



ALLEGATO 2- AREA TEMATICA DI RICERCA

“Data Governance and Data Protection (DGDP)”

A. Introduzione

I progressi nel campo delle Tecnologie dell'Informazione e della Comunicazione (ICT) e la disponibilità di una varietà di collezioni di dati offrono l'opportunità di accedere facilmente ed elaborare un'enorme quantità di informazioni, ottenendo così una descrizione estremamente dettagliata del comportamento delle persone e delle attività delle organizzazioni. La disponibilità di queste collezioni di dati, insieme alla grande capacità delle moderne infrastrutture di rete e di calcolo, offre chiaramente importanti vantaggi agli utenti, alle organizzazioni e alla società in generale. La raccolta, condivisione e analisi dei dati, con il contributo di soggetti diversi, è infatti un grande fattore abilitante per una società sempre più evoluta digitalmente.

Un chiaro ostacolo alla realizzazione di tale visione è rappresentato dalle preoccupazioni relative alla sicurezza e alla privacy. I dati possono essere sensibili o riservati e non possono essere condivisi apertamente. Più in generale, i dati possono essere soggetti a restrizioni di accesso e utilizzo (comprese quelle derivanti dalle normative sulla protezione dei dati) e la loro riservatezza, nonché la loro integrità, dovrebbero essere garantite anche quando i dati sono memorizzati o processati da terze parti. Senza una adeguata protezione dei dati esiste, infatti, un forte rischio di uso improprio, con minacce alla tutela della riservatezza e della privacy di cittadini, aziende e istituzioni. Per la piena realizzazione di una moderna società digitale, è quindi fondamentale fornire soluzioni efficaci per la governance e la protezione dei dati.

Spoke 10, organizzato in un singolo progetto (DGDP), risponde a questa sfida sviluppando soluzioni avanzate che consentono a diversi attori (ad es. individui, aziende e istituzioni) di mantenere il controllo sui propri dati in vari scenari di rilascio, condivisione e analisi. Le soluzioni permetteranno anche l'implementazione delle restrizioni dettate dalla legislazione e regolamentazione sulla protezione dei dati. Lo Spoke affronta il problema in modo olistico, esaminando diversi aspetti della governance e della protezione dei dati durante l'intero ciclo di vita dei dati e in vari scenari.

Il progetto DGDP mira allo sviluppo di nuove soluzioni per la gestione privata e sicura dei dati in scenari moderni ed emergenti, consentendo la condivisione dei dati e l'analisi collaborativa, permettendo ai proprietari dei dati di mantenerne il controllo. Il progetto è organizzato in quattro Work Package (WP), che affrontano diversi problemi di ricerca cruciali per la realizzazione di questo obiettivo.

Il presente documento è allegato al Bando pubblico per la selezione di proposte progettuali per lo svolgimento di tutti i task inclusi nel Work Package 4 (WP4) del progetto DGDP.

Dopo una sintesi degli aspetti principali del progetto, si passa alla descrizione in maggior dettaglio dell'area tematica, delle attività, della pianificazione e i dei deliverable da realizzarsi.

B. Struttura del Progetto DGDP in Work Package

Nell'ambito dello Spoke 10, il progetto DGDP è strutturato nei WP di seguito descritti.

a. WP1 - Data Governance Framework

WP1 si concentra sulla progettazione di un framework di governance per la gestione sicura e privata dei dati. Il framework includerà la definizione di soluzioni per la modellazione di dati e metadati per esprimere le proprietà che devono essere considerate per la regolamentazione dell'accesso e della condivisione dei dati,

per le relative specifiche ontologiche e ragionamenti correlati, e per l'identificazione e l'analisi dei requisiti di protezione. L'obiettivo è creare un framework completo in grado di esprimere e ragionare su requisiti provenienti da diverse parti, inclusi quelli imposti da normative sulla privacy (come il GDPR) e quelli derivanti da bisogni delle organizzazioni e dagli individui. Le attività del WP1 sono organizzate nei seguenti task:

Task 1.1 - Modellazione di dati e metadati (UNIMI, UNIROMA1)

Task 1.2 - Ontologie e ragionamento sui metadati (UNIMI, UNICA, UNIROMA1)

Task 1.3 - Requisiti di protezione dei dati (UNIMI, LEONARDO)

b. WP2 - Politiche di protezione dei dati

WP2 si concentra sulla definizione di soluzioni per supportare la specifica e l'applicazione dei requisiti di protezione dei dati, attraverso linguaggi e modelli per esprimere e ragionare su tali requisiti e per supportare la valutazione e l'amministrazione delle politiche. L'obiettivo è quello di offrire un modello e un linguaggio semplici, per consentirne l'uso effettivo da parte di non esperti del settore, ma anche flessibili, estendibili ed espressivi, quindi in grado di rappresentare i diversi requisiti di protezione dei dati. Le attività del WP2 sono organizzate nei seguenti task:

Task 2.1 - Modello e linguaggio delle politiche (UNIMI, UNIFI, UNIROMA1)

Task 2.2 - Implementazione delle politiche (UNIMI, UNIFI, UNIROMA1)

Task 2.3 - Composizione, ragionamento e amministrazione delle politiche (UNIFI, UNIROMA1)

c. WP3 - Wrapping e sanitizzazione dei dati

WP3 si concentra sulla definizione di tecniche per implementare la sicurezza e la privacy durante tutto il ciclo di vita dei dati. Tali tecniche includono soluzioni per sanitizzare i dati e garantirne la privacy e la riservatezza in base a diverse metriche e rischi per la privacy ed inoltre anche tecniche di wrapping che consentono la protezione dei dati nella loro memorizzazione e analisi. Il WP3 si occupa anche del supporto della funzionalità di accesso e analisi sui dati protetti con considerazione di diversi scenari di condivisione dei dati e di nuove applicazioni di Intelligenza Artificiale (AI) e Machine Learning (ML). Il WP3 sviluppa inoltre soluzioni per consentire la valutazione delle garanzie di protezione e la verifica del corretto funzionamento del sistema. Le attività del WP3 sono organizzate nei seguenti task:

Task 3.1 - Metriche per la privacy e modelli di privacy (UNIMI)

Task 3.2 - Tecniche di sanitizzazione dei dati (UNIMI, UNICA, UNISA)

Task 3.3 - Wrapping dei dati e protezione dei dati per condivisione e analisi efficiente ed efficace (UNIMI, UNISA)

Task 3.4 - Interrogazione di, e computazione su, dati protetti (UNIMI, UNICA, UNISA)

Task 3.5 - Scenari applicativi di AI/ML (UNIMI, UNISA, LEONARDO)

Task 3.6 - Verifica e valutazione (UNIMI, UNISA, LEONARDO)



d. WP4 – Integrazione con tecnologie e modelli di gestione dei dati emergenti

WP4 si concentrerà sulle tecnologie emergenti di gestione dei dati, studiando l'applicazione della protezione dei dati a nuovi modelli e moderne piattaforme di gestione su larga scala, di scenari collaborativi distribuiti, garantendo l'applicazione efficiente ed efficace - e quindi la scalabilità e applicabilità – delle misure di protezione, con l'obiettivo di ridurre l'impatto sulla funzionalità di utilizzo dei dati. Le attività del WP4 sono organizzate nei seguenti task:

Task 4.1 - Nuovi modelli di dati

Task 4.2 - Moderne piattaforme di gestione dei dati su larga scala

Task 4.3 - Applicazione scalabile ed efficiente della protezione

Task 4.4 - Scenari distribuiti e collaborativi

C. Ambito e attività previste

L'ambito di riferimento del Bando pubblico per la selezione di proposte progettuali a cui si riferisce il presente allegato riguarda lo svolgimento di tutti i task inclusi nel Work Package 4 (WP4) del progetto DGDP di Spoke 10 nel contesto del Partenariato Esteso SERICS. L'obiettivo generale del WP4 è lo studio di nuovi approcci e tecniche per la gestione dei dati nel rispetto dei requisiti di sicurezza e privacy in scenari caratterizzati dall'adozione di nuove tecnologie per la gestione dei dati. Il lavoro nel WP4 opererà a diversi livelli, dalla gestione di nuovi modelli di dati, con un'enfasi sulla rappresentazione e gestione delle politiche di controllo degli accessi, alla progettazione di tecniche di protezione per grandi collezioni di dati, considerando la compatibilità con il panorama tecnologico e requisiti di performance e scalabilità. Il lavoro sarà sviluppato puntando all'integrazione con gli altri WP. I quattro Task in ambito prevedono lo svolgimento di attività di ricerca fondamentale organizzate come segue.

a. WP4 Task 4.1 - Nuovi modelli di dati

L'obiettivo del Task 4.1 è sviluppare soluzioni di protezione dei dati in grado di operare su modelli di dati flessibili ed espressivi, supportando la realizzazione di politiche che regolano l'accesso e l'utilizzo dei dati. La ricerca nel Task 4.1 dovrebbe riguardare ricerche e applicazioni che vanno oltre le soluzioni basate sulla classica tecnologia relazionale. Le politiche che rappresentano i diversi requisiti di privacy e sicurezza delle moderne applicazioni possono ottenere un vantaggio significativo dall'adattamento di queste tecnologie. I risultati della ricerca in questa direzione possono trovare applicazione in molti domini, in particolare quando l'integrazione dei dati va oltre soluzioni rigide basate sulla tecnologia relazionale classica. Il dominio dei mercati dei dati digitali rappresenta un esempio di un dominio in cui c'è una forte necessità di politiche espressive. Le direzioni di ricerca possono includere la considerazione di sistemi di gestione dati a grafo, RDF e linked data per la modellazione dei dati, tecnologie e ontologie del web semantico, approcci basati sulla logica per supportare il ragionamento su collezioni di dati distribuite e su larga scala, nonché l'adattamento di soluzioni avanzate di modellazione e analisi del software.

b. WP4 Task 4.2 - Moderne piattaforme di gestione dei dati su larga scala

La quantità di dati gestiti dai moderni sistemi informativi è in continuo aumento seguendo una crescita esponenziale. Le applicazioni che non possono funzionare sui tradizionali sistemi centralizzati stanno diventando sempre più comuni, con molti utenti al di fuori delle grandi aziende. In questi casi è necessario

affidarsi a piattaforme di gestione dei dati su larga scala, spesso di proprietà di fornitori cloud, che attualmente offrono un supporto limitato per la rappresentazione e la gestione dei requisiti di sicurezza. L'obiettivo del Task 4.2 è adattare ed estendere a questi nuovi domini le soluzioni per il controllo degli accessi sviluppate nei tradizionali sistemi di gestione dei dati. Il lavoro dovrà tenere conto delle caratteristiche specifiche delle tecnologie utilizzate per la gestione delle raccolte dati in ambienti Big data e NoSQL. Di particolare interesse è l'integrazione con soluzioni Apache Software Foundation.

c. WP4 Task 4.3 - Applicazione scalabile ed efficiente della protezione

L'applicazione dei controlli di sicurezza è sempre stata guidata dalla compatibilità con i requisiti di performance. Le soluzioni tradizionali per reti e sistemi operativi hanno sempre considerato fondamentale la capacità di far rispettare le politiche di sicurezza introducendo un piccolo ritardo nelle risposte alle richieste di accesso. La natura distribuita di molte piattaforme di gestione dei dati moderne da un lato impone la rappresentazione di politiche più complesse, dall'altro consente l'utilizzo di risorse computazionali maggiori per la valutazione delle restrizioni di sicurezza. L'obiettivo del Task 4.3 è sviluppare soluzioni di protezione dei dati e di controllo degli accessi che pongano al centro il rispetto dei requisiti di performance e scalabilità, ad esempio rispetto alla considerazione dei requisiti di sicurezza nell'identificazione di strategie di esecuzione efficienti. L'efficienza può essere valutata non solo in termini di tempo di risposta, ma anche rispetto al costo economico e/o all'impatto ambientale dell'esecuzione di un determinato carico di lavoro di elaborazione dati. L'obiettivo è inoltre progettare nuovi approcci che considerino tutte le dimensioni che caratterizzano questo problema: tempi di risposta, requisiti di sicurezza e privacy, costo economico per l'uso delle risorse cloud, impatto ambientale. Ulteriori sfide vengono introdotte dalla considerazione delle opportunità offerte dal edge-cloud continuum.

d. WP4 Task 4.4 - Scenari distribuiti e collaborativi

I moderni sistemi di gestione dei dati richiedono l'integrazione di dati provenienti da una varietà di fonti, utilizzando una varietà di formati e vedendo la cooperazione di più parti. Obiettivo del Task 4.4. è lo sviluppo di soluzioni in scenari caratterizzati dal coinvolgimento di più attori. La molteplicità può essere considerata da diverse prospettive: lato utente/consumatore dei dati, lato proprietario/responsabile del trattamento, e/o lato fornitore di dati o di calcolo. Le possibili direzioni di ricerca possono anche considerare scenari in cui i dati non sono gestiti in modo centralizzato, ma sono prodotti e consumati in modo decentralizzato/peer-to-peer, con l'obiettivo di passare da un approccio per la protezione dei dati servizio-centrico a un approccio utente-centrico, in cui gli utenti possono avere il controllo sui propri dati in contesti distribuiti e collaborativi. Le direzioni di ricerca possono includere il supporto per approcci user-friendly per la condivisione controllata dei dati, la considerazione di diversi scenari distribuiti e le relazioni di fiducia tra i diversi attori (ad esempio, utenti e organizzazioni), nonché le blockchain. La ricerca può prendere in considerazione diverse soluzioni architetturali per i meccanismi di applicazione (ad esempio, cloud, edge, fog) per trovare un compromesso tra efficienza e garanzie di sicurezza/privacy.

D. Piano delle attività e milestone

Il piano di lavoro del progetto DGDP è in linea con il piano generale del progetto SERICS riportato nell'Allegato C "cronoprogramma di attuazione e piano dei pagamenti" del Decreto MUR nr. 1556 del 11.10.2022 e successive modificazioni (Cronoprogramma).

Coerentemente con il Cronoprogramma, il progetto DGDP ha una durata prevista di 36 mesi come rappresentato in Figura 1, in cui sono riportate unicamente le attività e milestone maggiormente rilevanti per l'ambito del presente Bando.

Il progetto DGDP è stato avviato il 1° gennaio 2023 (evidenziato con T0 in Figura 1).

Si noti che la pianificazione riportata in Figura 1 non mostra il dettaglio dei Work Package in quanto sebbene esista una naturale interdipendenza tra di essi (ad esempio, i linguaggi/modelli delle politiche previste si baseranno anche sulla modellazione di dati e metadati e potranno ricorrere a soluzioni di wrapping e sanitizzazione per garantire la protezione), tutti i Work Package e i loro Task mirano allo sviluppo di nuove soluzioni di ricerca nel loro ambito e saranno attive in parallelo durante tutta la durata del progetto.

| Milestone | Descrizione | T0 | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 | M13 | M14 | M15 | M16 | M17 | M18 | M19 | M20 | M21 | M22 | M23 | M24 | M25 | M26 | M27 | M28 | M29 | M30 | M31 | M32 | M33 | M34 | M35 | M36 | |
|-----------|------------------------|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| M1.0 | Reclutamento Personale | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| M11.0 | Ricerca Fase 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| M21.0 | Ricerca Fase 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Figura 1. Cronoprogramma SERICS (estratto)

Inoltre i Soggetti Beneficiari saranno tenuti a partecipare alle attività di coordinamento e controllo avanzamenti ed a effettuare la rendicontazione periodica delle spese ammissibili con la frequenza espressa nelle linee guida di rendicontazione di riferimento e comunque con cadenza almeno mensile non oltre il secondo giorno lavorativo successivo alla fine di ogni mese, sino alla completa rendicontazione di tutte le spese ammissibili di propria competenza. Contestualmente i Soggetti Beneficiari dovranno inoltre fornire una descrizione dell'avanzamento delle attività di ricerca condotte nel mese di riferimento che evidenzii le attività svolte nel mese, i risultati raggiunti, le pubblicazioni scientifiche realizzate (ove applicabile), secondo le modalità e standard definiti da UNIMI.

Le linee rosse verticali riportate nel Cronoprogramma di Figura 1 identificano i punti di controllo che separano i periodi di rendicontazione tecnico scientifica. In corrispondenza di ciascun punto di controllo, i soggetti che partecipano al progetto DGDP devono rendicontare i risultati tecnici e scientifici ottenuti nel periodo di riferimento. Allo stesso modo, i Soggetti Beneficiari dovranno fornire una relazione tecnica e scientifica per ciascuno dei punti di controllo riportati in Figura 1 (linee rosse) che ricada all'interno del loro periodo progettuale, oltre che alla conclusione delle proprie attività, nei tempi e nelle modalità che saranno richieste da UNIMI. La relazione tecnica e scientifica descriverà l'avanzamento delle attività nel periodo di rendicontazione, i risultati raggiunti, le pubblicazioni scientifiche o altri deliverable in corso di realizzazione, secondo le modalità e standard definiti da UNIMI allegando alla relazione tutte le informazioni utili e le relative evidenze documentali.

I Soggetti Beneficiari si impegnano a partecipare con proprio personale alle attività di coordinamento e controllo degli avanzamenti e ad adeguare su richiesta di UNIMI la pianificazione delle proprie attività coerentemente ad eventuali modifiche che dovessero intercorrere riguardanti il Cronoprogramma.

E. Deliverable

Le attività svolte nel WP4 produrranno soluzioni di ricerca avanzate e relativi strumenti per la governance e la protezione dei dati.

I deliverable da realizzare in ambito ricerca includono le pubblicazioni scientifiche che presentano i risultati di ricerca, ulteriore documentazione di ricerca ad essi afferenti oltre ad eventuali componenti software realizzati in tale contesto.



Le pubblicazioni e altra documentazione realizzata dovranno necessariamente presentare il testo dell'Acknowledge SERICS seguente: "This work was supported in part by project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU." e rispondere alle ulteriori specifiche presenti nel Bando.

Spoke10 promuove il concetto di open science: i risultati finali dovranno essere sotto forma di pubblicazioni scientifiche e documenti di ricerca pubblicati in modalità Open Access e nel caso di software dovrà essere Open Source.

Il livello target TRL (Technology Readiness Level) atteso è 4-5.

I Soggetti Beneficiari realizzeranno inoltre i rendiconti tecnico scientifici corrispondenti ai punti di controllo (per quelle linee rosse di Figura 1 che ricadono all'interno del loro periodo progettuale) e contribuiranno alla realizzazione dei deliverable complessivi del Progetto DGDP integrando in essi le informazioni afferenti ai risultati ottenuti nelle attività di propria competenza.