

ALLEGATO A

AZIONE 1.2.2 DELL'ASSE 1 DEL PR FESR 2021-2027. PROGRAMMA DELLE INIZIATIVE RELATIVE ALLO SVILUPPO DI UNA INFRASTRUTTURA REGIONALE PER L'ANALISI DEI DATI ATTRAVERSO IL RICORSO ALL'INTERNET OF THINGS (IOT) E AI BIG DATA.

Regione Lombardia intende mettere in atto un programma di investimenti con il fine di sviluppare una infrastruttura regionale per facilitare l'accesso e l'analisi del proprio patrimonio informativo consentendone la promozione e la condivisione con diverse tipologie di stakeholder. Il progetto prevede l'introduzione di una piattaforma trasversale e omogenea rispetto a tutti gli ambiti informativi gestiti da Regione Lombardia, al fine di ottimizzare le analisi sull'intero patrimonio digitale regionale.

L'iniziativa progettuale si pone i seguenti obiettivi:

- Favorire l'introduzione e il costante sviluppo di soluzioni tecniche a supporto dell'integrazione e valorizzazione del Patrimonio Informativo per migliorare la capacità programmatoria "Data-Driven" in coerenza con il Programma Regionale di Sviluppo Sostenibile (PRSS);
- Evolvere gli strumenti integrati di analisi dei dati per supportare le politiche istituzionali e le decisioni implementando tecnologie innovative (es. Big Data, Predictive Analytics, Artificial Intelligence, IoT, Blockchain, Intelligent Process Automation, ecc.);
- Potenziare le soluzioni organizzative e tecnologiche di governo del Patrimonio Informativo Regionale e locale per garantire la conoscenza, la sicurezza, la qualità e l'accessibilità del dato agli utenti;
- Favorire la realizzazione di soluzioni a supporto dell'interoperabilità tecnologica e lo sviluppo di Ecosistemi Digitali per la condivisione di servizi e dati con soggetti pubblici e privati;
- Realizzare interventi per ottimizzare i servizi infrastrutturali, soprattutto in ottica "cloud", e i servizi di connettività, per aumentare la qualità dei servizi e abilitare nuove modalità di lavoro.

Gli obiettivi descritti verranno perseguiti congruentemente alle esigenze di tutti gli stakeholder mettendo a disposizione dati facilmente accessibili e certificati, in conformità con le linee guida riguardanti la sicurezza dei dati e delle informazioni definite dagli organi europei e nazionali competenti.

Compatibilmente con gli obiettivi dell'azione 1.2.2 del PR FESR 2021-2027 ("*Sviluppo di una Infrastruttura regionale per l'analisi dei dati attraverso il ricorso all'Internet of Things (IOT) e ai Big Data*") e tenuto conto dei criteri di selezione dell'ASSE 1 - Un'Europa più competitiva e intelligente - Obiettivo specifico 1.2.) Permettere ai cittadini, alle imprese, agli organismi di ricerca e alle amministrazioni pubbliche di cogliere i vantaggi della digitalizzazione - azione 1.2.2. sviluppo di una infrastruttura regionale per l'analisi dei dati attraverso il ricorso all'internet of things (iot) e ai big data del PR FESR 2021-2027, Regione Lombardia si propone di sviluppare una Data Platform regionale e di realizzare i Digital Asset a supporto dei processi per la generazione di valore attraverso l'impiego di dati e informazioni.

La Data Platform (descritta in fig. 1) è una piattaforma tecnologica integrata che garantisce l'integrità, la coesistenza e l'interoperabilità di dati provenienti da differenti Data Source e permette ai Data User di ampliare le modalità di consumo e utilizzo dei dati. La piattaforma prevede la raccolta e messa a disposizione del dato attraverso i layer di **Data Ingestion e Data Provisioning** che comprendono funzionalità di Data Capture/Provisioning avanzate (es. real-time). L'archiviazione dei dati è garantita dal layer **Data Storage and Processing** che prevede la creazione di spazi comuni di dati eterogenei, cioè senza specifici vincoli sugli ambiti informativi gestiti. L'evoluzione delle funzionalità di **Data Visualization** abilita la trasformazione in data-driven company attraverso la possibilità di eseguire analisi avanzate sui dati raccolti e supportare il decision making. La Data Platform prevede soluzioni tecnologiche che hanno l'obiettivo di garantire un patrimonio informativo di qualità e consentire analisi rapide, rigorose e innovative utilizzando nuovi set di dati e strumenti analitici. L'accesso al patrimonio informativo da parte dei Data User è garantito dal layer di **Data Consumption** che supporta il programma di trasformazione digitale di Regione Lombardia attraverso lo sviluppo e la promozione degli ecosistemi digitali sul territorio. La Data Platform comprende infatti soluzioni dedicate alla

pubblicazione e condivisione di dati certificati, integrati e ottimizzati per consentire analisi avanzate da parte di diverse tipologie di stakeholder. La Data Platform prevede inoltre l'evoluzione del layer di **Knowledge Management and Compliance** attraverso l'introduzione di soluzioni tecnologiche dedicate alla gestione del dato in un ecosistema dati integrato.

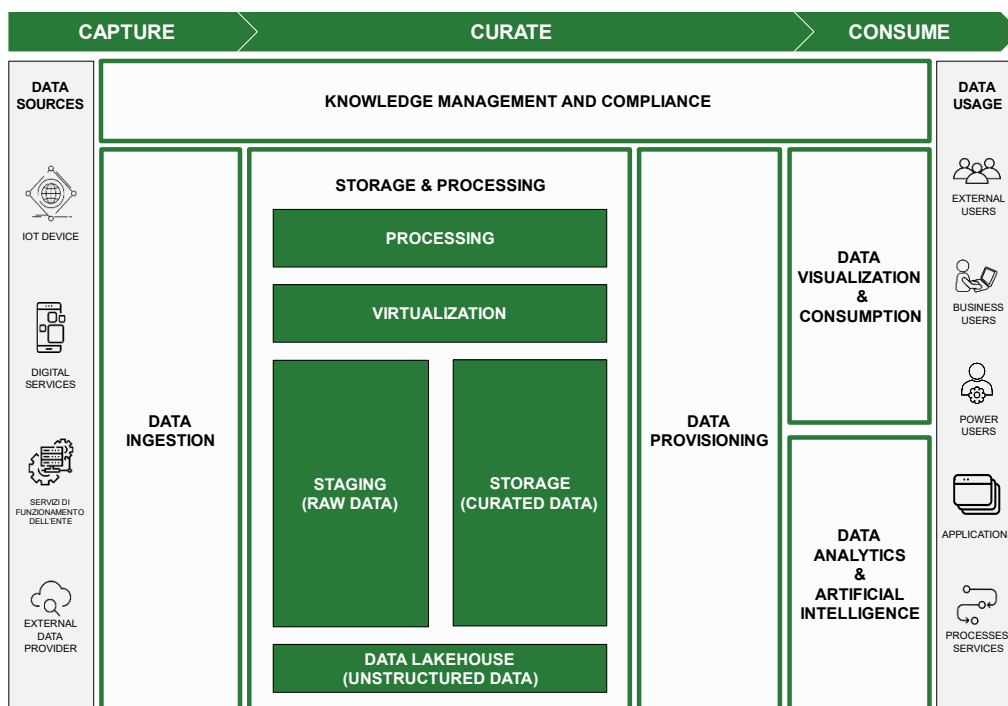


Fig. 1: Architettura logica della Data Platform regionale

Il programma di sviluppo di una infrastruttura regionale per l'analisi dei dati comprende le attività di realizzazione dei **Digital Asset** a supporto dei processi di generazione di valore attraverso l'impiego di dati e informazioni (Data Value Chain). Tali attività sono incluse nel programma poiché supportano l'adozione di un approccio Data-Driven che richiede di rispondere tempestivamente ed efficacemente, internamente ed esternamente, a nuove esigenze e scenari di business.

I Digital Asset consentono di:

- Migliorare il rendimento e ridurre il tempo di attraversamento della Data Value Chain riducendo gli sprechi di risorse e ottimizzando i passaggi tra le fasi della catena del valore in ottica agile e/o Lean;
- Supportare la raccolta, l'integrazione e la preparazione delle informazioni utili alle analisi che abilitano la valorizzazione del Patrimonio Informativo.

Il nuovo modello infrastrutturale dovrà capitalizzare e integrare gli investimenti pregressi e i nuovi investimenti andranno valutati in un'ottica di armonizzazione e omogeneizzazione dell'assetto tecnologico. Inoltre, l'implementazione di nuove capabilities tecnologiche richiede l'introduzione di nuove competenze organizzative all'interno di Regione Lombardia e di ARIA S.p.A., con investimenti in comunicazione e aggiornamento del know-how ad oggi presente.

Il programma dell'iniziativa

Gli obiettivi del programma di sviluppo di una infrastruttura regionale per l'analisi dei dati saranno definiti secondo un orizzonte temporale quinquennale e dettagliati all'interno di incarichi tra Regione Lombardia e ARIA S.p.A.. Il programma complessivamente prevede due filoni progettuali (fig. 2):

- Sviluppo ed evoluzione, nel rispetto della normativa in materia di privacy e data security, di una piattaforma tecnologica (Data Platform) che mette a disposizione strumenti e servizi a supporto della Data Supply Chain (Capture, Curate, Consume);
- Realizzazione dei Digital Asset a supporto della Data Value Chain, ovvero dei processi per la generazione di valore attraverso l'impiego di dati e informazioni, sia a supporto dell'attuazione della programmazione regionale che a beneficio di stakeholder ed utilizzatori esterni pubblici e privati.

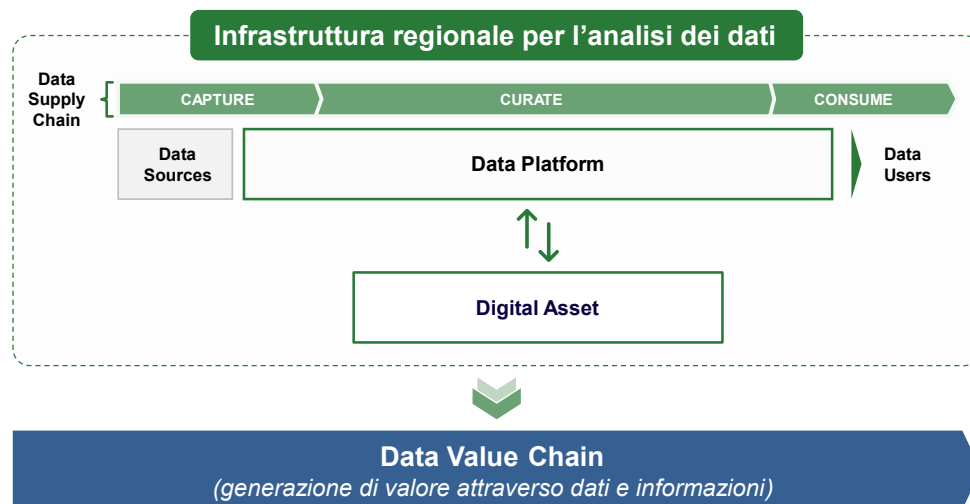


Fig. 2: L'infrastruttura regionale per l'analisi dei dati

SVILUPPO ED EVOLUZIONE DELLA DATA PLATFORM

Il filone progettuale comprende le seguenti attività:

- Acquisizione, configurazione ed evoluzione degli strumenti e dell'infrastruttura tecnologica;
- Governance complessiva delle iniziative progettuali;
- Comunicazione agli utenti e dissemination degli strumenti.

Il filone progettuale si suddividerà, indicativamente, in 4 “cantieri” tra loro integrati in una visione evolutiva comune ma ognuno con una propria autonomia, di seguito dettagliati.

1. **Data Ingestion, Storage, Processing e Provisioning:** il cantiere prevede l'evoluzione delle piattaforme/sistemi che consentono di archiviare, integrare, preparare e mettere a disposizione i dati in maniera rapida ed efficiente, con prestazioni, scalabilità e sicurezza ottimali.

Data Ingestion e Data Provisioning

Evoluzione delle modalità attualmente utilizzate per la raccolta e la messa a disposizione dei dati attraverso l'introduzione di strumenti di Data Streaming. La nuova piattaforma abiliterà la gestione di dati in real-time e permetterà di evolvere le funzionalità attualmente utilizzate per la Data Ingestion e il Data Provisioning in modalità batch.

Data Lakehouse

L'introduzione della nuova piattaforma prevede la realizzazione di un repository centralizzato che raccoglie e archivia grandi volumi di dati nella loro forma originale. Il sistema garantisce la disponibilità e l'integrità dei dati per essere elaborati e utilizzati come base per diverse esigenze analitiche.

Data Virtualization

Introduzione di una piattaforma che consente di eseguire query su data source distribuiti, per creare viste virtuali integrate di dati in-memory, mettendo a disposizione un repository di metadati (Data Catalog) e un query-engine per supportare la Data Consumption.

La piattaforma offrirà un approccio unico all'integrazione e alla gestione dei dati, agendo come un vero e proprio sistema di ricerca e scoperta sui dataset.

2. **Data Visualization & Consumption:** il cantiere prevede l'evoluzione delle piattaforme che consentono l'accesso al patrimonio informativo da parte di Data User esterni e degli strumenti che supportano le analisi sui dati e la visualizzazione dei relativi risultati.

Data Visualization

Ottimizzazione delle soluzioni attualmente presenti con l'obiettivo di ampliare l'utilizzo di strumenti innovativi per la Data Visualization e la self-service BI dismettendo i sistemi (o relativi moduli) non più rispondenti alle esigenze di business. La piattaforma supporterà la creazione di report, dashboard e infografiche e comprenderà i seguenti sistemi:

- Sistema che consente di analizzare grandi volumi di dati e facilita la realizzazione di grafici, dashboard e infografiche;
- Sistema che consente l'accesso ai dati e permette la creazione di report e dashboard interattive in modalità self BI;
- Sistema che abilita la realizzazione di query e report avanzati (saranno dismesse le funzionalità attualmente utilizzate per la creazione di dashboard);

Data Sharing & Publishing

Creazione di un ecosistema che supporta l'accesso e la condivisione del patrimonio informativo regionale verso utenti esterni. Si prevede l'evoluzione dei seguenti servizi:

- Servizio DaaS – sistema che permette l'accesso al patrimonio informativo regionale ad enti ricerca accreditati per finalità di analisi;
- Ecosistemi Digitali – sistemi, facilmente accessibile da diverse tipologie di utenti interni ed esterni all'ecosistema di Regione Lombardia, a supporto dell'interoperabilità tecnologica;
- Portale servizi DIH – sviluppo di un unico punto di accesso al patrimonio informativo per semplificare l'accesso a tutti i servizi offerti dal Digital Information Hub (es. DaaS, Open Data, API, ecc.).

3. **Data Analytics & Artificial Intelligence:** il cantiere prevede l'evoluzione delle piattaforme che consentono l'accesso al patrimonio informativo per finalità di analisi avanzate dei dati, introducendo anche tecnologie e strumenti di Artificial Intelligence.

Data Analytics, ML & AI

Evoluzione della attuale piattaforma al fine di abilitare la data exploration con funzionalità statistiche avanzate e permettere l'utilizzo di tecniche di data mining, intelligenza artificiale e machine learning. La soluzione offrirà la possibilità di effettuare analisi, elaborazioni e rappresentazioni del patrimonio informativo per creare valore per gli stakeholder.

4. **Knowledge Management and Compliance:** evoluzione delle piattaforme a supporto della gestione del ciclo di vita del dato e delle informazioni all'interno dell'organizzazione.

Information & Data Governance

Il cantiere ha l'obiettivo di introdurre o evolvere le piattaforme dedicate al governo dell'utilizzo dei dati al fine di migliorare la capacità decisionale dell'organizzazione "fluidificando" i processi di utilizzo dei dati. La piattaforma comprenderà specifiche componenti per la gestione dei seguenti ambiti:

- Master Data Management – per supportare i processi di creazione, gestione e condivisione delle anagrafiche aziendali;
- Information Architecture – con l'obiettivo di strutturare e organizzare ambiti informativi per supportare l'usabilità e la reperibilità delle informazioni, consentendo la creazione di un Business Glossary del patrimonio informativo regionale;
- Data Architecture – al fine di permettere la visualizzazione delle modalità di utilizzo dei dati, dove sono conservati e quale è il loro ciclo di vita;
- Data Quality – per misurare la qualità dei dati valutando fattori quali l'utilità per lo scopo specifico, la completezza, l'accuratezza, la tempestività, la coerenza, la validità e l'unicità.

Information Security & Operation

Il cantiere ha l'obiettivo di introdurre o evolvere gli strumenti dedicati a garantire la sicurezza di dati e informazioni, in conformità con le norme e le linee guida definite dagli organi europei e nazionali competenti. La piattaforma comprenderà le seguenti funzionalità:

- Data Protection – per proteggere i dati e le informazioni inattive crittografando dati sensibili e per garantire la sicurezza della rete realizzando appropriati sistemi firewall;
- Data Privacy & Policy – per classificare i dati al fine di anonimizzare / pseudoanonimizzare le informazioni sensibili o abilitare la creazione di dati sintetici ai fini di analisi, garantendo il rispetto di norme e direttive europee e nazionali;
- Identity & Access – per classificare le risorse informative gestite attraverso la Data Platform, al fine di assegnarne i privilegi di accesso introducendo meccanismi di Access Control;
- DevSecOps – per l'applicazione delle policy di sicurezza in tutte le fasi del ciclo di sviluppo di applicazioni e servizi.

Per tutti i cantieri, sarà valutata la possibilità di introdurre o migrare i sistemi sopra descritti in Cloud considerando eventuali riduzioni dei costi di gestione e/o incrementi nella qualità del servizio.

REALIZZAZIONE DEI DIGITAL ASSET A SUPPORTO DELLA DATA VALUE CHAIN

Il filone progettuale si suddividerà, indicativamente, in alcune linee di attività interconnesse con i cantieri relativi al filone progettuale di sviluppo ed evoluzione della Data Platform, di seguito dettagliate:

1. Sviluppo di conoscenza da dati e informazioni:

- Realizzazione di strumenti (es. Business Glossary, Data Catalog, ecc.) per comprendere e semplificare l'accesso al patrimonio informativo;
- Applicazione di metodi e strumenti di Data Architecture e Data Lifecycle per approfondire la conoscenza del patrimonio informativo;
- Implementazione di metodi e strumenti di Data Quality per verificare la qualità e le modalità di utilizzo del patrimonio informativo;
- Creazione, messa a disposizione e aggiornamento di Reference & Master Data (es. anagrafiche trasversali);
- Realizzazione delle appropriate strutture e procedure di stoccaggio dei dati (DataMart, Data Pool, ETL / ELT) per facilitare l'integrazione di diversi domini informativi;
- Creazione di dati sintetici attraverso algoritmi di Data Science per consentire l'utilizzo di dati tutelati;

- Analisi e predisposizione delle Data Source (es. banche dati esterne, fonti interne, ecc.) per l'ingestion e l'interconnessione dei dati all'interno della Data Platform.

2. Governo dei dati e delle informazioni:

- Definizione e implementazione di logiche privacy by design e data protection by design per l'evoluzione delle modalità di trattamento dati della Data Platform;
- Evoluzione del modello operativo di Data Management attraverso la valutazione della Data Maturity e la definizione del percorso per l'adoption e dissemination di processi e strumenti;
- Redazione e pubblicazione di linee guida e standard in ambito Data Management (es. Data Governance, Masterdata, Data Quality, Data Security, ecc.);
- Revisione e implementazione del modello organizzativo di Data Governance;
- Adozione degli approcci AnalyticOps, DataOps e MLOps per evolvere l'offerta di servizi data-driven in ottica Open Community e Open Science;
- Messa a disposizione di nuove modalità per la fruizione e l'analisi di dati provenienti da diverse fonti in ottica self-service BI e self-service Analytics.

I risultati attesi dell'iniziativa

L'iniziativa progettuale ha l'obiettivo di sviluppare una infrastruttura regionale per l'analisi dei dati per valorizzare il patrimonio informativo di Regione Lombardia e semplificarne l'accesso. L'introduzione o l'evoluzione dei principali componenti della Data Platform e la realizzazione di Digital Asset attraverso la piattaforma permetterà di ottenere i seguenti benefici.

DATA INGESTION, STORAGE, PROCESSING E PROVISIONING

- L'evoluzione dei sistemi di Data Ingestion, Data Lakehouse e Data Provisioning semplificherà l'attività di storage dei dati, eliminando la fase di preparazione e abilitando l'archiviazione di dati real-time senza limiti di formato. Eliminando la necessità di dover trasferire i dati in un sistema separato di Analytics, la soluzione agevolerà le analisi sui dati archiviati e darà la possibilità di accesso ad analisi avanzate (e.g. Machine Learning), attraverso la possibilità di utilizzare diversi linguaggi di analisi.
- L'introduzione della piattaforma di Data Virtualization permetterà agli utenti aziendali di fruire di dati sempre aggiornati all'interno delle loro applicazioni nascondendo le complessità di accesso ai dati dovute alla diversità delle Data Source, ai formati e alla struttura dei dati. La possibilità di connettersi in tempo reale alle fonti abilita viste di dati virtualizzati in grado di adattarsi in base ai cambiamenti delle fonti sottostanti (aggiunta, rimozione o modifica).

DATA VISUALIZATION & CONSUMPTION

- Le piattaforme di Data Visualization consentono di migliorare la produttività operativa abilitando il data-driven decision making e la self BI. I sistemi avanzati di Data Visualization riescono a trasmettere, attraverso elementi visivi, le informazioni rapidamente facilitando l'individuazione delle connessioni tra azioni e risultati e portando alla luce eventuali incongruenze o errori.
- L'evoluzione del servizio DaaS ha lo scopo di portare all'interno dell'ecosistema di Regione Lombardia, idee, competenze e conoscenze innovative, grazie alla messa a disposizione, ad enti di ricerca ed università, del patrimonio informativo regionale.
- L'evoluzione degli Ecosistemi Digitali ha lo scopo di migliorare la collaborazione e facilitare la capacità di innovazione ampliando il numero di fruitori dei servizi offerti da Regione Lombardia. Le nuove soluzioni tecnologiche svolgeranno un ruolo fondamentale nell'adempiere agli obblighi di amministrazione trasparente.

DATA ANALYTICS & ARTIFICIAL INTELLIGENCE

- L'evoluzione della piattaforma di Data Analytics & Artificial Intelligence consentirà un decision making più rapido ed efficace e supporterà la creazione di valore dai dati attraverso l'utilizzo di metodologie innovative di data preparation and analysis (es. Machine Learning, Artificial Intelligence). La soluzione abilita la realizzazione di Digital Asset orientati verso utenti interni / esterni a Regione Lombardia o finalizzati a supportare organizzazioni esterne nella generazione di valore attraverso l'impiego di dati e informazioni.

KNOWLEDGE MANAGEMENT AND COMPLIANCE

- L'evoluzione della piattaforma per la gestione della Information & Data Architecture è orientata verso un aumento della conoscenza dei dati aziendali, facilitando la Data Exploration e abilitando il controllo degli accessi ai dati con vari strumenti. Questi strumenti facilitano l'integrazione dei dati riducendo i tempi di data preparation e di creazione di Data Mart o Data Pool, permettono di uniformare il vocabolario e consentono di eliminare repliche e spostamenti di dati abilitando analisi in real-time.
- L'evoluzione della piattaforma per la gestione dei Master Data ha come scopo il miglioramento del decision making e la produttività delle analisi garantendo l'autorevolezza e l'integrità dei dati, grazie alla fruizione di dati accurati e aggiornati. Il posizionamento dei dati in un unico repository centrale semplifica il backup e la modifica dei dati e permette di assegnare l'accesso ai dati master sulla base di ruoli e responsabilità, in conformità con norme e direttive europee e nazionali.
- L'evoluzione della piattaforma per la gestione della Data Quality ha lo scopo di consentire agli utenti aziendali di rivedere, correggere e approvare le eccezioni alle regole di DQ e si pone l'obiettivo di creare un processo atto ad analizzare i dati per individuare e risolvere problemi di qualità. Questo sistema faciliterà il processo di creazione e test delle regole da parte degli utenti business, per ridurre tempi e costi di creazione.
- L'evoluzione degli strumenti di Data Security permetterà l'introduzione di funzionalità utili a garantire il rispetto delle linee guida riguardanti la gestione dei dati sensibili. Tali sistemi permetteranno di individuare e classificare i dati sensibili al fine di assegnare i corretti privilegi di accesso ai data user e di abilitare l'anonimizzazione e la pseudoanonimizzazione delle informazioni sensibili. Inoltre, saranno introdotte funzionalità per il monitoraggio degli accessi e la segnalazione di comportamenti anomali con l'obiettivo di evitare modifiche non autorizzate ai dati.