

ALLEGATO A – CAPITOLATO TECNICO

CONTESTO DI RIFERIMENTO

La tipologia di dati soggetti ad anonimizzazione sono i dati personali degli utenti che accedono ai servizi del Comune di Parma (es. cittadini) e quindi dati anagrafici, di contatto, sensibili e in minima parte giudiziari.

I dati sono gestiti tramite database relazionali proprietari:

- ORACLE 11.2.04;
- ORACLE 19.9;
- Microsoft SQL Server 2022;

e database open source:

- MySQL;
- PostgreSQL.

I database sono ospitati in parte presso il Data Center del Comune di Parma e in parte nei cloud dei servizi SaaS.

I database nel Data Center del Comune di Parma sono gestiti direttamente da It.City e sono accessibili senza vincoli o restrizioni.

I database dei servizi SaaS gestiti dai fornitori possono essere acceduti tramite servizi di comunicazione quali ad esempio: microservizi, API, web services, etc.

Il processo di anonimizzazione è applicato ai soli database che hanno ambienti di test/stage/sviluppo ospitati nel data center del Comune di Parma.

Non sono presenti strutture dati multilivello cross database, il processo di anonimizzazione è applicato a sottoinsiemi di dati (più o meno complessi) memorizzati in un singolo database.

OGGETTO DELLA FORNITURA

1. Servizio biennale INFORMATICA IDMC (360 IPU/anno)
2. Assistenza per il processo di data masking
3. Progetto pilota per l'analisi delle funzionalità della piattaforma

Servizio INFORMATICA IDMC

Oggetto della fornitura è il servizio di “Intelligent Data Management Cloud” completo delle seguenti caratteristiche di base:

- Disponibilità di connettori per l'acquisizione semplificata e continuativa di dati da molteplici tipi di database, in particolare: Oracle e Microsoft SQL Server;
- Capacità di eseguire task di anonimizzazione dati su input strutturati (e.g., subset di tabelle con relazioni o interi database, subset di righe o di colonne di una stessa tabella)

o semi-strutturati (e.g., JSON), producendo in output file dello stesso tipo, in risposta ad un servizio chiamante o tramite flussi batch;

- Pannello di amministrazione del sistema, che permetta:
 - Definizione dei modelli: business data modeling, data privacy model / profiling;
 - Esplorazione (input): applicazione di filtri e regole, visualizzazione dati (input);
 - Creazione dei flussi: configurazione del flusso per progetto, data mapping (output);
 - Analisi dell'accuratezza (output): misurazione rischio (e.g., probabilità di re-identificazione) e misurazione valore (utilità e integrità residua dell'informazione) dell'output;
 - La gestione delle utenze per l'accesso ai file di input e output in modalità sicura e controllata.
- Possibilità di applicare dinamicamente più algoritmi di anonimizzazione;
- Cruscotto di monitoraggio, con possibilità di verificare i flussi non andati a buon fine ed avere evidenza in tempo reale delle situazioni in cui l'analisi dell'output abbia fornito minore accuratezza;

il servizio INFORMATICA IDMC dovrà prevedere un minimo di 360 IPU/anno; le IPU (Informatica Processing Unit) esprimono le unità di calcolo del software così come descritto nel documento "price guide" reperibile sul sito "<https://www.informatica.com>"

Assistenza per il processo di data masking

Il Fornitore dovrà mettere a disposizione le risorse (personale e strumenti) per avviare un primo use case di DATA MASKING (anonimizzazione), su database ORACLE e Microsoft SQL, in affiancamento al personale It.City in modalità "training on the job".

Lo use-case dovrà prevedere l'anonimizzazione ed eventualmente il discovery dei subset del Data Warehouse Microsoft SQL Server e del database di Anagrafe ORACLE contenenti i dati personali e di contatto degli utenti.

Non è richiesta la presenza in loco presso gli Uffici di It.City.

Gli obiettivi dell'attività di anonimizzazione sono:

- Minimizzazione del rischio di re-identificazione (misura del livello di accuratezza), ad esempio: in funzione della numerosità degli attributi, della eterogeneità dei valori o della grandezza del campione;
- Massimizzazione del valore informativo del dataset originale (minimo utilizzo di generalizzazione).

A titolo esemplificativo, non esaustivo, si riportano alcune tecniche di anonimizzazione:

- Mascheramento o offuscamento: prevede la cancellazione o sostituzione parziale o totale del valore della variabile; di fatto mantiene la validità dal punto di vista formale, ma non da quello semantico;
- Pseudonimizzazione: consiste nel trattamento dei dati personali in modo tale che i dati personali non possano più essere attribuiti a un soggetto specifico senza l'utilizzo di informazioni aggiuntive;

- Permutazione: prevede il mescolamento di valori all'interno di un dataset in modo che risultino artificialmente collegati ai soggetti del dataset, in modo da rendere impossibile ricondurre con certezza tali valori ai soggetti originali ed allo stesso tempo mantenendo il valore statistico del dataset.

Il Fornitore dovrà descrivere nella proposta come il sistema utilizza tali tecniche, attraverso quali strumenti e soluzioni. Il Fornitore dovrà, quindi, mostrare come la strategia di anonimizzazione che ne deriva è qualcosa di già implementato nella piattaforma cloud.

Il Fornitore dovrà inoltre descrivere quale tecnica utilizza per calcolare il codice fiscale in coerenza con i dati anonimizzati.

Le attività di DATA MASKING (anonimizzazione) devono prevedere i seguenti deliverable:

- Piano attività di anonimizzazione, da mantenere aggiornato periodicamente, al più tardi ogni due settimane;
- Descrizione delle operazioni effettuate;
- Descrizione degli eventuali problemi riscontrati.

Progetto pilota Proof of Value

Il Progetto pilota ha l'obiettivo di fornire le informazioni minime necessarie per la valutazione delle principali funzionalità della piattaforma INFORMATICA IDMC in materia di DATA ARCHITECT.

Il progetto pilota (PoV) dovrà prevedere i seguenti deliverable:

- Documento con descrizione dell'intera piattaforma INFORMATICA IDMC (architettura di integrazione e funzionalità);
- Sessioni formative di base sulle principali funzionalità della piattaforma INFORMATICA IDMC;
- Opzionale: Valutazione/definizione e presentazione di un eventuale progetto evolutivo quale ad esempio il data fabric, un approccio all'architettura che garantisce una modalità unificata e coerente per l'accesso, la gestione e la governance dei dati all'interno dell'Ente.

Bologna, 07/03/2025

Il Responsabile del Procedimento
(F.to Piero Santovito)

Firmato digitalmente da: Piero
Santovito
Data: 07/03/2025 12:12:30