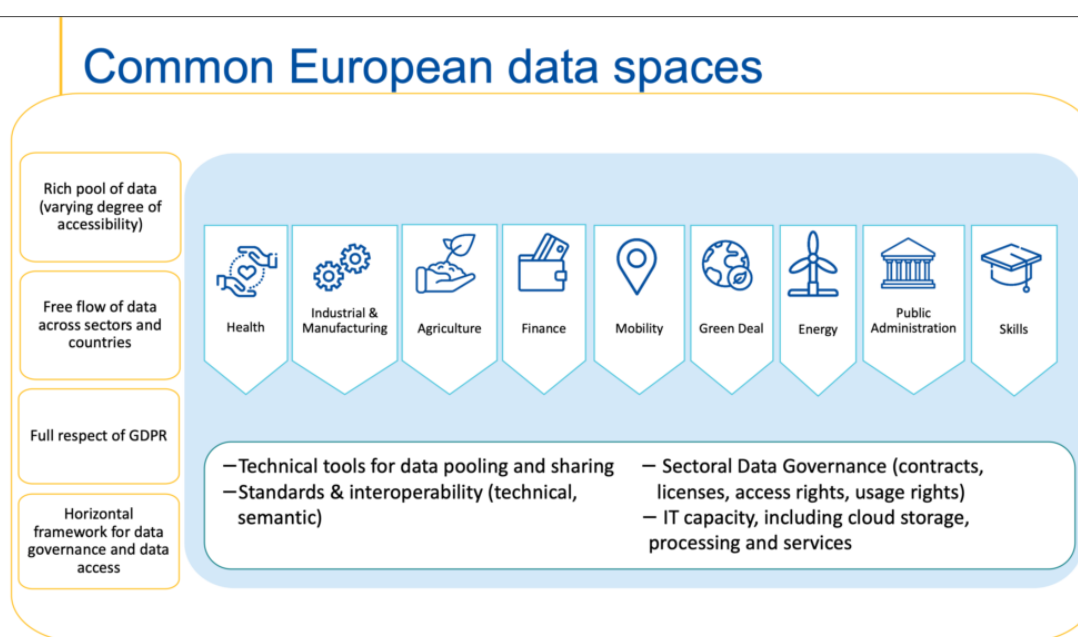


Figura 6. Illustrazione dei 9 Common European data spaces.

Lo Spazio europeo dei dati sanitari è il Common European data space nella fase più avanzata di sviluppo. Gli obiettivi del progetto sono²⁹:

- Assicurare l'accesso e la condivisione dei dati sanitari per scopi di assistenza sanitaria (utilizzo primario dei dati sanitari), nonché per il controllo degli individui sui propri dati.
- Facilitare l'accesso ai dati e il loro riutilizzo per scopi di ricerca, innovazione, salute pubblica e policymaking, in modo trasparente, sicuro e con una governance istituzionale appropriata (uso secondario dei dati sanitari).

²⁸ Commissione Europea, Commission Staff Working Document on Common European Data Spaces. SWD(2022) 45 final.

²⁹ Ibidem.

- Creare un mercato unico digitale per i servizi di sanità digitale, inclusi quelli abilitati dall'IA.

Le azioni chiave per il completamento del progetto sono:

- l'approvazione di una proposta legislativa che riguarderà la governance del Data space;
- un *grant* per un progetto pilota sull'infrastruttura europea per l'utilizzo secondario dei dati sanitari³⁰;
- due progetti all'interno del programma Digital Europe:
 - accessibilità transfrontaliera ai dati genomici³¹
 - infrastruttura europea federata per le immagini di tumori anonimizzate³²

È fondamentale, per il completamento dei progetti sui data spaces comuni, il coordinamento con le altre iniziative in corso a livello europeo. Tra le quali vi è Gaia-X. Tutte le iniziative hanno un obiettivo comune nella creazione di un'infrastruttura pan-Europea dei dati e un mercato unico dei dati. Per questo è importante assicurare l'interoperabilità utilizzando quanto più possibile gli stessi modelli di riferimento e processi³³.

3.1.2 Data Act

Un altro tassello fondamentale della Strategia Europea sui dati è il "Data Act". Si tratta di una proposta di regolamento che ha come obiettivo il migliorare l'equità nell'accesso ai dati (dati industriali, vale a dire dati non personali provenienti da oggetti connessi) e, allo stesso tempo, potenziarne l'utilizzo nel sistema produttivo. Già nel discorso sullo Stato dell'Unione del 2020 la Presidente Von Der Leyen esplicitava l'importanza di un intervento legislativo nell'ambito dei dati industriali:

The amount of industrial data in the world will quadruple in the next five years – and so will the opportunities that come with it. [...] And

³⁰ [Funding & tenders \(europa.eu\)](#)

³¹ [Funding & tenders \(europa.eu\)](#)

³² [C_2021_7914_1_EN_annexe_acte_autonome_cp_part1_v3_x3qnsqH6g4B4JabSGBY9UatCRc8_81099.pdf \(europa.eu\)](#)

³³ Commissione Europea, Commission Staff Working Document on Common European Data Spaces. SWD(2022) 45 final. s

*industrial data is worth its weight in gold when it comes to developing new products and services. But the reality is that 80% of industrial data is still collected and never used. This is pure waste.*³⁴

A ciò si aggiungeva la constatazione, nella comunicazione ufficiale sulla Strategia europea per i dati, dell'insufficiente livello di condivisione dei dati tra le imprese: *"in spite of the economic potential, data sharing between companies has not taken off at sufficient scale"*³⁵.

Infine, uno dei principi che hanno ispirato maggiormente la proposta è quello della Digital Sovereignty. La Commissione stima che circa il 90% dei dati prodotti in europa sono controllati da aziende oltreoceano³⁶, da qui l'obiettivo strategico di rafforzare la *data economy* europea anche con misure, come quelle che verrebbero introdotte dal nuovo Regolamento, che riducono il flusso di dati verso l'estero.

Perché il Data Act veda luce sarà necessario trovare dei compromessi: su una serie di fratture (ad esempio, *Digital Sovereignty* contro libero mercato e atlanticismo, *Big Tech* contro Start-up e PMI), per alcune filiere (su tutte probabilmente quella dell'automotive), e tra gli stati membri (Francia, Italia, Spagna da un lato, Irlanda e paesi est-europei dall'altro³⁷).

3.2 Gaia-X: Data driven Innovation

Gaia-X nasce dall'ambizione europea di essere leader nella *data-driven innovation*, specialmente nelle aree emergenti dell'Intelligenza Artificiale (AI), big data e cloud computing e, allo stesso tempo, dalla constatazione che per quanto ci siano dei provider regionali mancano in Europa degli Hyperscaler globali. Per questi, infatti si fa spesso riferimento ai provider di infrastrutture e piattaforme negli Stati Uniti o in Cina. Gaia-X sta lavorando alla creazione di una visione per un ecosistema digitale sicuro, trasparente e aperto, dove dati e servizi saranno offerti, raccolti e condivisi con la massima fiducia.

Gaia-X sta creando dei framework dove segmenti industriali e gruppi diversi possono incontrarsi per scambiare dati con alto livello di fiducia. Un esempio recente è Catena-X, il Network di Aziende dell'industria Automotive all'interno di Gaia-X.

³⁴ State of the Union Address by President von der Leyen at the European Parliament Plenary
https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_20_1655

³⁵ Ibidem

³⁶ <https://www.ceps.eu/the-data-act-six-impossible-things-before-breakfast/>

³⁷ C. Perarnaud e R. Fanni. The EU Data Act. Towards a new European data revolution? CEPS 2022

L'architettura di Gaia-X è progettata per soddisfare i seguenti requisiti:

- Interoperabilità di dati e servizi
- Portabilità di dati e servizi
- Sovranità sul dato
- Sicurezza e fiducia (security-by-design)

Il disegno architettuale è articolato sui principi di “federazione” (*federation*); “decentralizzazione” (*decentralization*) e “apertura” (*openness*). In altre parole, il disegno complessivo è basato alcuni principi generali di interoperabilità e sovranità quali:

- Nessuna integrazione fisica dei dati: i dati sono archiviati alla fonte
- Nessuno schema comune: l'integrazione avviene a livello semantico, attraverso vocabolari
- Invece di Data sharing: Data networking, Data visiting e Data co-existence
- Possibilità di sovrapposizioni e nesting: creazione di un ecosistema di data spaces
- Piena Sovranità del dato, controllo e tracciabilità
- Fiducia tra i partecipanti
- Cloud Neutral Portability
- Reversibilità
- Apertura e predisposizione per l'open source

Gaia-X quindi non è un cloud service provider. Gaia-X sviluppa i componenti software necessari a realizzare un sistema federato (i.e. API) che inter-connette vari partecipanti per lo sviluppo di nuovi servizi e prodotti innovativi. Tali ecosistemi consistono di dati e infrastrutture interconnesse, aggregate come “Federatori”. L'ecosistema di dati alimenta i data spaces che si basano sullo scambio di dati con regole condivise per ciascun segmento industriale.

L'obiettivo di un eventuale data space della Sanità Digitale è quello di creare un flusso di dati standardizzato all'interno dell'intera catena di valore della Sanità. Di creare inoltre benefici e valore aggiunto per ciascun partecipante e allo stesso

tempo mantenere la sovranità del dato in accordo con gli standard dell'Unione Europea.

Questo ecosistema basato sul Cloud dovrebbe essere aperto a tutte le aziende nell'industria sanitaria Europea e alla rete di partner globali che fanno parte della catena del valore. In questo scenario lo sviluppo di Gaia-X rappresenta un prerequisito e un abilitatore per lo scambio di dati e informazioni.

Un progetto aderente agli schemi architetturali di Gaia-X creerebbe vantaggi per il Polo Strategico Nazionale (Cloud Nazionale) e i Poli Strategici Nazionali Regionali, inoltre permetterebbe di:

- assumere un ruolo di International Data Space Connector e Cloud Service Provider in Italia
- potersi unire al network Europeo Gaia-X per usufruire dei servizi dati messi a disposizione degli altri membri del network stesso a scopo scientifico, culturale, economico e sociale
- contribuire con i propri dati alla crescita di data space utilizzabili (in totale sicurezza dal punto di vista dell'accesso e della privacy) dagli altri membri a scopi di progresso scientifico, culturale, economico e sociale da parte di cittadini, imprese e istituzioni.
- monetizzare, se lo si ritiene opportuno, l'accesso ai propri dati.

4. CONTESTO NAZIONALE

4.1 Fonti e flussi di dati sanitari

Il Servizio Sanitario Nazionale (SSN), tramite la SOGEI, elabora da anni una notevole mole di dati sanitari raccolti con il contributo delle Regioni e delle Aziende sanitarie e ospedaliere. I flussi sono analizzati in ottica di monitoraggio della spesa, controllo della produzione e calcolo dei rimborsi fra Stato e regioni e fra Regioni e Aziende ospedaliere. Tuttavia, è tutto evidente che questi dati potrebbero essere utilizzati in parte anche per elaborare strategie e politiche sanitarie, sia di tipo organizzativo (volte, ad esempio, alla gestione efficace ed efficiente delle organizzazioni sanitarie), sia, se integrati con il patrimonio informativo latente, di tipo clinico ed epidemiologico.

Nel SSN possiamo individuare due principali tipologie di dati:

- il dato dei flussi informativi ministeriali e regionali gestiti da SOGEI;
- il dato clinico proveniente dalle attività di prevenzione, diagnosi e cura.

Vanno inoltre considerati i dati recuperabili dalle attività del privato e dei comuni nell'ambito socioassistenziale, anch'essi parimenti utili per costruire un Healthcare Data Hub.

Un Digital Information Hub (DIH) è un polo di raccolta di tutti i dati di interesse e mette a disposizione questa conoscenza a tutti gli stakeholder (centri di ricerca, ospedali, università, ma anche industrie e startup), al fine di potenziare il sistema decisionale e governare sistemi complessi.

4.1.1 Flussi sanitari e flussi clinici

I flussi informativi sono un insieme di dati e indicatori utili a comprendere il funzionamento dei servizi del SSN. Ogni Azienda ospedaliera produce una grande quantità di dati e ognuna ha un proprio sistema informativo, variamente organizzato.

Il ruolo principale del sistema informativo è quello di fornire delle informazioni attendibili, aggregate e di facile gestione. La gestione delle informazioni è un'attività importante per tutti e particolarmente complessa per supportare chi svolge funzioni manageriali. Non esistono sistemi di documentazione validi per tutte le esigenze e situazioni e ogni strumento deve essere utilizzato per perseguire determinati obiettivi. Le informazioni raccolte servono a prendere decisioni appropriate al livello operativo nonché di pianificazione e di gestione.

Periodicamente le Regioni e le Aziende raccolgono i dati, fanno una verifica di coerenza sintattica e semantica del dato, inviano i relativi flussi a SOGEI mediante file di testo che vengono riversati in un grande database centrale che contiene i dati di tutte le Regioni italiane.

Vi sono inoltre diversi registri regionali, come il Registro Tumori e il Registro Cause di Morte. Questi registri, portati a un alto livello di interoperabilità e omogeneità semantica su base nazionale portano un notevole valore informativo permettendo di individuare in anticipo anomalie nelle tendenze di periodo e su base geografica sia per fenomeni di lungo periodo che per fenomeni a diffusione più rapida.

Ai flussi nazionali e regionali è possibile associare – con granularità sino al livello individuale – tutto il mondo dei dati clinici che vengono prodotti dalle cartelle cliniche, dai sistemi ambulatoriali, dai sistemi di refertazione, RIS/PACS, sistemi di telemedicina, ecc.

La quantità di dati è altrettanto notevole, si pensi che ad esempio un solo sistema di telemedicina può produrre migliaia di letture all'anno per singolo paziente (temperatura, pressione, ...).

Vediamo una elencazione dei principali tipi di dati da considerare:

- Dati di osservazione del paziente includendo i parametri vitali
- Dati di laboratorio (analisi del sangue, urine, ecc.)
- Dati anamnesici,
- Dati farmacologici (somministrazione del farmaco e osservazioni)
- Dati di diagnosi e referti di ambulatorio
- Procedure applicate al paziente

4.2 Necessità di un'infrastruttura

I dati sono presenti nei vari sistemi ma, nella maggior parte dei casi, non sono omogenei e non sono tra loro semanticamente collegati.

Manca un contenitore logico e fisico in grado di raccogliere i dati, trasformarli in informazioni e infine abiliare lo sfruttamento di questa conoscenza mediante l'uso di sistemi avanzati di analisi ed elaborazione (attraverso Intelligenza Artificiale e Machine Learning).

A livello tecnologico esistono già i repository clinici in grado di raccogliere questa enorme quantità di dati. Esiste lo standard FHIR che realizza la semantica dei dati sanitari. Il tema urgente è rendere omogenei e trasversalmente integrabili i diversi contenuti per poter estrarre conoscenza.

Nell'ambito di una *Data Driven Healthcare*, il paradigma della "Prossimità Territoriale" declinato dal PNNR presuppone l'esistenza di un'infrastruttura capillare che consenta:

- la raccolta dei dati prodotti dalle diverse sorgenti (Medici, Elettromedicali, Wearable Devices, ...)
- l'analisi e l'elaborazione dei dati sanitari collezionati (anche applicando logiche di AI, ML)
- la fruizione delle informazioni prodotte dall'analisi dei suddetti dati da parte dei soggetti destinatari (decisori politici, clinici, pazienti)

Con una metafora, si potrebbe considerare l'infrastruttura al pari delle fondamenta di un edificio, non si vedono ma sorreggono tutto quanto è visibile.

HIMSS ha sviluppato una metodologia di valutazione specifica per l'infrastruttura IT in sanità: l'Infrastructure Adoption Model (INFRAM)³⁸.

Le caratteristiche più rilevanti che un'Infrastruttura Digitale a supporto della Data Strategy, deve presentare sono:

- Un approccio modulare in logica "Multi-Cloud" che presenti alta scalabilità, affidabilità e resilienza
- Un criterio di definizione e sviluppo che preveda un'impostazione di "Security by Design". Un approccio "Zero-Trust" che tocchi tutti gli ambiti di un servizio sanitario digitale, non da ultimo quello organizzativo.
- L'integrazione "nativa" delle soluzioni di gestione dell'identità, per fornire al contempo sicurezza avanzata e l'applicazione automatizzata delle policy di accesso ai dati e alle risorse informatiche
- Il monitoraggio delle matrici di traffico per verificarne la conformità rispetto alle policy e alle prescrizioni dinamiche di Cyber Sicurezza

³⁸<https://www.himss.org/what-we-do/solutions/digital-health-transformation/maturity-models/infrastructure-adoption-model-infram>

- Visibilità “end-to-end”, in tempo reale, della qualità con cui i servizi sono forniti
- Strumenti di “Full-Stack Observability” che permettano di governare nel modo migliore la migrazione o l’adozione del cloud, mettendo nelle mani dei soggetti deputati alla governance dei sistemi le informazioni di Performance, Disponibilità, Carico e Costo di un’architettura Multi-Cloud Ibrida

4.3 Piattaforma Digitale Nazionale Dati (PDND)

Il Team per la Trasformazione Digitale prima e il Dipartimento poi, si sono impegnati dal 2017 nello sviluppo di protocolli informatici, un insieme di regole definite al fine di gestire la comunicazione tra sistemi, che permettano alle Pubbliche Amministrazioni di comunicare correttamente tra loro in Italia e – tramite servizi digitali transfrontalieri – in Europa.

Il Dipartimento per la Trasformazione Digitale sta lavorando per rendere possibile l’interoperabilità dei sistemi informativi e delle basi di dati delle pubbliche amministrazioni e dei gestori di servizi pubblici. Il nuovo Modello di Interoperabilità rappresenta un asse portante del Piano Triennale per l’informatica nella PA 2021-2023, necessario per il funzionamento dell’intero Sistema informativo della PA³⁹. Il modello è progettato in coerenza con i principi declinati nel nuovo European Interoperability Framework (EIF) oggetto della Comunicazione COM(2017) 134 della Commissione Europea adottata il 23 Marzo 2017⁴⁰. Il modello rende possibile la collaborazione tra pubbliche amministrazioni e tra queste e soggetti terzi, per mezzo di soluzioni tecnologiche che assicurano l’interazione e lo scambio di informazioni.

AgID, con Determinazione n. 547 del 1 ottobre 2021, ha adottato le Linee guida sull’interoperabilità tecnica delle Pubbliche Amministrazioni e le Linee guida Tecnologie e standard per la sicurezza dell’interoperabilità tramite API dei sistemi informatici che tutte le pubbliche amministrazioni devono adottare al fine di garantire l’interoperabilità dei propri sistemi con quelli di altri soggetti e favorire l’implementazione complessiva del Sistema informativo della PA.

³⁹ <https://www.agid.gov.it/it/infrastrutture/sistema-pubblico-connettivita/il-nuovo-modello-interoperabilita>

⁴⁰ [https://ec.europa.eu/transparency/documents-register/detail?ref=COM\(2017\)134&lang=it](https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2017)134&lang=it)

5. CAPITOLO QUINTO – PNRR

5.1 Azioni del PNRR

La pandemia ha reso ancora più evidenti alcuni aspetti strutturali critici del nostro sistema sanitario. E' fondamentale agire per risolvere tali criticità dato che in prospettiva potrebbero essere aggravate dalla crescita della domanda di cure derivante dalle tendenze demografiche, epidemiologiche e sociali in atto.

Il PNRR evidenzia l'importanza delle tecnologie, delle competenze – digitali, professionali e manageriali – per rivedere i processi di cura e per ottenere un più efficace collegamento tra ricerca, analisi dei dati, pianificazione e prevenzione.

Nel riassumere le spinte evolutive e i punti di debolezza dell'attuale sistema possiamo trovare anche le azioni da intraprendere per un futuro in cui il dato sanitario abbia un valore strategico determinate per il Sistema Paese.

Nella situazione attuale i dati – come abbiamo visto anche nel capitolo precedente – esistono ma non vengono utilizzati tempestivamente, la loro condivisione è solo parziale e il loro formato di archiviazione è eterogeneo (danneggiando quindi l'interoperabilità). Per cui:

- L'uso del dato non sempre è programmato
- Il sistema è centrato sui flussi e non sulla condivisione
- Il valore della condivisione del dato per gli attori coinvolti non è percepibile
- I sistemi coinvolti sono eterogenei e quindi procedono a diversa velocità nel percorso evolutivo

Già la Missione 1 in modo generale affronta i temi della sicurezza nella PA, dell'accessibilità e della cooperazione, ponendo le basi per una serie di interventi organici volti all'adeguamento dell'"impalcatura" tecnica. Gli interventi atti creare coinvolgimento attivo e fiducia nella sicurezza del sistema devono essere infatti visti come abilitanti e fondanti di qualsiasi successiva azione specifica.

La Missione 6, invece, si concentra maggiormente sugli aspetti legati alla sanità. In particolare la "componente 2", affronta la maggior parte dei temi finora evidenziati. I prossimi due paragrafi tratteranno più nel dettaglio l'investimento 1.3 della M1C1 e l'investimento 1.3 della M6C2.

5.2 Investimento 1.3 (M1C1) – Dati e interoperabilità

La logica “once only” descritta dall’obiettivo di investimento sposa perfettamente il bisogno di ottenere un punto di cooperazione che, partendo dal dato, supporti servizi e crescita. Nel contesto dell’*healthcare* avviene la standardizzazione si giova del lavoro svolto a livello internazionale.

IHE in ambito sanità già da tempo ha contribuito a definire i modelli di cooperazione e interoperabilità adottati già oggi in altre realtà nazionali, per quanto su differente scala e dimensione; questi modelli sono stati adottati e sperimentati con la creazione di strumenti incentrati sul dato e non sul documento.

Infatti già oggi il cittadino e le istituzioni possono accedere a documentazione in forma digitale ma il dato puntuale resta sparso, replicato o inserito in contesti di difficile fruibilità e sarà dunque questo il primo passo da compiere: la razionalizzazione delle informazioni e delle modalità di accesso.

Il successo di dell’investimento 1.3 della M1C1 dipende da quanto – e come – l’interoperabilità permetterà la fruizione e l’evoluzione di servizi di valore a partire dal patrimonio informativo che continua a crescere nel tempo.

Si vuole per questo raccomandare alcune priorità in merito:

- Adozione degli standard di dominio
- Adesione a consorzi di cooperazione internazionali pensando, al contempo, a una scala superiore a quella nazionale almeno in termini di apertura
- Decentralizzazione degli archivi con un modello che punti a costruire una rete capace di resistere a difetti di gioventù, debolezze dei singoli ed evoluzione della tecnologia
- Apertura verso i soggetti, anche in ambito industria e non solo della ricerca, possono contribuire alla crescita degli strumenti

Si vuole dare particolare enfasi agli ultimi due aspetti: se da una parte il dato e la sua gestione sono un valore comune, esistono oggi strumenti digitali per l’attribuzione della proprietà (ad esempio con soluzioni abilitate dalla Blockchain) con la possibilità quindi di affidare la conservazione anche a soggetti terzi senza il timore di perderne completamente il controllo.

Un sistema distribuito sarà più veloce nel reagire al cambiamento, più veloce nella fruizione, più rapido nell’attivazione, robusto in caso di incidenti e con le giuste regole e i giusti strumenti ugualmente sicuro e affidabile.

Un sistema in cui l'impresa può intervenire sarà poi motore accelerato rispetto alle sole attività del mondo accademico che non potranno che beneficiare reciprocamente della collaborazione.

5.3 Investimento 1.3 (M6C2) – Rafforzamento dell'infrastruttura tecnologica e degli strumenti per la raccolta, l'elaborazione, l'analisi dei dati e la simulazione

Perché il Fascicolo Sanitario Elettronico – con l'investimento del PNRR – passi da un semplice archivio di documenti a uno strumento in grado di abilitare analisi e simulazioni, è fondamentale passare da un approccio documentale a uno che preveda l'immediata sfruttabilità del dato strutturato.

Per abilitare la ricerca è indispensabile fornire informazioni che siano immediatamente consumabili, sfruttando soprattutto dati certificati e tempestivi. Si vuole quindi raccomandare l'implementazione di un sistema nazionale che, in linea con l'impostazione dell'interoperabilità per la PA, risponda a poche semplici esigenze:

- Digitale a livello nativo e quindi basato su dati e non su solo su documenti
- Tempestivo
- Aperto
- Sicuro
- Estendibile

Potrebbe essere utile partire dalle esperienze di realtà che hanno costruito un fascicolo nazionale basato su tecnologia FHIR, per quanto queste non siano necessariamente confrontabili con il contesto italiano. Ad ogni modo è indispensabile fare delle valutazioni su punti di forza e di debolezza di tali esperienze internazionali.

Se infatti lo standard HL7 fornisce un modo semplice per dare coerenza di rappresentazione e scambio del dato, la sfida è progettare un sistema fatto di regioni cooperanti e ottenere il coinvolgimento attivo e spontaneo di quanti più attori possibili.

Il FSE deve uscire dallo stato di contenitore per accumulo e rinascere come strumento di accesso al dato, accesso per il proprietario (il cittadino), accesso per chi deve erogare prestazioni a tutela della salute ma anche per chi fa ricerca e innovazione.

Il dato non deve più essere ostaggio di chi lo “colleziona” ma, con tutela dei diritti del cittadino, deve essere messo a disposizione della comunità.

Il sistema deve quindi essere una rete distribuita, supportata da porte di accesso standard, capace di accettare un insieme aperto di dati, capace di tenere traccia della titolarità del dato e dei diritti dell’interessato (compresi consensi e anonimizzazione ad esempio), agnostico rispetto all’origine tecnologica, in grado di espandersi, e in massima sostanza flessibile verso i diversi tipi di consumo.

Il sistema non può fondarsi come strumento universale per la fruizione del dato ma piuttosto essere la piattaforma su cui gli strumenti esistenti e futuri possano poggiare per finalità specifiche, siano esse il ritiro di un referto, la tracciatura dell’andamento di una epidemia o l’addestramento di un algoritmo per la diagnosi di una patologia.

L’investimento 1.3 tratta specificamente il tema del FSE insistendo su “diffusione”, “omogeneità” e “diffusione” prevedendo:

“(i) la piena integrazione di tutti i documenti sanitari e tipologie di dati, la creazione e implementazione di un archivio centrale, l’interoperabilità e piattaforma di servizi, la progettazione di un’interfaccia utente standardizzata e la definizione dei servizi che il FSE dovrà fornire; (ii) l’integrazione dei documenti da parte delle Regioni all’interno del FSE, il supporto finanziario per i fornitori di servizi sanitari per l’aggiornamento della loro infrastruttura tecnologica e compatibilità dei dati, il supporto finanziario alle Regioni che adotteranno la piattaforma FSE, il supporto in termini di capitale umano e competenze per realizzare i cambiamenti infrastrutturali e di dati necessari per l’adozione del FSE. ”

Approccio che senz’altro avvicina le criticità finora descritte ma ancora non declina un risultato di lungo periodo se non nella forma di LEA e sorveglianza e vigilanza sanitaria. Non possiamo però dimenticare che il progresso del settore è legato alla creazione, sulla stessa base, di una piattaforma di collaborazione internazionale che alimenti la ricerca sia accademica che industriale.

La creazione di un FSE che sia ancora una volta limitato a una consultazione occasionale dei contenuti negli episodi clinico-sanitari o anche una sua convergenza con i flussi informativi esistenti sarebbe un risultato parziale.

Il patrimonio informativo sanitario dell’FSE, ancor di più con il passaggio a scala nazionale, può essere motore per l’evoluzione scientifica e industriale, può allenare sistemi esperti e aprire le porte alla trasformazione del dato in informazione e conoscenza.

Possiamo dunque dire che per il mondo sanità la data strategy dovrà essere misurata sulle missioni 1 e 6, con la creazione di una piattaforma unica capace di aprirsi verso una dimensione europea, verso una fruizione quotidiana e nel contempo alimentare la crescita e l'evoluzione scientifica e industriale.

Questo ambizioso obiettivo deve essere perseguito con una strategia nazionale che conduca i diversi fascicoli esistenti (oltre a quelli ancora in fase di creazione) a un livello omogeneo e alla creazione di una rete; che, con gli opportuni paradigmi di sicurezza e tracciabilità, coinvolga tutti i soggetti capaci di trarre valore dal dato.

Il far sì che ogni “producer” contribuisca, passa dalla creazione di logiche di *rewarding* che solo i “consumer” possono dare in un ciclo virtuoso. In definitiva è indispensabile che quanto prodotto dalla missione 1 abiliti nell'ambito della missione 6 gli interventi verso i soggetti capaci di alimentare questo ciclo con servizi e prodotti.

6. CAPITOLO SESTO – DATA GOVERNANCE

È di vitale importanza che i dati su cui si basano i sistemi sanitari digitali siano incardinati in strutture di governance trasparenti, resilienti, solide e giuridicamente applicabili.

Le più innovative tecnologie sanitarie digitali si affidano in modo critico alla capacità di raccogliere, archiviare, accedere e condividere dati clinici strutturati e non strutturati relativi alla salute.

La qualità dei dati utilizzati per alimentare questi strumenti, compresi i dati raccolti attraverso il telemonitoraggio continuo – che molti potrebbero considerare intrusivo –, deve essere garantita al fine di facilitarne l'uso sicuro ed efficace.

Se allarghiamo poi l'ambito anche sui dati genomici e le biobanche, le sfide della governance dei dati sono particolarmente complesse a causa del legame biologico con la propria sfera parentale, l'individuo deve acconsentire la condivisione delle conoscenze ottenute attraverso test genetici sul proprio dna e, la donazione dei propri dati per scopi specifici, come la ricerca, possono entrare in conflitto se, gli appartenenti alla propria sfera parentale, non danno anch'essi il proprio consenso.

Senza un sistema giuridicamente applicabile ed efficace di governance dei dati che i pazienti possano considerare etico, rispettoso dei diritti, sicuro e affidabile, le promesse delle tecnologie sanitarie digitali più evolute come l'AI potrebbero essere vane.

La Dichiarazione universale dei diritti dell'uomo afferma il diritto di tutti di beneficiare della scienza: è imperativo che la condivisione dei dati debba essere responsabile ed etica, in modo che tutti possano beneficiarne in modo significativo, a partire dalla ricerca clinica.

I prossimi anni rappresentano un'opportunità enorme e unica per affrontare la governance dei dati e i problemi di cyber sicurezza, concordare quadri etici e sviluppare le competenze del personale delle organizzazioni per mettere in campo sistemi evoluti guidati dai dati. La complessità dei requisiti in materia di governance dei dati non dovrebbe essere motivo di immobilismo. La cosa più importante è che vi siano meccanismi per garantire che le nuove tecnologie non disumanizzino le cure. E mentre l'automazione migliorerà l'efficienza, l'interazione umana tra medici e pazienti non verrà meno e sarà ancora più importante perché permetterà al medico di occuparsi del paziente non solo dal punto di vista biologico, della malattia, ma anche da quello psicologico e relazionale.

In questa sede si è deciso di approfondire con alcuni spunti i temi della Data Access Architecture e della Data Classification.

6.1 Data Access Architecture

Dopo aver dedicato spazio – nei precedenti capitoli – ai temi dell’interoperabilità e della sicurezza dei dati, è utile soffermarsi sull’architettura tecnologica di accesso ai dati. Questa dovrebbe essere in grado di:

1. Fornire un unico punto di vista logico-gerarchica (Single Name Space) dei vari data lake come sintetizzato nella seguente figura:

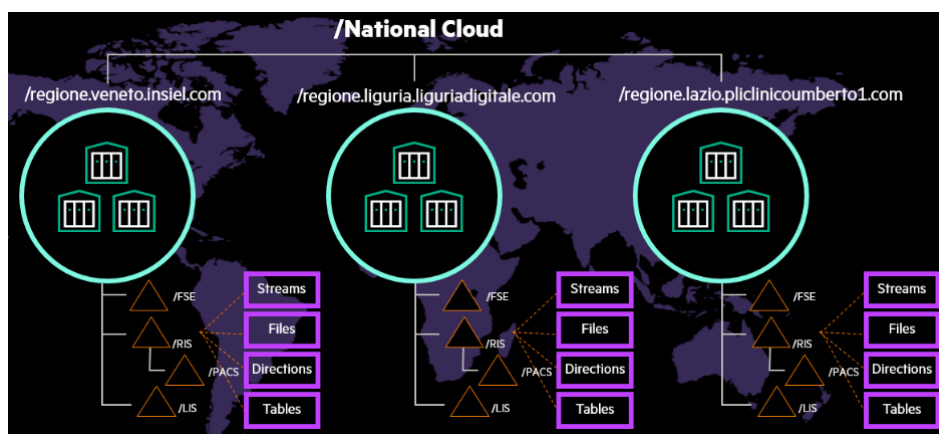


Figura 8. Data Access Architecture +

2. Poter accedere ai vari data lake distribuiti territorialmente con le stesse performance che si otterrebbero nel caso in cui i data lake fossero letti in locale, ossia scalare in modo distribuito indipendente dal numero di siti o regioni dove il data lake sono localizzati e si alimentano.
3. Mantenere durante le operazioni di lettura la consistenza e l'integrità del dato in tutti i data lake interrogati o scritti, nonché la proprietà di sovranità del dato stesso.

È da notare che in questa architettura i dati non vengono spostati ma rimangono localizzati presso i vari enti dei diversi SSR/SSN.

Risultati attesi

Il risultato è quello di mettere a disposizione del SSN uno spazio dati con le seguenti caratteristiche:

- Potenzialmente rappresentativo di tutte le informazioni cliniche, sanitarie, farmacologiche, di ricerca, etc., presenti sul territorio nazionale sulle quali poter effettuare analisi di qualsiasi sorta.
- Non vi è la necessità di dover creare un'apposita infrastruttura centralizzata in quanto i dati rimangono dove sono oggi.
- I dati e le informazioni sulle quali vengono svolte le analisi sono dati caldi ossia al livello massimo di aggiornamento possibile.

6.2 Data Classification e Code Systems

Lo sviluppo di un linguaggio globale per la salute e per la condivisione delle informazioni tra i professionisti sanitari, è un obiettivo da perseguire a livello globale.

I principali sistemi di classificazione e codifica per la sanità riconosciuti a livello internazionale e utilizzati in Italia per i dati e documenti del FSE (cfr Art.24, comma 2 del decreto del Presidente del Consiglio dei ministri n. 178/2015) e per il Nuovo Sistema Informativo Sanitario del Ministero della Salute, sono:

1. ICD

International Classification of Diseases (in particolare, International Statistical Classification of Diseases, Injuries and Causes of Death) è la classificazione internazionale delle malattie e dei problemi correlati, stilata dall'Organizzazione mondiale della sanità (OMS-WHO) contiene circa 55.000 codici univoci per lesioni, malattie e cause di morte. La nuova versione 11 entrerà in vigore dl 2022.

2. ICPC-2

International Classification of Primary Care 2nd edition è la classificazione dei dati del paziente e dell'attività clinica nell'ambito delle cure primarie, stilata dall'Organizzazione mondiale della sanità (OMS-WHO).

3. ICF

International classification of health and health-related domains, è la classificazione utilizzata per misurare il livello di salute a livello individuale e della popolazione, stilata dall'Organizzazione mondiale della sanità (OMS-WHO).

4. ATC

International classification of Anatomical Therapeutic Chemical (ATC) è la classificazione internazionale delle sostanze attive, stilata dall'Organizzazione mondiale della sanità (OMS-WHO).

5. SNOMED-CT

SNOMED International è l'organizzazione non profit che gestisce SNOMED CT (dizionario dei termini clinici) per renderlo il miglior standard per la terminologia medica.

SNOMED International e HL7 International nel 2019 hanno siglato un accordo per una licenza 'Free for Use' Set of SNOMED CT di terminologie (più di 8000) per l'implementazione dei documenti HL7 CDA R2 e FHIR nell'International Patient Summary (IPS è un documento elettronico contenente le principali informazioni sanitarie di una persona utilizzabile a livello internazionale così come a livello nazionale o locale. A livello internazionale è completamente allineato allo standard europeo CEN per il Patient Summary).

6. LOINC

Logical Observation Identifiers Names and Codes (LOINC) è un sistema di codifica universale per test, misure e osservazioni cliniche e di laboratorio. LOINC fornisce anche un vasto catalogo di misure, incluso test di laboratorio, dati clinici (parametri vitali e misure antropomorfe), formulari standardizzati per la raccolta dei dati.

A questi principali sistemi di classificazione e codifica se ne aggiungono tanti altri specifici per disciplina specialistica/figura professionale⁴¹.

6.3 HL7 FHIR®: lo standard per l'interoperabilità dei dati in sanità

Il settore sanitario sta trasformando rapidamente i dati sanitari verso lo standard HL7 Fast Healthcare Interoperability Resources (HL7 FHIR®). In particolare in Italia si basano su di esso i principali sviluppi del Fascicolo Sanitario Elettronico.

HL7 FHIR® offre un modello di interoperabilità con semantica standardizzata per consentire una collaborazione applicativa semplificata.

L'utilizzo delle risorse HL7 FHIR® consente l'accesso rapido a Sistemi Informativi Ospedalieri accelerando lo sviluppo di nuove applicazioni indipendenti dall'architettura dei diversi sistemi di origine. In questo modo si possono interfacciare applicazioni per ambiti e scopi differenti, come ad esempio:

⁴¹ a titolo di esempio, non esaustivo, si fornisce il link all'elenco delle terminologie usate da HL7 International: <https://terminology.hl7.org/codesystems.html>

- **IoT e sviluppo di app:** gli sviluppatori di APP per la salute e benessere per i cittadini e/o i loro provider possono sfruttare questo standard come servizio back-end con cui gestire e scambiare dati in un ambiente cloud sicuro progettato per i dati sanitari, sfruttando anche le linee guida di implementazione di SMART on FHIR;
- **Ecosistemi sanitari:** anche se il FSE arriverà ad essere il riferimento principale dell'informazioni cliniche dei cittadini, i sistemi informativi dei diversi provider e *payer* dei servizi sanitari possono usare questo standard per scambiarsi informazioni;
- **Enti di Ricerca:** il mondo della ricerca può sfruttare questo standard per ottenere dati normalizzati con terminologie internazionali e quindi più facilmente fruibili da parte dei sistemi di machine learning e in generale di intelligenza artificiale.

7. CAPITOLO SETTIMO – LIFE DATA

Il passaggio da Sistema Sanitario a Sistema Salute ha precise implicazioni in termini di Data Strategy.

Tradizionalmente i Sistemi Sanitari si sono concentrati sulla gestione degli episodi clinici relativi allo stato di malattia e sui dati da essi generati.

Tuttavia è riconosciuto che lo stato di salute e la prospettiva di vita in salute dipende da fattori determinanti che esulano dagli episodi clinici di malattia e che attengono ambiti diversi. In primis si può fare riferimento al nostro patrimonio genetico ma, in misura ancor più rilevante, sono determinanti una serie di fattori esogeni, quali condizioni socio-economiche, ambientali, abitudini, stili di vita e comportamenti.

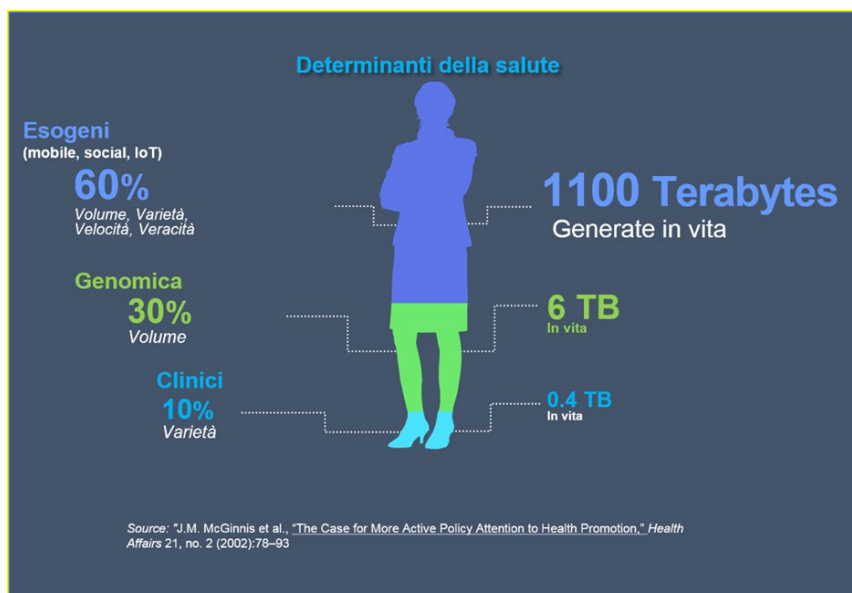


Figura 9. Dati e Determinanti della salute. Fonte J.M. McGinnis et al. "The Case for More Active Policy Attention to Health Promotion".

In questa prospettiva, l'evoluzione dei sistemi per la salute dovrà necessariamente estendersi ai determinanti non clinici della salute della popolazione. Si tratta di elementi di cui sarà fondamentale tenere conto per disegnare politiche sanitarie sempre più rivolte al mantenimento della salute piuttosto che al ripristino della salute a fronte di episodi clinici acuti.

Le implicazioni di questa evoluzione impattano chiaramente la natura e organizzazione dei sistemi sanitari.

Ai fini del presente documento, ci concentreremo esclusivamente sulle implicazioni in termini di Data Strategy. L'elemento da sottolineare è la crescita

esponenziale del volume e della varietà di dati a livello individuale e di collettività nel momento in cui si affiancano ai dati clinici i dati genomici e i dati “di vita” (per semplicità nei dati “esogeni”).

7.1 Dati genomici e dati esogeni

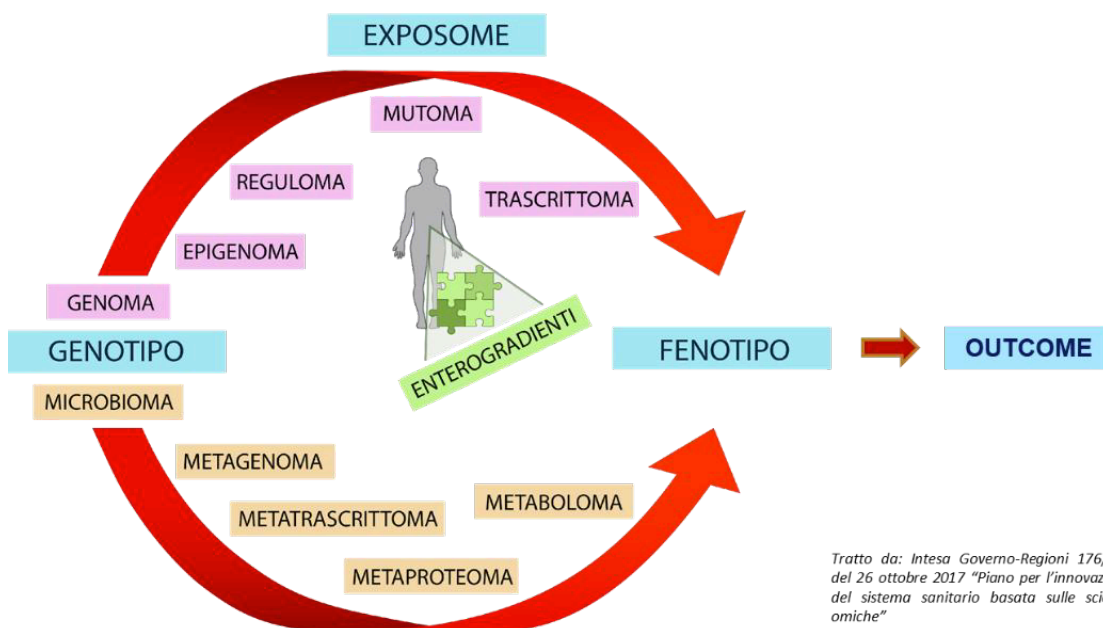


Figura 10. Dati genomici. Fonte: Augusto Ruggeri per DHS 2020.

È fondamentale riflettere sui modelli per la gestione dei “Big Data”, legati ai determinanti dello stato di salute di ciascun individuo, derivanti dallo stile di vita, dalle abitudini alimentari, dall’ambiente. I Big data sono forieri di importanti sfide, sia in termini di volumi e varietà di dati da gestire, sia in termini di garanzia della affidabilità delle informazioni raccolte sul campo attraverso sensoristica o registrazione da parte del paziente stesso.

Per poter comprendere questa ampia categoria, partiamo da una possibile classificazione:

- Ambiente (livelli di inquinamento, ...)
- Fattori socio-economici (livello di istruzione, attività lavorativa, ...)
- Stili di Vita (movimento, sonno, ...)
- Alimentazione

Negli ultimi anni abbiamo assistito a un proliferare di dispositivi IoT, app e mobile utilizzati direttamente dagli individui per svolgere le proprie attività quotidiane e

che raccolgono grandi quantità di dati sulle abitudini e sui comportamenti che hanno implicazioni in ambito salute.

Questa raccolta di informazioni attiene principalmente alla sfera individuale, ma è utile anche per finalità epidemiologiche e di ricerca. Talvolta giunge persino ad abilitare nuovi modelli di business (si pensi all'impatto degli stili di vita sui modelli assicurativi per la salute) e nuovi servizi (dalle digital therapy alle soluzioni di monitoraggio della propria salute)

È indubbio che questo processo pone grandi quesiti in termini di privacy e di possibilità di utilizzo dei dati.

È però ugualmente evidente che il passaggio dall'archiviazione dei dati dalle informazioni cliniche ai *life data*, apre nuove possibilità per la gestione della salute da parte degli individui stessi e, più in generale, per la prevenzione e la cura.

7.2 Prospettive per il futuro: un modello dati per il territorio

Come si è visto, la nozione di Sistema Salute è più ampia di quella di Sistema Sanitario e pertanto si fa riferimento a una varietà di dati molto più ampia. Una delle sfide del futuro prossimo è come gestire – in termini di governance, semantifica e infrastruttura – dati di tipo anche molto diverso. Si vuole proporre un modello che virtualizza i dati al fine di unificare informazioni provenienti da più origini in un singolo livello.

In questo modo, le applicazioni, gli strumenti di *reporting* e gli utenti finali possono accedere e manipolare i dati senza richiedere dettagli tecnici come l'origine, la posizione e le strutture. Due vantaggi chiave di questo modello sono quindi:

- Consentire l'accesso in tempo reale al sistema di origine per i dati
- Risolvere le differenze nei formati e nella semantica tra sorgente e utilizzo (*consumer*) utilizzando tecniche di astrazione e trasformazione

Il modello dovrebbe utilizzare tecnologie legate a Big Data e Data lake, in modo da creare un *repository* di dati che sia in grado di virtualizzare dataset di grandi dimensioni, strutturati e non strutturati, nel loro formato nativo. Verrebbe così garantito il recupero e l'organizzazione del dato nel momento della fruizione. L'architettura dovrebbe essere basata su "Produttori dei dati", "Contenitori dei dati" e "Fruitori (e analisti) delle informazioni (si veda la fig. 11).

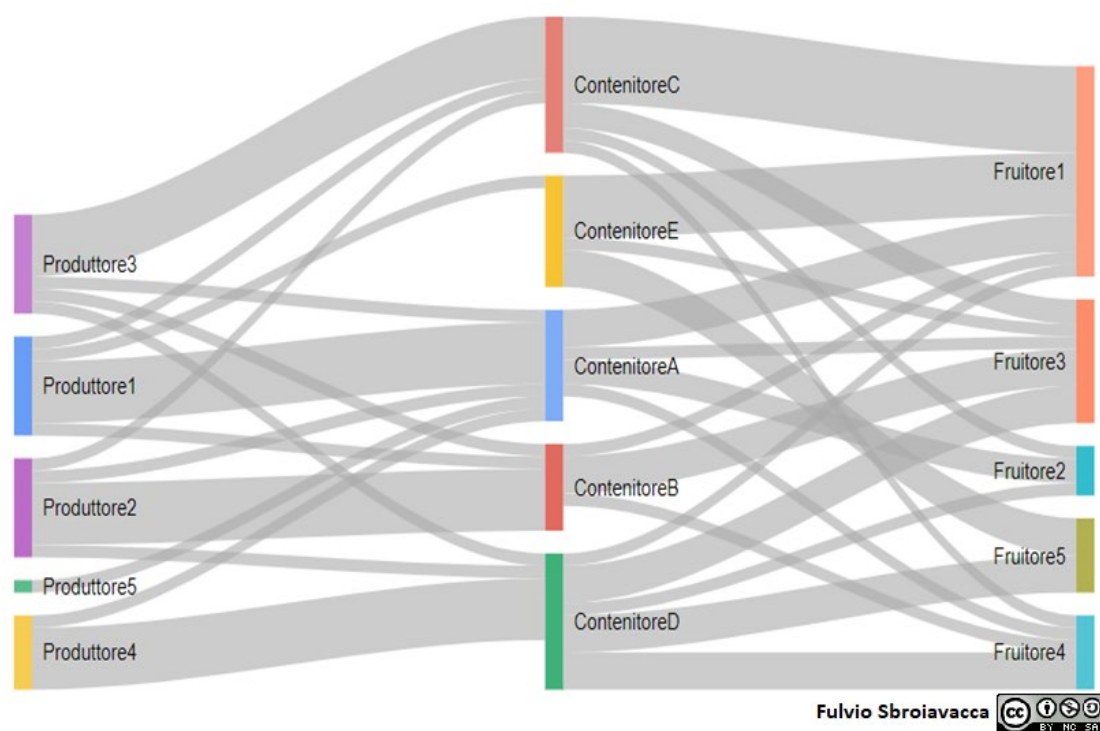


Figura 11. Schematizzazione dell'architettura del Modello Dati. Autore: Fulvio Sbroiavacca.

8. CONCLUSIONI – RACCOMANDAZIONI E PROPOSTE

Il mondo in cui viviamo è complesso e mutevole. Per questo la gestione del rischio deve fondarsi su decisioni *data-driven*, cioè basate su un insieme di informazioni sempre aggiornate, tracciabili, accessibili e trasparenti

Il nuovo paradigma nel mondo dell'interconnessione tra sistemi fisici e digitali, delle analisi complesse in real-time attraverso i Big Data, dell'Intelligenza aumentata, artificiale, del Digital Twin e Personal Digital Twin, pone il dato al centro.

Grazie alle informazioni e all'impiego di molteplici tecnologie digitali è possibile prendere decisioni più consapevoli e tempestive.

Cosa ci ha insegnato il COVID rispetto ai dati? È risultato chiaro che senza dati affidabili, immediatamente disponibili, si naviga alla cieca.

Gestire la transizione nell'era della *digital transformation* significa imparare a lavorare con le macchine intelligenti⁴². L'impiego su larga scala di tecnologie e algoritmi di intelligenza artificiale pone importanti quesiti in termini, non solo di *privacy*, ma anche di *trasparency*, *fairness*, *explicability*. Gli algoritmi di intelligenza artificiale sono fatti dall'uomo e prevedono una fase di addestramento che ne determinerà i risultati di applicazione. Più o meno inconsciamente, chi li sviluppa può introdurre *bias*. A questi si aggiungono gli algoritmi sviluppati in modo autonomo dalla IA, a esempio con tecnologia GAN, che pongono problemi di trasparenza ancora più complessi.

È necessario quindi che sia definito un quadro normativo e un protocollo di etica dell'intelligenza artificiale che indirizzi il profilo di fiducia prevenendo bias e quindi potenziali discriminazioni e in questa direzione si è mossa con grande attenzione e leadership l'Europa (Ethics Guidelines for Trustworthy Artificial Intelligence).

Cosa ci ha costretto a fare il COVID rispetto ai dati? È stato necessario raccogliere, con difficoltà e affannosamente, i dati da domini diversi per realizzare quadri via via più precisi della situazione e trattare i dati per rappresentarli. Significa utilizzare *dashboards*, rappresentazioni grafiche dinamiche, simulazioni, tutte attività che sono entrate prepotentemente nel quotidiano.

L'esperienza ci dice che c'è bisogno di una maggiore cultura dei dati. Senza dati non c'è progresso, non sono possibili digitalizzazione e innovazione.

L'attenzione ai dati sta cambiando in positivo, ma abbiamo bisogno che il valore dei dati sia riconosciuto da tutti gli attori del sistema della salute per sviluppare i paradigmi della *Data Driven Healthcare*.

⁴² Matt Beane, Learning to work with intelligent machines, Harvard Business Review. <https://hbr.org/2019/09/learning-to-work-with-intelligent-machines>

È urgente sviluppare un'infrastruttura nazionale dei dati, una riconosciuta semantica dei *Life Data*, adottare algoritmi trasparenti e condivisi, un'adeguata Data Governance per guidare i processi e le strategie nel nuovo ecosistema.

Portare i DATI da Volume a Valore è il primo passo.

Le raccomandazioni e le proposte di seguito illustrate possono costituire un fattore strategico di accelerazione a sostegno del percorso di digitalizzazione della Sanità e della messa in opera di servizi digitali per la salute dei cittadini.

8.1 Competenze

Il tema delle competenze rappresenta un nodo cruciale per il successo dei programmi di trasformazione digitale per il sistema paese, per la Pubblica Amministrazione e per la Sanità.

I temi di fabbisogno e di carenza di competenze digitali devono essere analizzati secondo due differenti punti di vista.

La prima chiave di lettura attiene alle competenze digitali degli utenti. Questo aspetto ha rappresentato a lungo uno dei principali inibitori dello sviluppo di servizi digitali.

Analogamente, la diffusione delle applicazioni di Data Science in ambito Sanitario e la fruizione di algoritmi di Machine Learning e artificial Intelligence, richiede a tutti i soggetti coinvolti la creazione di un bagaglio di conoscenze specifico.

Si ritiene di fondamentale importanza, pertanto, la costruzione di percorsi di formazione per consentire agli utenti di avere le competenze necessarie per trarre vantaggio dalla transizione digitale. Solo in questo modo sarà possibile ridurre l'inevitabile e connaturata resistenza al cambiamento che si sviluppa quando mancano la conoscenza degli strumenti e la consapevolezza dei benefici.

Bisogna evidenziare però come questa resistenza sia destinata a smorzarsi man mano che si affacciano al mondo del lavoro nuove generazioni fino ai nativi digitali.

Il secondo punto di vista riguarda la carenza di competenze per lo sviluppo e gestione dei servizi digitali, che diventa ancora più marcata nell'ambito Dati.

A fronte di questa carenza strutturale di competenze digitali le imprese hanno avviato iniziative, sia individualmente sia a livello associativo, finalizzate alla riqualificazione e alla creazione di nuove competenze allineate alla domanda di mercato. Lo sviluppo di competenze digitali è anche uno dei pilastri del PNRR, è importante riconoscere la centralità delle competenze relative ai dati per indirizzare gli investimenti pubblici.

A tal fine, come imprese del settore ICT, continuiamo nel nostro impegno a sviluppare programmi in collaborazione con le istituzioni, con le università e con il sistema della formazione secondaria.

In una situazione di oggettiva carenza di competenze e professionalità nel digitale e in particolare in ambito dati, è necessario evidenziare due punti di attenzione specifici per il settore sanitario: l'attrattività occupazionale del settore e l'esigenza

di coniugare le competenze digitali con competenze specifiche di natura medica, clinica e socio-assistenziale

Il mondo dei dati è profondamente connesso al dominio di applicazione: le competenze specifiche “di industria” sono necessarie a Data Scientists e Analysts per dare valore ai Dati e questo è ancora più vero per settori come quello medico-sanitario-assistenziale.

È necessario che le competenze digitali e quelle relative a dati e IA entrino a pieno titolo nel percorso formativo di medici e professionisti della salute. Ancora più interessanti sono le esperienze in corso anche in Italia per la definizione di programmi di confluenza fra percorsi di formazione STEM e medico-sanitari.

8.2 Cybersicurezza

Il dato sanitario è per sua natura sensibile.

La Cybersecurity è un elemento trasversale da considerare come componente essenziale di qualsiasi processo di trasformazione digitale in ambito sanitario.

È fondamentale che, nel tradurre la Data Strategy in progetti concreti, la sicurezza dei dati rivesta un ruolo centrale. Bisogna sfruttare servizi e infrastrutture per contrastare potenziali attacchi informatici, soddisfacendo i requisiti normativi e garantendo la privacy di operatori e pazienti.

Ogni progetto di digitalizzazione dei processi sanitari si deve sviluppare con una logica di “Security by Design” che contempli tutti gli elementi che compongono la catena di erogazione del servizio, non da ultimo quello organizzativo.

Sono fondamentali i CISO e la presenza nelle organizzazioni sanitarie di un'organizzazione dedicata alla Sicurezza con figure competenti dedicate e piani di resilienza informatica per gestire eventuali attacchi. La priorità dev'essere la la continuità operativa e la sicurezza di un patrimonio informativo sensibile come quello clinico-sanitario.

Inoltre, va considerato che i rischi di (cyber)sicurezza in Sanità travalicano i limiti dei progetti di digitalizzazione e si estendono a tutti gli apparati che si affacciano sulla rete

Le apparecchiature medicali oggi sono parte integrante dei sistemi sanitari, producono dati sensibili e sono vulnerabili quanto i sistemi informatici

Infine, occorre richiamare il tema delle Competenze (capitolo 8.1). I dipendenti sono, spesso involontariamente, una delle principali cause di breach di sicurezza, c'è bisogno di un cambio culturale che affermi come la sicurezza sia una responsabilità di tutti.

È essenziale creare una cultura diffusa della sicurezza che coinvolga l'intera organizzazione attraverso piani di formazione che consentano a tutti i professionisti della salute di difendere il patrimonio informativo, la privacy e l'operatività delle organizzazioni sanitarie.

8.3 Data Governance

A differenza dei dati amministrativi già organizzati e semanticamente standardizzati per verificare i costi del SSN, i dati clinici esistono ma sono molto frammentati e raccolti in modalità estemporanee, con semantiche che non seguono alcun standard o raccolti con software utili allo scopo transazionale ma che non consentono ai dati stessi di essere fruibili in tutto il loro valore da parte della ricerca scientifica o della pratica clinica.

L'esigenza di condividere dati tra attori diversi (ospedale, territorio, paziente e istituzione) è stata già colta a livello internazionale e in alcuni casi anche nazionale, tanto che sono oggi disponibili modelli di business (data model) e tecnologie per la raccolta e la condivisione sicura dei dati sanitari; tali modelli educano a una corretta concentrazione e organizzazione dei dati in un data lake sanitario.

È urgente normalizzare le prassi di utilizzo dei dati partendo dalla standardizzazione semantica dei dati clinici per arrivare a un Common Data Model nazionale. Il modello HL7-FHIR è un ottimo punto di partenza.

La classificazione giuridica del dato rappresenta un tema centrale e richiede esperti per le possibilità normate e legittime di utilizzo dei dati a fini di cura e di ricerca.

La capitalizzazione delle esperienze e la diffusione dei buoni risultati ottenuti dai progetti di utilizzo dei dati in maniera retrospettiva grazie a tecniche evolute di Deep Learning, intelligenza artificiale e Machine Learning, deve passare dall'accordo sulle possibilità di utilizzo e scambio di dati. Tale accordo costituisce il passo obbligato per la loro applicazione anche in ottica clinica.

Scalare a livello nazionale le esperienze già fatte in campo di IA, accresce il loro utilizzo su larga scala e permette di agevolare il dialogo con il garante della privacy. Nel dialogo con le istituzioni l'industria deve far emergere il valore creato dalle analisi predittive in termini di efficienza ed efficacia nella pratica clinica.

8.4 Infrastruttura

La salute sta diventando sempre più digitale. Si tratta di un processo di rinnovamento naturale che il COVID ha accelerato notevolmente fino a renderlo tangibile per chiunque. Al centro della digitalizzazione della sanità vi sono i dati.

Produciamo incessantemente dati e questa produzione aumenterà esponenzialmente in particolare con i sistemi IoT. Alcuni dati sono strutturati, altri invece hanno la forma di immagini, testi e audio.

Tutte le informazioni di cui disponiamo, sui nostri episodi di cura, ma anche di vita quotidiana, devono essere messe insieme per estrarne valore e supportare azioni a favore della salute e del benessere degli individui.

Una *Data Driven Healthcare* presuppone l'esistenza di un'infrastruttura capillare che consenta la raccolta, il trasferimento, l'elaborazione e la fruizione dei dati.

Si tratta di un'infrastruttura che rappresenta il contenitore fisico in grado di raccogliere questi dati, trasformarli in informazioni mediante analisi sofisticate e infine abilitare la conoscenza.

L'infrastruttura deve consentire l'interoperabilità dei dati attraverso una strutturazione logica di questi: occorre un modello semantico che, a partire dallo standard FHIR, consenta la comprensione del significato dei dati contenuti per tutti i possibili utilizzi.

È suggerita l'adozione di linee guida comuni, servono modelli maturi che aiutino le organizzazioni sanitarie a dotarsi di infrastrutture adeguate e affidabili che, una volta realizzate, dispongano di strumenti che permettano un attento monitoraggio funzionale. Il tutto a vantaggio della qualità dei servizi erogati all'utenza.

È necessario prevedere l'evoluzione della Piattaforma Digitale Nazionale Dati (PDND), verso l'analisi dei big data prodotti per l'elaborazione di politiche *data-driven*.

8.5 Interdisciplinarietà e Life Data

In un mondo sempre più complesso è necessario attuare decisioni *data-driven*, basate su insiemi di informazioni interdisciplinari, sempre aggiornate, tracciabili, accessibili e trasparenti

La pandemia ha evidenziato che nel fronteggiare eventi complessi e in rapida evoluzione i dati sono fondamentali per prendere decisioni rapide e quanto più precise. Senza dati si rischia di navigare alla cieca.

Lo stato di salute e la prospettive di vita in salute dipendono da fattori (determinanti della salute) che vanno oltre i soli episodi clinici e di malattia, ma riguardano ambiti diversi, come il nostro patrimonio genetico e, in misura rilevante, una serie di fattori esogeni quali: condizioni socio-economiche e ambientali, abitudini, stili di vita e comportamenti.

Si deve passare da una logica di uso verticale dei dati finalizzata al singolo progetto, a una strategia che valorizzi standard, collaborazione e riuso dei dati in senso trasversale, multidisciplinare, verso il paradigma dei Life Data. Bisogna arrivare a sfruttare tutti i dati che possono influenzare il nostro benessere e le nostre condizioni di vita.

Il passaggio al paradigma dei *Life Data* rappresenta un punto di riferimento e di svolta nel progresso scientifico che riguarda i dati: aggiungiamo ai dati clinici, i dati genomici e i dati “di vita”.

Il nuovo paradigma pone il dato al centro in un mondo caratterizzato dall'interconnessione tra sistemi fisici e digitali. Sono sempre più diffuse analisi complesse in real-time attraverso i Big Data, Intelligenza aumentata/artificiale, Digital Twin, senza contare le enormi moli di dati generati dai dispositivi IoT. Tutto ciò indica come i dati avranno un ruolo determinante nella strategia futura volta al miglioramento delle condizioni di vita

In questa prospettiva l'evoluzione dei Sistemi per la Salute non può che estendersi ai determinanti non clinici della salute della popolazione per sviluppare politiche sanitarie sempre più rivolte al mantenimento in salute piuttosto che al ripristino della salute a fronte di episodi clinici acuti.

I Life Data consentono di utilizzare algoritmi, Intelligenza Artificiale, Machine Learning per valorizzare queste informazioni, garantire una visione complessiva ma nello stesso tempo altamente dettagliata e specialistica per ogni necessità di approfondimento di cui chi partecipa al processo di cura ha bisogno in ogni momento.

In definitiva i life data abilitano la medicina personalizzata – o di precisione – aprendo la strada a nuove tecnologie come i digital twin personali.

9. APPENDICE – BEST PRACTICES

9.1 Mission Community Hospital – una singola piattaforma software unifica tutte le risorse di storage.

Contesto

Il Mission Community Hospital fornisce assistenza sanitaria di alta qualità a prezzi accessibili alla comunità della San Fernando Valley, negli Stati Uniti, da oltre 50 anni. La struttura, che è di proprietà e gestita da Deanco Healthcare, dispone di 145 posti letto.

Quando Dustin Fennell, CIO dell'azienda sanitaria, ha iniziato a lavorare presso l'ospedale, si è subito reso conto che server ormai obsoleti e uno storage DAS (Direct-Attached Storage) non avrebbero mai potuto sostenere l'aumento di produttività ed efficienza necessario per stare al passo con i tempi, nonostante la presenza di sistemi virtualizzati.

Così, sia pure con fondi limitati a disposizione, Fennell si è messo all'opera per trovare una soluzione. La priorità assoluta era però la salute, non l'IT, con solo due persone impegnate sul fronte tecnologico: un direttore IT e un amministratore di rete, con la responsabilità di gestire, implementare e mantenere aggiornati tutti i sistemi IT dell'ospedale, inclusi i server di back-end, lo storage associato, il layer di rete e molte altre funzioni amministrative.

Soluzione

La risposta è stata trovata in SANsymphony, grazie al quale la gestione dello storage è meno laboriosa, i sistemi sono più facili da gestire e la migrazione dei dati, quando necessaria, è molto semplice. L'economicità complessiva della piattaforma di software-defined storage e la sua capacità di rendere lo storage fisico completamente "agnostico", in modo che l'hardware sia intercambiabile, sono solo due dei grandi vantaggi per il team IT dell'ospedale.

"Non sono più costretto ad adottare storage Tier-1 ovunque per soddisfare tutte le esigenze", spiega Fennell. "Ora posso limitarmi ad acquistarlo per i sistemi mission-critical, mentre per lo storage secondario - per esempio in un ambiente di sviluppo, per memorizzare immagini o per eseguire backup da disco a disco - posso utilizzare uno storage meno performante. E poiché non siamo legati a un particolare fornitore, possiamo sempre procurarci lo storage migliore e più conveniente per ciò che ci serve".

Risultati

Grazie a SANsymphony, il Mission Community Hospital ha ottenuto anche elevata disponibilità (HA): due piattaforme ridondanti di virtualizzazione dello storage risiedono su server con mirroring sincrono. Questo è fondamentale per garantire sempre l'accesso al sistema di cartelle cliniche elettroniche (EHR) dell'ospedale e ad altri sistemi chiave che risiedono sulla piattaforma di storage. Il vantaggio più significativo è la continuità operativa, che in un ospedale garantisce un'assistenza ininterrotta ai pazienti.

“Qualsiasi downtime influenzerebbe negativamente la cura del paziente,” spiega Fennell. “Averli eliminati ha risolto quello che prima era un grande problema”. La quantità di storage gestita dalla piattaforma di virtualizzazione SANsymphony è di quasi 80 terabyte, un valore destinato a crescere continuamente. “Il nostro piano è mettere tutto lo storage sotto un unico ombrello.” aggiunge Fennell.

La migrazione dei sistemi legacy si è rivelata più semplice del previsto. Lo storage X-IO utilizzato nel back-end è stato riconfigurato impostando prima una serie di macchine virtuali e poi spostandole nella nuova infrastruttura. In pratica, si è preso ciò che era in esecuzione su un server fisico, lo si è spostato (migrato) su una macchina virtuale, si è disattivato il server, lo si è riavviato e quindi riutilizzato. Il risultato è un sistema unificato, dalle prestazioni più elevate, ridondante e facile da gestire.

9.2Uk Biobank – Enabling scientific discoveries that improve human health

Contesto

L'imaging fornisce informazioni strutturali e funzionali sull'anatomia interna e sui processi fisiologici. Il suo uso nella pratica clinica ha trasformato la diagnosi, la gestione e il trattamento della malattia. L'imaging è in grado di rilevare patologie asintomatiche prima dello sviluppo della malattia e quindi può essere utilizzato per monitorare popolazioni ad alto rischio per supportare la medicina di precisione e preventiva. In alcuni casi, la rappresentazione può fornire le comprensioni nei meccanismi biologici che sottendono le associazioni di esposizione-malattia.

Studi prospettici su larga scala basati sulla popolazione possono facilitare l'identificazione delle misure di imaging come obiettivi per la prevenzione o fornire una visione dei meccanismi di malattia. Sebbene alcuni studi epidemiologici abbiano incorporato misure di imaging, queste sono state di solito limitate a una specifica modalità di imaging o regione del corpo (come il cervello o il cuore), sono stati spesso limitati a sottogruppi selettivi della popolazione ad alto rischio per talune malattie e hanno incluso non più di poche migliaia di partecipanti. Ad

esempio, i primi studi di coorte a utilizzare la risonanza magnetica (MRI), come lo studio di Rotterdam e lo studio multi-etnico sull'aterosclerosi (MESA), hanno incluso <5000 partecipanti ^{1,2}.

Tuttavia, per valutare le associazioni moderate che possono esistere tra fattori genetici e di stile di vita e di imaging derivati fenotipi (IDP), o tra IDP e conseguente rischio di una vasta gamma di malattie, è necessario eseguire l'imaging in un numero molto elevato di individui sani, in quanto solo una percentuale relativamente piccola di essi svilupperà una particolare condizione durante il follow-up. Inoltre, nell'era dei Big Data, grandi coorti di fenotipi sono essenziali per massimizzare i recenti sviluppi dell'intelligenza artificiale (IA).

Soluzione

UK Biobank è una piattaforma biomedica per la ricerca clinica su larga scala, contenente informazioni genetiche e sanitarie e di imaging approfondite di mezzo milione di cittadini volontari del Regno Unito. La banca dati è regolarmente alimentata con dati aggiuntivi ed è accessibile a livello globale ai ricercatori autorizzati che intraprendono studi clinici vitali sulle malattie più comuni e pericolose per la vita. È un importante contributo al progresso della medicina moderna e per migliorare la nostra comprensione della prevenzione, diagnosi e trattamento di una vasta gamma di malattie gravi e mortali. Ha permesso diverse scoperte scientifiche il cui obiettivo è migliorare la salute umana.

Risultati

Le risorse della Biobank del Regno Unito sono disponibili "bona fide" per tutti i ricercatori in tutto il mondo. All'inizio del 2020, più di 1.750 progetti di ricerca erano in corso alla data con ricercatori registrati provenienti da oltre 85 paesi. Molti risultati di ricerca ottenuti grazie alla UK Biobank sono già noti. Ad esempio, il diabete è stato associato con struttura anormale e la funzione in tutte e quattro le camere cardiache, mentre in precedenza solo il ventricolo sinistro è stato in genere pensato per essere colpito. Sono state notate anche altre associazioni meno ovvie, come l'inquinamento atmosferico, la terapia ormonale in menopausa e la funzione polmonare.

La risorsa della Biobanca britannica ha iniziato a donare dati al tutto il mondo della ricerca sanitaria a partire dal 2006, e continua a donarli ancora oggi!

Attualmente UKB è impegnata a raccogliere dati di imaging dettagliati, di alta qualità multi-modale sul cervello, il cuore, composizione addominale, ossa, articolazioni e vasi sanguigni su 100.000 partecipanti. UKB ha anche iniziato il processo di esecuzione di immagini ripetute su almeno 10.000 partecipanti. La quantità di dati di imaging raccolti su un numero così elevato di partecipanti è davvero unica. Eppure, è la combinazione di questi dati con la ricchezza di altre

informazioni fenotipiche, genetiche e mediche disponibili in UKB che fornisce una risorsa potente per affrontare questioni di ricerca in precedenza senza risposta.

Tradizionalmente, i dati di rappresentazione potrebbero essere percepiti per essere di valore principalmente agli specialisti in una gamma ristretta di campi. Tuttavia, i ricercatori provenienti da molte discipline, tra cui, ma non solo, epidemiologi, neuroscienziati, statistici, genetisti e psicologi, possono e utilizzano gli IDP (imaging-derived phenotypes) già disponibili dalle scansioni di imaging per condurre salute-ricerche correlate per fornire nuove informazioni sulla prevenzione, la diagnosi e il trattamento delle malattie.

9.3 Mayo Clinic – Clinical Data Platform

Contesto

Mayo Clinic è un'organizzazione no-profit impegnata nell'innovazione nella pratica clinica, nell'istruzione e nella ricerca ed ha più di 150 anni di esperienza in medicina e una grande reputazione per la cura del paziente, sia in eccellenza e integrità⁴³. Valori e visione sono allineati per trasformare l'assistenza sanitaria attraverso l'innovazione. Mayo Clinic ha voluto lanciare una iniziativa di Clinical Data Platform in grado di sostenere i più alti standard di eccellenza nell'innovazione, mantenendo la privacy e mettendo al primo posto le esigenze del paziente.

Soluzione

La Clinical Data Analytics Platform è la prima iniziativa annunciata nell'ambito della Mayo Clinic Platform ancora prima dell'era Covid, un'iniziativa strategica per migliorare l'assistenza sanitaria attraverso approfondimenti e conoscenze derivanti dai dati. La piattaforma utilizza analisi avanzate su tutti i dati de-identificati di Mayo Clinic per promuovere la medicina e migliorare la salute dei pazienti; le aree principali sono legate alla analisi dei dati clinici per accelerare la scoperta e lo sviluppo dei farmaci in tutto l'ecosistema biofarmaceutico: si concentra infatti sull'identificazione degli obiettivi e dei biomarcatori per i nuovi farmaci, sulla corrispondenza ottimale dei pazienti con un regime terapeutico e real world data e applicazioni real world evidence, quali l'espansione delle label, la sorveglianza post-commercializzazione e l'uso di farmaci.

Risultati

Precision Phenotype and Digital Health

⁴³ <https://www.mayoclinic.org/about-mayo-clinic>

Le cartelle cliniche dei pazienti de-identificate sono ricche di informazioni necessarie: fenotipi, co-morbidità, risultati di cura, per migliorare il modo in cui curiamo i pazienti con nuovi farmaci, diagnosi migliori, ecc. Tuttavia, queste informazioni sanitarie che catturano la patologia biologica di base sono spesso presenti in una forma non strutturata (note cliniche, riassunti di scarico, rapporti di patologia, immagini di patologia, immagini di radiologia, ecc) normalmente destinato a un esperto (patologo/medico scienziato) da interpretare. Per consentire l'interpretazione e l'analisi di cartelle cliniche de-identificate su larga scala, le informazioni non strutturate nella Clinical Data Platform vengono prima strutturate e poi utilizzate. Poche altre iniziative e aziende stanno attualmente affrontando il problema utilizzando il data entry manuale, ma questa è stata una grande strozzatura che ha limitato la portata e l'impatto di queste importanti informazioni. La tecnologia sottostante la Clinical Data Platform di Mayo Clinic rende computabile la conoscenza non strutturata e consente la triangolazione senza soluzione di continuità con vari database strutturati (come parametri vitali, test di laboratorio, codici ICD10, varianti genetiche) che consentono la fenotipizzazione di precisione su una scala senza precedenti.

Molecular Characterisation and Digitization

Uno dei principali punti di forza della Mayo Clinic è il suo grande deposito di campioni. Biospecie, diapositive patologiche e altri dati donati dai pazienti stanno alimentando continuamente la piattaforma. Mayo intende condurre una caratterizzazione su larga scala di queste biospecie utilizzando un'ampia gamma di test come l'intero sequenziamento di esomi, RNA seq a singola cellula, Cytof... così come la digitalizzazione dei vetrini di anatomia patologica di Mayo stessa. Accanto ad altri importanti sforzi clinico-genomici in corso in tutto il mondo (come UK Biobank), Mayo Clinic è sulla via per generare il più grande e ricco database di dati clinico-genomiche de-identificate collegando i campioni molecolarmente caratterizzati di Mayo a fenotipizzazione di precisione guidata dalla raccolta automatizzata dei dati dalle cartelle cliniche.

9.4 National Health Service del Galles (NHS Wales) – National Data Resource

Contesto

Il National Health Service del Galles (NHS Wales) fornisce assistenza sanitaria universale a circa 3 milioni di residenti. Nel 2014, il ministro gallese per la salute ha lanciato una politica chiamata Prudent Health Care focalizzata su: co-creazione con i pazienti, equità (accesso alle cure prioritario per chi ha maggiori bisogni) e riduzione del ricorso eccessivo alle cure mediche (medicina difensiva). Per attuare il piano, NHS Wales ha affidato ad alcuni medici nei vari Health Board

locali un ruolo dedicato al Value Based Healthcare. Per il coordinamento a livello nazionale è stato creato un piccolo gruppo direttivo operativo. Negli anni gli sforzi in questa direzione si sono accentuati, assumendo un ruolo prioritario nell'agenda politica del Paese. Ad oggi, in ambito europeo, il Galles appare il caso più strutturato e organico di VBHC promosso da un sistema sanitario pubblico.

Soluzione

In una prima fase sono state lanciate diverse sperimentazioni sul campo, facendo leva su standard comuni e riconosciuti come quelli ICHOM. Un progetto di particolare interesse è stato condotto dall'Health Board ABUHB sulla Malattia di Parkinson. Attraverso la formazione dei team, il ridisegno dei processi e la ristrutturazione dei supporti informatici, ABUHB ha implementato un set di indicatori a supporto degli obiettivi di VBHC. A nove mesi dal lancio del progetto si è raggiunto l'obiettivo di una corretta raccolta della quasi totalità delle informazioni, principalmente attraverso questionari compilati dai pazienti. Le informazioni sono state poi digitalizzate e rese fruibile a pazienti e medici.

Dopo la fase iniziale caratterizzata da progetti su base locale, nel 2017 NHS Wales ha sviluppato un portale nazionale per la raccolta di PROM (Patient Reporte outcome Measures) standardizzati per 31 percorsi assistenziali di diverse patologie. NHS Wales ha anche avviato la costruzione della National Data Resource, un hub di interoperabilità di dati basato su API che consente un *benchmarking* degli operatori e la possibilità di consultare reportistica, ricerche e supporti clinico-operativi. I dati non sono stati confinati ai solo addetti ai lavori, ma sono rivolti all'intera popolazione in una logica di comunicazione e di sensibilizzazione. Ad esempio, i cruscotti di indicatori consentono ai cittadini gallesi di confrontare in modo trasparente operatori e risultati (i primi piloti si sono concentrati su tumore ai polmoni e insufficienza cardiaca).

Attraverso il dialogo con professionisti e associazioni mediche, NHS Wales ha raggiunto un forte consenso clinico sui risultati da considerare e sugli indicatori PROM. I costi sono misurati con il metodo TDABC (Time Driven Activity Based Costing), ove possibile, a livello di singolo paziente. L'approccio di lavoro "Plan-then-Do" è stato sostituito da quello "Decide-Do/Refine-Do", più orientato all'azione, agile e iterativo.

NHS Wales sta anche sviluppando alcune Learning Community per disseminare i principi e i contenuti del VBHC, facilitandone l'adozione sul campo. Ad esempio, ha lanciato un programma (Finance Academy), in cui esperti di finanza e di gestione aziendale lavorano a stretto contatto con i clinici per ideare e implementare progetti locali basati sul valore.

Risultati

Il percorso del Galles verso il VBHC è in corso, con il Piano 2019-2022 nel cuore della sua attuazione. Tra i principali prossimi passi, il governo intende rafforzare le partnership con operatori privati secondo logiche Value Based. Già in passato NHS Wales aveva definito nuove formule contrattuali in cui i fornitori erano remunerati al raggiungimento di obiettivi predeterminati; appare ora necessario un ulteriore passo per una più forte integrazione e collaborazione.

Inoltre, si intende proseguire nel rafforzamento delle attività di comunicazione e di coinvolgimento: si prevede ad esempio un Roadshow Value Based Healthcare indirizzato a tutte le amministrazioni sanitarie locali, oltre a conferenze annuali con respiro internazionale e a iniziative quali “Value in Health Week” che celebra l'adozione del VBHC in Galles.

9.5 Regione Lombardia – Alfabeto (ALI FASTER BETTER TOGETHER)

Contesto

La situazione epidemiologica in atto nel corso del 2020 nella Regione Lombardia ha richiesto fin da subito un potenziamento delle attività di sorveglianza e presa in cura, svolte nel territorio.

Regione Lombardia nell'ambito dei Fondi POR FESR 2014-2020 AZIONE I.1.B.1.3 ha emesso una misura a sostegno dello sviluppo di collaborazioni per l'identificazione di terapie e sistemi di diagnostica, protezione e analisi per contrastare l'emergenza coronavirus e altre emergenze virali del futuro. L'obiettivo della misura è stato quello di finanziare proposte progettuali volte a far progredire la conoscenza della SARS-CoV-2 della più ampia famiglia di Coronavirus, ovvero in generale allo scopo di contribuire a una gestione efficiente del paziente e/o alla capacità di preparazione e risposta a focolai attuali e futuri, come da orientamenti comunitari.

Il progetto ALFABETO si è collocato nell'ambito di intervento della misura, con l'obiettivo principale di coadiuvare il personale sanitario nella fase di triage iniziale presso l'abitazione del paziente, attraverso una valutazione oggettiva del grado di severità della patologia e un'analisi predittiva della possibile evoluzione nel breve-medio termine, fornendo quindi gli elementi essenziali per decidere il livello e la strategia di assistenza da attuare (cura a domicilio vs ospedalizzazione).

ALFABETO nasce dall'unione di tre Partner (Università di Pavia – Dipartimento di Scienze Cliniche Chirurgiche, Diagnostiche e Pediatriche, ICS Maugeri IRCCS e un'azienda socia di Anitec-Assinform) le cui specializzazioni e competenze hanno consentito una elevata sinergia in grado di fornire una soluzione concreta

e rapida, in linea con gli obiettivi dell'ambito di intervento e di attivare un percorso di sperimentazione retrospettiva del progetto e validazione dei risultati direttamente all'interno del progetto stesso.

Soluzione

Il progetto ha previsto lo sviluppo di una piattaforma software basata su componenti di Artificial Intelligence (Machine Learning e Deep learning), finalizzata al supporto alla diagnosi e alla condivisione delle informazioni cliniche in real-time con i diversi professionisti sanitari coinvolti nel processo e non necessariamente presenti al domicilio del paziente, l'integrazione di fonti eterogenee di dati, il supporto al personale clinico (Radiologo) nel definire il grado di priorità del caso per la diagnosi attraverso l'analisi di immagini RX con relativa misurazione dell'area polmonare interessata, l'utilizzo di modelli predittivi specifici per diagnosticare la possibile evoluzione della patologia, la condivisione real time con il Medico di Medicina Generale, Specialisti Ospedalieri e il personale al domicilio del paziente, costituiscono gli elementi qualificanti del progetto.

Alfabeto è una piattaforma basata su sistemi di Intelligenza Artificiale volta a supportare i processi decisionali dei medici nel contesto COVID ed è in grado di:

- Guidare in modalità data-driven i processi decisionali del medico integrando tutte le informazioni cliniche dei sospetti Covid-19, dall'anamnesi, ai dati FSE, alla cartella MMG;
- Indicare il grado di priorità al radiologo (che opera in remoto) per la refertazione della RX Torace effettuata al domicilio;
- Sviluppare un Triage "dinamico" mediante motori di A.I. applicati ai dati clinico- anamnestici e di contesto del sospetto Covid-19;
- Relazionare tutti gli operatori (USCA, MMG, Specialisti Ospedalieri) attraverso strumenti di collaboration integrati nella piattaforma e aderenti ai requisiti di privacy;
- Sviluppare analisi predittive sulla evoluzione della patologia supportando gli operatori nella decisione diagnostica (Ricovero vs Assistenza domiciliare).

Alfabeto è composto da due differenti moduli di AI: il primo, costituito da X-RAIS tecnologia proprietaria del partner tecnologico, in grado di estrarre informazioni da immagini radiologiche; la seconda, un modello previsionale, che sulla base delle variabili estratte da X-RAIS e dei dati di anamnesi e comorbidità del paziente è in grado di predire l'evoluzione della patologia.

Risultati

Il sistema è stato addestrato utilizzando tecniche di AutoML che hanno permesso di testare un numero considerevole di diversi modelli (più di 50). L'intero sistema è stato sviluppato secondo i paradigmi Cloud First, Once-only e Security by design e permette in modalità data-driven di guidare i processi di sistemi già esistenti nei vari ospedali quali Emergenza Urgenza, Enterprise Imaging, Radiology Information System e sistemi di Telemedicina e Telemonitoraggio per la gestione del paziente a domicilio.

La piattaforma è stata utilizzata in sperimentazione prospettica al Pronto Soccorso dell'IRCCS San Matteo di Pavia dopo che, i brillanti risultati delle fasi precedenti di studio retrospettivo hanno evidenziato una riduzione di circa il 47% dei pazienti in accesso al PS poiché correttamente individuati e trattati dal sistema direttamente a domicilio.

9.6 Agenzia Regionale per l'innovazione e gli acquisti (ARIA SpA) – Digital Information Hub (DIH)

Contesto

Il Digital Information Hub (DIH), è un polo unico di raccolta di tutti i dati, da quelli socioassistenziali e sanitari e quelli legati al mondo del lavoro, che mette a disposizione questa conoscenza a tutti gli stakeholder (centri di ricerca, ospedali, università, ma anche industrie e startup), al fine di governare al meglio le politiche della Regione Lombardia, migliorare il sistema degli acquisti, potenziare il sistema decisionale e ottimizzare la spesa pubblica⁴⁴.

Vengono valorizzati i dati raccolti, organizzati e correlati nel data warehouse di Regione Lombardia, con una profondità di 20 anni, e vengono messi a disposizione degli enti di ricerca, delle università italiane e internazionali e degli enti del sistema regionale, per progetti di ricerca di carattere scientifico.

La sfida del DIH è di spingere l'organizzazione verso il potenziamento della data governance in ottica integrata. Lavorare sui dati per una Pubblica Amministrazione comporta attuare azioni di *data security* e *regulation* che indirizzino sia l'accesso al patrimonio informativo in modo sicuro sia una revisione e un costante aggiornamento delle basi giuridiche regionali, in coerenza con la normativa nazionale ed europea. Nel percorso di *change management*, è fondamentale indirizzare un cambiamento culturale nell'organizzazione

⁴⁴Il caso studio è stato illustrato anche nell'articolo "Il caso del digital information hub di Aria SpA". <https://www.forumpa.it/pa-digitale/governance-by-data-il-caso-del-digital-information-hub-di-aria-spa/>

aziendale a livello di processi e servizi: lavorare insieme, sbagliare insieme, imparare sul campo come i dati possono quotidianamente aiutare a risolvere problemi e a creare nuove opportunità.

Soluzione

La piattaforma, definita DaaS, Data-as-a-service, rappresenta anche un modello organizzativo condiviso con la Regione Lombardia e oggetto di alcune delibere e di un'interazione con l'Autorità Garante della Privacy, grazie alla quale vengono messi a disposizione, in una sandbox segregata e ad accesso controllato, i dati sanitari presenti nel Data Warehouse e gli strumenti di data analysis per la loro elaborazione alle Università e agli Enti di ricerca, garantendo la compliance alla base giuridica e il rispetto delle misure di sicurezza ritenute più idonee.

In conformità con la normativa di privacy vigente, il DIH garantisce dati completi costantemente aggiornati e di qualità riguardanti più di 10 milioni di cittadini, con annessi strumenti di analisi e/o servizi di configurazione di ambienti di lavoro dedicati all'analisi degli stessi. Questi dati, provenienti da diverse fonti in stato ancora grezzo, vengono inizialmente raccolti in linea con la normativa, per poi essere successivamente inseriti in DATAKit che, su richiesta specifica dell'ente interessato, vengono caricati e messi a disposizione per analisi ed elaborazioni.

Per far questo ARIA ha coordinato team con differenti skill e responsabilità rispetto alle specifiche necessità, abbattendo barriere culturali e organizzative che impedivano la condivisione e comprensione del dato. Inoltre, in questo contesto, è stato rilevante il processo di responsabilizzazione delle figure che generano dati che con il tempo sono stati spinti ad attuare sempre più azioni di data quality.

La piattaforma è stata anche dotata, grazie a specifici investimenti a livello di piattaforme tecnologiche, per traguardare due obiettivi: l'introduzione di tecnologie innovative per l'elaborazione di dati e informazioni (intelligenza artificiale, machine learning...) al fine di poter effettuare studi e analisi sempre più mirate su svariati ambiti, e l'opportunità per gli enti del Sistema Regionale di accedere in modo sicuro al patrimonio informativo regionale, creando valore e conoscenza dal dato. In tal senso, ARIA ha messo a disposizione degli enti di ricerca accreditati dalla Regione Lombardia (Deliberazione Giunta Regionale 491/18) una piattaforma attraverso la quale è possibile accedere a dati regionali per realizzare progetti con finalità di monitoraggio e programmazione in maniera sicura.

Risultati

A valle delle scelte fatte, è possibile intravedere i primi risultati. Fino a questo momento, sono stati richiesti e valutati positivamente da Regione Lombardia 45

progetti di ricerca con + 20 enti e università all'interno del modello DIH. Di questi 13, già 7 sono stati attivati e avviati e sono state svolte.

Il servizio conta oltre 20 enti accreditati tra università ed enti sanitari pubblici e privati, oltre che essere utilizzato dall'Organizzazione Mondiale della Sanità. I progetti attivati sono oltre 40, di cui 15 in ambito COVID e diverse altre attività di analisi dei dati in collaborazione con l'Osservatorio Epidemiologico Regionale. Le risorse stanziare per la realizzazione dell'iniziativa arrivano da fondi europei e fondi regionali, per un totale di circa 5 milioni di euro all'anno fra investimenti, costi ricorrenti e attività professionali.

Tra questi, ci sono i progetti di Regione Lombardia di analisi predittiva dei bisogni di salute della popolazione regionale e quello che ha portato alla riorganizzazione dei modelli di presa in carico dei pazienti cronici e fragili.

Per supportare un progetto di questo tipo sono necessarie buone politiche di change management, a partire dai piani di formazione fino alla creazione di gruppi di lavoro specializzati, per dominare le difficoltà derivanti dal coinvolgimento di un numero di attori così ampio ed eterogeneo, oltre che dall'utilizzo di tecnologie avanzate.

In prospettiva, si guarda alla possibilità di replicare tale modello tecnico-organizzativo in altri territori o magari anche a livello nazionale, partendo dalla consapevolezza che l'evoluzione di un servizio come il DIH è legata alla disponibilità di un sempre maggiore flusso di dati, derivante da potenziamenti tecnologici o da un ulteriore allargamento degli enti coinvolti.

La declinazione della data governance in ambito istituzionale sarà una delle sfide più decisive della PA digitale del futuro. Contemperare i benefici delle analisi data-driven in ambito pubblico con il governo delle regole e dei processi alla base di tali analisi diventerà un asset sempre più essenziale. Il DIH va in questa direzione.

9.7 AI Score – Ospedale San Raffaele: una piattaforma di apprendimento autonomo per consentire interventi sanitari mirati e tempestivi e ridurre l'impatto sul sistema sanitario del COVID-19

Contesto

AI Score è un progetto che ha coinvolto l'IRCCS Ospedale San Raffaele, nato dall'esigenza di dover affrontare in tempi rapidi una patologia del tutto sconosciuta come il COVID-19. I medici osservavano sul campo forme lievi o più gravi e, nello stesso paziente, sviluppi della sintomatologia da blanda a grave

insufficienza respiratoria. Il virus si è presentato come una polmonite, ma in seguito ha dimostrato di poter colpire anche altri organi, e di portare a forme diffuse di trombosi a livello polmonare. Man mano, l'ospedale ha potuto raccogliere dati e parametri clinici che, col passare del tempo, andavano sempre più aumentando e stratificandosi. Una mole enorme di informazioni che, se confrontate tra loro, avrebbero permesso di identificare dei "valori standard" e di poter stabilire, nel momento in cui il paziente si presentava in ospedale, se avrebbe sviluppato una forma gravissima, una forma grave oppure una forma lieve di Covid-19; di conseguenza, sarebbe stato possibile programmare un'assistenza domiciliare, un'ospedalizzazione oppure il ricovero in terapia intensiva. Questo "schema" è parso chiaro fin dall'inizio.

Soluzione

Si è disegnata una soluzione per individuare, già in fase iniziale della malattia, quei parametri anamnestici, clinici e di imaging di un paziente Covid-19, correlabili a un indice prognostico. Per "indice prognostico" si intende la possibilità di stabilire, nel momento in cui il paziente si presenta in ospedale con i primi sintomi, quale sarà il suo decorso.

Sono stati individuati circa cento parametri, in parte derivanti dalla esperienza dell'IRCCS San Raffaele e in parte dalla letteratura ed è stato definito un algoritmo di Machine Learning che consentisse, tra tutti questi parametri, di identificare quali fossero i più predittivi del decorso della patologia. Tra i parametri individuati, la maggior parte sono relativi alle immagini diagnostiche, in quanto queste hanno la possibilità di misurare non solo i valori del polmone, ma anche valori in merito ad alterazioni cardio-vascolari, risultate fondamentali nel comprendere l'evoluzione dell'infezione da Coronavirus. Tra i parametri è inserito anche il profilo genomico dei pazienti, perché si pensava potesse esserci una predisposizione genetica, da parte di taluni soggetti, a sviluppare malattie correlate a un rischio maggiore di patologia grave. Individuati i parametri e definito il lavoro dell'algoritmo, sono stati raccolti – nel totale rispetto della privacy e in modo del tutto anonimo – dati da circa duemila pazienti, reclutati tra l'Ospedale San Raffaele, l'Ospedale Bolognini di Seriate e il Centro Cardiologico Monzino, consorziati nell'ambito del progetto.

Risultati

I dati ci confermano che l'applicazione di questa soluzione ha permesso di ridurre il tasso di ospedalizzazioni del 25%.

Il valore della soluzione è dato dalla scelta dei parametri che abbiamo deciso di analizzare e di correlare tra loro: parametri quasi omnicomprensivi. Per questo, si ritiene che lo schema di tale progetto, la sua metodologia, possano essere applicati allo studio di altre patologie e di altre eventuali forme epidemiche.

9.8 Swarm Intelligence – Unanimous AI

Contesto

Il settore sanità genera una enorme quantità di dati: gran parte di questi dati sono contenuti in silos, sia dal punto di vista “fisico” sia, più rilevante, dal punto di vista logico. Negli Stati Uniti l’adozione dell’EHR (Electronic Health Record) è una realtà ormai consolidata (introdotto a inizio secolo è arrivato a una penetrazione nel 2017 di oltre 85%, di cui oltre il 75% certificato che con la pandemia ora sfiora il 100%). I medici sono obbligati a trascrivere sull’EHR tutte le visite, esami, medicine dei loro pazienti (addirittura questo ha stimolato un mercato di trascrizione a distanza con il medico che “detta” diagnosi, risultati, prescrizioni a un sistema che viene utilizzato da personale specializzato in India per la trascrizione sull’EHR; oggi si sta andando verso sistemi di trascrizione automatizzata attraverso IA per riconoscimento del parlato).

Il contesto è quindi di elevata digitalizzazione ma la capacità di sfruttare i dati è ben al di sotto di quello che potrebbe essere il suo potenziale.

Questo è stato colto da Unanimous AI⁴⁵, un’azienda che ha sviluppato una tecnologia, chiamata “swarm intelligence” in grado di mettere insieme una molteplicità di “intelligenze” includendo in queste sia intelligenza artificiale sia naturale. Nel contesto specifico, quindi, siamo in uno scenario con migliaia di medici, ciascuno con la sua esperienza e capacità di analisi e decisione, milioni di articoli nel settore, migliaia di analisi di farmaci (applicazione, risultati e effetti collaterali).

Soluzione

Unanimous AI mette a disposizione la sua tecnologia per consentire a ciascun medico di:

- condividere le sue esperienze (nel momento esatto in cui queste si formano, quindi attraverso anamnesi del paziente, ipotesi di diagnosi, risultati esami, conferma diagnosi, cura, monitoraggio risultati inclusi effetti collaterali);
- accedere alla intelligenza distribuita, sviluppata attraverso Swarm AI, accendendo a una “sintesi” di esperienze (questo è l’elemento che differenzia Unanimous AI da un servizio di ricerca su una banca dati) che sembrano essere pertinenti al caso in oggetto;

⁴⁵ <https://unanimous.ai>

- mantenersi aggiornato sui risultati più recenti a livello di ricerca, genetica, diagnostica, farmacologia, ambiente e fattori di rischio;
- ricevere segnalazioni puntuali su possibili fattori di rischio di rilievo nel contesto in cui opera (contesto che viene continuamente, e automaticamente, aggiornato attraverso le interazioni del medico con il sistema).

L'interazione con il servizio avviene tramite linguaggio naturale, il medico "parla" e "ascolta" come in un normale consulto tra professionisti. È inoltre possibile farsi "leggere" i risultati degli esami dal sistema per avere indicazioni ulteriori (ad esempio vengono forniti servizi di lettura di radiografie⁴⁶, TAC, ...).

Le interazioni sono analizzate in automatico dal sistema consentendo un continuo affinamento del servizio.

Risultati

È stato riscontrato un deciso miglioramento in diversi aspetti:

- maggiore rapidità nella diagnosi, specialmente per i casi che escono dalla pratica usuale (il medico ha incontrato pochi/nessun caso ma tra le migliaia di medici che partecipano si sono avuti molti casi)
- maggiore focalizzazione della cura (per la possibilità di accedere a casistiche simili valutandone i risultati e la loro adattabilità al caso specifico)
- minori costi sia in termini di sistema sanitario (Medicare, Medicaid e assicurazioni private) sia per il paziente in quanto le cure risultano più mirate ed efficaci
- formazione continua "*on-the-job*" visto che ogni caso, in principio, può avvalersi di una consulenza automatizzata senza costi aggiuntivi (il medico sottoscrive un abbonamento)
- automatizzazione del processo di raccolta dati e uniformità della loro memorizzazione che consente di sviluppare ulteriori servizi minimizzando i costi (il medico in sostanza descrivendo -verbalmente- il caso in esame effettua anche l'input dei dati e in contemporanea riceve suggerimenti

⁴⁶<https://11s1ty2quyfy2qbmao3bwxyz-wpengine.netdna-ssl.com/wp-content/uploads/2021/11/Radiology-Case-Study-Swarm-AI.pdf>

instaurando quindi un circolo virtuoso: più uso il sistema più il sistema si arricchisce e più ricevo informazioni puntuali immediatamente utilizzabili)