Ministero della Giustizia

MECCANISMO DI VALUTAZIONE

RULE OF LAW 2023

Data Lake



Roma/Bruxelles, 20 marzo 2023

Datalake – Introduzione

Il progetto **Datalake** si inserisce nell'ambito della linea di **Investimento 1.6.2 "Digitalizzazione delle grandi amministrazioni centrali"** del PNRR ed è parte della realizzazione del progetto **"Digitalizzazione del Ministero della Giustizia".** Il programma ha l'obiettivo di apportare benefici in termini di «aumento della conoscenza» dei dati, documenti e dei processi e, attraverso di essa, un indirizzamento ed efficientamento (qualitativo e temporale) delle scelte strategiche di carattere organizzativo e normativo dell'Amministrazione.

GLI OBIETTIVI DEL PROGRAMMA

L'obiettivo del Programma è la realizzazione di **6 sistemi di conoscenza** che consentiranno un accesso immediato al patrimonio informativo del Ministero attraverso l'introduzione di funzionalità di:

- **Pseudoanonimizzazione** dei dati sensibili
- **Data Governance** dell'intero patrimonio informativo civile e penale
- **Analisi multidimensionali** in tempo reale
- **Business Intelligence**
- **Record Linkage (**correlazione tra dati strutturati e documenti)

La realizzazione dei sistemi apporterà benefici in termini di «**aumento della conoscenza**» sui dati e sui processi **consentendo un indirizzamento ed efficientamento (qualitativo e temporale) delle scelte strategiche di carattere organizzativo e normativo dell'Amministrazione**

I 6 SISTEMI DI CONOSCENZA

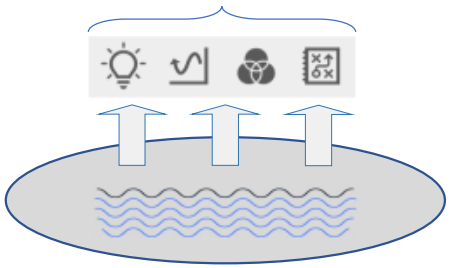
- Sistema di anonimizzazione sentenze civili e penali
- Sistema monitoraggio lavoro uffici giudiziari
- Sistema analisi di dati e documenti nei procedimenti in ambito civile
- Sistema analisi di dati e documenti nei procedimenti in ambito penale
- Sistema di statistiche avanzate su processi civili e penali
- Sistema automatizzato identificazione rapporto vittima-colpevole

Principali caratteristiche dell'iniziativa Datalake

FINALITÀ E DESTINATARI

Utenti interni
Uffici giudiziari e Uffici ministeriali

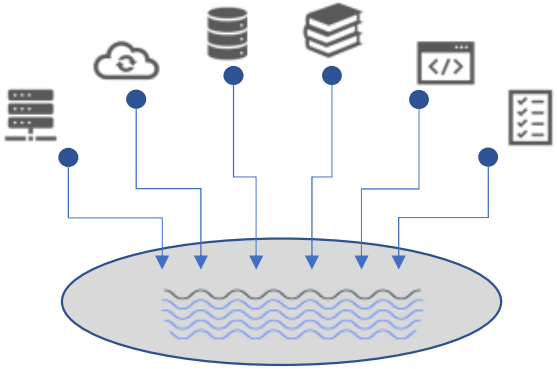
Utenti esterni
Interessati e abilitati



Data Lake

Data Discovery, Data Governance e **analisi multidimensionali** in tempo reale attraverso la **correlazione dei dati strutturati e documentali**, a beneficio delle scelte strategiche di carattere organizzativo e normativo

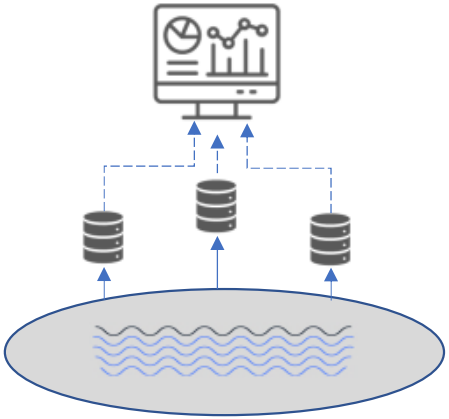
PATRIMONIO INFORMATIVO ESISTENTE



Data Lake

Sarà costituito un **nuovo repository** di dati strutturati utilizzando **il patrimonio** e le basi dati **distrettuali e nazionali**

PATRIMONIO INFORMATIVO ESISTENTE



Data Lake

Verranno realizzati **Datamart specifici** in base alle misure e dimensioni di interesse per lo sviluppo di cruscotti di **Business Intelligence** relativi a statistiche nazionali e distrettuali

I sei sistemi di conoscenza

P1 - Anonimizzazione sentenze civili e penali

Il Sistema prevede la pseudoanonimizzazione delle sentenze per consentirne un uso di studio e di analisi del giudicato nel rispetto dell'anonimato dei soggetti fisici e giuridici coinvolti.

L'obiettivo di tale progettualità consentirà la pubblicazione delle sentenze all'interno della nuova Banca Dati di Merito Civile

P2 – Monitoraggio uffici giudiziari

Il Progetto intende mettere a disposizione degli uffici giudiziari strumenti di supporto al monitoraggio e controllo dei tempi di lavorazione dei procedimenti, ai fini della migliore utilizzazione delle risorse umane e strumentali. Il Progetto riguarda sia il settore civile sia il settore penale ed è altresì funzionale alla misurazione degli obiettivi quantitativi per i procedimenti giudiziari e dei target previsti dal PNRR

P3 – Sistema di conoscenza civile

Il Sistema punta a fornire ai vari uffici strumenti utili a velocizzare il processo decisionale. Aumentando l'interoperabilità tra le fonti dati verranno fornite piattaforme per:

- Gestire la conoscenza del fascicolo del processo
- Analizzare le sentenze, i fascicoli e i registri del processo
- Massimare le sentenze civili
- Potenziare il DWH Civile

P4 – Sistema di conoscenza penale

Il Sistema punta a fornire ai vari uffici strumenti utili a velocizzare il processo decisionale. Aumentando l'interoperabilità tra le fonti dei dati verranno fornite piattaforme per:

- Gestire la conoscenza del fascicolo del processo
- Analizzare le sentenze, i fascicoli e i registri del proc. penale
- Realizzare un Lake House Penale
- Realizzare un Data Mart per il fenomeno della recidiva

P5 – Statistiche avanzate (civile e penale)

Il sistema di statistiche avanzate ha l'obiettivo di sviluppare una piattaforma in grado di fornire in tempi rapidi statistiche giudiziarie con elevata qualità del dato ed in grado di elaborare informazioni provenienti da fonti dato eterogenee, in ambito sia civile sia penale

P6 – Relazione vittima autore

L'obiettivo consiste nell'individuazione delle relazioni tra vittima e autore, con particolare riferimento alla violenza di genere, utilizzando tecniche di Natural Language Processing e Machine Learning.

Attraverso l'estrazione degli elementi salienti circa le diverse forme di violenza di genere che emergono dalle sentenze penali e civili, sarà possibile correlare tali dati con le informazioni provenienti dai registri per poter fornire un analisi completa del fenomeno

Principali avanzamenti progettuali

P1 - Anonimizzazione sentenze civili e penali

- Definizione e presentazione del **framework per la pseudonimizzazione** dei **dati non strutturati**, a supporto della Banca dati di Merito Civile
- Realizzato Engine di **Pseudonimizzazione dati strutturati**

P2 – Monitoraggio uffici giudiziari

- Effettuata la prima fase di analisi e definito il primo subset di use case per la gestione del monitoraggio
- In corso attività finalizzate all'integrazione dei dati del sistema del Personale con i dati dei processi civile e penale
- Creazione di nuovi **cruscotti per monitoraggio PNRR (arretrati civile, pendenze e disposition time)**

P3 – Sistema di conoscenza civile

- Definito primo subset di **Use case sulle** sentenze
- Analisi per **dashboard sul tema «Esecuzioni Immobiliari»**
- In corso attività di raccolta requisiti per la strutturazione di ulteriori use case in collaborazione con i Magistrati della **Corte d'Appello di Milano**

P4 – Sistema di conoscenza penale

- Definito primo set **Use Case in ambito Cognizione ed Esecuzione**
- Definito primo set di reporting per l'analisi del **fenomeno della Recidiva**
- Avviate attività per la creazione dei **DataMart per le statistiche del DAP** e del **Casellario**
- **Definiti dashboard per analisi area penale interna ed esterna**

P5 – Statistiche avanzate (civile e penale)

- Avviata fase di studio per la realizzazione dei cruscotti finalizzati all'analisi delle spese di Giustizia
- Realizzata prima versione dei **cruscotti sui Giudici di Pace** in ambiente pilota Giustizia

P6 – Relazione vittima autore

- In corso analisi di un subset di fascicoli forniti dai Tribunali di Roma e Napoli, finalizzato allo studio del fenomeno e alla realizzazione di un modello di analisi dedicato

Macropiano di riferimento

Allo stato attuale sono state avviate in parallelo tutte le iniziative inerenti alla realizzazione dei 6 sistemi di conoscenza, gestendo dei «quick win» finalizzati a fornire dei risultati tangibili in breve tempo all'Amministrazione

